Universiteit van Amsterdam

Function of creak in Korean: A perception study

rMA Thesis

Patrik Hrabánek

Supervised by Silke Hamann and Michaela Watkins

Faculty of Arts and Humanities
University of Amsterdam
The Netherlands

June 19, 2024

Abstract

This study investigates the influence of creaky voice on the perception of the fortis-lenis word-initial stop distinction in Standard Seoul Korean. Given that creaky voice often accompanies vowels following fortis stops, we examined whether creak serves as an additional cue for the perception of a fortis stop. Thirty two (n=32) native Korean speakers participated in an online forced-choice ABX task, where they determined whether an artificially creaky lenis token with ambiguous values of VOT and F0 resembled a natural fortis or lenis token more closely. While creak did not predominantly cause the perception of fortis stops, it significantly influenced the categorization of stops, suggesting that creak may serve as an additional cue alongside VOT and F0. Interestingly, the presence or absence of creak in the natural fortis comparison sounds did not significantly alter the participants' choices, implying that listeners likely categorize sounds mentally and then compare these mental representations, rather than compare purely auditory information. These findings open avenues for future research into the developmental aspects of creak perception and its role in language acquisition.

Contents

1. Introduction

1.1 Korean Laryngeal Contrast

Korean is known for having a typologically uncommon three-way laryngeal contrast among word-initial stop consonants. In Standard Seoul Korean (SSK), these stops are separated into the categories of aspirated, fortis, and lenis (see Table 1). Aspirated stops are often described as strongly aspirated; fortis stops as tense, forced, or constricted, while the lenis stops are referred to as lax, weak, or slightly aspirated (Kang et al., 2022). Importantly, they are all voiceless in word-initial position (but see Kim & Duanmu, 2004 for an opposing account), although the lenis stop type generally becomes voiced word-medially (Kang et al., 2022).

Table 1: Overview of Korean word-initial stops.

	Aspirated	Fortis	Lenis
Labial	ㅍ /$\mathrm{p^h}$/	ㅃ /p*/	ㅂ /p/
Coronal	ㅌ /$\mathrm{t^h}$/	ㄸ /t*/	ㄷ /t/
Velar	ㅋ /$\mathrm{k^h}$/	ㄲ /k*/	ㄱ /k/

Phonetically, the three stop types display several differences between each other. Early literature describes mainly differences in voice-onset time (VOT), wherein the fortis stop produces shortest VOT and the aspirated stop produces the longest VOT, with the lenis stop lying somewhere in the middle (Abberton, 1972; Lisker & Abramson, 1964; Weitzman & Han, 1966). Quite early on, though, differences in fundamental frequency (F0) of the following vowel were also described (Kagaya, 1974; Kim, 1965; Weitzman & Han, 1966). The results of these studies showed that aspirated and tense stops lead to a higher F0 of the following vowel because they instigate a raised glottal position which lengthens and stiffens the vocal folds (Kim et al., 2010), while the lenis stop leads to a lower F0. Other studies focused more closely on the behavior of the glottis during the pronunciation of the different stop types, showing that the glottal aperture (size of the opening between vibrating vocal folds) was similar for aspirated and lenis stops, with both being much wider compared to the nearly complete contact of the vocal folds of the fortis stop (Abberton, 1972; Kagaya, 1974; Kim, 1965). It was the latter studies that also first mentioned the tensity of the glottis associated with the tense stop type, which might lead to the presence of creaky voice in the directly following vowels.

The problem with much early literature on the topic of Seoul Korean stop types is the predominant focus on VOT. Initially, many researchers considered VOT as the central cue used by listeners to distinguish among the stop types, which led to research centered on VOT manipulations (as summarized in Lee et al., 2020). However, this approach presented its shortcomings, as the modulation of VOT could only influence the perception of the aspirated vs. lenis categories; the perception of a fortis stop type could not be reliably incited just by shortening VOT (Han & Weitzman, 1970). Despite this, not many studies investigated the influence of other cues until the start of the new century (although see Han, 1998 and Lee, 1990).

In 2000, Lee & Jung conducted a perception study in which they manipulated VOT and F0 as cues in direct opposition, with the results pointing towards F0 of the vowel being a more important cue than VOT. Around the same time, an ongoing sound change in Standard Seoul Korean (SSK) was starting to be noticed and was pointed out a few years later through production experiments (Silva, 2006). Lenis stops produced by younger speakers (typically born after 1965) were found to differ from lenis stops of older speakers, in that they no longer displayed VOT values between fortis and aspirated stops, but rather their VOT values totally overlapped with those of the aspirated stops (Kang, 2014; Silva, 2006). Thus, younger speakers' aspirated stops were found to contrast with lenis stops primarily through their higher F0 values. In clear speech, younger speakers also tended to increase the F0 difference between the stop types, whereas older speakers (born before 1965) chose to modulate the VOT differences instead (Kang & Guion, 2008). This shift in cues between older and younger speakers was additionally investigated through perception studies which demonstrated that listeners of all age groups tended to prioritize F0 over VOT in the lenis vs. aspirated contrast (Kang, 2010; Kim et al., 2002; Kong & Lee, 2018). It was also found that the acoustic information encoded in the vowel itself was a sufficient cue to a correct identification of the preceding stop type (Cho, 1996), especially for the identification of the lenis stop (Kim et al., 2002). Furthermore, when listeners were presented with conflicting information for the stop type, they were more likely to make a decision based on the information encoded in the vowel rather than the information encoded in the consonant portion of the syllable (Kim et al., 2002).

This still ongoing sound change in production and perception of Korean word-initial stops has been interpreted in various ways, with the consensus currently pointing towards transphonologization of F0 over VOT as the primary phonetic cue for a phonological contrast between aspirated and lenis stop types; the so called (quasi-)tonogenesis of Seoul Korean (Bang et al., 2018; Jun, 2020; Kang, 2014; Ren & Mok, 2018). However, it is still a matter of debate whether the ongoing sound change is true transphonologization, or whether the use of F0 as the main cue is an effect of prosodic position and prominence (Jun, 2020). Choi et al. (2020) investigated Korean stops both in phrase-initial and phrase-medial positions as well as in focused vs. unfocused prominence conditions and found that speakers differ in their use of VOT and F0 as cues based on the prominence and prosodic position. The authors proposed that the sound change is rather a prosodically conditioned change in the manner of VOT usage as a cue, wherein listeners merely take use of the post-lexical tones already existing in the intonational phonology of Seoul Korean.

Current research into the word-initial stop type distinction acknowledges the importance of both VOT and F0, and has been focusing more closely on the interplay of the two (e. g. Bang et al., 2018; Choi et al., 2020; Kong & Lee, 2018; Lee & Katz, 2016). For instance, Bang et al. (2018) found that in younger speakers, the VOT contrast reduction in combination with the F0 contrast enhancement are stronger in more frequently used words, which is a sign of articulatory reduction as well as predictability. Studies have also demonstrated that the change in cue importance is led by female speakers (Kang, 2014; Yu, 2018), which is in accordance with sound change in general (Labov, 1990). In addition to VOT and F0, some studies have also decided to examine the role of voice quality, although mostly only in production (Cho et al., 2002; Kang et al., 2022; Yu, 2018; see Kim et al., 2002 for a perception study). These studies have generally shown that vowels following tense stops often display a creaky voice quality whereas lax and aspirated stops often produce breathy voice. For instance, it was shown that not only F0 differs between stop types, but that voice quality differences in tandem with F0 differences occur from the onset of the vowel and extend into at least its midpoint (Yu, 2018). Younger speakers of SSK were also found to use breathy voice more often and more strongly than older speakers after both aspirated and lenis stops. When it came to the fortis stops, these were most saliently distinct from the other two types through their often creaky voice quality (ibid).

