

Mary E. Beckman\*, Fangfang Li, Eun Jong Kong and Jan Edwards

# Aligning the timelines of phonological acquisition and change

**Abstract:** This paper examines whether data from a large cross-linguistic corpus of adult and child productions can be used to support an assumed corollary of the Neogrammarian distinction between two types of phonological change. The first type is regular sound change, which is assumed to be incremental and so should show continuity between phonological development and the age-related variation observed in the speech community undergoing the change. The second type is dialect borrowing, which could show an abrupt discontinuity between developmental patterns before and after the socio-historical circumstances that instigate it. We examine the acquisition of two contrasts: the Seoul Korean contrast between lax and aspirated stops which is undergoing regular sound change, and the standard Mandarin contrast between retroflex and dental sibilants which has been borrowed recently into the Sōngyuán dialect. Acquisition of the different contrasts patterns as predicted from the assumed differences between continuous regular sound change and potentially abrupt dialect borrowing. However, there are substantial gaps in our understanding both of the extent of cross-cultural variability in language socialization and of how this might affect the mechanisms of phonological change that must be addressed before we can fully understand the relationship between the time courses of the two.

---

\*Corresponding author: **Mary E. Beckman:** Ohio State University.

E-mail: mbeckman@ling.osu.edu

**Fangfang Li:** University of Lethbridge. E-mail: fangfang.li@uleth.ca

**Eun Jong Kong:** Korea Aerospace University. E-mail: ekong@kau.ac.kr

**Jan Edwards:** University of Wisconsin–Madison. E-mail: jedwards2@wisc.edu

## 1 Introduction

This paper explores the extent to which data on young children’s productions of sounds that are involved in a phonological change in progress can illuminate an assumed consequence of the Neogrammarian distinction between regular dialect-internal sound change and dialect borrowing. As noted by Hoenigswald (1978), among others, this distinction can be viewed, at its root, as merely “a heuristic technique . . . for comparative research” (Hockett 1965: 188) – i.e., a

practicable partitioning of the data when the comparative method is applied to analyze correspondences between word forms used by divergent speech communities in the aftermath of phonological change. However, the Neogrammarians' understanding of regular sound change was that it is a distinct mechanism from borrowing, which targets a different level of linguistic knowledge.

Specifically, the Neogrammarians understood sound change proper as a series of mutations that directly target the phonological grammar of pronunciation norms for sublexical elements; if such a change splits a previously unitary word class into two phonemically differentiated ones, the split falls along some fault line in the phonetic distribution that can be identified as the valley between two modes in some pattern of segmentally or prosodically conditioned allophony at a stage before the split. By contrast, phonological change by borrowing always begins with a replacement of one word form by another in a motley set of mutations that target the lexical items themselves. The accumulation of such word-specific mutations can lead to a systemic change in the phonological grammar by splitting a previously unitary word class, but in that case the split falls along a fault line that is defined by the phonemic contrasts in the lending dialect rather than by some phonetically conditioned allophony in the borrowing dialect.

Another difference that has been posited involves the timelines for the two types of change. Labov (2007) describes this difference as a corollary of the difference in mechanism. Regular sound change ('change from below') is a continuous and gradual shift in pronunciation norms that is "generated by the process of INCREMENTATION, in which successive cohorts and generations of children advance the change beyond the level of their caretakers and role models, and in the same direction over many generations" (Labov 2007: 346; see also Hockett 1950, 1965). Changes should continue in the same direction over successive generations when they involve "variables which have been evaluated in the same way by the speech community over a considerable period of time" (Weinreich et al. 1968: 146). That is, regular sound change progresses by a generation-by-generation ratcheting of values along some gradient of pronunciation variation relative to an associated scale of social evaluation. Where a child's pronunciations fall along this gradient can evolve over the course of development as the phonological system is transmitted from previous generations of speakers. Specifically, the range of variation commanded by the child can expand as the child's social network expands beyond the immediate family circle so as to expose the child to pronunciations by the current leaders of the change. But this trajectory should be continuous with the longer-term trajectory of the sound change along the phonetic dimensions that are mutating.

By contrast, change through borrowing ('change from above') is typically ascribed to a specific historical event that brings adult speakers of two divergent

systems into contact. This is independent of the regular line of transmission of pronunciation norms, so children may learn a distribution that is discretely different from the system that was transmitted to the immediately previous generations. Moreover, given the circumstances that can induce this kind of change to the phonological system, the social evaluation of phonetic variation in the aftermath of the contact event often will be discontinuous with the social evaluation of the ‘same’ variables prior to the change in the system. For example, phonetic variation that marked different styles or differences in social identity prior to the contact event might come to instead differentiate two word classes, or vice versa. Moreover, this shift in interpretation can be very abrupt, occurring in the space of one generation instead of over the course of three or more generations as in the typical cases of sound change by incrementation that have been observed.

In this paper, we will explore this issue of different timelines using data from two portions of the Paidologos corpus gathered in speech communities where there are phonological changes affecting some of the lingual consonants that are targeted in this corpus. In one community (Seoul, South Korea), there is an ongoing shift in the pronunciation of lax and aspirated stops that began more than fifty years ago. This shift is a ‘change from below’ – i.e., a Neogrammarian sound change by incrementation. In the other community (Sōngyuán City, PRC), there is a change to the pronunciation of sibilant consonants that has split the apical series in two, resulting in a contrast between retroflex and dental word classes, as in the Běijīng dialect that is the basis for the standard language. This split is clearly a ‘change from above’ – i.e., a pattern of pronunciations from a different dialect of Mandarin Chinese that was borrowed into the Sōngyuán variety within the last 30 years. In Sections 4 through 7, we will present the evidence for the phonological changes in each of the two communities and describe the relationships between this evidence and the patterns of phonological development that are evident in the corpus. First, however, we describe the Paidologos corpus (in Section 2) and review the model that we will assume for the role of children in Neogrammarian sound change (in Section 3).

## 2 The Paidologos corpus

The Paidologos corpus is a database of speech productions recorded in the same way from native speakers for the target language variety in day care centers, nursery schools, or participant homes in a half dozen different locations for as many different languages. That is, all of the Cantonese speakers for the Cantonese portion of the corpus were recorded in Hong Kong, all of the Greek speakers for the Greek portion were recorded in Thessaloniki, and so on. The portions of the

corpus that are used in the current paper are the Mandarin Chinese recordings made in 2006 in Sōngyuán City, PRC, by the second author, and the Korean recordings made in 2007 in Seoul, South Korea, by the third author.

In each portion of the corpus, the raw data are productions of picturable words beginning with target lingual obstruents in prevocalic position that are likely to be familiar to children. For example, the Korean materials include 9 words such as [taŋkin] ‘carrot’ that target the lax dental and velar stops and 9 words such as [tʰaktʃa] ‘table’ that target the contrasting aspirated stops (see Kong et al. [2011] for the full list of stop-initial words used in the current analyses). Similarly, the Mandarin materials include 16 words such as [ʃa<sup>55</sup>fa<sup>55</sup>] ‘sofa’ that target the retroflex sibilant fricative, 17 words such as [san<sup>55</sup>tɕao<sup>213</sup>] ‘triangle’ that target the dental, and 14 words such as [ɕan<sup>55</sup>tɕao<sup>55</sup>] ‘banana’ that target the alveolo-palatal (see Li [2008] for the full list of sibilant fricative initial words used in the current analyses). These productions are elicited in a picture-prompted word-repetition task, from at least 100 monolingual speakers of each target language variety, in gender-balanced groups (at least 10 males and 10 females) for each of 5 age bands (2-, 3-, 4-, and 5-year-old children and college-age adults). This task was designed so that the target word forms can be elicited in the same way from all of the participants, even from the youngest 2-year-olds, whose vocabularies may not yet include all the target words (Edwards and Beckman 2008).

The productions are analyzed using various measures of pronunciation, beginning with one measure that can be calculated for all child productions – namely, the transcribed accuracy of the target sound. That is, the initial consonants (and the following vowels) in all of the children’s productions are transcribed by a trained phonetician who is a native speaker of the target variety, with a subset of at least 10% of the tokens re-transcribed by a second native speaker to calculate an inter-transcriber reliability rate. For example, the Korean children’s productions that are analyzed in Section 5 were transcribed by the third author of the current paper, with 10% re-transcribed by another native speaker of Seoul Korean (see Kong et al. [2011] for more details).

Other measures of pronunciation are specific to particular research questions and consonant types. For example, the measure of degree of aspiration in the Seoul Korean stops that are the target of the phonological change discussed in Sections 4 and 5 is voice onset time (VOT), calculated in the usual way by subtracting the time of the stop burst from the time of the first glottal pulse in the following vowel. The measure of an associated pitch pattern on the following vowel is the fundamental frequency ( $F_0$ ), calculated by taking the reciprocal of the glottal period located at 20 ms after voice onset. The primary measure of a more retroflex versus more dental pronunciation of the Sōngyuán sibilants discussed in Sections 6 and 7 is the centroid frequency (first spectral moment) calcu-

lated from a spectrum over a 40 ms Hamming window centered at the mid-point of frication. A second measure is the second formant (F2) frequency measured at the first appearance of a clear band of spectral energy that is continuous with the F2 in the following vowel.

### 3 The role of children in Neogrammarian sound change

There is long-standing interest in the nature of the relationship between phonological acquisition and phonological change. For example, Grammont (1902) noted similarities between some of the stereotypical misarticulations observed in early diary studies and some widely attested sound changes, a theme that was picked up by others, including Jakobson (1941), who accounted for the similarities by positing the axiomatic existence of substantive universals of the phonological grammar that control both development and change. This axiom was incorporated into many of the phonological frameworks developed in the 1960s and 1970s (e.g., Kiparsky 1968; Stampe 1979 [1973]). Moreover, it has persisted as a cornerstone of some current accounts, such as that of Kiparsky (2008), despite the mounting evidence against it. Indeed, already by the late 1970s, Vihman (1980) could conclude that any account of the role of children in phonological change that is predicated on positing such axiomatic homologies between children's characteristic errors and common sound changes is untenable. (See Foulkes and Vihman [in press] for an updated presentation of more data from Vihman's own work that supports her original arguments, and Edwards et al. [2012] for a review of some of the large body of subsequent work that also argues against invoking substantive universals to explain phonological development.)

The 1960s and 1970s also saw the development of accounts of variability, including work documenting variability in acquisition, both within and between children (e.g., Ferguson and Farwell 1975; Menn 1975; Clark and Bowerman 1986). A common theme in this work is that variability in phonological acquisition may reflect different strategies for resolving articulatory difficulties in early development. For example, Clark and Bowerman (1986) describe omission of coda voiced stops, substitution of voiceless stops with some lengthening of the preceding vowel, and substitution of nasal consonants as three possible 'strategies' for resolving the articulatory difficulty that voiced stops pose. As Guy (1980), Roberts (1994), and others point out, we need to disentangle this kind of child-internal source of variability from the type of socially structured variation that drives sound change to understand either. Another corollary conclusion is that we need

better models of the social forces driving both phonological development in an individual learner and phonological change in a speech community before we can begin to understand the relationship between the two.

