Universiteit van Amsterdam

Production of Korean Stops: A Corpus Analysis of the Lenis-Aspirated Stop Distinction in Standard Seoul Korean

Thesis

Degree: rMA Linguistics and Communication

Michaela Watkins (Student ID 13220861)

Supervised by Dr. Suki Yiu

Faculty of Arts and Humanities
University of Amsterdam
The Netherlands
June 24, 2022

Abstract

Standard Seoul Korean (SSK) is typologically unique, in that all laryngeal consonants remain voiceless in word initial position, despite a three-way contrast. In younger speakers (born after 1965), F0 appears to be replacing VOT as the main acoustic cue for distinguishing between the stops, in particular in distinguishing between lenis and aspirated types. This research evaluated two groups of speakers (N = 20) from the open-access corpus *The Korean Corpus of Spontaneous Speech* (Yun et al., 2015) to evaluate two different age groups all under the age of 60 years in their F0, VOT, and H1-H2 production across stop types. The results confirmed that all participants preferred F0 as an acoustic cue in production, but that young speakers within the post-1965 group were more consistent in use and showed more overlap in VOT across the lenis-aspirated stop distinction. This suggests that VOT has undergone further redundancy in distinguishing lenis-aspirated stops in younger speakers of the post-1965 group, with F0 strengthening as a dominant acoustic cue, entailing that younger speakers are further ahead in the process of tonogenesis compared to older speakers.

Contents

1 Introduction

For many languages if a three-way laryngeal contrast for stop consonants is observed, then typically it is categorized by voicing and aspiration: “voiced", “voiceless unaspirated", and “aspirated" (Cho et al. 2002). Standard Seoul Korean (SSK), however, is typologically unique and does not follow this pattern, as instead the stops all remain voiceless in word-initial position (Cho et al. 2002). These three stops in Standard Seoul Korean (SSK) are termed aspirated, fortis, and lenis (cf. table 1 below, also table 14 Appendix A).

Previous analyses demonstrated an important role of voice onset time (VOT), amplitude differences between first and second harmonics (H1-H2), as well as fundamental frequency (F0) in distinguishing between the stops (Cho et al., 2002; Kang & Guion, 2008, 2014; Silva, 2006). Silva (2006) demonstrated that in older speakers (born pre-1965) VOT appeared to be the primary cue for distinguishing the stop types, with F0 being largely redundant. For these older speakers, aspirated stops were described as having a long VOT (in ms), lenis intermediary but closer to fortis than aspirated, and fortis approaching 0ms in duration (Silva, 2006). Speakers born after 1965, however, appear to be shifting towards preferring F0 as a primary cue, with VOT playing a lesser part in distinguishing the stop types, particularly regarding differentiating between aspirated and lenis stops (Kang & Guion, 2008; Silva, 2006). This is due to the fact that lenis and aspirated stops appear to be shifting closer together in their VOT values to the point of overlap, whereas fortis stops remain marked due to their short VOT lengths and therefore readily distinguishable from the other two stop types by VOT alone (Cho et al. 2002; Kang & Guion, 2008; Silva, 2006).

This implies that, at least for the lenis-aspirated stop distinction, younger speakers may well be undergoing tonogenesis, the process by which a non-tonal contrast is replaced by a tonal contrast (Kang, 2014; Michaud & Sands, 2020). In the case of SSK, this would be F0 becoming the primary acoustic contrast in place of VOT. This is because a shift towards F0 preference also implies a shift towards reliance on pitch, as pitch can be defined as our perception of frequency as being placed on a scale of high or low (Ladefoged & Johnson, 2014 p.25), thus resulting in a reliance on tonal or pitch contour features to distinguish lexical items rather than other acoustic cue such as duration (Ladefoged & Johnson, 2014 p.269). Kang (2014) summarises the stages described by Maran (1973) in figure 1 below:

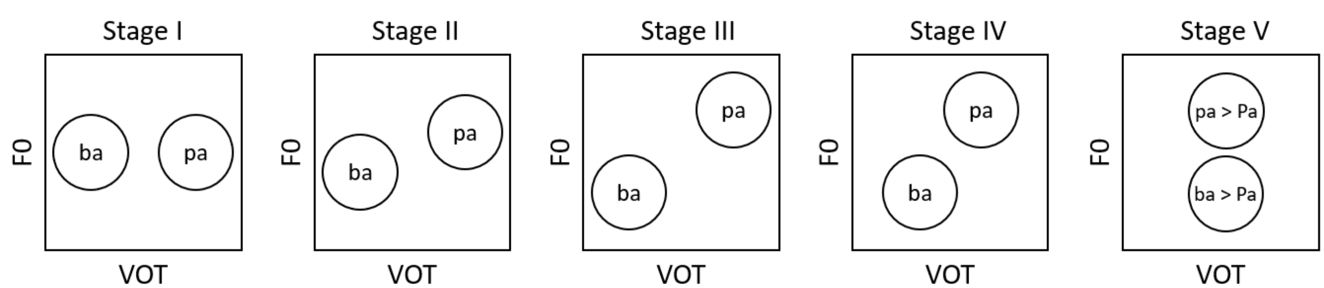

Figure 1: Stages of tonogenesis. (Kang, 2014, p.2)

The above figure describes the following: in the initial stage, there is no F0 distinction, only VOT. In stage 2, the emergence of F0 distinction is seen, however F0 remains redundant. In stage 3, full redundancy of the F0 cue is realised, followed by stage 4 wherein there is a further enhancement of the F0 distinction resulting in a weakening of the VOT distinction. The final stage then describes the full transition to a loss of VOT distinction and a full reliance on F0 as the acoustic cue for differentiating between two tokens (Kang 2014).

To summarise the above information, SSK can be hypothesised to have the following acoustic cue differences and tonogenesis stages in the pre- and post-1965 generations, based on the results of the aforementioned studies:

	Pre-1965 Generation	(Hypothesised) Post-1965 Generation
Lenis	VOT: Intermediary between aspirated and fortis F0: Low **Tonogenesis stage 1-2 (F0 redundant)**	VOT: Longer, approaching values seen in aspirated stops. F0: Intermediary **Tonogenesis stage 3-4 (F0 reliant, VOT not yet fully redundant)**
Fortis	VOT: Extremely short (approaching 0) F0: Higher than other stop types **Tonogenesis stage 1-2 (F0 redundant)**	VOT: Extremely short (approaching 0) F0: High **Tonogenesis stage 1-2 (VOT remains as a distinguishing cue)**
Aspirated	VOT: Extremely long F0: High, lower than fortis **Tonogenesis stage 1-2 (F0 redundant)**	VOT: Shorter, approaching values seen in lenis stops. F0: Higher than previous analyses, possibly approaching fortis values. **Tonogenesis stage 3-4 (F0 reliant, VOT not yet fully redundant)**

Table 1: *Summary of pre- and post-1965 groups in terms of VOT and F0 values based on previous analyses. Also includes hypothesised tonogenesis stages for both groups. Created using information extracted from Bang et al. (2017), Choi et al. (2020), Kang & Guion (2006:2008), Silva (2006).*

Naturally, it is important to note that tonogenesis was not a focus in the majority of these studies for either group, and so the above is a hypothesis based on the results of the aforementioned studies: the post-1965 group is a hypothetical scenario of what may be observed in modern speakers, as is hoped to be demonstrated in the current study.

It is interesting that SSK has not progressed further to stage 5 with tonogenesis (the completed development of a tonal or pitch accent system with full VOT redundancy) despite a seemingly strong preference for F0 among younger speakers as observed in prior experiments (e.g. Bang et al. 2020; Kang & Guion, 2008). This is because if the tonogenesis model can be applied to SSK, then it would be expected that in SSK there be some degree of tonal distinction between lexical items, especially between lenis and aspirated stops. However, this does not seem to yet have been reported in previous studies: lexical items still do not seem to be distinguished at to any degree based on pitch accent or tone as is observed in linguistic systems like Japanese (partial tonality) or Mandarin Chinese (fully tonal).

Previous studies (cf. section 2) have primarily focused on the pre- and post-1965 generations, but few have looked within-group to evaluate if the shift has progressed further within younger speakers that should in theory all show some degree of F0 preference in distinguishing between lenis and aspirated stops. This demonstrates a need to revisit and assess what is occurring within this generation of native speakers, that in theory all possess a strong preference towards F0 as their primary distinguishing cue - in other words, has tonogenesis progressed or stabilised within this generation? Are speakers at the older end of this post-1965 generation more reliant on VOT than those at the younger, or is the difference between younger and older speakers minimal in their acoustic cue production?

An additional reason to re-investigate is due to the fact that previous experiments, e.g. Silva, (2006), Kang & Guion (2008), are problematic for three main reasons.

Firstly, many of the studies demonstrating this shift towards F0 as a primary cue have largely been done on Korean diaspora, namely Korean students in the USA (Silva, 2006; Kang & Guion, 2008). This will be discussed in further detail in section 2. Secondly, the sample sizes of these studies are often small, with the result that individual variation is likely to greatly influence the outcome of the results. Finally, as mentioned above, previous studies use macro distinctions such as 'old' and 'young', with little within-group or individual analyses. This coupled with small sample sizes could result in individual differences skewing the results greatly.

The research here will focus purely on the previously described 'young' group (i.e., born after 1965), and evaluate internal group asymmetries in F0 production. The individuals analysed this this study are born roughly between the years 1967 to 2000, with two groups selected for analyses at opposite ends of the age range: older (born between 1967 and 1970) and younger (born between 1995 and 2000). The exact years of birth and year each individual recording was taken are not specified in the corpus, but instead calculated here from given ages and median of date range of corpus data collection (data was collected between between 2012-2015).

This research therefore proposes to evaluate the shift from VOT to F0 as the primary cue, particularly when distinguishing between lenis and aspirated stops in speakers of SSK born after 1965. The amplitude difference between the first and second harmonics (H1-H2) will also be assessed, to further evaluate voice quality. This analysis will be done through the use of an open-source spoken speech corpus, the *Korean Corpus of Spontaneous Speech* Yun et al. (2015), which contains recordings of native speakers of SSK, whose parents were also born and raised in the Seoul (Gyeonggi) area. Additionally, this research looks within the 'younger' post-1965 group as defined by Kang & Guion (2008), to evaluate if the process of tonogenesis has progressed or stagnated across the two age groups as indicated by an increased F0 preference.

Section 2.1 of this paper will present the research background, in which the typological uniqueness of SSK consonants will be explained in further depth regarding their unique F0, VOT, and H1-H2 properties as well as previous experiments that evaluated the shift towards F0 as a primary contrasting cue. In section 2.2 the *Korean Corpus of Spontaneous Speech* (Yun et al. 2015) will be introduced. This thesis will then present the methodology of data analysis as well as corpus analysis (section 3). Results will be given in section 4. Finally this paper will conclude with a discussion (section 5) of the results and give a conclusion of said results (section 6).

The research questions are therefore the following:

To what extent has the shift in preference towards F0 (i.e., tonogenesis) progressed in younger speakers (born after 1965)? Are the youngest of this group similar to the oldest, or does the older group show less clear preference for F0 to contrast between lenis and aspirated stops? If so, is the difference significant? Have aspirated and lenis stops began to merge, i.e. overlap in their VOT values, suggesting a redundancy of this acoustic cue? If so, to what extent (if any) does this differ across the two age groups? Finally, what possible stage of tonogenesis is observed, and are the two groups in the same or different stage (stagnation or progression)?

In terms of hypotheses, the corpus evaluated here contains individuals born after 1965, and so it is expected that neither group selects VOT as the primary acoustic cue for distinguishing between lenis and aspirated stops. That is to say, lenis and aspirated stops should not show a significant effect of VOT, with possible overlapping of values between the two stop types in both groups. VOT is however hypothesised to overlap more in the younger group than the older group, due to an increased redundancy in the use of VOT to distinguish between lenis and aspirated stops. F0 is expected to be significant as a cue in both groups, however, it is expected that this is clearest in the youngest group, with F0 being most distinctive, whereas the older group showing a less clear distinction between their F0 values (i.e. more overlap and greater inconsistency in use between lenis and aspirated stops).