1.2 Voice quality and Creak

Let us now focus more closely on voice quality and explain the associated physiological and phonetic features. Voice quality refers to the manner in which the vocal folds vibrate. It is an important characteristic as it can be used contrastively in some languages in order to signal differences in meaning (Garellek, 2019; Keating et al., 2010). The three voice qualities found in typical Seoul Korean speech are modal, breathy, and creaky, although none of them are phonologically contrastive. Whereas breathy phonation, during which the vocal folds are more open compared to modal voice, can be found in aspirated and lenis word-initial stops, creaky phonation is most usually found in fortis word-initial stops (Abberton, 1972; Cho et al., 2002; Kang et al., 2022). Breathy phonation is often described as 'slack', 'murmured', or 'aspirated' (Garellek, 2019). In contrast, creaky voice ('creaky phonation' or 'creak') is an umbrella term for different types of often irregular voicing, with its characteristics most typically being low F0, irregular F0, and constricted glottis (Garellek, 2019; Huang, 2023b; Keating et al., 2023). When creaky voice is perceived, it is often described as 'rough', 'tense', 'pressed' or 'harsh' (Garellek, 2019). The high glottal constriction in creaky voice causes low airflow due to the vocal folds being generally close together and due to a long closed phase of each vibration cycle (Keating et al., 2023). The acoustic parameter most relevant to the degree of vocal fold constriction is spectral tilt, which is the difference in amplitudes between different harmonics (e. g. H1-H2 or H2-H4) (Garellek, 2019; Keating et al., 2015). High glottal constriction typically produces low values of H1-H2, which is often considered a correlate of creaky voice (Huang, 2023b; Keating et al., 2015, 2023). In comparison, breathy voice would show high H1-H2 values. The aforementioned phonetic characteristics might appear either on their own or in tandem (e. g. low F0 + irregular F0), but not all singular characteristics or their combinations lead to a perception of creaky voice (Keating et al., 2023). For instance, Slifka (2006) mentions that creaky voice was perceived even if the glottis of the speakers was generally unconstricted but other features were present. Analogously, just the presence of a constricted glottis (specified as 'tense voice') did not instigate the perception of creak (Keating et al., 2023). The presence of low F0 is a key correlate of creaky voice in Hmong and Mixtec (Keating et al., 2023), as well as in some tones of Mandarin (Huang, 2020), but it is unclear if low F0 on its own can lead to a perception of creak (Keating et al., 2023). On the other hand, irregular F0 on its own is capable of producing a creaky percept (Keating et al., 2023), likely because in this instance the vocal fold vibration varies from one pulse to another ('pulse-to-pulse jitter') and the fundamental frequency is consequently difficult or impossible to be determined (Keating et al., 2015, 2023). As evidenced by the above-mentioned findings, creaky voice is a highly complex voice quality that is known to differ from one language to another in terms of its phonetic qualities but also perceptually (Keating et al., 2010; for a comprehensive overview of different types of creak see Keating et al., 2015).

1.3 The present study

As previously mentioned, creak is known to have been mentioned already in very early literature on the Korean stop type distinction (Abberton, 1972; Kagaya, 1974). The first way in which creak found after the fortis stops of the SSK dialect differs from 'prototypical creak' (Keating et al., 2015), such as the creak found in Mandarin (Huang, 2020, 2023a), is that low F0 does not necessarily cue creaky voice, because lenis stops consistently instigate lower F0s than fortis stops (Bang et al., 2018; Kim, 1965; Silva, 2006). What is more, vowels following the fortis stops of the SSK dialect have been found to show aspects of multiply pulsed voice (Watkins et al., 2024), which involves a particular irregularity in F0, wherein the vocal folds alternate in creating longer and shorter pulses (Huang, 2020; Keating et al., 2015). In turn, this establishes two simultaneous periodicities which contribute to an indeterminate pitch in production and a rough voice quality in perception (Huang, 2022). Acoustically, this typically manifests as two F0s, one low and one about one octave higher. In some vowels following the fortis stops of SSK, the two periodicities were found to co-occur from the onset of the vowel at least up to the first third of the vowel (Watkins et al., 2024). Afterwards, the creaky phonation typically fades, giving way to modal phonation. It remains unclear whether the presence of two concurrent periodicities affects the distinction of word-initial stop types in SSK or whether it has any perceptual significance (ibid).

It is known that the presence of the aforestated multiply pulsed voice might cause problems for some automatic pitch trackers (e. g. those relying on raw autocorrelation pitch detection algorithms, such as the pitch tracker implemented into Praat; Boersma & Weenink, 2024), which rely on establishing multiple F0 candidates at each point in time and then determining an optimal global F0 curve running through those candidates (Boersma, 1993). If the automatic tracker encounters pitch tracking problems caused by creaky voice qualities (e. g. irregular F0, multiply pulsed voice etc.), the pitch curve might show pitch jumps (i. e. instances where the pitch curve suddenly jumps up or down by roughly an octave). Consequently, tokens containing pitch jumps might be discarded or ignored from F0 analyses, as seen in previous literature (Holliday & Kong, 2011; Kang et al., 2022; Lee & Jung, 2000). For example, Kang et al. (2022) mentioned that about 7% of the total tokens produced by their participants had to be omitted from acoustic analysis due to the presence of very creaky or breathy voice. Out of those tokens, 82% were vowels following fortis stops. However, a recent study by Watkins et al. (2024) proposed that an adjustment of settings in Praat might help mitigate the number of tokens where pitch tracking was found to be problematic.

Seeing as the issue of creak in SSK has not been properly studied, the present study aims to fill in this gap. Previous studies on Korean word-initial stops have predominantly been focused on production, with only a handful of studies investigating perception (Cho & Keating, 2001; Kim et al., 2002; Kong & Lee, 2018; Lee et al., 2013; Schertz et al., 2015, 2019). Crucially, most of the aforementioned perception studies have mainly explored the cue-weighting of VOT and F0 (Cho & Keating, 2001 and Schertz et al., 2015 also investigated closure duration), but only few have examined effects of voice quality (Schertz et al., 2019; Kang, 2010 only investigated voice quality differences between aspirated and lenis stops). Importantly, no previous perception studies have mentioned the presence of creaky voice in their experimental stimuli, although it is possible that the fortis vowel used in the study of Schertz et al. (2019) did have a creaky quality since it was described as having low H1-H2. It is known from research on Mandarin that creak does in fact play a role in tone identification (Huang, 2020, 2022; Li et al., 2023; Yu & Lam, 2014), suggesting that the presence of creak changes the perceived F0 contour. In the case of SSK fortis stops, where creaky voice in following vowels is known to be present, it is plausible that the presence of creak might help with the identification of this stop type. There is mostly anecdotal evidence that fortis and lenis stops sometimes get confused with each other. For instance, stop type labeling in Yun et al.'s (2015) corpus of Korean spontaneous speech (which was done by native speakers of Korean) was sometimes found to be done incorrectly as the fortis and lenis stops were mixed up. In contrast, the fortis and aspirated stops were almost never confused with each other in the labeling of this corpus. This is also supported by the findings of Schertz et al. (2019), wherein at low VOT values perceptual stop identification was substantially ambiguous between fortis and lenis stop types, but fortis and aspirated stops were clearly differentiated by VOT. In such a situation, would creaky voice be a meaningful cue in the fortis-lenis distinction, leading to a more consistent percept of the fortis stop type? In order to respond to this inquiry, the present study aims to answer the following research question:

(1) Does the presence of creak influence fortis-lenis stop type discrimination for listeners of Standard Seoul Korean (SSK)?

We hypothesize that since creaky voice is only found amongst SSK fortis stops, it will serve as a strong enough cue to incite a consistent fortis percept. In this way, creak as a potential cue is hypothesized to override other cues, especially in contexts where VOT and F0 are ambiguous.

2. Method

Because the purpose of the present study was to investigate stop type perception (i.e. whether a stop is perceived as fortis or lenis when creak is present), an ABX listening experiment was chosen. In an ABX task, the listener hears three sounds in succession (A, B, X) and then has to make a choice whether sound X more closely resembled sound A or sound B. The reason for choosing an ABX task is that such a task does not require any meta-knowledge of Korean phonology and it allows for omission of any orthographical labelling. Different manipulations were done for sounds A and B, and for sound X. The following subchapters describe the relevant stimuli manipulations and experimental design used in this study.

2.1 Stimuli

In order to avoid the influence of semantic processing, nonce word stimuli were used as our experimental tokens and fillers. The relevant nonce words were created according to Korean phonology and each of them was two syllables long. Each word was produced with each stop type according to the laryngeal contrast (i.e. aspirated, fortis, lenis). Words for all three places of articulation (i. e. labial, coronal, velar) were used, however, it was decided to only use words starting with a coronal stop (/tʰ/, /t*/, /t/) as the experimental stimuli for the sake of eliminating the effect of place of articulation (Abberton, 1972). Words starting with labial or velar stops were instead used to create filler items. The first syllable of each word always ended with the vowel /a/, while the second syllable was always either /sja/ (phonetically realized as [ɕja] or [ɕa]) or /dʑut/, except for the labial stops, where the second syllable /nu/ was used in order to avoid the word 파샤 /pasja/, which did have a dictionary term (a Turkish general's title). See Table 2 for the nonce words used as experimental tokens.

Our nonce words were recorded in a sound-attenuated booth of the Speech Lab at the University of Amsterdam. They were produced by a female native speaker of the Standard Seoul Korean dialect (24 years old), who had lived in Seoul from birth until 11 years of age but had since then lived elsewhere. Nevertheless, she reported that she used Korean on a daily basis. Each of the nonce words was recorded in three sentence contexts. The first sentence context involved producing the word in isolation. The second context consisted of the word produced at the beginning of a carrier sentence, and the third sentence context involved producing the word sentence-medially (see Table 3 for an example with translations). The speaker read each word from a sheet of paper lying on the desk in front of them. Each nonce word was repeated at least five times in every sentence context. See Table 7 in the Appendix for the list of all recorded stimuli.

Table 2: Experimental tokens used in the present study.

ㅌ **/tʰ/**	ㄸ **/t*/**	ㄷ **/t/**
타샤 /tʰasja/	따샤 /t*asja/	다샤 /tasja/
타즛 /tʰadʑut/	따즛 /t*adʑut/	다즛 /tadʑut/

Table 3: Sentence contexts used for each nonce word. The example nonce word is in bold.