A prescient early model of these social forces is Hockett's (1950) account of age grading and its relationship to Neogrammarian sound change. Hockett borrowed this term "age grading" from the anthropology literature, where it refers to culturally-specified expectations related to "the tendency for the members of a community to fall into various subgroups on the basis of similar age, with differentiation of economic function, ceremonial activity, and so on" (Hockett 1950: 453). Hockett identified childhood as the "age grade" during which sound change might occur in any generation of speakers. That is, prefiguring Roberts (1994, 2002), Eckert (1997), Cameron (2010), and others, Hockett argued that a sound change in progress can be advanced in any cohort of talkers only after the children's social networks become larger than their immediate families, when the next set of small mutations can be driven by "the fires of childhood competition and conflict which forge the . . . speaker's idiolect" (Hockett 1950: 452). In other words, Hockett suggested that the generation-by-generation ratcheting of incrementation happens in social interactions with age peers that begin only after the "cradle stage" – i.e., after an initial transmission stage when the child develops enough of a phonology to grow a toddler-sized lexicon in interactions with adult caretakers.

As in some more recent accounts, such as Chambers (1992), Hockett identified the end point for the incrementation process as puberty, the "age at which a child's moral sensibilities become fixed" with "a concomitant loss of linguistic flexibility" (Hockett 1950: 450). What Hockett did not anticipate in his account was the growing body of work showing that late adolescence (post-puberty) also is a time of potentially even greater change in an individual's production patterns, as teenage men and women begin to move into the larger societal 'linguistic marketplace' (see, e.g., work summarized in Eckert 1997 and Sankoff 2004).

Labov (2001) describes his close study over several decades of sound changes in progress in Philadelphia and reviews other similarly long-term studies, which, when integrated with the work described by Eckert, together support the following model of the relationship between phonological development in an individual and the incrementation of a Neogrammarian sound change in that individual's generation of speakers. Development begins with the transmission of the phonological system of the child's primary caretaker. This system will be more or less conservative, depending on the caretaker's age. Then, at age 4 or 5 or 6 (whenever the child's social network expands beyond the immediate family), the child can begin to learn to parse and produce a wider range of variation, includ-

ing more advanced pronunciations to which the child is exposed during social interactions with other adults who are younger than the child's caretaker and with other children who are older than the child. Through these interactions, social leaders will emerge who might produce even more advanced pronunciations as they differentiate themselves from more conformist members of their own generation as well as from their parents' generation of speakers. In this way, the range of pronunciations produced by any cohort can expand and the center of the distribution can shift in the same direction as that of the sound change. The initial range expansion and subsequent shift can progress until the phonological system for that generation of talkers stabilizes as the cohort transitions into the age grade of young adulthood.

This stable phonological system is what these adults will transmit to their own children. In the meantime, though, other intervening cohorts will be advancing the sound change further, so that the system that adults in any cohort transmit to their children will be more conservative relative to the system to which their children will become exposed when they enter the age grade of childhood. The social forces that drive sound change, then, are ones that can produce this generation-by-generation ratcheting in the distribution of productions.

These social forces must include an associated generation-by-generation ratcheting in listeners' interpretation of the resulting age-graded phonetic distribution – that is, in the perceptual grammar that allows each cohort to “align the variants heard in the community with the vector of age, that is, they grasp the relationship: the younger the speaker, the more advanced the change” (Labov 2007: 380). Thus, the model is predicated on there being the kinds of effects of perceived speaker age on the phonological parse that are demonstrated in perception studies such as Warren et al. (2007) and Koops et al. (2008). The model also predicts that the generation-by-generation incrementation will have a logistic shape as the sound change progresses from its beginning to its end.

We apply this logistic incrementation model in reviewing the evidence for a gradual Neogrammarian sound change in progress in Seoul Korean, which differs from the much more abrupt change in Sōngyuán Mandarin. We assume this model also in predicting that the distribution of phonetic values in young children's productions will match the somewhat more conservative patterns of young men rather than the more advanced patterns of young women, and that perception experiments will show that listeners expect this. Moreover, we predict that the developmental pattern for children in Sōngyuán will differ. Specifically, because of the break in transmission, the distribution of phonetic values in Sōngyuán children will match instead that of the more advanced talkers in the more abrupt change from above.

## 4 Shift from VOT to $F_0$ in the Korean lax/aspirated contrast

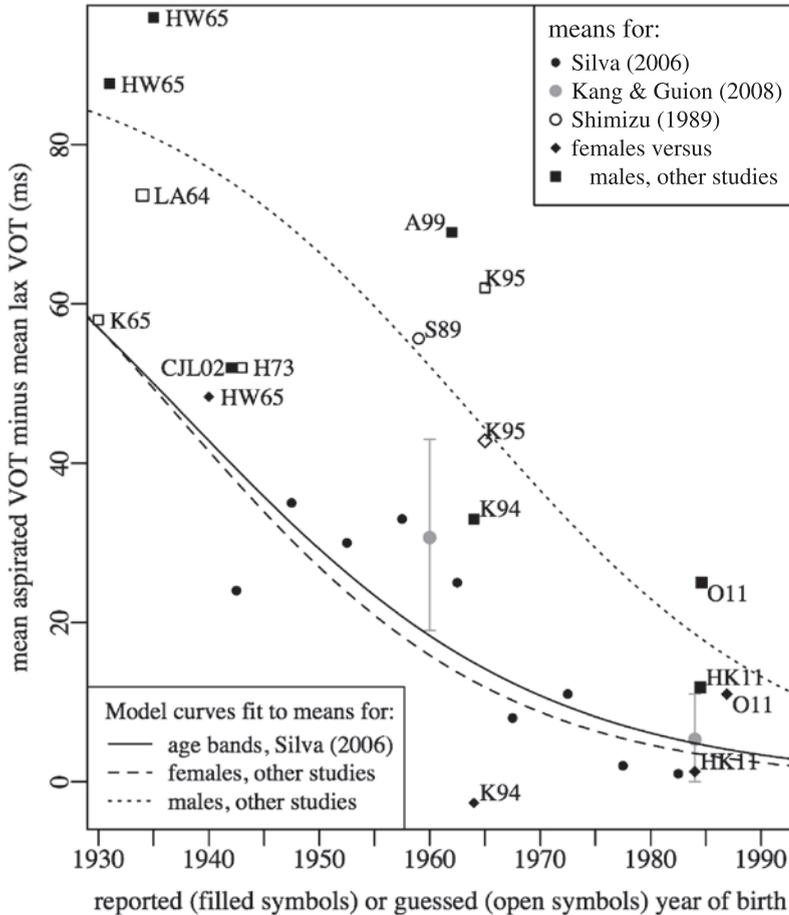
The first change that we examine is a shift in the phonetic basis of the Korean contrast between lax and aspirated stops from being primarily a contrast in degree of aspiration in phrase-initial position to becoming primarily a contrast in the tone pattern that marks the beginning of what Jun (1998) termed the Accentual Phrase. This is a female-led ‘change from below’ that began in the Seoul speech community.

The larger socio-historical context for this change is that the variety of Korean spoken in and around Seoul was the original basis for the development of the written language in the 15<sup>th</sup> century, and “Seoul speech now generally used in middle-class society” (translation in Lee and Ramsey 2012: 291) was designated as the national standard in the orthographic reforms of the early part of the 20<sup>th</sup> century. Relocations in the war after the 1945 partition and the subsequent industrialization and concomitant urbanization that began in the 1970s have only reinforced this status of the Seoul dialect in South Korea. Today, the majority of South Koreans live in large cities such as Busan (population 3.5 million in the 2005 census) and Daegu (2.5 million), with an especially high concentration in the Seoul metropolitan area, which includes Incheon (2.5 million) and Seoul proper (9.8 million). The situation is much like that in Italy, where differences among *dialecti* (the morphological and lexical variables that can make regional varieties mutually unintelligible) have given way to differences among *accenti* (the phonological variables that can mark regional variants of the standard language). As Lee and Ramsey (2012: 292) describe the situation, the Seoul dialect is now “universally understood, the primary medium of communication everywhere; it is rapidly displacing all regional dialects and usages, especially among the young.”

What this “displacement” has meant for the Seoul community is that even first-generation migrants from other regions use standard morphological endings and vocabulary, but some features of accent have become salient phonological variables. For example, some other dialects, such as Gyeongsang, were more advanced than Seoul in a set of sound changes that simplified historic diphthongs and clusters with [w]. There is now variation in the pronunciation of words such as [tɕwasək] ~ [tɕasək] ‘seat’ and [sakwa] ~ [saka] ‘apple’ in the Seoul speech community, which Kang (1997) showed to be stratified by class, age, and the formality of the elicitation task. Since this [w]-deletion can be represented transparently in writing by using a non-standard Hangeul spelling, Kang was also able to probe awareness using a written survey. He found that the likelihood of self-reporting of [w]-deletion also was stratified by class and age.

By contrast to [w]-deletion, the sound change that we examine here is not reflected in mismatches between the pronunciation and the written form, and it is very far below the level of conscious awareness that the orthographically transparent variable has. It is a shift over the last half century or more in the phonetic realization of the contrast between the ‘lax’ stops (transliterated as ‘b’, ‘d’, ‘g’ in the new official Romanization system that the Ministry of Education adopted in 2000) and the aspirated stops (transliterated as ‘p’, ‘t’, ‘k’) that had escaped the notice of even most linguists until fairly recently. In word-medial position, the lax stops have very short closures, and are often voiced, whereas the aspirated stops have intermediate-length closures followed by measurable aspiration – i.e., voice onset time (VOT) values in the long-lag region. In initial position, on the other hand, both types are voiceless, as is the third type, the ‘tense’ stops (transliterated as ‘pp’, ‘tt’, ‘kk’). Writing at a time much closer to the beginning of the sound change of interest, Martin (1951: 525) characterized the lax stops as “slightly aspirated” in this position, as opposed to the much stronger aspiration of the phonemically aspirated stops. This characterization is consistent with the VOT values reported for initial stops produced by male talkers in instrumental studies in the 1960s. However, the size of the difference in VOT values reported in subsequent studies has diminished decade by decade. Silva (2006), Wright (2007), and Kang and Guion (2008) all report apparent-time evidence (i.e., systematic variation across talkers of different ages recorded in the same year) that strongly suggests a change in the contrast from being primarily a difference in degree of aspiration to becoming a difference in voice quality and pitch level on the vowel of the initial syllable of the Accentual Phrase.

Figure 1 plots data from two of these apparent-time studies (unlabeled data points) and from a subset of the other studies over the last five decades that together constitute the real-time evidence for a change in the role of aspiration (see Table 1 for a key to the labels). Each data point shows the difference in mean VOT values (i.e., the mean VOT of the lax stop subtracted from the mean VOT of the aspirated stop), either for an individual talker [points LA64, HW65, K65, and H73] or for means averaged over a group of talkers [all other data points], as a function of the individual’s year of birth or of the group’s mean year of birth. A diamond-shaped plotting symbol indicates data from a female or a group of females, a square indicates data from male or a group of males, and a circle indicates data from a study that either does not report talker gender [open circle for S89] or that reports only means averaged over talkers of both genders [small black filled circles for Silva (2006) and large gray filled circles for Kang and Guion (2008)]. Also, an open plotting symbol indicates that the study providing the data point does not report talker age, so that the year of birth value is based on our guess that the talker was the author himself [K65] or that the talker was about 30



**Fig. 1:** Difference in VOT (mean for the aspirated stop minus mean for the lax stop) as a function of year of birth, calculated from values for the two stop types reported for Seoul speakers in 13 studies (see first two parts of Table 1), along with curves from two regression models (see first two parts of Table 2). The gray lines extending upward and downward from the large gray data points indicate the range of means for the three different speaking styles elicited in Experiment 1 of Kang and Guion (2008).