However, in neither group is VOT expected to be completely redundant for distinguishing between lenis and aspirated stops (i.e. it is still used to some degree as a contrasting cue). However, it is expected that VOT has progressed in redundancy in the younger group more so than in the older group. Fortis stops are expected to show little change in VOT, however, it is not yet known if or how the F0 will be affected due to the salience of VOT in this stop type. It is also hypothesised that fortis stops retain their creaky voice property, a remnant of the generation prior to 1965 in which lenis and fortis stops were mostly differentiated by creak (cf. section 2.). However, it is expected that creaky voice is more common and possibly more extreme in older speakers than younger speakers, especially for fortis and lenis stops, as a remnant of this previous acoustic distinction.

Finally, it is hypothesised that both groups are in at least stage 3 of the tonogenesis process (Figure 1, partial redundancy of VOT). However, it is hypothesised that the youngest speakers may have progressed further than older speakers, with a further weakening of the VOT distinction (redundancy) in comparison to the older group. As no partial or full tonal system is yet observed in SSK, it is unlikely that tonogenesis is completed.

2 Research Background

This section provides an overview into the three-way stop distinction inn SSK, regarding F0, VOT, and harmonics. This section aims to help the reader understand the reason for evaluating the lenis-aspirated stop distinction, in particular for evaluating the shift from VOT to F0, as well as discussing why tonogenesis may be an interesting theoretical area to consider. Additionally, this section also provides information on previous analyses, as well as problematic features of said previous analyses. Finally, this section aims to provide the reader a comprehensive summary of the *Korean Corpus of Spontaneous Speech* (Yun et al. 2015).

2.1 Previous Analyses

As mentioned earlier, consonants in SSK are typologically unique, as usually a three-way laryngeal contrast is distinguished by voicing and aspiration (Cho et al. 2002); however, in word-initial position in SSK all plosives remain voiceless. This therefore raises the question of how exactly do speakers distinguish between the three cues? This is particularly relevant to the lenis-aspirated stop distinction, as fortis has an extremely short VOT making it the most marked out of the three stop types.

The first full acoustic study of the unique three-way laryngeal contrast in Korean began with that of Lisker and Abramson (1964), in which it was demonstrated that the VOT of Korean fortis stops was short lag, aspirated long lag, and lenis stops intermediary but closer to fortis than aspirated (Lisker & Abramson, 1964). Similar findings were soon reported in consequent studies, such as that of Kim (1965), Han & Weitzman (1970), and Hardcastle (1973). All of these initial studies demonstrated that aspirated stops were well differentiated from the other two stop types with a significantly longer VOT duration. Lenis was more similar in length to fortis, even at times overlapping in speakers born pre-1965, however fortis was characterised by an extremely short VOT coupled with larnygeal muscular tenseness during oral closure (Hirose, Lee, & Ushijima, 1974), consequently also resulting in a creakier voice compared to lenis stops (Abberton, 1972).

The results of the preliminary studies into this three way contrast can therefore be summarised with F0 having a hierarchy of fortis > aspirated > lenis, and VOT as aspirated > lenis > fortis (e.g. Hardcastle, 1973; Cho et al. 2002). A summary of the the information given here concerning the pre-1965 generation is given below:

	FORTIS	LENIS	ASPIRATED
VOICE ONSET TIME (VOT, ms)	Extremely short (approaching 0)	Intermediary	Longer
FUNDAMENTAL FREQUENCY (F0, Hz)	High	Low	High (lower than fortis)
HARMONICS 1-2 (amplitude, dB)	Not breathy. Lower than other stop types	Breathy, less than aspirated	Breathy
ENGLISH EQUIVALENT	No equivalent	/d/ in die	/t/ in tie

Table 2: Summary of Fortis, Lenis, and Aspirated stops in word-initial position for SSK speaker born before 1965. Information adapted from Cho et al. (2002).

After these initial studies, research began to focus more on prosody and discussions on suprasegmentals such as F0 were given more attention; consequently, it was observed that the previous VOT near-overlap between fortis and lenis stops was no longer present, suggesting that a sound change had begun in which the VOT difference between aspirated and lenis stops was decreasing (Silva, 1991, 1992). This then leads into the most recent research of this shift, with both Silva (2006) and Kang (2014) demonstrating that lenis stops in speakers born after 1965 no longer bore any overlap with fortis but rather were fully distinct from fortis, instead now overlapping with aspirated stops instead, suggesting a neutralisation of VOT between lenis and aspirated stops (Kang, 2014).

Kang & Guion (2008) then used clear speech studies to evaluate the lenis-aspirated distinction between older and younger speakers (pre- and post-1965). Clear speech is usually given as speech that is designed by speakers to "minimize perceptual confusions when they estimate potential speech perception difficulty on the listener's part" (Kang & Guion, 2008, p. 3909): for example, native speakers explaining a word to a non-native speaker. Their study demonstrated that younger speakers emphasised the F0 cue in clear speech while older speakers emphasised VOT, suggesting a generational difference in cue preference for differentiating between stop types.

Using clear speech is attractive, as it is thought that acoustic cues are most readily distinguished from one another in contexts where the speaker is making an effort to produce clearly articulated cues; one theory of this being the Hypo- and Hyper-articulation Theory (H & H Theory; Lindblom 1990). The H & H Theory states that any minimisation of effort drives a "weakening" of consonants, resulting in the consonants becoming similar to neighbouring sounds through processes such as assimilation, lenition, and deletion, all of which reduce clarity (Melguy, 2018). The theory also states the opposite to be a maximisation of clarity, i.e. consonantal "strengthening" in which target segments become more distinct, rather than less so (Melguy, 2018). The H & H Theory expands the above further by applying the idea of minimizing and maximizing speech to phonetic gradience, claiming that speech occurs "along a continuum with more forcefully articulated 'hyper' forms at one end, and less energetic 'hypo' forms at the other" (Lindblom, 1990, p. 1687). Hyper-articulation can therefore be thought to be the process in which 'extreme' productions of sounds are created; resulting in sounds that are maximially distinct from their neighbours. In contrast, hypo-articulation is the situation in which sounds overlap to some degree with one another (Melguy, 2018).

Bang et al. (2017) provided more evidence that the VOT contrast reduction and F0 enhancement was still in progress, suggesting the process of tonogenesis (figure 1) was occurring, stating that the "vowel intrinsic pitch difference is attenuated as contrastive F0 emerges, possibly due to the combined effect of controlled and automatic mechanisms" (Bang et al., 2017 p.141).

However, although the studies by Silva (2006) and Kang & Guion (2008) are highly influential, both are problematic in that the participants used were Korean diaspora currently living in the USA. Language contact studies have shown that diaspora speakers are more conservative than those remaining back home for two main reasons; firstly, interference due to language contact resulting in phonetic traits unique to those in diaspora, and secondly limited exposure to a sound change due to distance and immersion in a different linguistic environment (Nützel & Salmons, 2011). A study on Chadic languages by Pearce (2009; 2013) demonstrated that the degree of tonogenesis depended on how much the individual came into contact with French, i.e. different speakers used different weightings of F0 and VOT depending on their exposure to a different linguistic environment. Bang et al. (2017) also used the same data from Kang & Guion (2008), and thus their results potentially face a similar problem.

Finally, Kang (2014) used corpus data to evaluate the previous findings, namely the *Speech Corpus of Reading-Style Standard Korean*, made by the National Institute of the Korean Language (NIKL, 2005). This corpus contained 120 speakers reading out short stories in their natural speech rate, with an age of roughly between 19 to 71, or born between 1932 to 1984, at time of recording (in 2003). Although this corpus did not have the previous problems of Korean diaspora, due to using native monolingual Seoul speakers, it did not have even groupings of speakers across ages categories. That is to say, for example, only 6 speakers were used for the 1930 age group and then 38 for the 1970 age group. Gender was not equal either, as for example the 1950 age group had 4 male speakers and 25 female (Kang 2014). This unequal distribution of both age and gender makes cross-group comparison difficult, as it entails that in smaller groups a high amount of skewing may be present if just one individual in the group presents extreme values in their production of (e.g.) VOT or F0. Nevertheless, Kang (2014) stated that F0 distinction in younger speakers is enhanced more in younger speakers than in older speakers, and to a lower extent in the fortis-lenis contrast. The results suggested that instead of redundant tonal patterns being reinterpreted as a primary contrast and becoming phonologised first, followed by a stage where the redundant VOT difference is lost, rather more likely was that the two processes may actually occur (to some degree) simultaneously (Kang 2014). However, as mentioned before, group differences were unequal in this study and so the results should be taken with a degree of caution.

2.2 *The Korean Corpus of Spontaneous Speech*

The Korean Corpus of Spontaneous Speech, known also as the 'Seoul Corpus' is a fully transcribed, open-source corpus funded in 2015 by the Korea Research Foundation (Yun et al., 2015). The corpus consists of 40 speakers of Seoul Korean were interviewed, with 10 speakers (5 male and 5 female) each from the age ranges 10-19, 20-29, 30-39, and 40-49 (Appendix B: table 16). The recordings were made in a recording studio at Hanyang University in South Korea using a TASCAM HD-P2 recorder and AKG C420 headworn microphone, with .wav files sampled at 44kHz, 16-bit. Seven topics in total were covered, ranging from the personal details of the candidate, hobbies, work, and family. The amount of data collected totals 40 hours in length. The corpus is fully transcribed orthographically as well as phonemically in Hangeul (Korean orthography) as well as romanisation in PRAAT software (Boersma & Weenink, 2022). For this study the two oldest and youngest groups were selected with a total of 20 participants (see Table 15, Appendix B for the participants selected for the current study).

3 Methodology

This section begins with a discussion on the extraction of tokens from the corpus, followed by an explanation of the data analysis methods used.

3.1 Corpus Analysis Procedure

For this study, the oldest and youngest groups were selected due to the fact that these two groups were expected to show the biggest difference and thus the easiest to cross-compare. This resulted in 20 participants total across 2 groups (10 participants per group: table 15, Appendix B).

The corpus provided six recordings per participant, and for the current research the third and fourth recordings of each participant in each group were selected for analysis. This was due to participants being more nervous in the initial recordings and speaking less, and in later recordings becoming more at ease with the interview process and less likely to mumble and give incomplete answers.

The collected tokens were either CVC or CV, with the consonant to be evaluated in word-initial position. Word-internal consonants in Korean are often subjected to rules such as elision, lenition, and assimilation, altering their phonological properties and making them unsuitable for this research (Ladefoged & Johnson, 2014). In terms of word position, both sentence initial and internal words starting with the required consonant were collected. This was due to the fact that some consonants (e.g. aspirated) were extremely rare in sentence-initial and word-initial position, resulting in too few tokens should only sentence-initial be permitted.

Sentence-internal words were only considered if preceded by a plosive, as it was seen that often if preceded by an open vowel, a nasal, or a fricative the token was subject to merging with the previous word to the extent that extraction of the /a/ vowel was impossible[1]. The only exception this was if there was sufficient distance between the two words so as to not show signs of assimilation (e.g. speaker pausing).

Assimilation was also possible with following vowels (e.g. 차이아 /tɕʰɐiɐ/ pronounced by some speakers as 재이 /tɕʰɛi/) so careful listening was required to assess if the initial vowel was clear enough both auditorily as well as visually (i.e. in the PRAAT spectogram) for measurement to be taken. Consonants were also required to have a following /a/ vowel: this was for comparison across tokens as well as with prior research (e.g. Kang & Guion, 2008). Tokens in sentence-final position were also omitted due to the possibility of decreasing F0 towards the end of the speech signal distorting values (Ladefoged & Johnson, 2014).

As the corpus is already fully transcribed, tokens were selected and listed in an *Excel* data sheet by comparing the provided transcription with the audio recorded. Tokens were discounted if the speaker whispered, coughed, or laughed, as well as any other irregularities in the speech signal or mumbling.