	Original Korean text + Romanization + Transcription	English Translation
Isolation	**다샤** (Dasya) /tasja/	**[NONCE WORD]**
Initial	**다샤** 는 우리가 갈 곳이다. (Dasya neun uriga gal gosida.) /tasja nɯn uɾiga kaɭkoɕida/ **[NONCE WORD]** topic.particle us-subj.particle go-future place is	**[NONCE WORD]** is where we are going.
Medial	나는 **다샤** 에 갔다. (Naneun dasya e gatda.) /nanɯn tasja e katta/ I (informal) topic.particle **[NONCE WORD]** locative.particle go-past.	I went to **[NONCE WORD]**.

2.2 Stimuli manipulation

Out of the three sentence contexts, the sentence-initial tokens were deemed to be ideal for the creation of experimental stimuli because they reflected real speech behavior more than the isolated words, and because they were comparably much longer and more clearly pronounced than sentence-medial tokens. Thus, all the sentence-initial lenis and fortis tokens of the nonce words 다샤, 따샤 (/tasja/, /t*asja/) and 다줏, 따줏 (/tadʑut/, /t*adʑut/) were extracted from the recordings. A great portion of fortis tokens was naturally creaky, therefore we decided to create a balanced set of creaky and non-creaky tokens in our experiment. In order to reach this balance, two additional creaky fortis tokens were extracted from a recording in which the speaker produced the nonce words in isolation. Additionally, all lenis tokens selected as experimental tokens were deemed to be pronounced with a breathy quality by the authors based on informal listening. Every experimental token differed in the intonation of the second syllable, therefore, a clearly pronounced second syllable from one of the fortis tokens of each word (/sja/, /dʑut/) was chosen and pasted over the second syllable of all relevant experimental tokens. In other words, the second syllable of all experimental tokens was the same /sja/ or /dʑut/ syllable depending on the nonce word. Additionally, the intensity of all tokens was set to 70dB. These were the only manipulations done to stimuli used as sounds A and B in our experimental scheme.

Table 4: Mean values and standard deviations of VOT and F0 of the first syllable in the unmanipulated fortis (‘F’) and lenis (‘L’) tokens.

	/t*asja/ (F)	/tasja/ (L)	/t*adʑut/ (F)	/tadʑut/ (L)
VOT				
Mean (ms)	12.15	69.01	9.34	58.98
SD (ms)	5.23	5.88	1.04	7.07
F0				
Mean (Hz)	224	226	230	243
SD (Hz)	7	7	6	5

For the stimuli used as sound X of our scheme, lenis tokens were taken and multiple manipulations upon them were made. In addition to the above mentioned second syllable standardization, VOT was manipulated in order to mitigate its influence as a cue. VOT was set to 40ms for every token via a Praat script (Winn, 2020). This value was chosen based on previous literature which reports VOT values between ∼30-70ms for lenis stops and ∼10-50ms for fortis stops, depending on place of articulation and sentence context (Kim et al., 2002; Kim & Lotto, 2002; Lisker & Abramson, 1964). VOT of this value was deemed to be an ambiguous cue for the distinction between lenis and fortis stops. The F0 of all tokens was left at its original value. Mean values for VOT and F0 were measured in Praat, with the pitch settings set to an octave-jump cost value of 0 and an octave cost of 0.1. Interestingly, our speaker produced slightly higher F0 values for the lenis tokens than for the fortis tokens. The aforementioned VOT and F0 values of our experimental tokens can be seen in Table 4.

Another parameter possibly relevant to our stimuli was the burst of the stop. It is known from research on other languages that voiceless stops differ from voiced stops in multiple aspects, such as center of gravity (CoG), aspiration amplitude, or F0 contour (Chodroff & Wilson, 2014; Haggard et al., 1970; Liberman et al., 1958; Repp, 1979). To our knowledge, there is only one study that investigated the differences between the bursts of the word-initial voiceless stop types of Korean, which reported that the intensity, skewness and CoG does not differ across stop types significantly, although high between speaker-variation was observed (Park, 2004). Nevertheless, we might expect the burst of our X stimuli, which was taken from lenis tokens to serve as a potential secondary cue pointing towards the lenis stop type in our fortis-lenis distinction.

While searching for an optimal way of artificially creating perceptible creaky voice, two parameters were found to be important. Huang (2023a) showed that artificially created tokens can be perceived as creaky if at least one of two manipulations to the soundwave of the vowel are made. The first manipulation is frequency modulation, whereby the duration of every even pulse is shorter compared to the duration of every odd pulse. The second manipulation is amplitude modulation, which involves reducing the amplitude of every even pulse as compared to every odd pulse. After informal listening to example tokens which were amplitude-modulated only, it was concluded that amplitude manipulation on its own would be sufficient to create perceptibly creaky tokens. The creaky percept was also confirmed by the native SSK speaker who recorded the original stimuli.

For the stimuli used in sound X, amplitude manipulation was done via a combination of manual manipulation and a Praat script. The initial step involved making a copy of the sound file, and then manually setting the sound wave of every second periodical cycle of the vowel from its onset up to the fourth even cycle (typically around 50-60% of the vowel's duration) to zero. Interestingly, this manipulation already instigated the perception of creaky voice during informal listening by the authors. Next, the original sound file and its copy were combined to create a new sound file using a script. During this process, the amplitude of the original sound file was reduced to 60% of its original level, while the amplitude of the copy was set to 40% of its original level. Since these percentages add up to 100%, the resulting sound file matched the original sound file in all areas except where the amplitude was set to zero in the copy. In these zero-amplitude sections, only the original sound file at 60% of its original amplitude was included. See Figure 1 for illustration of the respective steps taken.

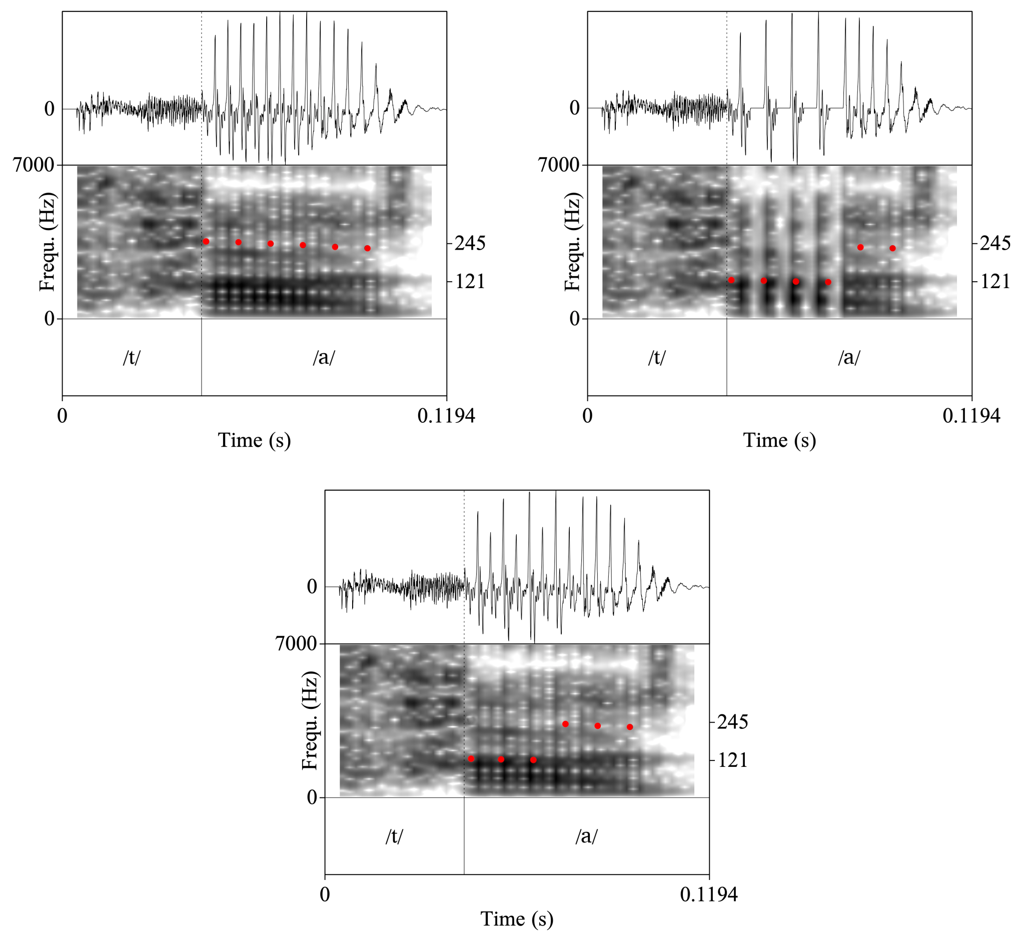

Figure 1: Lenis stop /t/ followed by the vowel /a/. Original audio file (top left). A copy of the original sound file with every other cycle of the vowel portion set to zero (top right). The original audio file and its copy overlaid on one another with a ratio of 60% amplitude (original sound file) and 40% amplitude (copy).