[LA64, H73], or that the talkers were on average about 30 [S89, K95] at the time of recording.

As the figure shows, the VOT difference (or the mean VOT difference) tends to be quite large for talkers (or groups of talkers) who were born in the 1930s, at the left edge of the plot. For example, the two male talkers in Han and Weitzman (1965) [squares, HW65], who were reported to be 33 and 29 years of age at the time

**Table 1a:** Apparent-time studies that provided the unlabeled data points (and model for solid curve\*) in Figure 1. (The Kang and Guion study also provided data points for Figure 3.)

Reference for the paper describing the study that provided the data	Number of talkers (number by gender)	Speech materials and elicitation task(s)
Silva 2006*	36 (21f, 15m)	words read in frame sentence
Kang & Guion 2008, Exp. 1	11 (6f, 5m) in each of a younger & an older group	words said or read in tasks to elicit different speech styles

**Table 1b:** Other studies that provided labeled points in Figure 1. The data points in the first group also served as the real time data for the model that provided the dashed and dotted curves in this figure. Some of these studies also provide data points for Figure 2\* or Figure 3†.

Reference for the paper describing the study that provided the data	Label	No. of talkers (by gender)	Speech materials and elicitation task(s)
Han & Weitzman 1965†	HW65	3 (1f, 2m)	word list, read
Kim (1965)	K65	1 (1m)	word list, read
Cho, Jun, & Ladefoged 2002*†	CJL02	4 (1m)	word list, recited/read
Kim 1994*†	K94	6 (3f, 3m)	words read in frame sentence
Ahn 1999†	A99	6 (6m)	words read in frame sentence
Oh 2011	O11	38 (19f, 19m)	words isolation & frame sent.
Holliday & Kong 2011*	HK11	12 (6f, 6m)	word list, read
Lisker & Abramson 1964	LA64	1 (1m)	subject-devised sentences
Hardcastle 1973†	H73	1 (1m)	word list, read
Shimizu 1989†	S89	3 (unknown)	word list, read
Kim 1995†	K95	5 (3f, 2m)	word list, read

**Table 1c:** The apparent-time study\* that provided the solid curve for Figure 2, and the three other studies that provided the labeled data points that are plotted on this figure.

Reference for the paper describing the study that provided the data	Label	No. of talkers (by gender)	Speech materials and elicitation task(s)
Holliday & Kong 2013*		11 (11f)	same task as HK11
Holliday & Kong 2011: Daegu	HK11	12 (6f, 6m)	see HK11 from Figure 1
Holliday & Kong 2011: Jeju Island	HK11	12 (6f, 6m)	
Cho et al. 2002: Jeju Island	CJL02	8 (8m)	see CJL02 from Figure 1
Kim 1994: Busan	K94	6 (3f, 3m)	see K94 from Figure 1

**Table 1d:** The study of the Paidologos corpus that provided the values for the gray diamonds and squares in Figure 3 and for all of the data points in Figure 4.

Reference for the paper describing the study that provided the data	Number of talkers (by gender)	Speech materials and elicitation task
Kong, Beckman, & Edwards 2011: adults in Figures 3 & 4	20 (10f, 10m)	picture-prompted word repetition task
children in Figure 4	67 (35f, 32m)	

of recording, have differences of 88 and 96 ms, consistent with Martin's characterization. The differences for the male talkers in the Lisker and Abramson (1964) and Kim (1965) studies [LA64, K65], who we guessed to be born in the same decade as the two males in Han and Weitzman (1965), are not as large, but both data points lie well above 50 ms.

At the right edge of the plot, by contrast, differences for talkers and groups of talkers born after 1980 are very small, ranging from a 25 ms mean difference for male talkers in Oh (2011) [square, O11] to virtually no difference for the youngest age band in Silva (2006) [rightmost small black circle], for the female talkers in Holliday and Kong (2011) [diamond, HK11], and for the spontaneous speech productions by the younger group of talkers in Experiment 1 of Kang and Guion (2008) [lower edge of line extending downward from the large gray circle at 1984].

In between these two groups, there is a general trend of a decreasing difference with each decade. For example, whereas the differences for the four males who were reported or guessed to be born in the 1930s range from about 60 ms to more than 90 ms, the difference for the female talker in Han and Weitzman (1965) [diamond, HW65], who was reported to be 24 years of age at the time of recording, so born in the 1940s, is only 48 ms. This is about the same as the difference calculated from the means for the four Seoul dialect talkers in Cho, Jun, and Ladefoged (2002) [CJL02], who were reported to be "in their late 50s and early 60s" (Cho et al. 2002: 199) at the time of recording, and larger than the difference calculated for the oldest talker in Silva (2006) [the left-most filled circle, for a single talker born in 1943].

A second trend that is apparent in the figure is that, whenever a study reports values for females and males separately, the mean difference for the females is smaller than the mean difference for the males born in the same decade. For example, whereas the mean difference for the females in Kim (1994) [diamond, K94] is slightly less than 0 ms, the mean difference for the males [square, K94] is more than 30 ms. Also, whereas the differences for the females in Oh (2011) and Holliday and Kong (2011) [diamonds, O11 and HK11] are 11 ms and 1 ms, respec-

tively, the differences for the males [squares, O11 and HK11] are 25 ms and 11 ms. Note also that the size of the apparent gender effect is smaller for the two more recent studies, which is exactly what we would expect if the contrast in VOT between lax and aspirated stops is collapsing; difference values for both genders should asymptote to a mean of 0 ms.

Moreover, the same two trends are evident in the productions of several other groups of talkers that Holliday and Kong (2011), Kim (1994), and Cho et al. (2002) recorded, as shown in Figure 2. This figure is like Figure 1 in plotting mean VOT differences against mean year of birth. Indeed, the five black data points for

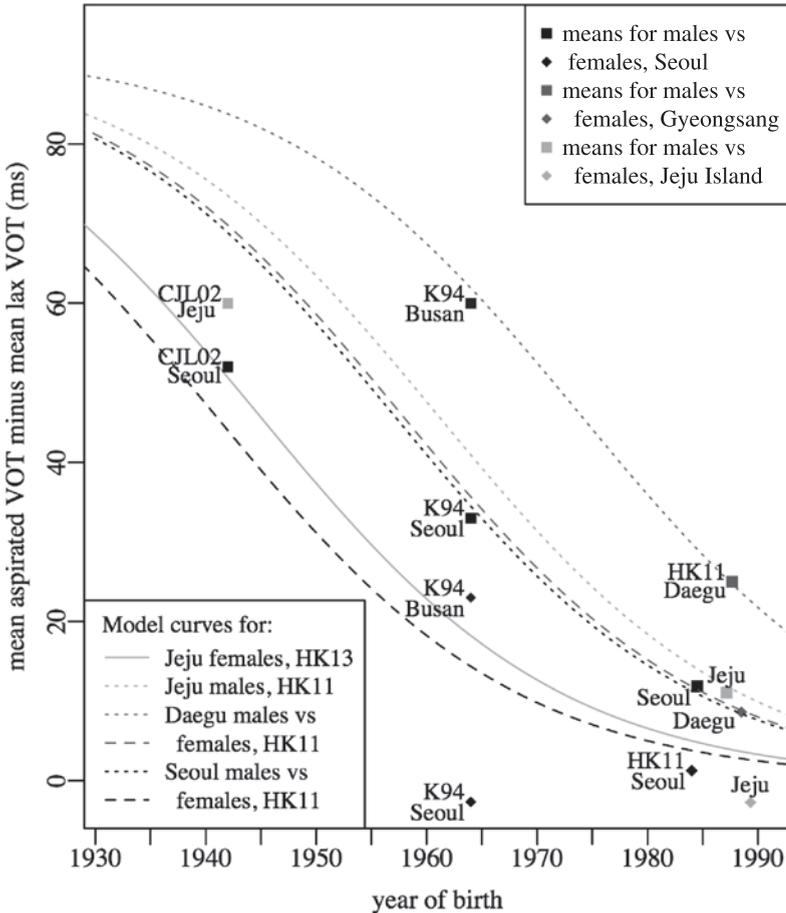


Fig. 2: The data for Seoul speakers from three of the studies in Figure 1, along with data points calculated from values reported for Jeju Island speakers or for Gyeongsang (Busan or Daegu) speakers in these studies. (See third part of Table 2 for the model curves.)

speakers of the Seoul dialect are the same data points shown on Figure 1 for these three studies. The light and dark gray data points show values reported for speakers of other dialects, namely Jeju Island and Gyeongsang (Daegu and Busan). Looking across the graph, we can see that the Jeju Island men born around 1940 [light gray square, CJL02] have a much larger VOT difference than the ones born in the 1980s [light gray square, HK11], and that the Gyeongsang speakers born in the 1960s [dark gray symbols, K94, from Busan] have a much larger VOT difference than the speakers born in the 1980s [dark gray symbols, HK11, from Daegu].

Note that the size of the generational effect in the Gyeongsang dialect is smaller for women than for men. Also, in both Kim (1995) and Holliday and Kong (2011), the Gyeongsang females [dark gray diamonds] have a smaller mean VOT difference than do the Gyeongsang males [dark gray squares], although the size of the gender effect is smaller for the more recent study. Again, this reduction in the size of the gender effect is what we should see if the VOT contrast is collapsing and being replaced with some other cue. Also, the size of the generational effect should be smaller for each gender when comparing a study that was conducted closer to the peak rate of change for that gender to a study that was conducted more recently.

In short, these two trends together – i.e., the reduction of the VOT difference across the decades in Figures 1 and 2, together with the large gaps between values for men versus women in Kim (1994) and the closing of these gaps in Oh (2011) and Holliday and Kong (2011) – suggest a female-led sound change that was well underway by the 1970s, and is now close to completion in the Seoul dialect.

In addition, Figure 2 shows a third trend, a difference among dialects such that the speakers from Seoul tend to have a smaller VOT difference than the speakers from Busan and Daegu. There is also a small difference between the older Seoul and Jeju Island speakers in Cho et al. (2002), but not between the younger speakers in Holliday and Kong (2011). This pattern suggests that the sound change began in Seoul, that it spread almost immediately to Jeju Island, and that it began spreading more recently to urban centers in Gyeongsang.

We evaluated these three trends by fitting regression models to two subsets of the data points in Figure 1 and to a third dataset that combines the values reported for young adults in Holliday and Kong (2011) with the values for older speakers in an apparent-time study that will be reported in Holliday and Kong (2013). The model results are summarized in Table 2, and the coefficients are also displayed as model curves in Figure 1 (first and second model) and Figure 2 (third model).