The values extracted from each token were F0, VOT, and amplitude difference between the first and second harmonics (H1-H2). VOT was manually measured by using the definition as the duration from the onset of stop release to the onset of the first periodic cycle of the following vowel (Ladefoged, 2003). H1-H2 and F0 were measured on the vowel following the target consonant. The F0 was measured by using the inbuilt pitch tracking function in PRAAT (Boersma & Weenink, 2022), with the F0 taken from the midpoint of the vowel. The F0 range was set to 75-300 Hz for male speakers, and 100-500 Hz

[1] For example, 거 같아요 is not pronounced /kɔkatajo/ as expected but more like /kɔtɔjo/ (elision to one word instead of two).

for female speakers. H1-H2, the amplitude (dB) difference between the first and second harmonics, was measured after the first full glottal pulse of the waveform using the fast fourier to spectrogram function in PRAAT (Boersma & Weenink, 2022).

The final number of tokens are given below, first by gender and group and then group alone:

Group	Gender	Aspirated	Fortis	Lenis	Total
Older (N = 10)	M	44	59	168	271
Older (N = 10)	F	47	43	124	214
Younger (N = 10)	M	33	42	178	253
Younger (N = 10)	F	26	31	141	198
Total per Stop Type		150	175	611	936

Table 3: Breakdown of stop type tokens observed per group as well as by gender within group.

Group	Aspirated	Fortis	Lenis	Total
Older (N = 20)	91	102	292	485
Younger (N = 20)	59	73	319	451
Total per Stop Type	150	175	611	936

Table 4: Breakdown of stop type tokens observed per group

3.2 Data Analysis Procedure

All tokens were evaluated using linear regression analysis through the *lmer* function in the *lme4* package in *RStudio* (Rstudio Team, 2021).

Firstly, three models were built to evaluate the overall interaction effects between the three stop types and VOT, F0, and H1-H2. The predictor variables for the three models were F0, VOT, and H1-H2 respectively. Contrast coding was applied to gender (F and M, -1/2, +1/2), group (O and Y, -1/2, +1/2) and stop type (A, L, F: +2/3, -1/3, -1/3). All data was centred on the median.

The three models were further tested to see if the position of the word in the utterance (at the start of the utterance or in the medial position of the sentence) was significant and if including this variable improved the model. This was done through firstly applying contrast coding to initial or internal (+initial, -internal: +1/2,-1/2) and then running an *ANOVA* between the model that included the position and the model that did not include the position of the word. The result indicated that for F0 and H1-H2 there was no improvement when including position, whereas a significant improvement was found for VOT when including sentence position. The final models therefore all contained stop type, gender, and group as fixed variables and participant as a random variable to control individual variation. All models also tested for the interaction between stop type and group. The model for

VOT included sentence position as well due to the improvement of the model upon its inclusion. The final overall models are given below:

F0:

$$lmer(F0.c \sim TypeStop * Group + Gender + (1|Participant), data = data)$$

VOT:

$$lmer(VOT.c \sim TypeStop * Group + Gender + Sposition + (1|Participant), data = data)$$

H1-H2:

$$lmer(Harmonic.c \sim TypeStop * Group + Gender + (1|Participant), data = data)$$

Additionally, separate models were also built to evaluate each stop type individually, to assess further if there are significant effects for the stops themselves without interference from other stop types. This was done by subsetting the data using the *subset*() function in *RStudio* (Rstudio Team, 2021) to create three smaller datasets for aspirated, lenis, and fortis respectively. These all included the above contrast coding, and only VOT included sentence position as a fixed variable. The nine formulae used for F0, VOT, and H1-H2 for all three stop types can be seen in Appendix C.

4 Results

This section summarises and compares F0, VOT, and H1-H2 both graphically as well as statistically. Initially this is done with the overall models to assess by acoustic cue, then the individual stop analysis is provided. By assessing by acoustic cues first, this research aims to see overall patterns in change of F0, VOT, and H1-H2 generally across groups and stop types, to see overall group differences. Afterwards, by analysing individually by stop type this research aims to provide a more detailed investigation into the individual stop types themselves to determine if it is the aspirated or lenis stop adjusting for F0 and VOT, and if consequent changes in fortis are detected.

4.1 F0

As shown in Figure 2, both groups mirror a similar pattern of aspirated stops having the highest F0, fortis intermediary, and lenis lowest for F0 values. This is confirmed in statistical analyses as being significant, with a significant effect found for the type of stop (an average of -25 Hz difference across stop types). For this sample, a significant effect for gender is observed as well as a clear difference graphically, with female participants of this study having on average 98 Hz higher F0 than males participants. However, given the biological differences between males and females and consequent differences in F0 frequency this is not surprising (Ladefoged, 2003), and does not therefore indicate any difference in the use of acoustic cues between male and female participants.

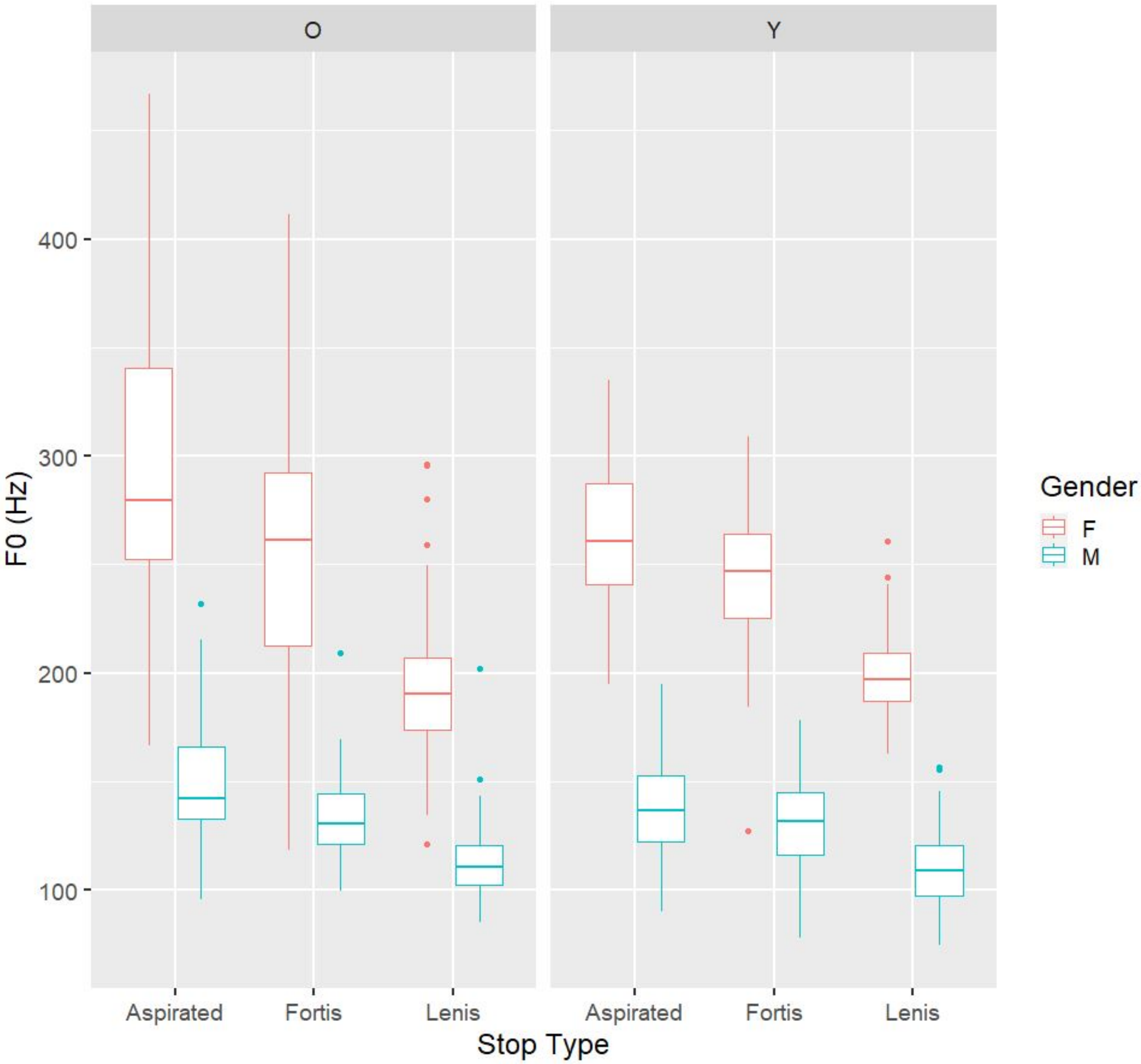

Figure 2: Comparison F0 values (Hz) by Gender and Group ('O' = Older, 'Y' = Younger)

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	40.567	2.706	19.065	14.992	<.001***
TypeStop+aspirated-fortis-lenis	39.552	2.716	915.597	14.565	<.001***
TypeStop	-25.232	1.776	917.989	-14.205	<.001***
Group-O+Y	-10.785	5.411	19.046	-1.993	0.0608
Gender-F+M	-98.563	5.238	16.746	-18.818	<.001***
TypeStop+aspirated-fortis-lenis: Group-O+Y	-21.879	5.431	915.590	-4.029	<.001***
TypeStop:Group-O+Y	4.945	3.552	918.038	1.392	0.1643

Table 5: Fixed and Interaction Effects for F0: $lmer(F0.c \sim TypeStop * Group + Gender + (1|Participant), data = data)$

Looking at the graphical representation, it is interesting that fortis is intermediary, given that prior research stated that fortis should show the highest F0 (cf. table 2). Of note here as well is that the younger group appear to have a fairly small distribution of their F0 values, with the median unskewed in each gender and few outliers. In contrast, older speakers appear to show much greater variation in their F0 values, with a much larger distribution, more skewing, and more outlying values, especially for older female speakers in aspirated and fortis stops. This may suggest that F0 is not as defined or consistent in older speakers, entailing that it may not be a major acoustic cue for this group. Interestingly, both groups and genders show relatively stable lenis stop values (seen clearer in Appendix E, Figure 5), possibly suggesting that for the lenis-aspirated stop distinction lenis stops are more stable than aspirated in their F0, implying potentially that it is the aspirated stops that are being subjected to change more so than the lenis stops.

This appears to be supported statistically, given that aspirated stops are significant, with on average 39.6 Hz higher value than other stop types. This again suggests that younger speakers are distinguishing aspirated stops from lenis by increasing F0 in aspirated stops alone to compensate for a reduction in VOT, which will be further analysed by stop type in Section 4.4.

Although it could be argued that not enough tokens were collected, as will be seen with the VOT analysis (cf. 4.2), the same large distribution is not observed in older speakers despite the same number of tokens collected (c.f. Figure 3). An alternative visualisation of this is given through a scatter plot in the Appendix (Appendix E, figure 5).

That group was not significant is also interesting, as this would suggest that there is no effect of stop type across groups and therefore no significant difference across groups in their F0 usage. However, there is a significant interaction effect between aspirated stops and group, with older participants having on average 21.9 Hz higher aspirated stops than younger speakers. This is likely due to the large amount of variation between the speakers as was observed in Figure 2, and suggests further when compared with the graphical results that younger speakers are comparatively stable in their aspirated F0 values, thus showing less extreme variation, especially when considering the stability seen in the median values of younger speakers.

The above results suggest that the hypothesis that both groups prefer F0 as the primary cue, particularly for the lenis-aspirated stop distinction, is correct. The results also indicate that this has progressed further in younger speakers than in older speakers due to stabilisation of F0 across speakers, supporting the original hypothesis of this paper.

4.2 VOT

Looking now at VOT, graphically (figure 3) there appears to be little difference visually between the two groups, with both groups showing roughly similar distributions and medians of VOT across their stop types. This is confirmed by statistical analyses (table 6), with no significant effect for neither group alone nor an interaction effect between stop type and group, suggesting that the two groups are not significantly different from each other in their VOT production.