The aforementioned process was applied to our lenis tokens, creating eight lenis + creak tokens for each of our experimental nonce words, /tasja/ and /tadʑut/, resulting in a total of 16 tokens. The list of tokens used in the experimental condition can be found in Table 8 in the Appendix.

2.3 Experimental design

As previously mentioned, the scheme chosen for our experiment was a forced-choice ABX task, in which a set of three sounds (A, B, and X) is played in succession. In each set, the sound A was a natural fortis token, the sound B was a natural lenis token (i.e. A and B were tokens with a natural first syllable, second syllable was normalized for all tokens as described in 2.2), and the sound X was an artifically created lenis + creak token. It should be noted that creak was also naturally present in 50% of the fortis tokens used as sound A. Each set of recordings was played twice in its natural order (ABX), and twice in a reversed order (BAX). This yielded 64 (2 x 2 x 16) total experimental tokens. An illustration of a possible experimental set can be found in Figure 2.

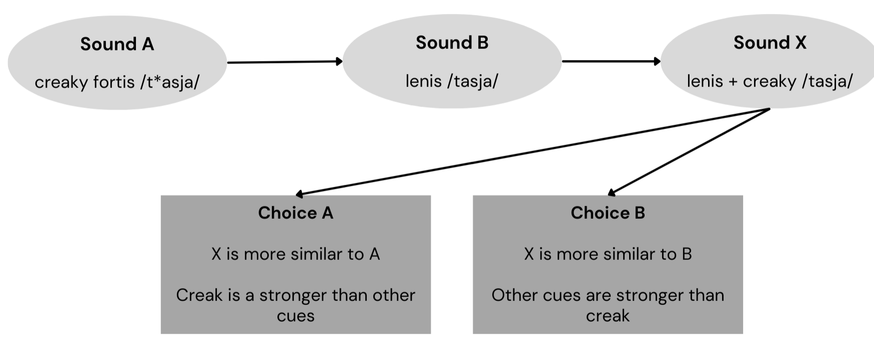

Figure 2: An illustration of a possible experimental set of the ABX scheme used in the present study with possible answers.

In addition to our experimental tokens, 192 (3 x 64) filler tokens were also created. These fillers were taken from various tokens in our recordings, no matter the stop type or word (refer to Table 7 in the appendix). In the case of fillers, sounds A and B were either different words or same words of a different stop type (e. g. aspirated vs. fortis), while sound X was always of the same word or stop type as one of the previous sounds. Moreover, 28 of our fillers were catch trials constructed to check whether participants were paying attention throughout the experiment. These involved ABX sets containing noticeably different words taken from the carrier sentences (e. g. /kadʑut/ – /nanɯn/ – /kadʑut/). If a participant scored correctly on the catch trials less than 75% of the time, their results were rejected and not used in the analysis.

2.4 Participants and experiment

Our participants were 33 (23 female, 10 male) native speakers of Korean (mean age = 30.48 years, SD = 8 years) recruited either through the *Prolific* website (*prolific.com*) or through personal relations. Most of them lived in the Netherlands, although a great deal of them also lived in South Korea. One participant was excluded because they only scored correctly on 43% of our catch trials, suggesting that they did not pay enough attention during the course of the experiment. All other participants scored more than 90% correctly on the catch trials (mean accuracy = 97%). Furthermore, one participant completed only 46 out of the 64 experimental sets, although their results were still used in the final analysis. All participants were financially compensated for their participation.

Table 5: Distribution of participant's self-reported region of origin.

Region	Number of participants
Chungcheong	4
Gangwon	1
Gyeonggi	7
Gyeongsang	8
Jeolla	1
Seoul	11

The experiment was carried out in an online environment, created by the Speech Lab at the University of Amsterdam, to which participants were redirected from *Prolific* and began with demographic questions about the participants' nationality, age, gender, and region of origin (regional distribution can be seen in Table 5). As Seoul Korean was the focus of this study, there was also a question about whether the participants had spent more than two years of their life in Seoul. Out of the total number, 26 participants responded 'Yes' to this question. Lastly, there was a question whether the participants learned Mandarin, and if yes, then for how long. This question investigated the possible sensitivity of the participant to the presence of creak as covariate of F0, as creaky voice has been found to affect tone perception in Mandarin (Huang, 2020). Six of our participants admitted to having learned Mandarin (P_01, P_03, P_21, P_22, P_24, P_27), all of them for a duration of about 1 year.

After the participants completed the demographic questions, they were given instructions on the design of the experiment. If the last word played (sound X) resembled the first word (sound A) more, the participant was instructed to press the 'Z' key on the keyboard. If the last word played (sound X) resembled the second word (word B) more, they were instructed to press the 'M' key. They were also told to wear headphones in order to reduce distraction from background noise.

This was followed by a training phase, in which five ABX sets of filler tokens were played. All participants were played the same training phase stimuli in the same order. These training sets contained stimuli similar to the catch trials. They were deemed by the authors to be very easy to correctly match, and they served as a way to learn to understand the experimental design. In order to pass the training phase, each participant had to score correctly at least three times. If a participant failed the training phase, they could not attempt the experiment again.

Then, the actual experiment began. A total of 256 ABX sets (64 experimental sets + 192 filler sets) were played. All sounds were played in stereo. Before each set, a fixation cross appeared on the screen, which was followed by an interval of 900ms. Then the first word of the set was played. Between each word, there was an interval of 900ms. Only after all three sounds of the set were played, the instructions to either press 'Z' or 'M' appeared on the screen. Each participant could take as much time as needed to make a choice as the next set only started playing after the answer key was pressed. However, they could only hear each set once. The order in which the sets were played was random for each participant. There was a self-timed pause half-way, in which participants were instructed to take a short break. The experiment lasted for about 30 minutes (mean = 30.4 minutes, SD = 6.9 minutes), although the completion time largely depended on how long the participant's self-timed break was.

Once all participants completed the experiment, the data was compiled into a spreadsheet using Microsoft Excel. The resulting spreadsheet was then loaded into RStudio (RStudio Team, 2022). Plots were made using the ggplot2 package (Wickham, 2016). For the purposes of statistical analysis, the lme4 package (Bates et al., 2015) was used.

3. Results

3.1 Key press bias

First, we checked whether participants were biased towards pressing one key more frequently than the other (e. g. more presses of 'Z' than 'M'). The data from the fillers was used and the results of a simple analysis showed that our participants had a slight overall preference for pressing the 'M' key (mean ratio of 'M' over 'Z' = 1.11, SD = 0.12).

3.2 Model building

As the first step, Pearson's Chi-squared test was used to show whether our potential predictors were subject to correlation. For this reason, *Age* was divided into age groups (21-30, 31-40, 41-50, 51-60). A significant correlation was found between *AgeGroup* and *Region* ($p < 0.001$), in addition to a correlation between *SeoulYears* (more than 2 years spent in Seoul) and *Region* (also $p < 0.001$). Due to this reason, *Age* and *SeoulYears* were not used in building our model. Another reason for the omission of *Age* as a predictor was that all our participants were born after 1965. According to previous literature, speakers of SSK born after 1965 should have theoretically already made the change in primary cue from VOT to F0 (Bang et al., 2018; Silva, 2006). Additionally, we were more interested in the by-region variation rather than the by-age variation. Furthermore, as some regions contained very few participants, it was decided to split regions into three groups – Seoul, Other, and Gyeonggi. This three-way split was done in order to see whether the Gyeonggi group behaved closer to the Seoul group or to the Other group. In this aspect, the Gyeonggi group was expected to behave more closely to the Seoul group due to Gyeonggi's geographic proximity to Seoul.

To achieve our results, the data corresponding to experimental tokens was submitted to analysis using a generalized linear mixed-effect model, with the participants' responses ('fortis' or 'lenis') as the dependent variable (*fxlChoice*; 'fortis' coded positively), and with *Creak* (sound A contained creak, sound A did not contain creak), *Order* (ABX = Normal; BAX = Reversed), *SeoulRegion* (Seoul, Gyeonggi, Other), *Mandarin* (knowledge of Mandarin; Yes or No), *Gender* (Male or Female), and *Word* (/tasja/ or /tadʑut/). Interactions effects between *Creak* and all other predictors were included. The random-effects structure was built with random intercepts for *Participant*, but no random slopes were included due to convergence problems. The resulting model structure was:

fxlChoice ~ *Creak * (Order + SeoulRegion + Mandarin + Gender + Word) + (1 | Participant)*

3.3 Model results

Table 6: Results of the mixed-effects model with 'fortis' coded positively as the dependent variable. Symmetric contrasts were set for all predictors.