The raw data for the first model are the 9 black circles in Figure 1, which plot VOT differences estimated from the mean values shown by Silva (2006: 292) for

**Table 2a:** Summary of the regression models for the curves in Figures 1 and 2. Regressing VOT differences against center of birth year band for the 9 data points from Silva’s (2006) apparent-time study – unlabeled black circles and the solid line in Figure 1.

Coefficients	Estimate	Std. err.	t-value	p	Values predicted by model
intercept	0.55028	0.53325	1.032	0.34	58 ms difference in 1929
year of birth	-0.06259	0.01769	-3.538	<0.01	peak change of -14 ms/decade, for cohort born in 1938

Dispersion parameter for quasibinomial family taken to be 0.05648677  
 Null deviance: 1.24187 (df = 8)  
 Residual deviance: 0.42101 (df = 7)

**Table 2b:** Regressing the VOT differences against year of birth and gender for the 10 data points in the meta-analysis of sufficiently informative studies in part 2 of Table 1 – labeled diamonds and squares and the dashed and dotted model curves in Figure 1. Reference talker is male.

Coefficients	Estimate	Std. err.	t-value	p	Values predicted by model
intercept	2.39736	0.70787	3.387	<0.01	84 ms for Seoul males in 1929
year of birth	-0.06859	0.01841	-3.725	<0.01	peak change of -15 ms/decade, for males born in 1968
gender = female	-1.83828	0.75123	-2.447	<0.05	for females born in 1937

Dispersion parameter for quasibinomial family taken to be 0.1717941  
 Null deviance: 6.6283 (df = 10)  
 Residual deviance: 1.4620 (df = 8)

**Table 2c:** Regressing the VOT differences against year of birth, dialect, and gender for the 47 talkers in the combined data set of Holliday and Kong (2011) and Holliday and Kong (2013) – the six model curves in Figure 2. Reference talker is Jeju Island female.

Coefficients	Estimate	Std. err.	t-value	p	Values predicted by model
intercept	1.49495	0.26850	4.281	<0.001	70 ms for Jeju females in 1929
year of birth	-0.07281	0.00792	-9.198	<0.001	peak change of -17 ms/decade, for Jeju females born in 1945
dialect = Seoul	-0.28977	0.32765	-0.884	0.382	for Seoul females born in 1941
dialect = Daegu	0.94250	0.29136	3.235	0.002	for Daegu females born in 1958
gender = male	1.17771	0.26100	4.512	<0.001	for Jeju males born in 1961

Dispersion parameter for quasibinomial family taken to be 0.05514172  
 Null deviance: 9.3074 (df = 46)  
 Residual deviance: 2.3610 (df = 42)

groups aggregated into 5-year bands. The bands were designed to be evenly spread over time, and so are not balanced by the number of talkers contributing to the mean. For example, there are 9 talkers in the band from 1975 to 1979, but only 1 in the band from 1940 to 1944.

The raw data for the second model are the 10 labeled solid diamonds or squares in Figure 1, which plot the VOT differences for the seven studies listed in the second section of Table 1. Here we chose all studies that reported age or birth year and that also did not aggregate across genders when reporting the mean values for the aspirated and lax stops. Again, the data points are more or less evenly spread across the plot, and represent anywhere from 1 talker (for each of the three HW65 data points) to 19 talkers (for each of the two O11 data points).

The raw data for the third model are individual VOT differences for the 36 talkers in Holliday and Kong (2011) [gray and black data points, HK11] and for two groups of older Jeju Island females who were born in the 1970s or the 1940s. (Since descriptive statistics for the apparent-time data for Jeju Island are reported in a manuscript that is currently under review [Holliday and Kong 2013], Figure 2 shows the model curves that are based on the full data set, but not the as yet unpublished data points for the two older groups.)

For all three models, we first divided the VOT differences by 92 ms in order to convert them into an estimated ‘proportional distance from completion’ of the sound change. That is, we assumed that the initial community norm before the sound change was a VOT difference value equal to the mean for the two male talkers in the Han and Weitzman (1965) study, and that the sound change is taking the VOT difference toward a final community norm that is equal to 0 ms. We converted the VOT differences to proportional distances in this way so that we could fit a general linear model (using the `glm` function in R) with the empirical logit as the link function (i.e., setting the ‘family’ argument to ‘quasibinomial’). All 9 data points from Silva (2006) were within the range from 1 to 0 (proportions corresponding to 92 ms and 0 ms), and so could be included in the first model as is. However, we had to adjust several values in the other two data sets to include them in the other models. Specifically, for the second model, we replaced a value of 1.02 (corresponding to the 96 ms difference for square HW65 at 1935) with 1, and a value of  $-0.03$  (corresponding to the difference of  $-3$  ms for diamond K94) with 0. And for the third model, we replaced six similarly small negative values with values of 0 (for 2 Seoul females and 4 Jeju Island females born in the late 1980s or early 1990s).

In all three models, then, we used this adjusted proportional distance from the asymptote value at the end of the sound change as the dependent variable. We used year of birth as an independent variable, after ‘centering’ the variable by subtracting 1929 so that the intercept value could be converted easily into a pre-

dicted VOT difference for the left edges of Figures 1 and 2. In the first model, which was based on means aggregated over the 21 female and 15 male Seoul dialect speakers in Silva's (2006) apparent-time study, this was the only independent variable. In the second model, we added talker gender as a second independent variable. (We used a treatment-style contrast with 'male' as the reference level, since we had more data points for males than females, so the intercept corresponds to a projected difference value for Seoul males born in 1929, and the slope corresponds to a projected peak incrementation rate for Seoul males born in some later year.) And in the third model, we entered both talker gender and dialect as additional independent variables. (We again used treatment-style contrasts with 'female' and 'Jeju' as the reference levels, since the apparent-time data for evaluating the effect of year of birth were from the 21 Jeju Island women in Holliday and Kong [2013], so the intercept and slope are for Jeju females.)

One noteworthy result is that, although the three models are based on different data sets, with the female-dominated means in Silva's (2006) apparent-time study providing the data points for the first model, but Seoul males providing the majority of the data points for the real-time analysis in the second model, and Jeju Island females providing the time depth for the apparent-time analysis of the third model, all three models estimate about the same drop of about 15 ms per decade around the year with the peak rate of change. A second noteworthy result is that the year in which the rate of change was projected to have peaked for Seoul females by the third model (1941) is very close to the year in which the rate of change was projected to have peaked for the Seoul females by the second model (1937). Both are also close to the year of maximum change projected for the Seoul speakers by the first model (1938). Here we note that, while this first model is based on data points that aggregate over males and females, there were 21 females to 15 males in the study, and there were only 3 talkers in the first two bands. So, it is likely that female values dominated in the part of the data set that was closest to the peak rate of change. These points of agreement among the models then add further support to our interpretation of the data in Figures 1 and 2 as evidence of a female-led sound change which was already well underway by the time of the first studies reporting VOT values for Korean and which is now nearing completion in Seoul.

It is important to keep in mind that the measure of the sound change in these models is the size of the contrast between lax and aspirated stops in the VOT dimension. The fact that the change in VOT values is now nearing completion means that, if degree of aspiration were to remain the primary correlate of the Korean contrast between lax and aspirated stops, there would be a merger or near-merger (i.e., a loss of contrast or a reduction in the robustness of the contrast) in the Seoul speech community today. However, there is no evidence that

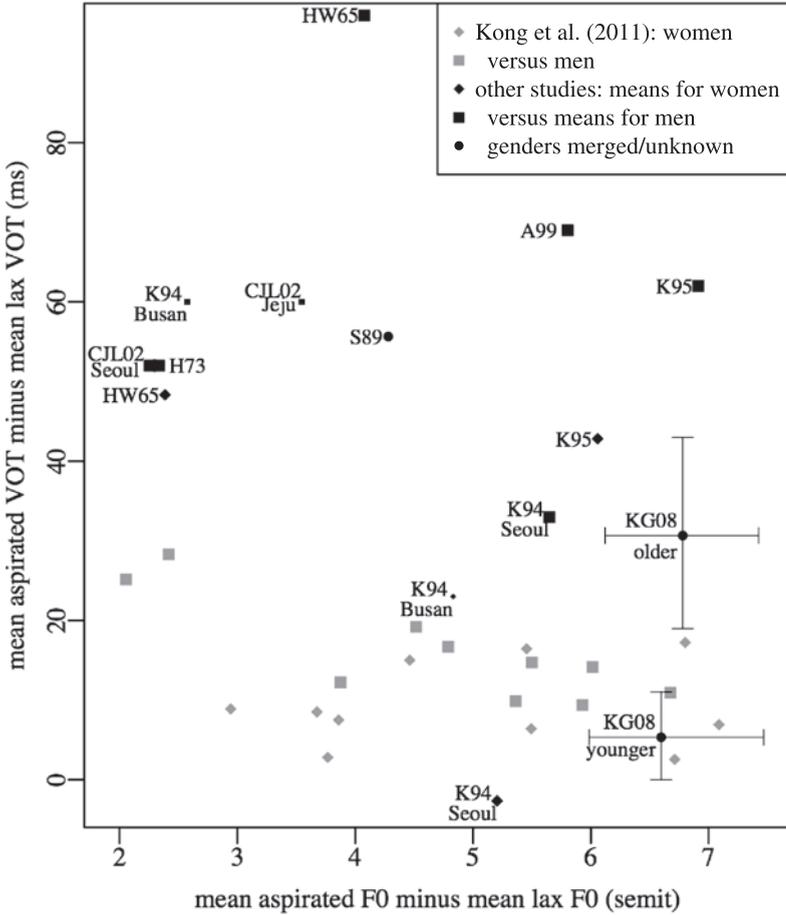
the contrast is any less robust today than it was in the 1960s. Instead, there is evidence of corollary changes, which can be characterized as the ‘phonologization’ of differences in the voice quality and in the pitch pattern. That is, many studies have noted that the aspiration in a lax stop is ‘weaker’ and has a more gradual transition into an interval of ‘breathy voice’, and many studies have also noted that an Accentual Phrase which has an initial lax stop begins on a lower pitch than one which has an initial aspirated stop.

These corollary changes cannot be documented using the kind of model we applied to the VOT differences in Figures 1 and 2, for several reasons. First, the studies that quantify voice quality use measures that are too disparate to track across decades. Also, they focus on the contrast between lax and tense stops and do not consistently report values for aspirated stops. Moreover, while fundamental frequency ( $F_0$ ) has been a consistent measure of the pitch pattern, the change in this cue to the contrast is not a dramatic shift in the mean values associated with the two stop types. Instead, it is a change in the function of a pitch difference that was noted even in the earliest instrumental studies, as shown in Figure 3.

This figure again plots the mean VOT differences from older studies that we plotted as a function of year of birth in Figures 1 and 2, but this time as a function of the mean difference in  $F_0$  measured on the vowel following the two stop types.<sup>1</sup> As the figure shows, even the speaker with the largest difference in VOT, the male speaker in Han and Weitzman (1965) who was reported to be 29 years old at the time of recording, has a difference of more than 2 semitones on the vowels following the initial lax versus aspirated stops.