Interestingly, here older speakers do not show the large variation observed for their F0 values, even though the number of tokens remains the same. No significant effect or visual difference was observed between genders for VOT, barring a slightly longer median VOT for male speakers in aspirated and fortis stops (in both groups), with a slightly longer median VOT for female speakers in lenis stops, although this difference is not seen to be significant (Table 6).

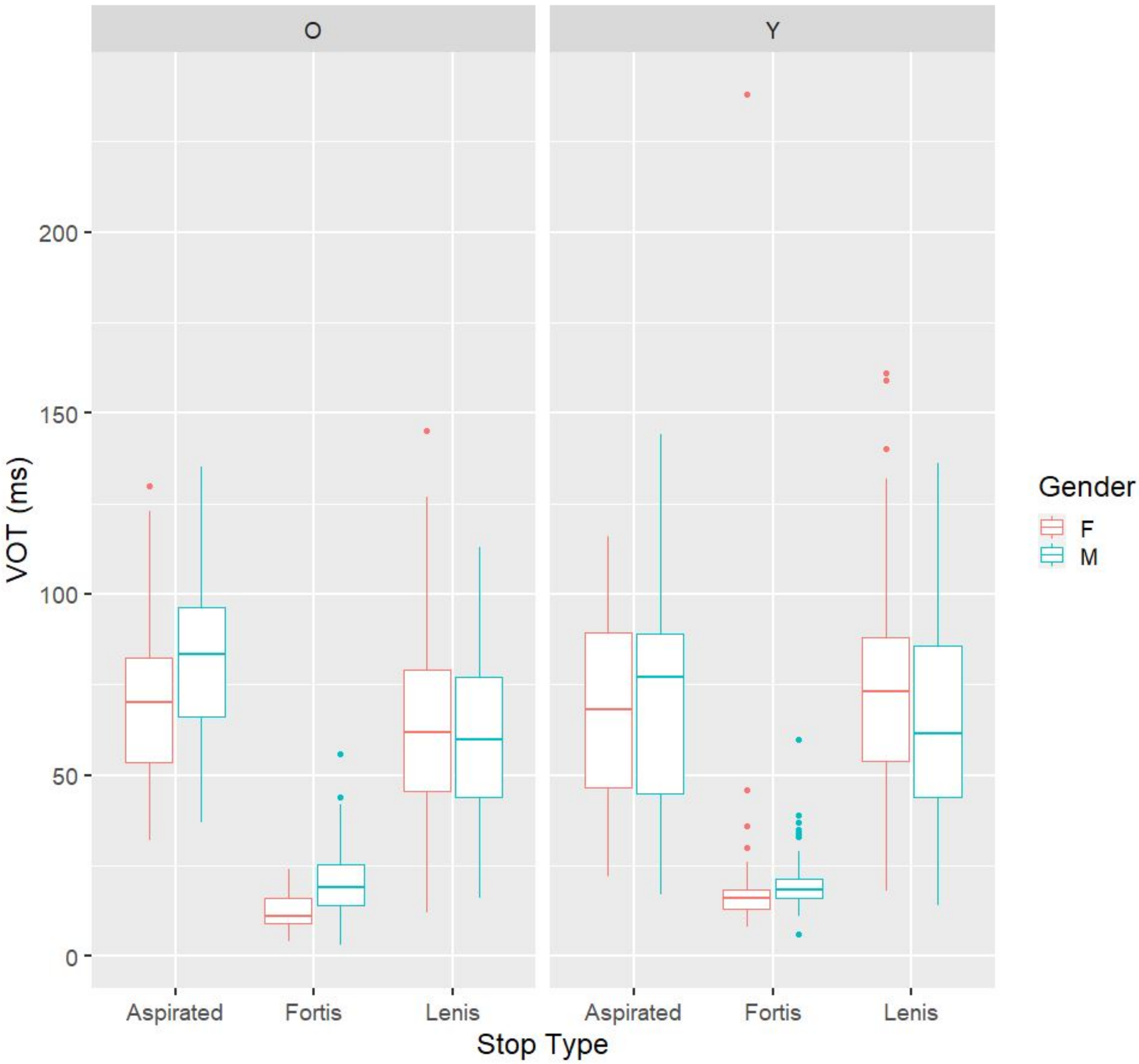

Figure 3: Comparison of VOT values (ms) by Gender and Group (Older vs Younger)

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-3.0059	1.8141	22.2872	-1.657	0.111534
TypeStop+aspirated-fortis-lenis	31.5785	2.3093	916.4078	13.674	<.001***
TypeStop	30.9447	1.5341	921.9754	20.172	<.001***
Group-O+Y	3.2887	3.5284	19.9315	0.932	0.362462
Gender-F+M	-2.1999	3.3366	16.0158	-0.659	0.519052
Sposition+initial-internal	6.0389	1.7636	921.6777	3.424	<.001***
TypeStop+aspirated-fortis-lenis: Group-O+Y	-11.7648	4.6289	916.8529	-2.542	0.11198*
TypeStop:Group-O+Y	0.4025	3.0155	920.6525	0.133	0.893856

Table 6: Fixed and Interaction Effects for VOT: $lmer(VOT.c \sim TypeStop * Group + Gender + Sposition + (1|Participant), data = data)$

That the position of the word is significant (with initial stops showing on average 6ms longer VOT than internal) makes sense for VOT, given various factors such as preceding words, speech rate, assimilation and so on possibly leading to a decrease in VOT for sentence medial words.

Aspirated stops are on average significantly longer for this sample, with an average of 31.6ms longer in their VOT than other stop types. This fits in with previous research that states that aspirated stops have the longest VOT (cf. Table 2).

Neither group or gender was seen to have a significant effect. This suggest that overall there is no significant difference between male and female participants in their VOT lengths, nor is there a significant difference between groups.

Most importantly here is the fact that there is significant overlap between lenis and aspirated stops for both groups: that is to say, the graphical results suggest that lenis and aspirated stops are no longer differentiated solely by VOT (cf. table 2). This was expected of all participants in both groups, as all were born after 1965 and all are expected to show some degree of VOT redundancy across the lenis-aspirated stop distinction (cf. Table 1).

However, this is especially true for the younger group where the range of the lenis-aspirated stop distinction is virtually the same, whereas in the older group male speakers do show an interquartile range higher than female speakers for aspirated stops. This is supported by a significant interaction effect between aspirated stops and group, with younger speakers on average having 11ms shorter aspirated stop VOT than older speakers. This is not immediately clear on the initial graphical results, and the effect size is also small, suggesting that this is not as strong an indication as was seen for F0. Nevertheless, the fact that a significant interaction effect exists (however small), gives support to the theory that aspirated stops are normalising (i.e. shortening) with lenis stops in their VOT in younger speakers, while remaining slightly more distinct in VOT across the two stop types in older speakers. This suggests that younger speakers are demonstrating an increased redundancy in the use of VOT as a distinguishing cue for lenis and aspirated stops compared to older speakers.

The above results support the hypothesis that neither group is using VOT as a primary cue due to being born after 1965 is correct, with both groups showing a redundancy in VOT as an acoustic cue. However, this appears to have progressed further with younger speakers than with older, due to the increased overlap between lenis and aspirated VOT in younger speakers.

4.3 Harmonics (H1 - H2)

As shown in Figure 4, the female participants in both groups show overall breathier stops (indicated by a higher db) than males. This is supported by a statistically significant effect as seen in Table 7, with female participants demonstrating on average a 3.9 db increase compared to male participants. Overall, aspirated stops are seen to be highly significant, with on average 2.1 db higher than other stop types for these participants. This suggests a strong breathy voice compared to the other stops.

However, it is also observed that older female speakers demonstrated an even lower db than their older male counterparts for fortis stops, suggesting a creakier voice for fortis stops in older women more so than any other age group and gender.

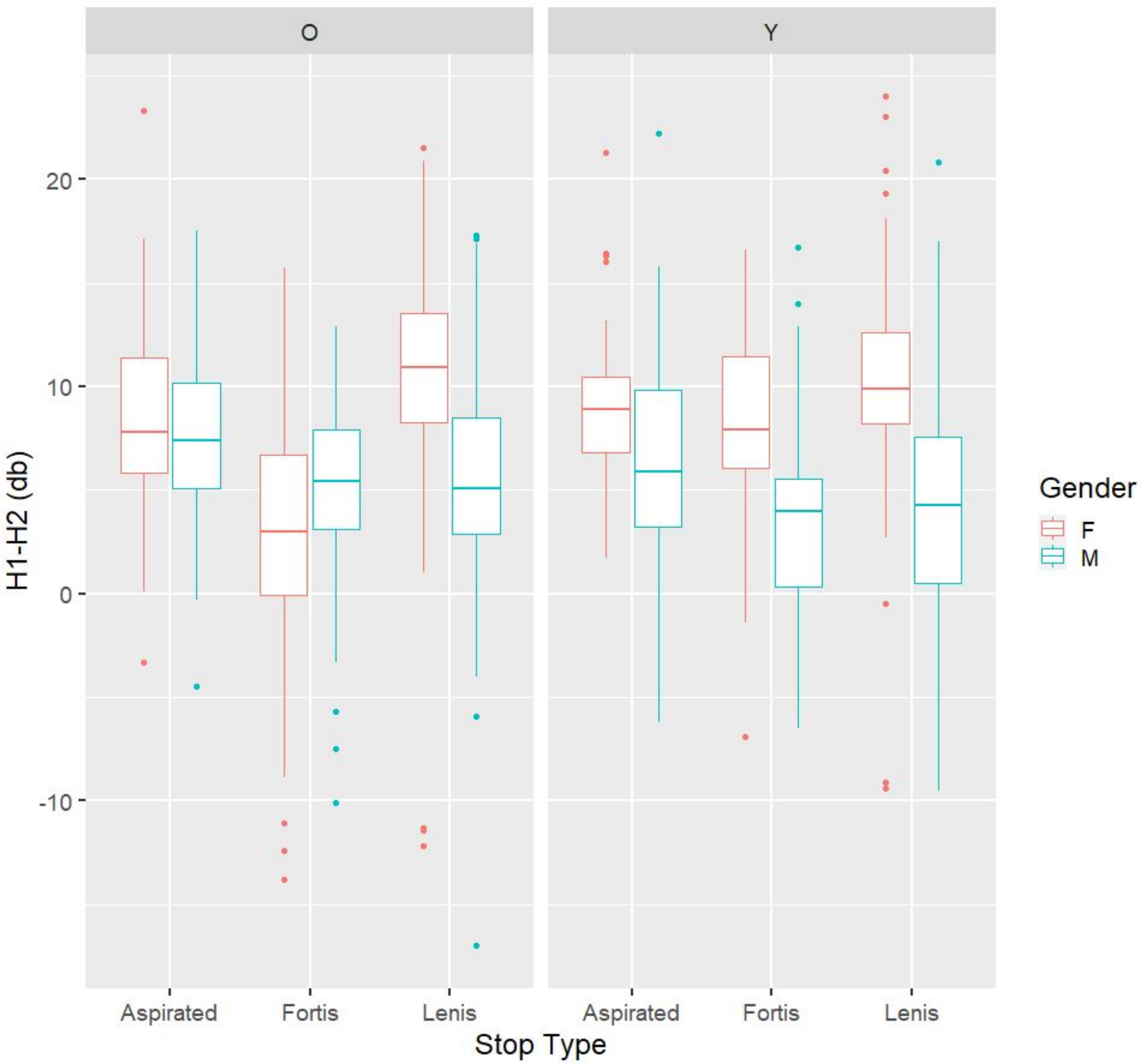

Figure 4: Comparison of H1-H2 values (db) by Gender and Group (Older vs Younger)

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-0.01303	0.50590	18.81417	-0.026	0.979719
TypeStop+aspirated-fortis-lenis	2.10737	0.45802	915.02556	4.601	<.001***
TypeStop	1.92474	0.29964	916.99158	6.423	<.001***
Group-O+Y	0.41428	1.01172	18.80164	0.409	0.686815
Gender-F+M	-3.88674	0.98537	16.93567	-3.944	0.001053 **
TypeStop+aspirated-fortis-lenis: Group-O+Y	-0.23699	0.91605	915.01689	-0.259	0.795918
TypeStop:Group-O+Y	-2.10132	0.59928	917.03594	-3.506	<.001***

Table 7: Fixed and Interaction Effects for Harmonics, H1-H2 (db): $lmer(Harmonic.c \sim TypeStop * Group + Gender + Sposition + (1|Participant), data = data)$

For male speakers, both groups appear to follow roughly the same pattern, with both showing the highest (breathiest) values for aspirated, followed by fortis and lenis, the latter two being extremely similar in their median value for both groups. Interestingly, younger male speakers show an increased distribution in their minimum and maximum values, suggesting that creaky voice is more common in younger males than in older males as the db approaches 0 and/or becomes negative more frequently. All three stop types in younger males show values below 0db, entailing that all stop types show some level of creaky voice, the only group to do so. This likely is also the cause for the significant interaction effect between stop type and group, with older speakers having on average 2.1 db higher amplitude than younger speakers, likely due to the fact that younger male speakers frequently went below 0db in their H1-H2 values compared to their female counterparts as well as the older speakers.