	Estimate	Std. Error	z-value	Pr(>|z|)
(Intercept)	-2.578153	0.202325	-12.743	< 2e-16 ***
Creak-No+Yes	0.101475	0.229451	0.442	0.658
Order-Normal+Reversed	1.982180	0.193207	10.259	< 2e-16 ***
SeoulRegion-Seoul+Other	0.101134	0.338987	0.298	0.765
SeoulRegion-SeoulOther+Gyeonggi	-0.140576	0.384323	-0.366	0.715
Mandarin-No+Yes	-0.034124	0.409205	0.083	0.934
Gender-F+M	0.306559	0.321768	0.953	0.341
Word-Tajut+Tasja	-0.058101	0.149741	-0.388	0.698
Creak-N+Y:Order-Normal+Reversed	-0.123138	0.382863	-0.322	0.748
Creak-No+Yes:SeoulRegion-Seoul+Other	-0.002102	0.350096	-0.006	0.995
Creak-No+Yes:SeoulRegion-SeoulOther+Gyeonggi	0.600510	0.421952	1.423	0.155
Creak-No+Yes:Mandarin-No+Yes	-0.046536	0.440027	-0.106	0.916
Creak-No+Yes:Gender-F+M	0.199970	0.324581	0.616	0.538
Creak-N+Y:Word-Tajut+Tasja	0.379097	0.299545	1.266	0.206

First of all, the results of the statistical model can be seen in Table 6. The intercept (when transformed from log-odds to odds) showed that the odds that an average participant would say that sound X resembled the fortis token more closely than the lenis token ('fortis choice') were 13.17 times smaller (95% c.i. = 8.79 … 19.74; $p < 0.001$) than the odds for a 'lenis choice'. In other words, the odds for a lenis choice were 13.17 higher than the odds for a fortis choice. The choices made by each participant are demonstrated in Figure 3. We can notice that every participant made a proportionately higher amount of lenis choices, with fortis choices being quite rare for many listeners, although some participants (P_04, P_10, P_15, P_21, P_23, and P_28) produced fortis choices more than 15% of the time.

Male participants were overall slightly more likely to make a fortis choice compared to female participants (estimated odds ratio = 1.36; 95% c.i. = 0.71 … 2.59), however, the effect of *Gender* was not found to be statistically significant (p = 0.34). The effect of having learned Mandarin was very minor and also non-significant (p = 0.93). Furthermore, the effect of *Word* (/tasja/ vs. /tadʑut/) was small and statistically non-significant (p = 0.7).

As for Regional variation, no statistically significant difference was found between Seoul and the rest of the regions (p = 0.77). A secondary contrast was included, which compared the Gyeonggi region to the average of Seoul and Other regions. This difference was also non-significant (p = 0.72).

The effect of *Creak* (the effect of presence of creak in sound A of the ABX or the reversed BAX set) was not found to be significant as well (p = 0.66). The interaction effects in our model concerned *Creak* and other within-participant and between-participant predictors. In this aspect, no interaction effects were found to be significant (all $p > 0.05$; refer to Table 6 for individual p-values).

Surprisingly, the only statistically significant effect was *Order*. The results of the model show that the reversed order of stimuli (BAX order) still caused the participants to predominantly make lenis choices, although the odds of making a fortis choice were significantly higher than in the ABX order (estimated odds ratio = 7.26; 95% c.i. = 4.93 … 10.68; $p < 0.001$). This shows that participants were significantly affected by the order of the presentation stimuli. Known problems with ABX tasks will be addressed in the Discussion (section 4.2).

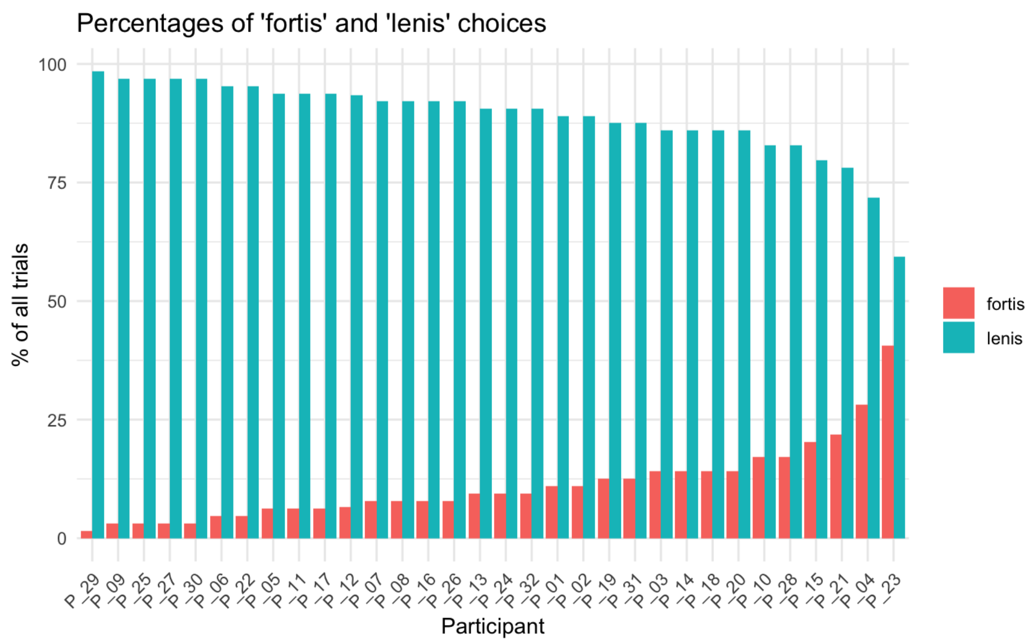

Figure 3: Percentages of 'fortis' vs. 'lenis' choices for all participants in all trials. Ordered from lowest to highest percentage of 'fortis' choices.

The individual lenis + creak stimuli were investigated to check whether some stimuli instigated different response rates of the 'fortis' choice. The choices for individual stimuli can be seen in Figure 4. It is evident from the figure that no individual stimulus substantially differed from other stimuli.

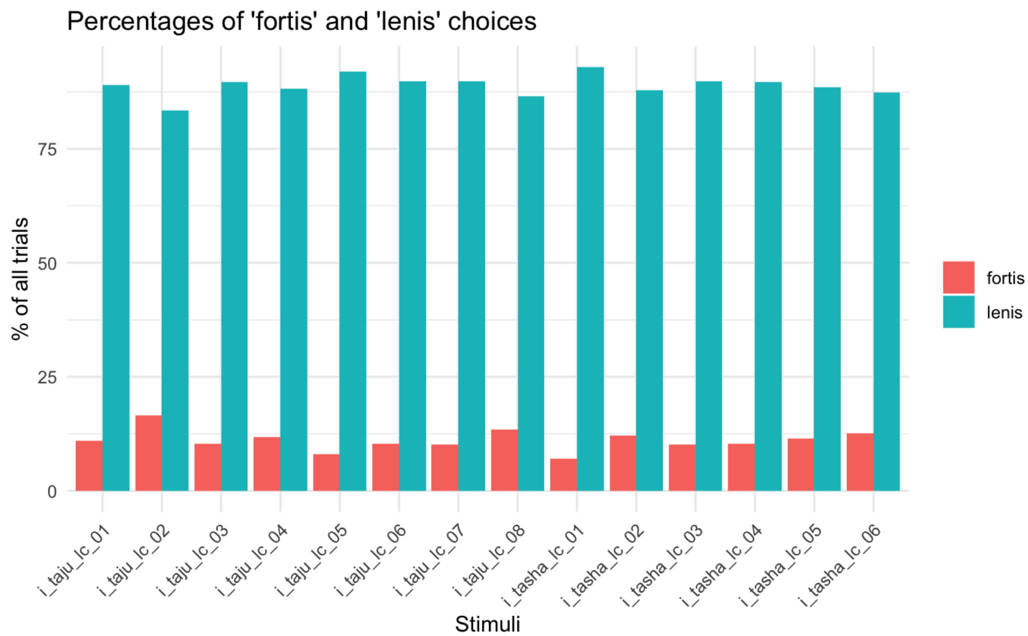

Figure 4: Percentages of ‘fortis’ vs. ‘lenis’ choices for each of the experimental stimuli (all participants combined).

4. Discussion

Our results show that the presence of creak did not cause a dominant fortis perception of the word-initial stop. The rates of fortis choices were overall very low, at least for the majority of our speakers. However, the central question of this study was whether the presence of creak influences the fortis-lenis distinction, leading to a consistent perception of fortis stops. In this aspect, our results show that the presence of creak definitely influences listeners’ choices between the two categories, as many of our participants have made a ‘fortis’ choice more than 10% of the time, and some even well over 20% of the time. It is very likely that their choices were not a result of chance but of a real effect of perceiving creaky voice and associating this voice quality with the fortis stop type.