What then is the evidence for the corollary phonologization? One type of evidence comes from a small shift in the mean  $F_0$  values associated with the aspirated stops as the sound change spread from Seoul females to Seoul males and Gyeongsang females. That is, the data points for studies reporting values for Seoul speakers born before 1950 [HW65, H73, CJL02] and for Busan males in Kim (1994) [small square K94] all lie in the left part of the figure, with values between 2.2 and 4.1 semitones. By contrast, the data points for Busan females in Kim (1994) [small diamond K94] and for studies reporting values for Seoul speakers born between 1950 and 1970 (in cohorts after the change had affected productions by Seoul men and was beginning to spread to Gyeongsang women, according to the

<sup>1</sup> A data point is shown in Figure 3 only if a mean  $F_0$  value can be matched up with the mean VOT value. Thus, there is no data point for the older male in Han and Weitzman (1965) because they report  $F_0$  values only for the female and the younger male. There is no data point for Kim (1965) because he reports mean durations of the first period of the following vowel for the lax stop versus the aspirated and tense stops combined. And there are no data points for Silva (2006) because he grouped the talkers into just two age bands in the figure showing mean values for  $F_0$ .



**Fig. 3:** Difference in VOT (defined as in Figures 1 and 2) as a function of the difference in  $F_0$  (mean for vowel after aspirated stop in semitones relative to mean for vowel after lax stop) for the 20 young adult talkers in the Paidologos corpus whose productions were measured by Kong et al. (2011) (gray plotting symbols) and for each of the data points in Figures 1 and 2 where an associated  $F_0$  difference value could be calculated (black plotting symbols). As in Figure 1, the horizontal and vertical lines extending outward from the two data points labeled KG08 indicate the range of means for the different speaking styles elicited in Experiment 1 of Kang and Guion (2008).

second and third models in Table 2) all lie in the right part of the figure, with values higher than 4 semitones.

The small size of this shift in  $F_0$  production values relative to the shift in VOT production values can be appreciated by comparing the range of values for

speakers born more recently. To allow this comparison, Figure 3 plots values individually for the 20 young adults from the Korean portion of the Paidologos corpus in gray in the background. The shift in the VOT dimension across the generations has been dramatic enough that all of these gray data points lie well below all of the black data points for speakers born before 1950. Moreover, there is still a significant gender effect even for these young adults who were born close to the end of the sound change in the aspiration dimension (mean VOT difference values of 16 ms for males versus 9 ms for females,  $t = -2.6$ ,  $df = 17.4$ ,  $p = 0.02$ , by a Welch's two-sided  $t$ -test).

By contrast, while the overall mean  $F_0$  difference for the adults in the Paidologos corpus is 4.9 semitones (thus in the right half of the graph, along with the data points for the Seoul speakers born between 1950 and 1980), the overall range of values for these 20 talkers spans the full width of the graph, from 2 semitones for the male with the smallest  $F_0$  difference to 7.1 ms for the female with the largest  $F_0$  difference. Moreover, while the 10 men have, on average, a smaller difference than the 10 women, the gender effect is not significant (mean  $F_0$  difference values of 4.7 semitones for males versus 5.0 semitones for females,  $t = 0.46$ ,  $df = 18.0$ ,  $p = 0.65$ ). The smaller sizes of both the generational shift and the gender effect along this dimension of the figure are in keeping with our understanding that the corollary change is an alteration in the role of variation in values for an already existing  $F_0$  contrast rather than a more major reorganization that would split the distribution neatly in two.

A second type of evidence for the phonologization of the pitch contrast comes from a difference in the within-individual distribution of  $F_0$  values reported for older versus younger speakers in Kang and Guion (2008), in two experiments where communicative function was systematically manipulated to assess the interaction between talker age and speaking style. Figure 3 shows the ranges of  $F_0$  differences as well as the ranges of VOT differences from their first experiment, where target stop-initial words were recorded in three styles: spontaneous speech (elicited by asking participants to define each target word), 'citation form' speech (elicited by asking participants to read each target word in a fixed frame sentence), and 'clear speech' (elicited by asking participants to "imagine they were teaching Korean to second language learners of Korean" as they read minimal pairs on flash cards).

Recall that the two VOT difference ranges were shown also in Figure 1, where we could compare the size of the style effects to the size of the gender effect across the graph. This comparison suggests that the style effects are larger for the older speakers because they came of age at a time when the diminution of the VOT difference was close enough to its peak incrementation rate that these talkers learned to produce (as well to interpret) a larger overall range of variation in de-

gree of aspiration.<sup>2</sup> For the  $F_0$  dimension of the contrast, on the other hand, the two age groups show a comparable range of variation, and how they differ is instead in the distribution of the mean  $F_0$  differences across the three styles. Specifically, whereas the productions with the largest mean  $F_0$  difference for the younger speakers are the ones elicited in the clear speech task, with the spontaneous speech and citation form sentences both showing smaller  $F_0$  differences, the productions with the largest mean  $F_0$  difference for the older speakers are the ones elicited in the spontaneous speech task, with the clear speech and citation form sentences both showing smaller  $F_0$  differences. Kang and Guion (2008) also confirmed that the difference in how clear speech affects the  $F_0$  is not an artifact of the speech materials by doing a second experiment where they asked participants only to read the target words in a frame sentence, but had them read the list of sentences twice, once using a “comfortable” speaking rate and volume (for the citation form utterances) and the other time reading “in a ‘clear’ way, as if speaking to a ‘foreigner’ audience who needs greater linguistic-phonetic resources to have full access to the linguistic information” (3915).

Kang and Guion (2008) describe these effects of clear speech as follows. Whereas both groups enhance the phonation type contrast in the clear speech task by increasing the difference in average VOT between the lax and aspirated stops, only the younger speakers also increase the difference in average  $F_0$  on the following vowel. They interpret these clear speech effects as indicating a change in progress in the role of  $F_0$  from being a mere secondary correlate of the contrast to becoming a primary acoustic correlate of the contrast in this position in the prosodic hierarchy.

In summary, then, this section has reviewed evidence from production studies going back to Lisker and Abramson’s seminal paper on VOT in 1964 which strongly supports the claim of Silva (2002), Wright (2007), and others that there has been a gradual shift in the phonetic realization of Korean stops in phrase-initial position. The originally moderately long VOT values for the “mildly aspirated” lax stops have lengthened a great deal and the originally very long VOT values in the aspirated stops have shortened so as to considerably reduce the VOT difference. The gradual collapse of the contrast in the VOT dimension is a female-led change from below that originated in the Seoul community, and is drawing near to completion there, but it is not resulting in a near merger. Instead, other

---

<sup>2</sup> Further support for this suggestion comes from the fact that the style with the largest VOT difference for the older group is not the clear speech but instead the read sentences even though (1) the style with the smallest VOT difference is the spontaneous speech, (2) the clear speech style involves an enhancement of the VOT difference by about 10 ms for both groups, and (3) this is the style with the largest VOT difference for the younger group.

correlates are taking over the role of degree of aspiration in cueing the contrast. Given this picture of the change, we predict that in perception experiments, listeners will attend differently to the VOT and the  $F_0$  pattern depending on the age of the talker and also (if the talker is an adult) whether the talker is a man or a woman. These predictions are supported by a perception study reported in Kong et al. (2011), which we will describe in the next section, after evaluating a prediction about the distribution of VOT and  $F_0$  values in young children's productions.

## 5 Acquisition of the Korean lax/aspirated contrast

Given the evidence presented in the previous section, we might predict that at the age when young children begin to reliably produce the contrast between lax and aspirated stops, their VOT and  $F_0$  patterns will show a strong continuity with values seen in the adult talkers who are their caretakers. That is, once variation across children that is due to the normal developmental trajectory for phonation type contrasts is accounted for, the distribution of values observed in children's productions will be more like that of women who are somewhat older than the college-age women in the Paidologos data [gray diamonds in Figure 3], hence more like the distribution for the men in that study [gray squares in Figure 3]. To disentangle such effects of caretaker input from the effects of the developmental trajectory, then, we need to first consider how stereotypical misarticulations observed in very young children might affect the two measures of the progress of the sound change presented in Figures 1 and 3.

Kewley-Port and Preston (1974) note that successful production of a long-lag VOT requires precise temporal coordination between a glottal opening gesture and the oral stop release. Kong et al. (2011) cite this paper in reviewing literature showing that, in English and other languages that contrast a short-lag with a long-lag category, such as Cantonese and Hindi, children are transcribed as producing long-lag stops accurately at a later age than short-lag stops. The literature on Korean reviewed in Kong et al. (2011) suggests that, at its earliest stages, acquisition of the three-way contrast in Korean more or less fits this cross-linguistically common developmental pattern for VOT. Specifically, the majority of Korean-speaking children are perceived as producing the phonation type correctly for tense stops first, and early errors for the other two stop types are transcribed as being substitutions of the tense phonation for the target lax or aspirated phonation. However, this stereotypical substitution is a characteristic only of very young children, and after about 3 years of age, most children are perceived as

producing the aspirated and lax phonation types correctly. Three years is somewhat later than the ages at which children master the aspiration contrast in languages such as English and Cantonese, but this discrepancy might be due to the contrastively low pitch on phrases beginning with the Korean lax stops. That is, the bulk of the literature on languages such as Mandarin Chinese suggests that children tend to master tone contrasts relatively early, but accurate production of low tone targets can be as late as 3 years of age (see Wong [2013] and the literature reviewed there).

Figure 4 shows evidence for this developmental trajectory in productions by children from the Korean portion of the Paidologos corpus. It combines information about the transcribed accuracy of the phonation type with information about the VOT and  $F_0$  values measured for all of the productions of aspirated and lax targets that had measurable plosive bursts elicited from the children in the corpus. The figure shows a data point for each of these two types for each child, with the target type indicated by the shape of the plotting symbol (triangles for lax versus circles for aspirated) and the transcribed accuracy indicated by using an open black symbol when the phonation type was transcribed as being on target in at least 80% of the child's productions and a filled gray symbol otherwise. The VOT for each stop type is plotted against the  $F_0$  difference on the following vowel (i.e., the same measure that is plotted on the x-axis of Figure 3 for the adults), and the two data points for a child are connected by a line, so that the length of the line indicates the VOT difference for that child. The data are shown separately by age, from the 2-year-olds in the left-most plots to the 5-year-olds in the right-most one, with data for girls in the upper row of plots and data for boys in the lower row. For reference, the mean values averaged over all 10 adult females and over all 10 adult males are plotted using open gray symbols in the panels for the oldest girls and oldest boys.