The female participants do not match their male counterparts in terms of H1-H2. Younger female speakers both have their breathiest stop type as lenis, followed by aspirated, and finally fortis (going by the median value). The distribution between the three seems relatively small, and only fortis stops cross the 0 db threshold, implying that creaky voice is only observed in this stop type whereas the other two stop types are breathy. Older female speakers also show the same patterns as younger female speakers (lenis > aspirated > fortis) but the distribution is overall larger, particularly for fortis stops, which have a much lower median value and a Q1 that sits at 0 db, indicating a greater presence of creaky voice in older female speakers compared to younger female speakers.

However, for both groups and genders there is considerable overlap between the H1-H2 values and the stop types, as can be additionally seen in Appendix E (figure 7). The reason for younger males having a strong creaky voice presence in all stop types is unknown, and requires additional investigation.

4.4 Results by Individual Stop Type

This section will evaluate each stop type individually, to further increase understanding of the more general results observed in the overall models given above and evaluate how the stop types differ individually.

4.4.1 Aspirated

Turning firstly to F0, it is unsurprising that a significant effect of gender is found, with female participants on average 135.2 Hz higher than male participants, due to the previously mentioned biological differences, and again this does not therefore entail that women are further progressed in the shift towards F0 preference as a distinguishing cue than men.

Aspirated stops also show a significant effect for group, with older speakers having on average 22.5 Hz higher F0, although the effect size is small (table 8). The large amount of variation seen for older female speakers (figure 2) seems to be supported by this outcome and is possibly the cause for this significant effect.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	66.633	4.077	16.926	16.416	<.001***
Group-O+Y	-22.549	8.160	17.132	-2.763	0.0132 *
Gender-F+M	-135.261	8.057	15.397	-16.788	<.001***

Table 8: Significant effects for F0 in aspirated stops only: $lmer(F0.c \sim Group + Gender + (1|Participant), data = aspiratedStop)$

As for VOT and H1-H2, no significant effects were observed for any variables (see summary tables of statistics in Appendix D). This entails that there were no significant differences between group or gender for either VOT or for H1-H2.

4.4.2 Fortis

Regarding F0, only gender demonstrated a significant effect, with female speakers having on average 115 Hz higher F0 than male speakers:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	45.724	4.867	16.103	9.395	<.001***
Group-O+Y	-6.241	9.721	15.947	-0.642	0.53
Gender-F+M	-115.878	9.724	15.962	-11.917	<.001***

Table 9: Significant effects for F0 in fortis stops only: $lmer(F0.c \sim Group + Gender + (1|Participant), data = fortisStop)$

Again, this result is unsurprising, given the biological differences between male and female speakers. Similarly to aspirated stops, no effect was seen for VOT or H1-H2 for any variable, again suggesting no significant difference between either gender or group for fortis stops for these two predictors (see summary tables of non-significant statistics in Appendix D).

4.4.3 Lenis

In terms of F0 for lenis stops, like fortis stops only a significant effect was found for gender, with female speakers on average 84.6 Hz higher than male speakers:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.8529	2.3560	17.1134	3.758	<.001***
Group-O+Y	0.6272	4.7118	17.1066	0.133	0.89565
Gender-F+M	-84.6385	4.7120	17.1125	-17.962	<.001***

Table 10: Significant effects for F0 in lenis stops only: $lmer(F0.c \sim Group + Gender + (1|Participant), data = lenisStop)$

In terms of VOT, an effect of sentence position was found for this stop type. Lenis stops in initial position were on average 7.1 ms longer than sentence medially positioned words:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.765	1.792	17.569	4.890	<.001***
Group-O+Y	7.328	3.557	16.973	2.060	0.055033
Gender-F+M	-5.825	3.559	17.064	-1.637	0.120026
Sposition+initial-internal	7.114	2.158	606.431	3.296	0.001037 **

Table 11: Significant effects for VOT in lenis stops only: $lmer(VOT.c \sim Group + Gender + Sposition + (1|Participant), data = lenisStop)$

Finally, H1-H2 showed a significant effect for gender, with female speakers having on average 5.3db higher than male speakers, implying a breathier vocal quality for this stop type than male speakers:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	0.7532	0.5523	17.3059	1.364	0.190160
Group-O+Y	-1.0722	1.1045	17.2939	-0.971	0.345064
Gender-F+M	-5.2853	1.1046	17.3042	-4.785	<.001***

Table 12: Significant effects for H1-H2 in lenis stops only: $lmer(Harmonic.c \sim Group + Gender + (1|Participant), data = lenisStop)$

The reason for lenis stops alone showing a significant effect for H1-H2 is not clear from this study alone, and requires possible further investigation and comparison to assess why this is the case. That VOT here is significant for position is difficult to interpret, as interestingly aspirated was not significant (despite previously to have a very long VOT, table 2). This possibly may be linked again to VOT normalisation, with lenis potentially also compensating marginally for the increase in F0 values for aspirated stops, which is more clearly seen in sentence-initial position. However, this requires a more comprehensive study to confirm this possibility, and cannot be interpreted through these results alone.

4.5 Overall Summary of Results

The overall results suggest that there is a shift in aspirated stops that results in a much higher aspirated F0 and a lower VOT value than previously reported and expected (tables 1 and 2). That is to say, aspirated stops have now overtaken fortis stops as having the highest F0 values, and as a compensatory measure may have also reduced in their VOT, resulting in an overlap with lenis stops.

This appears to be true for both groups, but especially so for younger speakers due to a significantly lower VOT for aspirated stops compared with older speakers. It was hypothesised that neither group uses VOT as a primary cue, although it is not likely to be redundant, and the results seem to support this. The results also suggest that lenis stops are more even across both groups in their distribution, while aspirated stops, particularly for older speakers, show a much larger range and therefore are possibly less stable as a primary distinguishing cue in the older group compared to younger speakers. The fact that lenis stops appear to be more stable also gives support to the previous statement that it is the aspirated stops undergoing a change, rather than other stop types.

A summary of the most noticeable outcomes between the two groups is given below:

	Group 1 (Younger)	Group 2 (Older)
F0	Stable for all stop types. Consistent across speakers. Aspirated > Fortis > Lenis	Unstable in aspirated and fortis stops. Does not appear to be consistent across speakers. Aspirated > Fortis > Lenis
VOT	Overlap between lenis and aspirated stops. Significant reduction in aspirated stop lengths Continuation in VOT redundancy: F0 becoming primary cue (progression of tonogenesis)	Overlap between lenis and aspirated stops. Significantly longer aspirated stops. Some degree of VOT redundancy: Suggests less progression of tonogenesis.
H1-H2	Younger males showing creaky voice in all stop types. Little creaky voice in younger females.	Creaky voice in older female fortis stops. Creaky voice in older male lenis and fortis stops.

Table 13: Summary comparison of results for older and younger speakers regarding F0, VOT, and H1-H2.

That fortis now demonstrates a lower F0 than aspirated stops may be due to a compensatory increase in F0 for aspirated stops, to better create a distinction between lenis and aspirated stops. This may also explain the 'instability' of F0 in aspirated stops compared to lenis stops in older speakers, as the shift towards a higher F0 in aspirated stops as a compensatory measure for a lower VOT may not yet complete in this age group. If this is correct, then it would also imply that lenis does not need to alter its current values, thus remaining relatively stable. This is further supported by the fact that fortis, due to its markedly short VOT, is already salient and thus needs no change in F0 to distinguish it from the other two stop types. This suggests then that aspirated stops alone may be changing, as in both groups lenis stops are roughly comparable for F0 and VOT.

Regarding the individual stop analyses, only aspirated stops showed an effect for group across all stop types in F0. This coupled with the fact that no significant effect was seen for VOT and group suggests again supports the above overall results that suggest aspirated stops are increasing in their F0 value further in younger speakers to compensate for increased redundancy in VOT.

Finally, the harmonic results suggest that speakers are roughly comparable in voice quality, although older female speakers demonstrate a creakier voice in their fortis stops compared to other participants (both younger and older males) in this study. Fortis stops used to use creakiness as a distinguishing measure from lenis stops (c.f. section 2), and it is therefore possible that some of the older female speakers in this study maintained the creaky voice quality, potentially from generational inheritance, although this requires further investigation. Why younger male speakers show creakier voice across all stop types is not clear from the results given here, and so a full investigation of voice quality across stop types would be beneficial to evaluate why this may be occurring.

The results therefore support the hypothesis that both groups are showing F0 as their main acoustic cue preference when differentiating between lenis and aspirated stops, as well as lending support to the extended hypothesis that younger speakers are further along in VOT redundancy for aspirated and lenis stops than older speakers, i.e. that aspirated and lenis stops are merging further in their VOT lengths in younger speakers, with the result of aspirated stops compensating for this normalisation by increasing in F0.

Therefore the null hypothesis that both groups would show a similar (non-significant) value and similar graphical patterning for the F0 and VOT differences (i.e, a stabilisation of tonogenesis) can be rejected and the hypothesis here that younger speakers are further along in VOT redundancy and increase in F0 reliance as a distinguishing cue (progression in tonogenesis) is supported.

5 Discussion

5.1 F0 and VOT

As mentioned in section 2, previous research had some problematic aspects, especially given the use of Korean diaspora. Nevertheless, the results observed here match the overall results of past research, in that speakers born after 1965 all seem to show a clear preference for F0 rather than VOT when contrasting lenis and aspirated stops. This was clearly seen for both age groups in the results, with VOT showing considerable overlap between lenis and aspirated stop types (Figure 3), and F0 being more distinct (Figure 2). This therefore supports previous research despite the possible problematic features of their participant selection.

That aspirated is now higher than fortis for F0 in both groups on average, (an unexpected result, cf. table 1), suggests that F0 may be increasing as a compensatory measure in aspirated stops to better distinguish between aspirated and lenis stops due to the increase in VOT redundancy, while fortis remains stable in F0 due to an already markedly short VOT.

The research question of this paper asked if F0 preference had progressed in younger speakers when compared to older speakers, and what the difference between the two groups is should progression be observed. The hypothesis was that younger speakers show a progression in their use of F0, implying a progression in the process of tonogenesis compared to older speakers.

The results here confirm the above hypothesis: although both groups are using F0 as a primary cue, younger speakers show less variation between their F0 values for both genders (figure 2) - i.e., there is a suggestion that their F0 values are stable and roughly consistent across speakers for the lenis-aspirated stop distinction. In contrast, older speakers show more instability in their F0 values, particularly for their aspirated stop values. This suggests that F0 is not yet fully 'stable' and not always used consistently as the primary contrasting cue for distinguishing between lenis and aspirated stops in older speakers. The statistical results for F0 (table 5), suggest a significant interaction effect between aspirated stops and group: older speakers show significantly higher aspirated stop values than younger speakers (by 21.9 Hz on average), matching the instability and wide variation in older speakers as observed on the graphical results in Figure 2. This suggests that aspirated stops are currently unstable in older speakers, and possibly that the compensatory F0 increase for the reduction in VOT of aspirated stops is either not complete (or present) in some older speakers, or that older speakers are less consistent in using it as a contrasting cue.