One important finding is that the presence of natural creak in sound A of the relevant experimental sets did not significantly affect the odds of a fortis choice. As the effect was not found to be significant, nothing can be concluded, but it opens room for several interpretations. If we assume that the lack of orthographic labels in our ABX task led to choices informed by purely auditory information, then one may hypothesize that the artificially created creaky voice in sound X was perceived as closer to the modal voice found in the non-creaky natural tokens than to the breathy voice found in the natural lenis tokens. Here it should be noted that no measurements of voice quality were made (e. g. H1-H2) but the ‘creaky’ fortis tokens in sound A were deemed to be creaky by informal listening of the authors, in the same way the natural lenis tokens in sound B were all deemed to be breathy. Continuing further with this hypothesis, if natural creak in sound A would have led to more fortis responses, it could have been argued that participants are simply perceiving creak in sound A and creak in sound X, and thus comparing only the auditory information associated with creaky voice. However, this is likely not the case, as evidenced by the lack of a significant difference between participants' choices when creak was and was not present in sound A. Another interpretation of the non-significant difference between the presence or absence of creak in sound A considers the auditory matching of modal voice as opposed to creaky voice. If we consider that the naturally creaky tokens, as well as the artificially creaky tokens, contained modal voice in the final ∼40% of the vowel, participants might have made their fortis choices based on the presence of modal voice in sounds A and X as opposed to the presence of breathy voice in sound B.

Lastly, the interpretation which we find most compelling posits that participants appear to be mentally categorizing the respective ABX sounds as fortis/lenis first, and only then they compare the mental representations across phonological categories. This interpretation is easily capable of explaining why there was no apparent difference in participants' performance whether or not creak was present in sound A. In this interpretation, participants would have categorized sound A as fortis no matter the presence of creak, because the cue of VOT would be used as a primary cue. Similarly, sound B would be categorized as lenis based on its long VOT and breathiness. Finally, participants would categorize sound X as fortis or lenis based on the combination of ambiguous cues (VOT, F0), the cue of the lenis burst, and the cue of creak. After this mental categorization was finished, the categorical representations would be compared and a choice would be made.

4.1 Creak as a standalone cue

The results of our experiment spark the question whether creak plays a role as an additional cue in the fortis-lenis distinction. First of all, it is known that at least two cues, VOT and F0, help listeners in stop type determination. In our experiment, the VOT of our experimental tokens was set at a value of 40ms which was deemed to be ambiguous as it lay between the prototypical VOT values of both fortis and lenis stops. As a result, VOT of this value would not be expected to be a useful cue. Analogously, F0 was also an ambiguous cue in the sense that the /a/ vowels of the artificially creaky tokens in sound X were the same vowels as in the natural lenis tokens (sound B) but with creaky voice present from the onset to about 50-60% of their duration. Furthermore, the natural fortis tokens (sound A) actually had slightly lower F0 values than the natural lenis tokens.

It should be noted that the presence of creak may have affected the perception of F0 by our listeners, as evidenced by previous literature on the covariation between voice quality and perceived pitch (e. g. Huang, 2020, 2023a; Kuang, 2017; Kuang & Liberman, 2016). It has been observed that some subtypes of creaky voice cause a lower perceived pitch (Kuang, 2013 on Mandarin and Black Miao; Zhang & Kirby, 2020 on Cantonese), and even a lower produced F0 in a pitch-imitation task (Huang, 2023a on Mandarin). Even though a great deal of the aforementioned studies investigated speakers of Mandarin, this effect was found to be crosslinguistic (ibid). Because creaky voice has commonly been associated with low F0, and with low tones in tonal languages, it is possible that listeners naturally utilize the presence of creak to cue low pitch, and that maybe this utilization is part of a natural psychoacoustic mechanism (Kuang & Liberman, 2016), at least for speakers of languages where creaky voice occurs.

Specifically, the amplitude manipulation which was used to create a creaky lenis token is quite similar to the manipulation done by Huang (2023a), whereby the amplitude of every other glottal cycle was decreased to 60% (ratio of ∼1/1.67) of its original amplitude. Huang's study employed a stepwise decrease in amplitude of even cycles, and at a comparable level (between the ratios of 1/1.6 and 1/1.8) there was already a noticeable increase of responses corresponding to a low tone, as opposed to a nearly consistent high tone response at the previous step (ratio of 1/1.4). It should be noted that Huang's experiment concerned both Mandarin and English speakers, with both groups performing comparably. Therefore, it is probable that our amplitude modulation produced the perception of lower F0, leading to more frequent lenis choices, as lenis stops are known to be followed by vowels with a lower F0 compared to the other two stop types (e. g. Kang, 2010; Kim et al., 2002). Due to these reasons, we can hypothesize that amplitude modulation might have had an effect on the perceived F0. However, testing the perceived F0 was not the focus of our experiment, as this effort would require a tone identification task, perhaps with an artificial tonal language.

If the pitch of the vowel in our experimental tokens really was perceived as low due to the presence of creak, then this would aid the perception of a lenis stop. However, as evidenced by the results, fortis choices were still made consistently by some speakers, leading us to believe that creak might be a standalone cue in the fortis-lenis distinction. What is more, this cue would be strong enough to override the F0 cue, at least some of the time (the potential strength of F0 as a cue in our data is discussed more closely in section 4.3.). As a result, creakiness could theoretically stand as another dimension of the stop type distinction. This is a very important finding which essentially answers our research question, as it is clear that some listeners do rely on creak in the fortis-lenis stop contrast. It is also important to note that fortis choices were consistently made, even though a lenis burst was present in every X stimulus. This further demonstrates that creak can override this cue in a situation where other cues are potentially ambiguous.

4.2 ABX scheme

What cannot be ignored is the statistically significant effect of order of stimuli presentation. In the reversed order (BAX), the odds for a participant to respond with a fortis choice were significantly higher than in the normal (ABX) order. Some kind of response bias was to be expected, as previous literature (e. g. Gerrits & Schouten, 2004; Schouten et al., 2003) suggests that (categorical) perception in experiments with sound stimuli depends on the discrimination task. In the case of the ABX task, there seems to be a bias towards the 'X is closer to B' response, as also displayed by the results of our experiment. This is likely due to the time distance between the 'A' and the 'X' stimuli. In the present experiment, the interval between individual stimuli was 900ms, therefore the interval between stimulus 'A' and stimulus 'X' was 1800ms plus the duration of stimulus 'B', which increases the chances that by the time that 'X' is presented, the auditory traces might have faded away in the listener's memory (Pisoni, 1973). This notion would further support our interpretation of the results, wherein participants do not compare purely auditory information but rather the mental representations of the 'fortis' and 'lenis' categories which the sounds in the ABX scheme activate.

It should be noted that the preference for saying that 'X is closer to the sound directly preceding it' might also relate to the results of our bias check, considering that in general our participants preferred to press the 'M' key (i.e. the key associated with the B sound in ABX order and with the A sound in BAX order) over the 'Z' key. It should be kept in mind that the bias check was performed only for the filler tokens, as the preference for one or the other sound in the experimental tokens was the subject of investigation, and therefore it would not make sense to perform such a bias check for these tokens.

Looking back at the experiment's design, the confounding effect of order could have theoretically been eliminated by employing a variation of the ABX scheme, the AXB scheme. In this scheme, the 'X' sound is presented at equal time differences from both of the other sounds. However, this scheme was deliberately not chosen because of two reasons. Firstly, in the AXB scheme participants often seem to ignore the third ('B') stimulus and a bias towards 'A' can arise as a result (Van Hessen & Schouten, 1999). Secondly, it would introduce a new problem in our experiment. In some of our 'easier' fillers and catch trials, the 'A' and 'B' sounds were different words, which would in the AXB scheme inevitably create situations where if two of the same words were heard in succession, the participant would immediately come to the conclusion that 'X is more similar to A' even before the third sound was played. Analogously, if the first sound and the second sound differed, it would be clear that 'X is more similar to B' even before the third sound was played. If this happened multiple times in succession, which is possible since the order of the sets was random, the participants could have learned to employ this rather easy way of discrimination. This would cause a problem in the experimental tokens, in which all three sounds were the same word, whereby the participant would hear two words in succession, leading to the assumption 'X must be closer to A' after which they could think they would not have to pay much attention to the third sound. Thus, it was decided to use the ABX experimental scheme.

4.3 Limitations of the present study

The exploratory nature of the present study brought about its inherent limitations. First, one should be careful about the interpretation of our results regarding the 'fortis' and 'lenis' choices. It was a conscious decision to omit any labelling of the sounds in our ABX task. Consequently, any result described above as a 'fortis' choice is really only a choice to say that an artificially creaky sound seems to be closer to a natural fortis sound rather than a natural lenis sound. In this way, our experiment attempted to ensure that participants' judgments were unaffected by orthography. It is very much possible that participants would stray away from 'fortis' choices if labels were provided because they could likely be influenced by preconceived notions about stop types.

Another limitation of this study was the lack of a fine-grained manipulation of VOT values in the experimental sets, so that a wider range of VOT's could have been compared in relation to the strength of creak as a potential cue. Furthermore, there were no manipulations of F0, whereby the F0 of the lenis tokens was not consistently lower than the F0 of the fortis tokens. A lower post-lenis F0 would be expected based on previous literature (e. g. Cho et al., 2002; Silva, 2006), although other literature displays mean F0 values not very far from each other (160 Hz for fortis stops, 148 Hz for lenis stops in male speakers in Kagaya, 1974) or essentially overlapping (216 Hz for fortis stops, 221 Hz for lenis stops in female speakers in Kim & Lotto, 2002). In this way, the potential strength of F0 as a cue in our experiment remains unclear. The values could be seen as ambiguous since the F0 differences between post-fortis and post-lenis vowels were not prominent in the recordings produced by our speaker. This ambiguity would be desirable in the context of this experiment. Alternatively, as the F0 values were quite high for post-lenis vowels, these could be considered closer to those expected for post-fortis vowels, which would be undesirable. This remains a limitation of this study.