Adapting conventions used in clinical norming studies, we will deem a child who was transcribed as being on target in at least 80% of the tokens for each of the two types as having 'mastered' the contrast. These children can be identified by having black rather than gray connecting lines as well as by having black open plotting symbols for both types. As the figure shows, only 5 of the 2-year-old children had mastered the contrast. Of the 16 others, 10 produced an inappropriately short VOT (i.e., making the stereotypical substitution error) for one or both of the stops, and four others produced inordinately long VOT values for the lax stop (making an aspirated for lax substitution). Also, one 2-year-old girl who was transcribed as not having mastered the contrast despite producing appropriate VOT values for both stops produced an  $F_0$  difference of  $-0.31$  (i.e., failed to produce an appropriate pitch difference). By contrast, all but 7 of the 20 three-year-olds and all but 6 of the 26 children in the two older groups had mastered

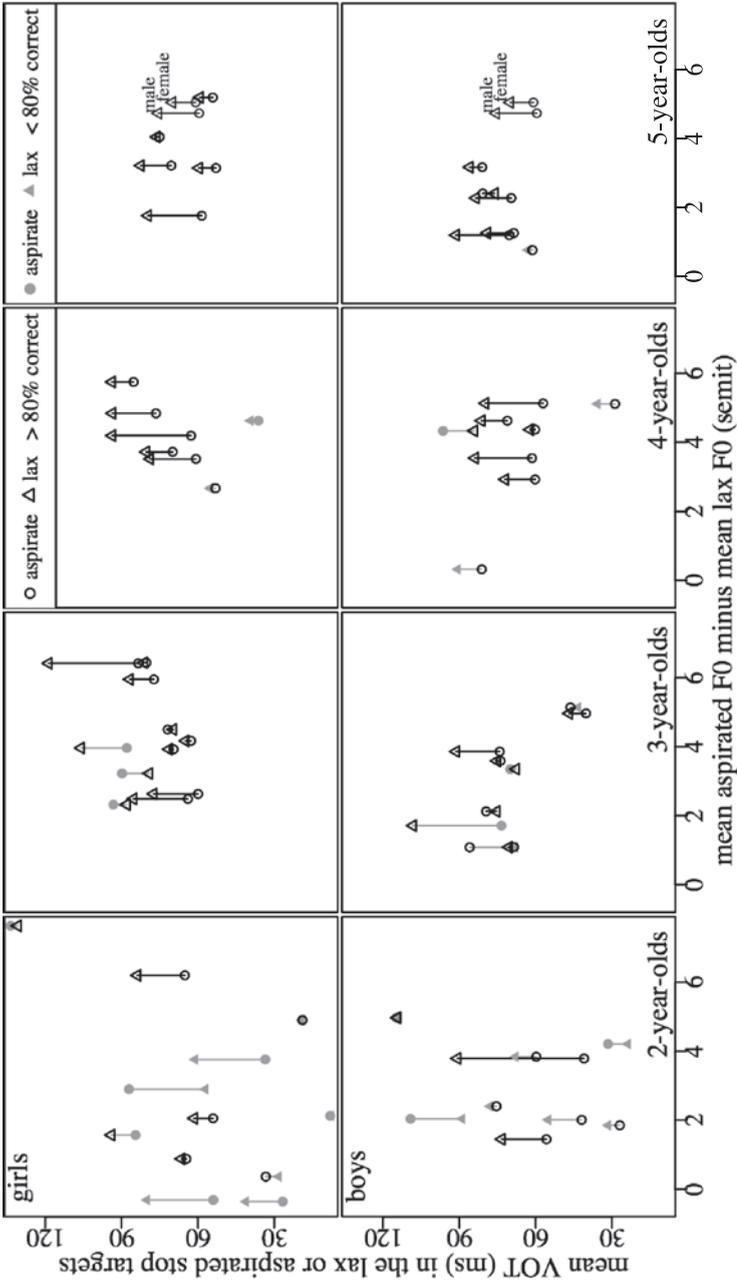


Fig. 4: Mean VOT values for each of the lax and the aspirated stop types plotted as a function of the difference in mean  $F_0$  between the two stop types for each of the Korean children in the Paidologos corpus. For reference, the mean VOT values and  $F_0$  difference values for the adult males and females are shown in gray open symbols in the plots for the 5-year-olds. (See text for a detailed description of the open black versus filled gray plotting symbols and the black versus gray lines connecting them.)

the contrast, and of those who had not, only two each of the 3- and 4-year-olds produced inappropriately short VOT values. Also, three 3-year-old girls and one 4-year-old boy produced inordinately long VOT values for the lax stop, and one 4-year-old boy who was transcribed as not having mastered the contrast despite producing appropriate VOT values for both stops showed an  $F_0$  difference of only 0.32 semitones.

Including the VOT difference values and  $F_0$  difference values for such children who are not yet producing appropriate values for one or both of the stop types could lead to spurious results in a test evaluating our prediction that the lax and aspirated productions of the children in the Paidologos corpus will be more like those of the young men than like those of young women. To make a more conservative test of our predictions, therefore, we include values only for the 38 children who were transcribed as having mastered the contrast.

These children tended to produce extremely long-lag ('strongly aspirated') values for their aspirated stops, with 30 of them having a VOT value in their aspirated stops that was larger than the mean for the women's aspirated stops, with an associated significant difference in the means (79 ms for children vs. 70 ms for women,  $t = 2.52$ ,  $df = 21.5$ ,  $p = 0.02$ ; vs 76 ms for men,  $t = 0.80$ ,  $df = 18.0$ ,  $p = 0.4$ ). However, the VOT values in the children's lax stops were quite variable, spanning a range larger than that for the adults' lax stops (40–85 ms versus 46–76 ms). Their VOT differences also then spanned a larger range, and their 12 ms mean difference value was midway between the means for the two adult groups, and not significantly different from either ( $t = 1.07$ ,  $df = 32.6$ ,  $p = 0.3$ , comparison to women;  $t = -1.52$ ,  $df = 25.7$ ,  $p = 0.1$ , comparison to men). These children also tended to produce small  $F_0$  difference values by comparison to the adults. All but 7 of them had an  $F_0$  difference that was smaller than the mean value for the women, and the mean  $F_0$  difference for the children was 3.6 semitones, a value that is significantly smaller than the 5.0 semitone mean for the women ( $t = -2.67$ ,  $df = 14.5$ ,  $p = 0.02$ ) and substantially (if not significantly) smaller than the 4.7 semitone mean for the men ( $t = -2.0476$ ,  $df = 14.25$ ,  $p = 0.06$ ).

We interpret these results as suggesting that, when children first begin to differentiate the aspirated and lax stops from each other as well as from the short-lag tense stops, they can do so by producing a more conservative 'strongly aspirated' VOT value for the aspirated stop in contrast to a possibly more 'weakly aspirated' lax stop, even if they have not yet acquired a robust pitch difference. In other words, restating this interpretation in terms of the measure of community norms in the figure, native-speaker transcribers attend to the aspiration cues and do not need to hear an unambiguously high pitch after the aspirated stop in contrast to an unambiguously low pitch after the lax stop to perceive the intended phonation type.

We tested this interpretation against the results of an experiment that explored how the patterns in production play out in perception. In the experiment 20 college students at a university in Seoul listened to the initial consonant-vowel sequences extracted from a subset of the stop-initial words in the corpus, and identified each token as tense, lax, or aspirated before rating the goodness of the token as an exemplar of the identified type. The tokens were chosen to sample the full ranges of VOT values for male adults, for female adults, and for children. (See Kong et al. [2011] for further details about the stimuli, methods, and results.)

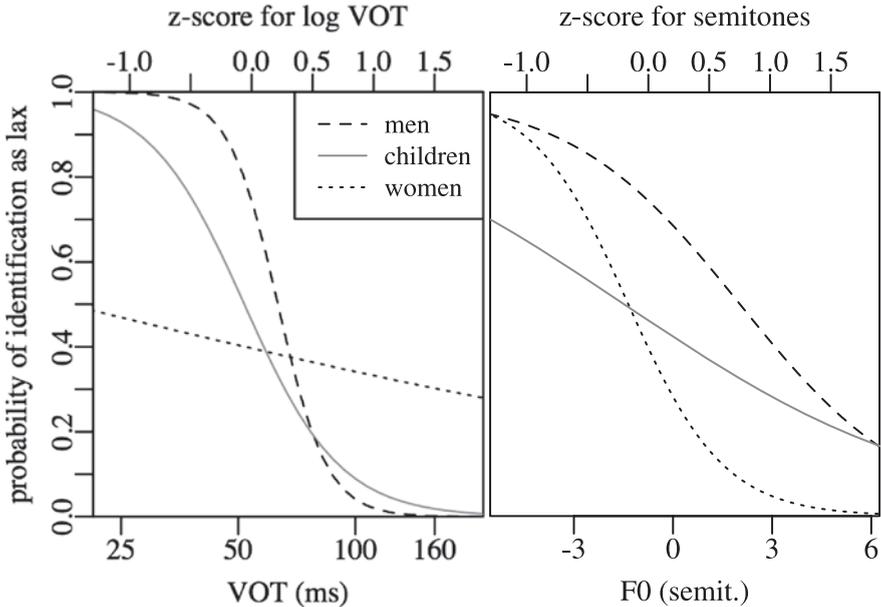
We did a reanalysis of the categorical response data using a single simple model to look only at the responses that identified the stimulus as being lax or as being aspirated, taking ‘lax’ as the ‘hit’ response in a logistic regression. This reanalysis looked only at the effect of VOT and  $F_0$  (i.e., excluding a third acoustic measure, of voice quality, that Kong et al. included in the three logistic regression models that they report). As in the earlier analyses, the VOT values were transformed to log ms, to better reflect what is known about perception of duration. Both the VOT and the  $F_0$  measures were ‘normalized’ to the range of the data by taking the  $z$ -score, but the  $F_0$  values were first converted to semitones relative to the mean  $F_0$  for all of the lax and aspirated stops produced by children, by women, or by men, so that all three talker groups could be included in the same mixed-effects logistic regression (using the `lmer` function in R). The model included the two continuous acoustic variables both as fixed effects and as random (listener-level) slopes. It also included a categorical variable for the type of talker who produced the stimulus (woman vs. man vs. child), using a treatments-style contrast with the women as the reference stimulus talker category, as well as interaction terms for the categorical variable with both acoustic variables.

Table 3 shows the model results, and Figure 5 plots the model curves for the fixed effects. There was a significant main effect of  $F_0$  (evident in the steep curve for the women in the right panel of the figure), but no main effect of VOT (evident in the very shallow curve for the women in the left panel), indicating that listeners attended primarily to the  $F_0$  values in identifying the women’s stops as lax versus aspirated. Moreover, all of the interaction effects were significant (evident in the steeper slopes of the curves for the men and children relative to the slope for the women in the left panel and the shallower slopes for the men and children relative to women in the right panel). The magnitude of the interaction with VOT was substantially larger than the magnitude of the interaction with  $F_0$  (a bigger difference between the men’s curve and the reference women’s curve in the left panel than in the right, and similarly for the children’s curves), indicating that listeners did attend to the VOT when identifying men’s or children’s stops as lax versus aspirated in a way that they did not when listening to women. In fact,

**Table 3:** Fixed effects and goodness-of-fit for a mixed-effects logistic regression model with the listener’s judgment of the stimulus as lax (rather than aspirated) for 5,089 trials distributed over 20 listeners.