Further comparison with VOT (figure 3) also supports the above, given that VOT is virtually identical for younger speakers but less exact for older speakers; a small interaction effect was found here (table 6) with younger speakers having a slightly shorter aspirated VOT length (-11.7ms) compared to older speakers. This further suggests that aspirated stops are increasing in VOT redundancy, and as a consequence shortening and normalising with lenis stops (i.e. overlapping) in their VOT lengths.

This study therefore supports the hypothesis that younger speakers have progressed in their use of F0 as a primary cue compared to older speakers, consequently suggesting further progression in the tonogenesis process (Figure 1) compared to older speakers. This is due to the fact that the F0 cue, for this sample of speakers, appears to have resulted in a much higher F0 value and a decrease in VOT length for aspirated stops, which is most stable and consistent in younger speakers. The statistical results support this, with a significant decrease in aspirated VOT lengths in younger speakers suggesting a normalisation of the VOT contrast in lenis and aspirated stops, resulting in further overlap in VOT lengths between the two stop types. The exact stage of tonogenesis is difficult to place, given the complex nature of the phenomena (cf. Kang, 2014), but this study provides evidence that the process has progressed within the younger generation.

5.2 Harmonics

As mentioned in section 2.1, speakers born before 1965 seemed to rely on creaky voice coupled with VOT to distinguish between lenis and fortis stops, due to frequent overlap between fortis and lenis stops resulting in VOT alone not being enough to contrast the two (Abberton, 1972; Hirose et al. 1974; Lisker & Abramson, 1964). This is relevant to the study here, as it is observed in the results that older female speakers still seem to show a creakier voice quality in their fortis stops compared to their lenis stops, noticeably more so than younger female speakers. This is possibly due to the female speakers of this particular sample being influenced by their parent‘s generation and thus inheriting the creaky voice distinction between lenis and fortis stops. However, it is unclear why older male speakers do not show the same pattern, as their lenis and fortis stops are very similar in H1-H2 values both in distribution as well as median (Figure 4).

What is more surprising perhaps, is that of the results of younger male speakers. Younger male speakers show a large distribution in their H1-H2 values across all stop types, but interestingly are the only set of participants to show clear creaky voice values across all stop types, dipping far below 0 db in even aspirated and lenis stops. This is the opposite of female speakers in their age group, who only cross 0 db in their fortis values, and then only at the more extreme ends of their distribution. Older male speakers show much more conservative values, with a more even distribution and more centred medians (Figure 4).

The hypothesis that fortis stops retain some amount of creaky voice quality appears to be supported in these results, with both groups and genders showing values at or below 0 db for fortis stops. However, the results reject the hypothesis that creaky voice is more extreme in older speakers across the lenis-fortis distinction due to a remnant of a previous acoustic feature: this cannot be accepted due to the values observed in younger male speakers.

It is not clear from the current experiment why creakier voice is observable in younger males across all stop types compared to the other groups. It does not seem like a generational effect, as could be interpreted in the older female case, as neither older males nor the younger female participants show a similar result. Due to this, further research is needed to analyse the voice quality of the stop types in younger speakers, to assess if this is simply due to individual differences or if voice is emerging to play a role as a contrasting cue.

5.3 Future Research

This study proposes two main directions for future research: firstly, the unclear results observed in harmonics warrant further attention, to assess whether or not voicing has an impact on contrasting cues or if this was simply a feature of the participants in this study. The fact that this was especially prominent across all stop types in younger men is interesting: there is no reason to suggest that this is generational (as could be the case for older female speakers), especially given that this phenomena is not seen in older males speakers or younger female speakers.

The other avenue of research would be to investigate other varieties of Korean, to see if this has progressed outside of the Seoul area. If the shift in preference is seen in other parts of Korea, comparing with SSK would provide insights into possible progression of the shift towards F0 preference, both for SSK as well as other areas (e.g., if the SSK is further progressed than a regional variant then it may indicate the direction that the regional variant is going towards). It may also indicate instead further divergence from the standard, if the shift is not accepted or manifests differently in other regions of Korea. If observed, this could be further contrasted with perception experiments, to see if the acoustic cue preferences are internalised differently in different areas of Korea despite a similar production.

6 Conclusion

This study has further given support to the idea that within the post-1965 age group all speakers show a preference for using F0 as an acoustic cue, especially for the lenis-aspirated stop distinction. Both groups show a strong overlap in their VOT values regarding the lenis-aspirated stop distinction, suggesting that redundancy of VOT is occurring for both age groups. However, this is especially prominent in younger speakers, who show a robust and stable range of F0 values, whereas older speakers show more irregularities in their F0 use, especially for aspirated stops. This suggests that in younger speakers, the shift towards F0 reliance is more complete and consistent than in older speakers.

This was further supported through statistical analyses: younger speakers use shorter VOT in their aspirated stops, resulting in further overlap between lenis and aspirated stops values and implying further VOT redundancy. That older speakers demonstrated a higher F0 is likely due to the wide variation in older speakers and not so much that older speakers use F0 more than younger speakers.

Fortis stops (previously the highest F0 value stop type) have now been significantly overtaken by aspirated stops, suggesting an increase in F0 as a compensatory measure for VOT redundancy. This research cannot give an exact tonogenesis stage, as others have suggested this may be non-linear (e.g. Kang, 2014), but it is likely given the results observed here that younger speakers are further progressed than older speakers. Therefore all hypotheses are supported, except for the statement on creaky voice - while older females showed creakier voice in fortis stops, younger males showed a large amount of creaky voice across all stop types. This was not observed in older males nor younger females, and so the matter requires further investigation.

Appendix

Appendix A: Further Details on Korean Stops

Lenis		Fortis		Aspirated	
방 /paŋ/	"room"	빵 /p*aŋ/	"bread"	팡 /pʰaŋ/	"bang"
달 /taɭ/	"moon"	딸 /t*aɭ/ or /d*al/ (intermediary)	"daughter"	탈 /tʰaɭ/	"mask"
자다 /t͡ɕada/	"to sleep"	짜다 /t͡ɕ*ada/	"to squeeze"	차다 /t͡ɕʰada/	"to kick"
개다 /kɛda/	"to fold"	깨다 /k*ɛda/	"to break"	캐다 /kʰɛda/	"to dig up / delve"
사다 /sada/	"to buy"	싸다 /s*ada/	"to wrap"	-	-

Table 14: Adapted from Cho et al. 2001, p.194. "*Minimal Contrasts for Korean Obstruents in word-initial position*"

Appendix B: Full Participant Details from the Seoul Corpus

Participant	Group	Age	Gender
01	1	16	Male
02	1	16	Male
03	1	15	Male
04	1	15	Male
05	1	16	Male
06	1	18	Female
07	1	16	Female
08	1	16	Female
09	1	17	Female
10	1	18	Female
31	2	43	Male
32	2	43	Male
33	2	44	Male
34	2	47	Male
35	2	43	Male
36	2	43	Female
37	2	46	Female
38	2	46	Female
39	2	43	Female
40	2	43	Female

Table 15: Details of the participants used in production analysis (with original participant numbers from Yun et al. 2015)

Speaker	**Age**	**Gender**	**Age Group**	**Speaker**	**Age**	**Gender**	**Age Group**
s01	16	M	10-19	s21	31	M	30-39
s02	16	M	10-19	s22	37	M	30-39
s03	15	M	10-19	s23	36	M	30-39
s04	15	M	10-19	s24	36	M	30-39
s05	16	M	10-19	s25	32	M	30-39
s06	18	F	10-19	s26	32	F	30-39
s07	16	F	10-19	s27	32	F	30-39
s08	16	F	10-19	s28	34	F	30-39
s09	17	F	10-19	s29	37	F	30-39
s10	18	F	10-19	s30	38	F	30-39
s11	25	M	20-29	s31	43	M	40-49
s12	23	M	20-29	s32	43	M	40-49
s13	26	M	20-29	s33	44	M	40-49
s14	23	M	20-29	s34	47	M	40-49
s15	22	M	20-29	s35	43	M	40-49
s16	22	F	20-29	s36	43	F	40-49
s17	24	F	20-29	s37	46	F	40-49
s18	27	F	20-29	s38	46	F	40-49
s19	24	F	20-29	s39	43	F	40-49
s20	24	F	20-29	s40	43	F	40-49

Table 16: Details of all participants in the original corpus (Yun et al. 2015)

Appendix C: Formulae Used for Statistical Analysis

A. Aspirated stops:

F0:

$$lmer(F0.c \sim Group + Gender + (1|Participant), data = aspiratedStop)$$

VOT:

$$lmer(VOT.c \sim Group + Gender + Sposition + (1|Participant), data = aspiratedStop)$$

H1-H2:

$$lmer(Harmonic.c \sim Group + Gender(1|Participant), data = aspiratedStop)$$

B. Fortis stops:

F0:

$$lmer(F0.c \sim Group + Gender + (1|Participant), data = fortisStop)$$

VOT:

$$lmer(VOT.c \sim Group + Gender + Sposition + (1|Participant), data = fortisStop)$$

H1-H2:

$$lmer(Harmonic.c \sim Group + Gender(1|Participant), data = fortisStop)$$

C. Lenis stops:

F0:

$$lmer(F0.c \sim Group + Gender + (1|Participant), data = lenisStop)$$

VOT:

$$lmer(VOT.c \sim Group + Gender + Sposition + (1|Participant), data = lenisStop)$$

H1-H2:

$$lmer(Harmonic.c \sim Group + Gender(1|Participant), data = lenisStop)$$

Appendix D: Table Results for non-significant outcomes in Statistical Analysis

A. Significant Effects for VOT in aspirated stops:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	17.480	2.860	18.900	6.112	<.001***
Group-O+Y	-6.038	4.808	9.881	-1.256	0.238
Gender-F+M	8.704	4.789	9.016	1.817	0.102
Sposition+initial-internal	4.434	5.448	141.852	0.814	0.417

Table 17: Significant effects for VOT in aspirated stops only: $lmer(VOT.c \sim Group + Gender + Sposition + (1|Participant), data = aspiratedStop)$

B. Significant Effects for H1-H2 in aspirated stops:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	1.07046	0.49151	5.49031	2.178	0.0765
Group-O+Y	0.05917	0.98327	5.53257	0.060	0.9541
Gender-F+M	-1.48930	0.97449	5.10862	-1.528	0.1858

Table 18: Significant effects for H1-H2 in aspirated stops only: $lmer(Harmonic.c \sim Group+Gender+(1|Participant), data = aspiratedStop)$

C. Significant Effects for VOT in fortis stops:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-36.5727	2.8036	15.7862	-13.045	<.001***
Group-O+Y	6.9501	5.1449	11.2555	1.351	0.203
Gender-F+M	1.2118	5.1381	11.1930	0.236	0.818
Sposition+initial-internal	0.9845	3.4878	158.8519	0.282	0.778

Table 19: Significant effects for VOT in fortis stops only: $lmer(VOT.c \sim Group+Gender+Sposition+(1|Participant), data = fortisStop)$

D. Significant Effects for H1-H2 in fortis stops:

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	-2.4419	0.5915	18.0638	-4.129	<.001***
Group-O+Y	1.6800	1.1801	17.7299	1.424	0.171926
Gender-F+M	-1.3034	1.1805	17.7219	-1.104	0.284299

Table 20: Significant effects for H1-H2 in fortis stops only: $lmer(Harmonic.c \sim Group + Gender + (1|Participant), data = fortisStop)$

Appendix E: Further Graphical Results

Scattergraph of F0 (Hz) by group and gender:

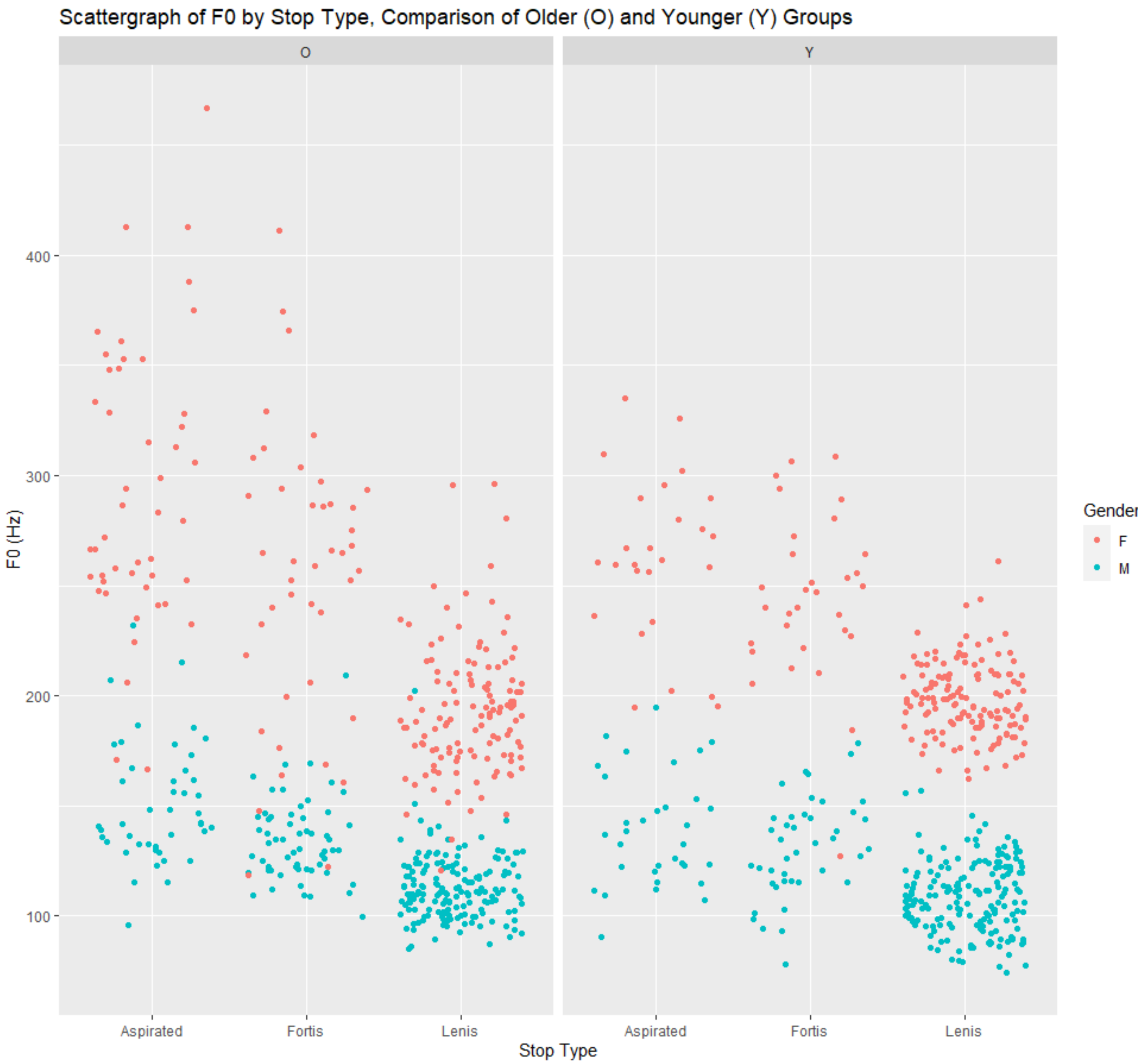

Figure 5: Scattergraph of F0 values (Hz) by Group (Older vs Younger) and Gender

Scattergraph of VOT (ms) by group and gender:

Figure 6: Scattergraph of VOT values (ms) by Group (Older vs Younger) and Gender

Scattergraph of H1-H2 (db) by group and gender:

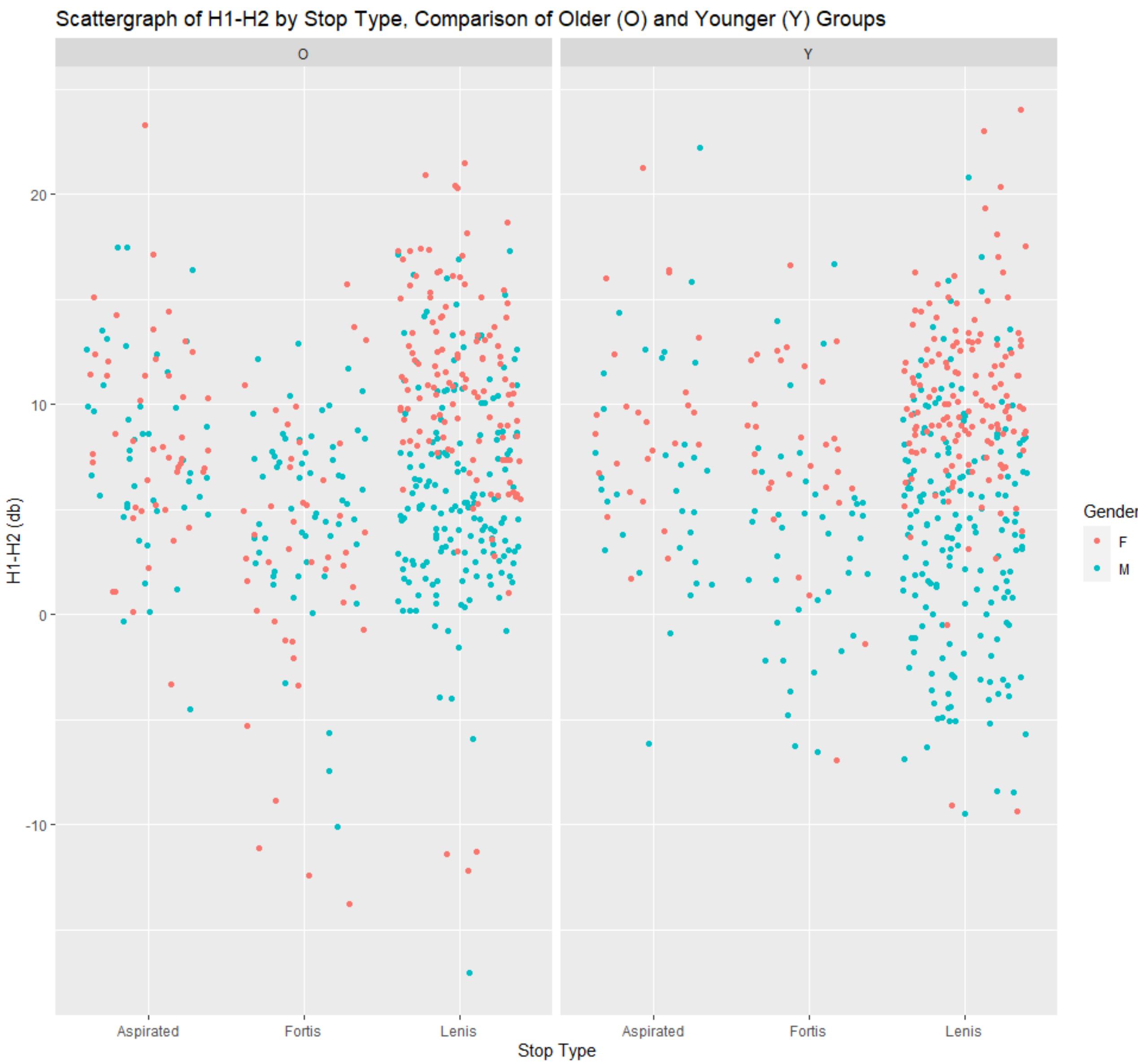

Figure 7: Scattergraph of H1-H2 values (db) by Group (Older vs Younger) and Gender

Overall Results

Overall Results of F0 by group:

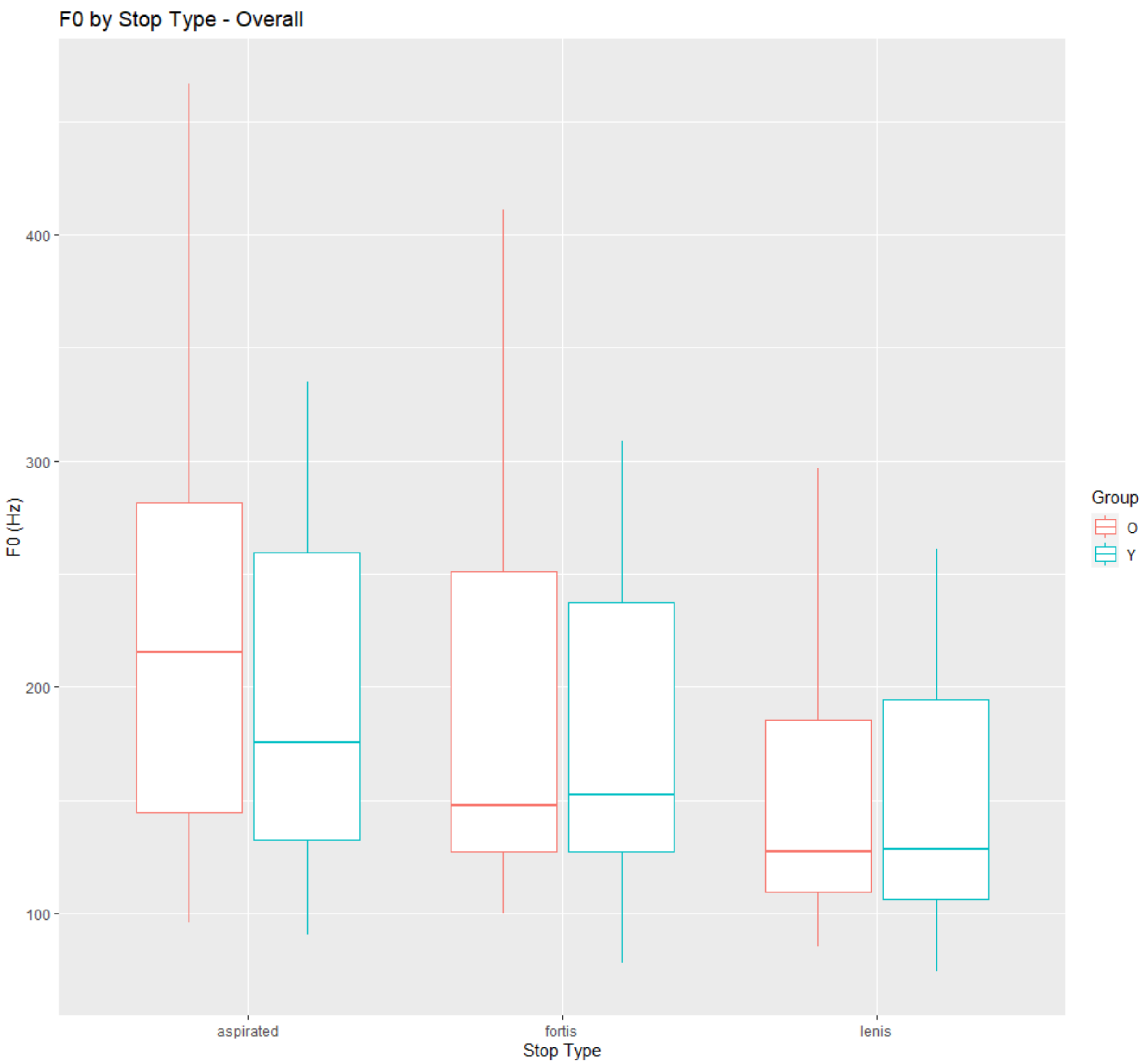

Figure 8: Comparison F0 values (Hz) by Group (Older vs Younger)

Overall Results of VOT by group:

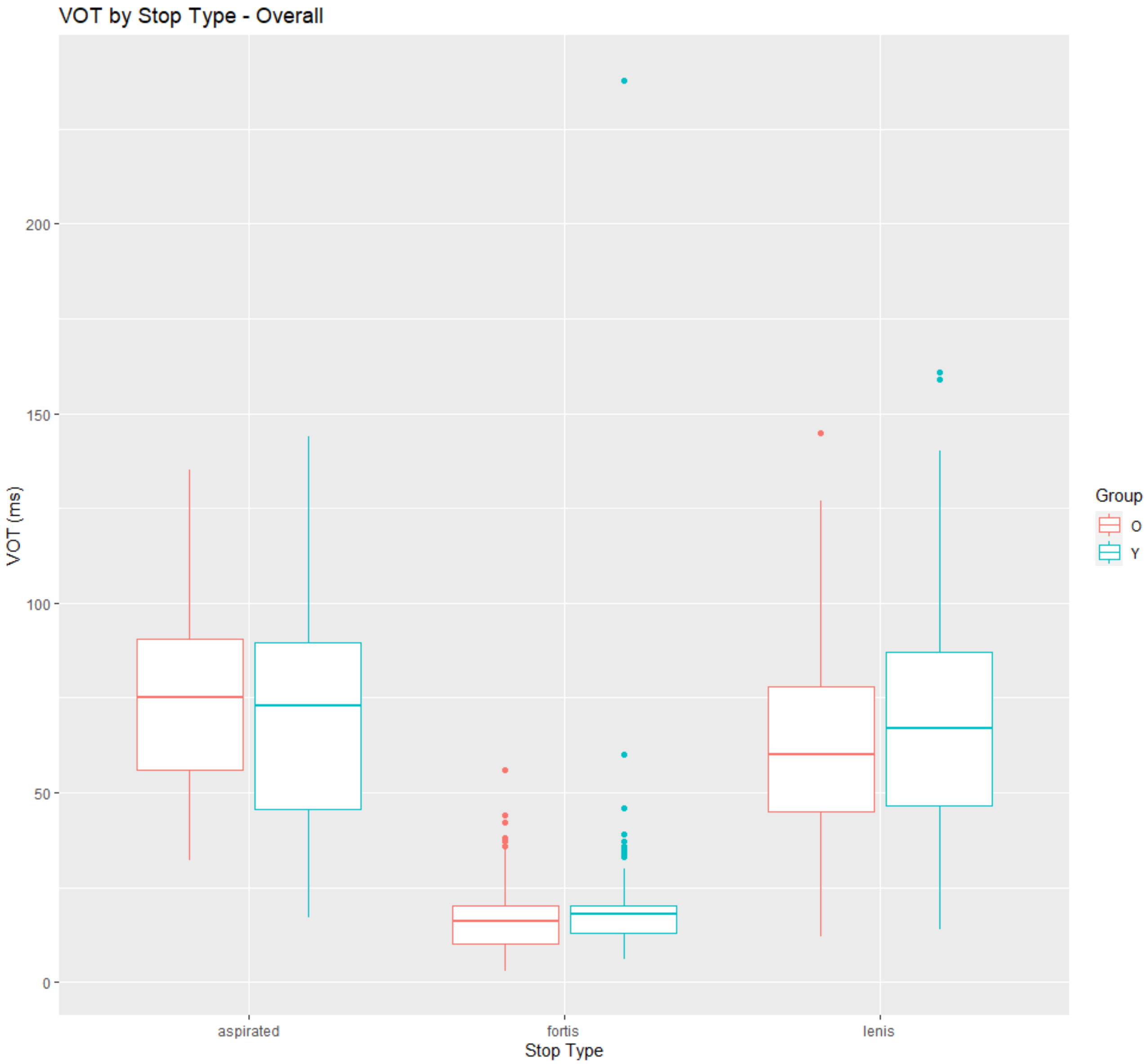

Figure 9: Comparison VOT values (ms) by Group (Older vs Younger)

Overall Results of H1-H2 by group:

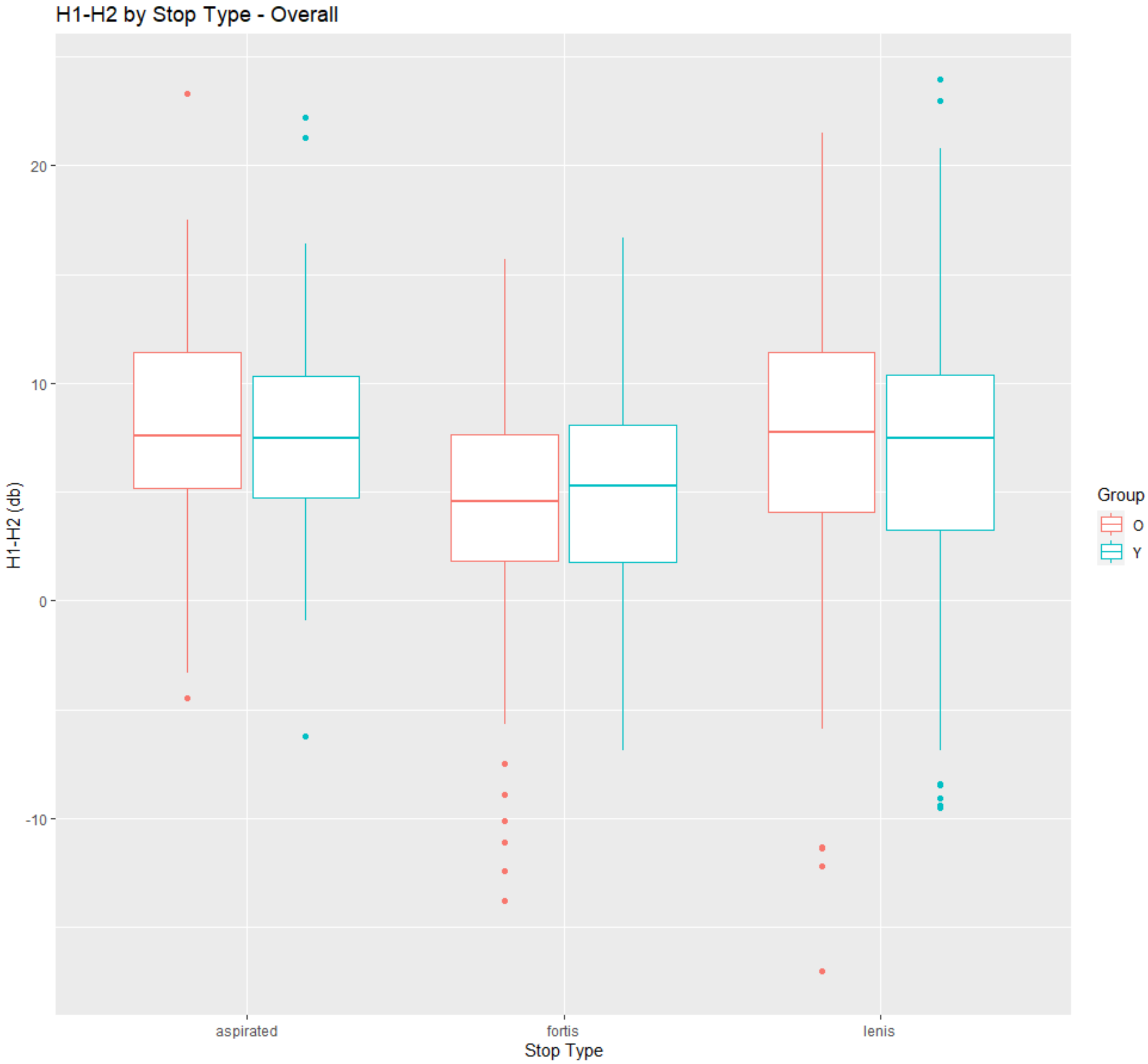

Figure 10: Comparison H1-H2 values (db) by Group (Older vs Younger)

Bibliography

Abberton, E. (1972) “Some laryngographic data for Korean stops", *Journal of the International Phonetic Association, 2*, pp.67-78

Bang, H-Y., Sonderegger, M., Kang, Y., Clayards, M., Yoon, T-J. (2017) “The emergence, progress and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis". *Journal of Phonetics* 66, pp.120-144

Boersma P. & Weenink, P. (2022). *PRAAT: Doing Phonetics by Computer* (computer program). URL: <http://www.Praat.org>. Last accessed 08/04/2022

Cho, T., Jun, S-A., Ladefoged, P. (2002). “Acoustic and aerodynamic correlates of Korean stops and fricatives“. *Phonetics* 30, pp.193-228

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, pp. 1-29

Hardcastle, W.J., (1973). “Some observations on the tense-lax distinction in initial stops in Korean". *Journal of Phonetics 1*, 264-272

Han, M., & Weitzman, R. (1970). “Acoustic Features of Korean /P,T,K/, /p,t,k/ and /ph,th,kh/". *Phonetica, 22*, pp.112-128.

Hirose, H., Lee, C.Y., & Ushijima, T. (1974). “Laryngeal control in Korean stop production", *Journal of Phonetics 2*, pp. 145-152

Kang, K-H., & Guion, S.G. (2008). “Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies" *Acoustical Society of America 6*(124) pp.3903-3917

Kang, K-H. & Guion, S.G. (2006), “Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals". *Acoustical Society of America* 119, pp. 1672-1683

Kang, Y. (2014). “Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study" *Journal of Phoetics, 45*, 76-90

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

Ladefoged, P. & Johnson, K., (2014). *A Course in Phonetics: Seventh Edition.* Cengage Learning: Boston

Ladefoged, P. (2003). *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*, Wiley-Blackwell: Oxford

Lindblom, B., (1990). *Explaining phonetic variation: A sketch of the H and H theory* in Speech Production and Speech Modelling eds. W.J. Hardcastle and A. Marchal, pp. 403-439. Kluwer: Dordrecht

Lisker, L., & Abramson, A.S., (1964). “A cross-language study of voicing in initial stops: Acoustical measurements." *Word 20*(3), 384-422

Maran, L.R. (1973). On becoming a tone language: a Tibeto-Burman model of tonogenesis. In L. M. Hyman (ed.), *Consonant Types and Tones (Southern California Occasional Papers in Linguistics 1.*

1-95. Los Angeles: University of Southern California

Michaud, A., & Sands, B. (2020). "Tonogenesis", eds. Aronoff, M., *Oxford Research Encyclopedia of Linguistics*, (Oxford University Press).

Melguy, Y. (2018). "Strengthening, Weakening and Variability: The Articulatory Correlates of Hypo- and Hyper articulation in the Production of English Dental Fricatives" *UC Berkeley PhonLab Annual Report 14* (1), pp.221-255.

National Institute of the Korean Language (NIKL), (2005). *The Speech Corpus of Reading-Style Standard Korean*. Seoul, Korea: The National Institute of the Korean Language. URL: https://www.korean.go.kr/ (Last accessed 24/06/2022).

Nützel, D., & Salmons, J. (2011). "Language Contact and new dialect formation: evidence from German in North Amserica" . *Language and Linguistics Compass, 5*, pp. 705-717

Pearce, M. (2013). *The interaction of tone with voicing and foot structure*. Stanford: Center for the Study of Language and Information.

Pearce (2009). "Kera tone and voicing interaction" . *Lingua 119*, pp. 846-864.

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Silva, D.J. (1992). *The phonetics and phonology of stop lenition in Korean.* (Doctoral Dissertation), Cornell University

Silva, D. (2006), "Acoustive evidence for the emergence of a tonal contrast in contemporary Korean". *Phonology* 23, pp.287-308

Silva, D.J. (1991). "A prosody-based investigation into the phonetics of Korean stop voicing". In Susumu Kuno, *Harvard Studies in Korean Linguistics IV*, pp. 181-195 (Seoul: Hanshin).

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.* Daegu: Industry-Academic Cooperation Foundation, Keimyung University (Distributor). https://doi.org/10.13064/ksss.2015.7.2.103