4.4 Future aspects

The findings of this study showed that in addition to VOT and F0, the presence of creaky voice can potentially serve as another cue in the fortis-lenis distinction. The need to manage two, or possibly three distinct acoustic cues to produce and perceive the three-way stop contrast prompts further inquiries into how children learn this skill and how adults assess children's use of these cues. Previous studies into children's language acquisition have shown that word-initial fortis stops are mastered by children usually around 2;6 (years;months), which is much earlier than in the case of word-initial lenis and aspirated stops, with those typically becoming mastered by age 4;0 (Kim & Pae, 2005). Other studies have demonstrated that children initially produce phonologically undifferentiated tokens with short-lag VOT, and only later they gradually learn to produce long-lag VOT for lenis and aspirated stops, thereby effectively distinguishing these from the short-lag fortis stops (Kong et al., 2011; Son, 2017). Comparatively, the use of F0 as a cue has been observed later in life (3;8 and older), although mostly for the lenis-aspirated contrast (Son, 2017). The fortis-lenis distinction was not, to our knowledge, yet investigated but it can be hypothesized that this distinction for children is still very much dependent on VOT. This raises a question whether the presence of creak already influences perception at such an early age. Such issue could be explored by an apparent time study.

Another aspect relevant to the presence of creak is child-directed speech (CDS). Ko (2018) investigated CDS of Korean mothers, and found that they did not neutralize the VOT difference between lenis and aspirated consonants, but rather employed both VOT and F0 in this contrast. It was suggested that mothers might aim to support children's development by paying attention to secondary cues, such as VOT for the lenis-aspirated contrast, rather than emphasizing innovative forms of stop categories with an enhanced F0 cue. Again, fortis stops were found to differ from the lenis and aspirated stops mainly through VOT, however, creaky voice might have been present in at least some of the fortis tokens produced by the mothers. It should be noted that the author did not specify from which region the investigated mothers came from, therefore, these findings should be interpreted carefully. However, what influence creaky voice has on how children perceive their mothers' speech is another possible area of investigation.

There also remains the question of what role the presence of creak plays for L2 learners of Korean. As the three-way stop contrast found in Korean is typologically very rare, L2 learners need to learn how to use VOT and F0 to effectively distinguish between the individual stop types, especially because all three are phonetically voiceless in word-initial position. A study into whether new learners of Korean are already sensitive to creak could bring valuable insights. It would also be particularly interesting to investigate how learners from L1 backgrounds where creaky phonation has a contrastive fuction differ in their sensitivity to creak when learning Korean as their L2 compared to learners whose L1 does not use creak contrastively.

Above all, it is generally important to consider voice quality when investigating contrasts such as the one found in Korean. As is obvious from the present study, the instigation of creaky voice, or at least an audio manipulation that instigates the percept of creaky voice, is likely to change the perception of the stop. Production studies have already employed measurements of spectral slope (e. g. H1-H2) and other parameters of voice quality (e. g. Harmonics-to-Noise Ratio (HNR), jitter, flutter, period doubling), however, perception studies where the effect of different voice quality parameter manipulations was investigated have only been conducted sparsely in the context of Korean (e. g. Kang, 2010; Schertz et al., 2019). In this way, voice quality is quite under-researched compared to other acoustic parameters, especially in research of Korean. Further investigations into the strength of creak would also be beneficial because the ratio of amplitude manipulation present in each of our experimental tokens was kept at a constant ratio of 60% of the original amplitude. An experiment investigating the amount of amplitude modulation (or other types of modulation) needed for successful perception of creak would be yet another step towards thorough investigation of creak in the context of Korean.

Lastly, the results of our perception study spark the question whether stronger reliance on creak as a cue (as displayed by some of our participants with higher amount of fortis choices) in the fortis-lenis distinction actually translates into reliance on creak in production. A follow-up study examining the production data of listeners who appear to use creak as a cue (such as participants P_23 or P_04) in distinguishing stop types would be highly valuable.

5. Conclusion

The present study investigated the influence of creaky voice on the perception of word-initial stops in Korean, focusing on the fortis-lenis distinction. Our findings reveal that while the presence of creak alone does not overwhelmingly lead to the perception of a fortis stop, it significantly influences listeners' choices between fortis and lenis stops. A notable portion of our participants associated creaky voice with fortis stops, suggesting a real perceptual effect rather than random chance. This indicates that listeners may be integrating creak as an additional cue in distinguishing between fortis and lenis stops, alongside already established cues like VOT and F0. Interestingly, presence of natural creak in the experimental stimuli did not significantly alter the odds of a fortis choice compared to the absence of creak, indicating that listeners might first categorize stops as fortis or lenis and then make choices based on comparing these mental representations rather than comparing purely auditory information. Future research should explore the interplay between creak and other acoustic cues in greater detail, potentially incorporating a broader range of VOT and F0 values and more sophisticated voice quality measurements. Importantly, research into creaky voice in Korean is still very new and a solid body of work is yet to be built through studies examining creak in both production and perception. As of now, this study is only the first step towards understanding the function of creak in Standard Seoul Korean, although the results display that it might be a worthwhile area to investigate.

6. Bibliography

Abberton, E. (1972). Some laryngographic data for Korean stops. *Journal of the International Phonetic Association*, *2*(2), 67–78. https://doi.org/10.1017/S0025100300000517

Bang, H.-Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T.-J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120–144. https://doi.org/10.1016/j.wocn.2017.09.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/doi:10.18637/jss.v067.i01.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sample sound. *Proceedings of the Institute of Phonetics, University of Amsterdam*, *17*, 97–110.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer.* (6.4.07) [Computer software]. https://www.praat.org.

Cho, T. (1996). *Vowel correlates to consonant phonation: An acoustic-perceptual study of Korean obstruents.* [M.A. thesis]. University of Texas at Arlington.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*(2), 193–228. https://doi.org/10.1006/jpho.20.0153

Cho, T., & Keating, P. A. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*(2), 155–190. https://doi.org/10.1006/jpho.2001.0131

Chodroff, E., & Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *The Journal of the Acoustical Society of America*, *136*(5), 2762–2772. https://doi.org/10.1121/1.4896470

Choi, J., Kim, S., & Cho, T. (2020). An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLOS ONE*, *15*(10), e0240682. https://doi.org/10.1371/journal.pone.0240682

Garellek, M. (2019). The phonetics of voice. In *Routledge Handbook of Phonetics* (pp. 75–106).

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376. https://doi.org/10.3758/BF03194885

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America*, *47*(2B), 613–617. https://doi.org/10.1121/1.1911936

Han, J.-I. (1998). VOT in the surface distinction of Korean plain and tense stops in initial position: A perception test. *Speech Sciences*, *3*, 109–117.

Han, M. S., & Weitzman, R. S. (1970). Acoustic Features of Korean /P, T, K/, /p, t, k/ and

/ph, th, kh/. *Phonetica*, *22*(2), 112–128. https://doi.org/10.1159/000259311
Holliday, J. J., & Kong, E. J. (2011). *Dialectal variation in the acoustics correlates of Korean stops*. ICPhS XVII, Hong Kong.
Huang, Y. (2020). Different attributes of creaky voice distinctly affect Mandarin tonal perception. *The Journal of the Acoustical Society of America*, *147*(3), 1441–1458. https://doi.org/10.1121/10.0000721
Huang, Y. (2022). Articulatory properties of period-doubled voice in Mandarin. *Speech Prosody 2022*, 545–549. https://doi.org/10.21437/SpeechProsody.2022-111
Huang, Y. (2023a). *Perception and imitation of period doubling*. ICPhS 2023, Prague.
Huang, Y. (2023b). *Phonetics of period doubling* [PhD. thesis, University of California San Diego].
https://escholarship.org/uc/item/33n789f2
Jun, S.-A. (2020). *Studies in Korean prosody: Current issues and future prospects.*
Kagaya, R. (1974). A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics*, *2*(2), 161–180. https://doi.org/10.1016/S0095-4470(19)31191-X
Kang, K.-H. (2010). Generational differences in the perception of Korean stops. *Phonetics and Speech Sciences*, *2*(3), 3–10.
Kang, K.-H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategiesa). *Journal of the Acoustical Society of America*, *124*(6).
Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76–90. https://doi.org/10.1016/j.wocn.2014.03.005
Kang, Y., Schertz, J., & Han, S. (2022). The phonology and phonetics of Korean stop laryngeal contrasts. In S. Cho & J. Whitman (Eds.), *The Cambridge Handbook of Korean Linguistics* (1st ed., pp. 215–247). Cambridge University Press. https://doi.org/10.1017/9781108292351.009
Keating, P. A., Garellek, M., Kreiman, J., & Chai, Y. (2023). Acoustic properties of subtypes of creaky voice. *The Journal of the Acoustical Society of America*, *153*(3), A297–A297. https://doi.org/10.1121/10.0018918
Keating, P., Esposito, C., Garellek, M., Khan, S., & Kuang, J. (2010). Phonation contrasts across languages. *UCLA Working Papers in Phonetics*, *108*, 188–202.
Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. *In Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, 2–7. https://doi.org/10.1121/10.0018918
Kim, C.-W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). WORD, *21*(3), 339–359. https://doi.org/10.1080/00437956.1965.11435434
Kim, H., Maeda, S., & Honda, K. (2010). Invariant articulatory bases of the features [tense] and [spread glottis] in Korean plosives: New stroboscopic cine-MRI data. *Journal of Phonetics*, *38*(1), 90–108. https://doi.org/10.1016/j.wocn.2009.03.003
Kim, M.-J., & Pae, S.-Y. (2005). The percentage of consonants correct and the ages of aonsonantal aquisition for "Korean-Test of Articulation for Children(K-TAC)." *Speech Sciences*, *12*(2), 139–149.