Coefficient	Estimate	Std. err.	z value	Pr(> z )
intercept (talker = woman)	-0.4199	0.624	-0.673	0.5
stimulus talker = man	1.4879	0.668	2.229	0.03
stimulus talker = child	0.2666	0.535	0.498	0.6
VOT (z-score of log ms)	-0.2775	0.958	-0.290	0.8
$F_0$ (z-score of semitones)	-2.5491	0.303	-8.403	<0.001
VOT: man	-4.6882	1.274	-3.679	<0.001
VOT: child	-2.2715	0.948	-2.398	0.02
$F_0$ : man	1.1301	0.380	2.976	0.003
$F_0$ : child	1.7787	0.304	5.847	<0.001

AIC: 4462; BIC: 4560; log-likelihood: -2216; deviance: 4432



**Fig. 5:** Model curves from a mixed-effects logistic regression analysis applied to all of the responses in Kong et al. (2011) other than those that identified the stimulus as a tense stop. See Table 3 for model details.

each of the men's and children's curves in the left panel is also steeper than the corresponding curve in the right panel, indicating that listeners attended more to the VOT than to the  $F_0$ .

These patterns are consistent with the predictions from the incrementation model. That is, the significant main effect for  $F_0$  but not for VOT is in keeping with the predicted effect on perception of the role of women as the leaders in this Neogrammarians shift in the phonetic basis for the contrast. The larger effect of VOT than of  $F_0$  for men, too, is in keeping with the predicted effect on perception of men's generally more conservative values in a sound change led by women, and the larger effect of VOT than of  $F_0$  for children could also be interpreted in the same way. On the other hand, the slopes for the responses to the children's productions were shallower than those for the responses to the men's productions in both panels, suggesting that the children's values were not as reliable as the men's in differentiating the two phonation types for the listeners. The perceptual effects, then, seem to reflect not just the generational shift in the perception of the age grading of VOT versus  $F_0$  as this female-led sound change nears completion, but also listener expectations about the unreliability of both measures as phonetic cues to the intended phonation type in productions by very young children.

## 6 Borrowing of Běijīng features in the Sōngyuán sibilant system

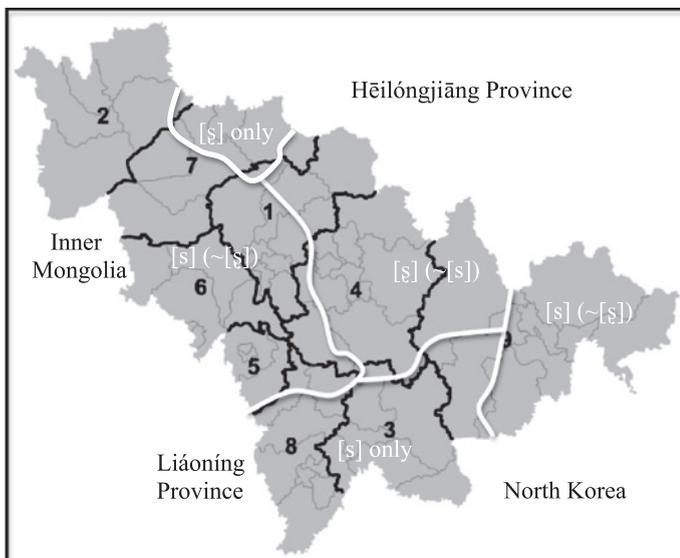
The other speech community we examine is Sōngyuán City, a municipal county in the western part of Jílín Province, at the center of the Dōngběi ('east north') region of China. The phonological change of interest here affects the pronunciation of the sibilant consonants, in a way that makes the Sōngyuán variety of Mandarin sound more like the Běijīng dialect. Specifically, it is a set of phoneme splits. The Běijīng dialect, which is the basis for the Pǔtōnghuà ('common language') standard, has a 3-way contrast among an alveolopalatal series ([ç, tç, tçʰ], Pīnyīn 'x', 'j', 'c'), an apical or 'retroflex' series ([ʂ, tʂ, tʂʰ], Pīnyīn 'sh', 'zh', 'ch'), and a dental series ([s, ts, tsʰ], Pīnyīn 's', 'z', 'c'). Until the 1980s, the Sōngyuán variety differed in having only a two-way contrast between an alveolopalatal series (in words that have standard cognates with alveolopalatal sounds) and a single more apical series (in words that have standard cognates with either retroflex or dental sounds). The phonetic value of this single 'apical' series varied across Jílín Province, with a more retroflex quality in Sōngyuán City and neighboring rural areas. However, Li (2005) provides evidence that this 'apical' series

has now split in a way that projects the standard contrast between retroflex and dental onto what had been a single set of word classes, in what is clearly a ‘change from above’ – i.e., a borrowing of the system of lexical contrasts from the standard variety.

The larger socio-historical context for this change begins with the recent settlement history of Jílín Province. Today, Han Chinese are the majority, far outnumbering the Manchu, Korean, Mongol, and Hui ethnic groups. However, for most of the Qīng dynasty, this part of the Dōngběi region was the sparsely populated official homeland of the Manchu rulers, and ethnic Han were allowed to enter the area only as itinerant peddlers, servants in border garrisons, serfs on Manchu fiefdoms, and the like. The present-day Han dominance began only when the Qīng government lifted these restrictions to promote the Chuǎng Guāndōng (‘Guandong Rush’), a huge influx of farmers, miners, and then railroad workers, which swelled the population of Jílín Province from fewer than 350,000 in 1850 to nearly 4 million in 1910 (see, e.g., Gottschang and Lary 2000; Reardon-Anderson 2005). A primary source of the Han immigrants in the Guandong Rush was Shāndōng Province, so Dōngběi Chinese is a variety of the Northern Mandarin dialect group that includes Shāndōng Mandarin.

Because the Han dominance in the region is so recent, the Dōngběi Chinese seem to lack the very strong local identity that has led to the richly variegated village-by-village differences that are characteristic of other major dialect groups such as Jiāngsū Mandarin. However, there was apparently enough variation in the settlement history, in addition to possible adstrate influences across the Dōngběi, to have resulted in some relevant regional variation. Specifically, Sūn, Lù, and Lǐ (1986) map the realization of the more apical sibilant fricative as varying across Jílín Province, as shown in Figure 6, with predominately [s] in the counties that border Korea, Liaoning Province, and Inner Mongolia to the south and west but predominately [ʃ] in the Sōngyuán metropolitan district and other neighboring districts that are closer to the center of the historically Manchu-speaking area.

Sūn and colleagues chose the variables to survey for their maps on the basis of a more thorough three-volume dialect survey that was conducted in the late 1950s, after the government of the newly established People’s Republic of China held a Symposium on the Standardization of the Modern Chinese Spoken Language at which the standard language was defined as Pǔtōnghuà, “a common speech with pronunciation based on the Beijing dialect” – a definition that has been maintained in the current law (Jiang 2000). The Ministry of Education then issued a set of Directives Regarding the Promotion of the Common Language in Primary, Secondary, and Normal Schools, which mandated the use of Pīnyīn Romanization in teaching Pǔtōnghuà in all five years of



**Fig. 6:** Distribution of pronunciations of the anterior sibilant fricative in the Sūn, Lù, and Lǐ (1986: 45) dialect survey map superimposed on a map of administrative divisions in Jílín Province ([https://en.wikipedia.org/wiki/File:jilin\\_prfc\\_map.png](https://en.wikipedia.org/wiki/File:jilin_prfc_map.png)). Zone 4 is the division of the provincial capital, Jílín City, and zone 7 is the Sōngyuán City administrative division. The metropolitan area proper is the Níngjiāng district, the part at the northern edge of zone 7 that is at the center of the ‘[ʃ] only’ region on the dialect survey map.

primary school and included provisions for establishing teacher-training courses in pronunciation as well as pedagogy at government-funded centers such as Jílín Teachers’ College (see Cheng 1979; Yin and Baldauf 1990; Chen 1999: Chapter 2). Although there was a hiatus when the “Cultural Revolution from 1966 to 1976 dramatically ruined the education system” (Wang 2012: 60), the Ministry of Education then implemented a wide-scale curricular reform in the 1980s, which included a heavy re-investment in local teacher training and local documentation efforts such as the Sūn et al. survey. As a result of this reinvestment, Yin and Baldauf (1990: 285) could state a decade later, “Most students graduating from primary schools can speak *Putonghua* fluently, and they will continue to speak *Putonghua* if they go to secondary schools or universities for further study.”

Note that in non-Mandarin regions, speaking Pǔtōnghuà meant the spread of societal bilingualism into rural areas and the emergence (or reinforcement) of regional urban L2 Mandarin accents, which could take on the same social meanings that speaking the regional language has always had. For example, Shànghǎi-

accented Mandarin, like the neighboring Nánjīng dialect, has dental [s, ts, tsʰ] for both the dental and the retroflex series, and use of the retroflex variant is now associated with middle class, younger speakers, as well as with a less strongly Shanghainese cultural identity (see, e.g., Zhu 2012). By contrast, in already Mandarin-speaking regions, speaking Pǔtōnghuà typically means something more like eschewing local words that do not have standard cognates and adopting more standard pronunciations for those that do. While many sociolinguistic studies of what it means to speak Pǔtōnghuà today have focused on the first case, there are a few studies of the second. For example, Zhang (2005: 453) reports that in a study of the Tiānjīn dialect that she did in 1996, “many respondents contended that the ability to speak MSM [Mainland Standard Mandarin] indicated that the speaker was educated or cultivated.”

This ‘educated’ or ‘cultivated’ social evaluation of Běijīng-based ‘standard’ pronunciations in the 1980s and 1990s is the basis for the phonological change that we have noted in the Sōngyuán sibilants. Figure 7 shows data from Li (2005), documenting the split. Li recorded a list of words exemplifying the Pǔtōnghuà contrast between dental [s, ts, tsʰ] and retroflex [ʂ, tʂ, tʂʰ] as produced by three groups of speakers, each containing 3 men and 3 women. The older Sōngyuán dialect speakers were aged between 45 and 50 years at the time of the recording (in 2002), so they are in the generation who came of age during the Cultural Revolution. The younger Sōngyuán dialect speakers were aged between 18 and 25 years, so they began primary school after the curricular reforms of the 1980s had begun. And the Běijīng dialect speakers that Li recorded as a control were also young adults. We reanalyzed these productions for the current paper, using two measures that Li (2008) found to differentiate among all three sibilant types in fricative productions that she elicited from a larger group of young Sōngyuán adults for the Paidologos corpus: the centroid frequency of a spectrum estimated over the middle 40 ms of the portion of turbulence and the frequency of the second formant at the onset of energy in that formant in the following vowel.