Kim, M.-R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, *30*(1), 77–100. https://doi.org/10.1006/jpho.2001.0152

Kim, M.-R. C., & Lotto, A. (2002). An investigation of acoustic characteristics of Korean stops produced by non-heritage learners. *The Korean Language in America*, *7*, 177–187.

Kim, M.-R., & Duanmu, S. (2004). "Tense" and "Lax" stops in Korean. *Journal of East Asian Linguistics*, *13*(1), 59–104. https://doi.org/10.1023/B:JEAL.0000007344.43938.4e

Ko, E.-S. (2018). *Mothers would rather speak clearly than spread innovation: The case of Korean VOT*. 16th Conference on Laboratory Phonology, Lisbon, Portugal.

Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, *39*(2), 196–211. https://doi.org/10.1016/j.wocn.2011.02.002

Kong, E. J., & Lee, H. (2018). Attentional modulation and individual differences in explaining the changing role of fundamental frequency in Korean laryngeal stop perception. *Language and Speech*, *61*(3), 384–408. https://doi.org/10.1177/0023830917729840

Kuang, J. (2013). *Phonation in tonal contrasts* [PhD thesis]. University of California Los Angeles.

Kuang, J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *The Journal of the Acoustical Society of America*, *142*(3), 1693–1706. https://doi.org/10.1121/1.5003649

Kuang, J., & Liberman, M. (2016). The effect of vocal fry on pitch perception. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5260–5264. https://doi.org/10.1109/ICASSP.2016.7472681

Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, *2*(2), 205–254. https://doi.org/10.1017/S0954394500000338

Lee, H., Holliday, J. J., & Kong, E. J. (2020). Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass*, *14*(7), e12374. https://doi.org/10.1111/lnc3.12374

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, *41*(2), 117–132. https://doi.org/10.1016/j.wocn.2012.12.002

Lee, K.-H., & Jung, M. S. (2000). Acoustic characteristics and perceptual cues for Korean stops. *Speech Sciences*, *7*(2), 139–155.

Lee, S. (1990). Korean lenis and fortis stops: Synthesis and categorial speech perception task. In *G. Lee and W. Cowart (Ed.), Papers from the linguistics laboratory.* (pp. 105–120). Department of Linguistics, Ohio State University.

Lee, S., & Katz, J. (2016). Perceptual integration of acoustic cues to laryngeal contrasts in Korean fricatives. *The Journal of the Acoustical Society of America*, *139*(2), 605–611. https://doi.org/10.1121/1.4926435

Li, A., Lai, W., & Kuang, J. (2023). Creaky voice identification in Mandarin: The effects of prosodic position, tone, pitch range and creak locality. *The Journal of the Acoustical Society of America*, *154*(1), 126–140. https://doi.org/10.1121/10.0019941

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153–167. https://doi.org/10.1177/002383095800100301

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. WORD, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Park, H. (2004). Intensity and Spectral Characteristics of the Release Burst in Korean Alveolar Stops. 언어학 0.39: 117–128.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*(2), 253–260. https://doi.org/10.3758/BF03214136

Ren, X., & Mok, P. (2018). Tonogenesis in Seoul Korean and L3 production of Korean stops by Cantonese- English bilinguals. *6th International Symposium on Tonal Aspects of Languages (TAL 2018)*, 192–196. https://doi.org/10.21437/TAL.2018-39

Repp, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, *22*(2), 173–189. https://doi.org/10.1177/002383097902200207

RStudio Team. (2022). *RStudio: Integrated Development for R.* (2022.12.0+353) [R]. PBC.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, *52*, 183–204.
https://doi.org/10.1016/j.wocn.2015.07.003

Schertz, J., Kang, Y., & Han, S. (2019). Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *10*(1), 13. https://doi.org/10.5334/labphon.67

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71–80. https://doi.org/10.1016/S0167-6393(02)00094-8

Silva, D. J. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(02), 287–308. https://doi.org/10.1017/S0952675706000911

Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, *20*(2), 171–186. https://doi.org/10.1016/j.jvoice.2005.04.002

Son, G. (2017). *Interactive development of F0 as an acoustic cue for Korean stop contrast* [PhD. thesis]. University of Pennsylvania.

Van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, *56*(1–2), 56–72. https://doi.org/10.1159/000028441

Watkins, M., Boersma, P., & Hamann, S. (2024, accepted). Revisiting pitch jumps: F0 ratio in Seoul Korean. *Proceedings of Interspeech 2024*.

Weitzman, R. S., & Han, M. S. (1966). Acoustic features in the manner differentiation of Korean stop consonants. *The Journal of the Acoustical Society of America*, *40*(5), 1272–1273. https://doi.org/10.1121/1.2143467

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* [Computer software]. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852–866. https://doi.org/10.1121/10.0000692

Yu, H. J. (2018). Tonal development and voice quality in the stops of Seoul Korean. *Phonetics and Speech Sciences*, *10*(4), 91–99. https://doi.org/10.13064/KSSS.2018.10.4.091

Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320. https://doi.org/10.1121/1.4887462

Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., Hahn, H., & Kim, J. (2015). The Korean Corpus of Spontaneous Speech. *Phonetics and Speech Sciences*, *7*(2), 103–109. https://doi.org/10.13064/KSSS.2015.7.2.103

Zhang, Y., & Kirby, J. (2020). The role of F0 and phonation cues in Cantonese low tone perception. *The Journal of the Acoustical Society of America*, *148*(1), EL40–EL45. https://doi.org/10.1121/10.0001523

7. Appendix

Table 7: Nonce words used in the present study. Coronal stops were used as experimental tokens, the rest was used as fillers.

	Aspirated	Fortis	Lenis
Labial	**ㅍ /pʰ/**	**ㅃ /p*/**	**ㅂ /p/**
	파즛 /pʰadʑut/	빠즛 /p*adʑut/	바즛 /padʑut/
	파누 /pʰanu/	빠누 /p*anu/	바누 /panu/
Coronal	**ㅌ /tʰ/**	**ㄸ /t*/**	**ㄷ /t/**
	타샤 /tʰasja/	따샤 /t*asja/	다샤 /tasja/
	타즛 /tʰadʑut/	따즛 /t*adʑut/	다즛 /tadʑut/
Velar	**ㅋ /kʰ/**	**ㄲ /k*/**	**ㄱ /k/**
	카샤 /kʰasja/	까샤 /k*asja/	가샤 /kasja/
	카즛 /kʰadʑut/	까즛 /k*adʑut/	가즛 /kadʑut/

Table 8: Experimental triads used in the present study. All tokens were taken from sentence-initial contexts except for two creaky fortis tokens taken from words produced in isolation (individual_fortis+creak_01, individual_fortis+creak_02).

Word	A	B	X
/tadʑut/	fortis_01	lenis_05	lenis+creak_01
/tadʑut/	fortis_02	lenis_06	lenis+creak_02
/tadʑut/	fortis_03	lenis_07	lenis+creak_03
/tadʑut/	fortis_04	lenis_08	lenis+creak_04
/tadʑut/	fortis+creak_01	lenis_01	lenis+creak_05
/tadʑut/	fortis+creak_02	lenis_02	lenis+creak_06
/tadʑut/	fortis+creak_03	lenis_03	lenis+creak_07
/tadʑut/	fortis+creak_04	lenis_04	lenis+creak_08
/tasja/	fortis_01	lenis_06	lenis+creak_01
/tasja/	fortis_02	lenis_05	lenis+creak_02
/tasja/	fortis_03	lenis_04	lenis+creak_03
/tasja/	fortis_04	lenis_03	lenis+creak_04
/tasja/	fortis+creak_01	lenis_02	lenis+creak_05
/tasja/	fortis+creak_02	lenis_01	lenis+creak_06
/tasja/	individual_fortis+creak_01	lenis_04	lenis+creak_05
/tasja/	individual_fortis+creak_02	lenis_03	lenis+creak_02