Figure 7 shows the values we measured for all of the productions of retroflex and dental targets in the Li (2005) recordings, with the circles versus triangles differentiating the two series. We fit a mixed-effects logistic regression model to these data, with random intercepts by talker, to see how well the two measures predicted whether the target was dental rather than retroflex (i.e., the independent variable was the categorical variable of word class in Pǔtōnghuà with ‘dental’ as the ‘hit’ value). The fixed effects also included two categorical variables, for talker gender (female versus male) and for dialect/age group (Běijīng versus younger Sōngyuán versus older Sōngyuán), with a treatments-style contrast that made the Běijīng speakers the reference category. The model showed

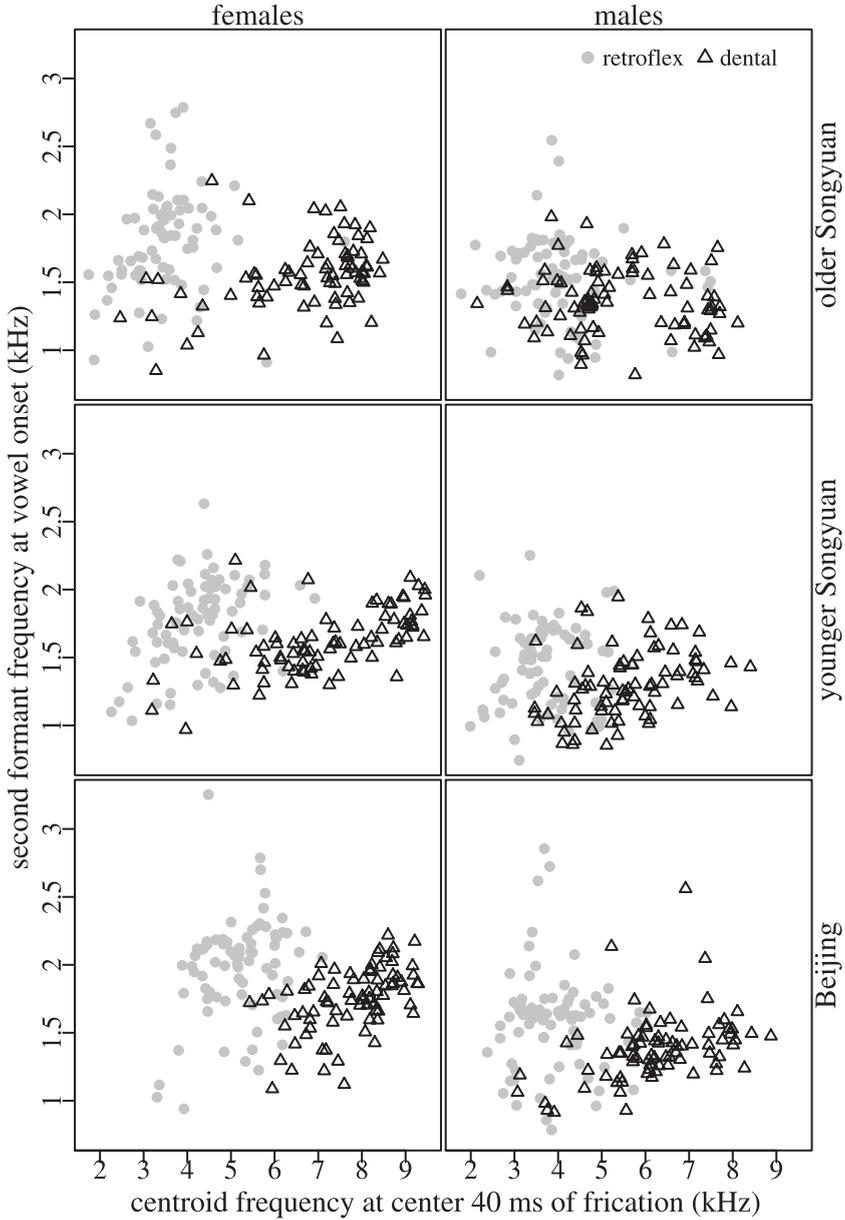
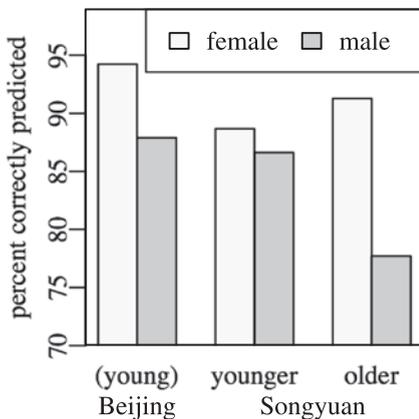


Fig. 7: Frequency of F2 at vowel onset as a function of centroid frequency in all of the productions recorded by Li (2005).

**Table 4:** Fixed effects and goodness-of-fit for a mixed-effects logistic regression model predicting the Pǔtōnghuà word class as dental (rather than retroflex) for 935 tokens distributed over 18 talkers.

Coefficient	Estimate	Std. err.	z-value	Pr(> z )
intercept (talker = Běijīng)	-1.3541	0.3921	-3.453	<0.001
talker = Sōngyuán younger	1.4170	0.4888	4.926	0.004
talker = Sōngyuán older	2.3449	0.4760	2.899	<0.001
centroid (kHz)	1.9118	0.1569	12.188	<0.001
F2 at vowel onset (kHz)	-3.6574	0.5087	-7.190	<0.001

AIC: 596; BIC: 649; log-likelihood: -287; deviance: 574



**Fig. 8:** Prediction accuracy by talker group from the mixed-effects logistic regression model shown in Table 4, which predicts the target category from the acoustic measures plotted for these 6 talker groups in Figure 7.

significant effects of both acoustic variables and significant differences between both Sōngyuán groups and the Běijīng group. However, there was no significant effect of gender, possibly because there were so few talkers in each cell defined by the two categorical variables, so we replaced this initial model with a simpler model that did not include talker gender. Table 4 shows the results for this second model, and Figure 8 shows its prediction accuracy.

As Figures 7 and 8 show, the two acoustic measures differentiated the dental from the retroflex targets with fairly good accuracy for the Běijīng dialect speakers' productions. Specifically, the dental targets had somewhat lower F2 values and considerably higher centroid values, which separated them fairly robustly from the values for the retroflex targets. The same was true of the female Sōngyuán speakers' productions, for both age groups. The younger male Sōngyuán

speakers' productions also showed a fairly clear separation between the two targets, although the dentals showed somewhat lower centroid frequencies than those in the male Běijīng speakers' productions, suggesting a somewhat less anterior, more alveolar place of constriction. The older male Sōngyuán speakers' productions, on the other hand, showed considerable overlap, with many dental targets having centroid frequencies as low as or even lower than the median value for the retroflex sounds and some retroflex targets having high centroid values, suggesting hypercorrection, so that the model predicted the correct target in fewer than 80% of their productions. We interpret these results as evidence that, between the two generations of talkers plotted here, the educated urban Sōngyuán dialect borrowed standard pronunciations for enough words to split the single non-alveopalatal series into the two series of phonemically contrasting retroflex versus dental sibilants.

Li's (2008) analysis of 10 of the adults in the Sōngyuán Mandarin portion of the Paidologos corpus provides further evidence for this split. Li measured the centroid frequency and F2 onset frequency in these adults' productions of the target sibilant fricatives elicited in the picture-prompted word-repetition task. There was no overlap at all between the values for the retroflex and for the dental fricatives in the centroid dimension for any of the 5 men and 5 women whose fricatives were analyzed. We redundantly confirmed this lack of overlap by applying a mixed-effects logistic regression model that predicted whether the target was a dental as opposed to a retroflex (i.e., the same dependent variable as in the model in Table 4), with centroid frequency and talker gender as fixed effects and random intercepts and centroid frequency slopes at the individual talker level. The model resulted in a false convergence warning because of this complete lack of overlap between the two categories, and prediction accuracy was at 100% for all 10 talkers.

In summary, then, there is strong evidence that there have been changes in the pronunciations of words that have cognates in the dental word class in the standard Běijīng-based lexicon. These changes have effected a switch from what had been a two-way contrast between alveopalatal and just one other, more apical (retroflex), category to a three-way contrast among alveopalatal, retroflex, and dental. Moreover, this switch was very abrupt. It was accomplished in the space of one generation, at least for educated speakers like the ones that Li recorded in 2002 for her apparent-time study and in 2006 for the Paidologos corpus. In the next section we describe how the transmission of this abrupt 'change from above' is reflected in the phonological development of children whose grandparents are in the generation before the change (same age as the older speakers in Li 2005) but whose parents are in the generation after the change (same age as the younger speakers in Li 2005).

## 7 Acquisition of the Sōngyuán retroflex versus dental fricative contrast

By comparison to the wealth of studies on the acquisition of different stop phonation types across languages, the literature on the acquisition of different sibilant consonants is very small. At the same time, the little information available suggests that acquisition of sibilant place contrasts is more or less late relative to the acquisition of aspiration contrasts and also highly variable across different languages. For example, both English and Japanese contrast a more posterior post-alveolar or alveopalatal [ʃ] with a more anterior alveolar or dental [s]. However, whereas most English-learning children make acceptable productions of [s] before they are 5 years of age, many Japanese-learning children master [s] at a later age. Also, whereas a stereotypical error for English-learning children is the transcribed substitution of an [s] for target [ʃ], the stereotypical error for Japanese-learning children is the substitution of an alveopalatal fricative or affricate for a target [s] (see Li et al. [2009] and the earlier literature reviewed there).

Moreover, the literature on acquisition of Mandarin phonology suggests that there can be comparably large differences even across different dialects of the same language. For example, Zhu's studies of phonological development in Běijīng show a relatively early mastery of all three sibilant fricatives. In one large cross-sectional study (Zhu and Dodd 2000; Zhu 2000: Chapter 3), for example, Zhu reports that 75% of the children had "stabilized" [ç] by age 3, and 75% had "stabilized" [ʃ] and [s] before age 5, with transcribed substitutions of [ʃ] and [ç] for [s] as well as of [s] for [ʃ] in the interval before the majority of children have mastered the later contrast. This pattern differs substantially from the developmental progression and substitution patterns that Shih (2012) documents for children in Kaohsiung (in the south of Taiwan). In Kaohsiung, children simultaneously acquire a Southern Min Chinese dialect that has only one sibilant phoneme (a dental that alternates with an alveopalatal in front vowel environments) and a variety of Mandarin that is like the Shāndōng, Nánjīng, and Tiānjīn varieties in not having a robust retroflex-dental contrast. In their Mandarin productions, these children are transcribed as being correct for the alveopalatal and dental fricatives first, and they substitute [s] for target [ʃ] until after they enter primary school, where they begin to be explicitly taught retroflex pronunciations for words that are in the retroflex class in the Běijīng dialect. That is, they do not begin to produce [ʃ] until they are at least 6, at which time they begin to make some [ʃ] for [s] substitutions as they learn to read Mandarin (in textbooks written using hànzì logographs with zhùyīn fúhào symbols as an interlinear annotation of the consonants, vowels, and tones).

Shih interprets this [ʂ] for [s] substitution pattern as a hypercorrection. This interpretation is supported by the substitution patterns that she observed in the older two of the three groups of adults that she recorded. That is, Shih combined her developmental study with an apparent-time study of young college-age adults, middle-aged adults from the first generation of Taiwanese to have compulsory primary school education in Mandarin, and older adults from the generation that came of age just after the Japanese surrender in 1945 that transferred sovereignty to the Republic of China and so were taught Japanese rather than Mandarin as the language of literacy if they attended primary school. (Many of these oldest speakers were, in fact, illiterate.) The oldest group produced only [ʂ] for both dental and retroflex targets, but the middle-aged group was more variable, producing some tokens of [ʂ] for dental targets as well as many tokens of [s] for retroflex targets.

Note that Shih observed these substitution patterns even though the productions were repetitions of accurately produced targets. That is, she used the same picture-prompted word repetition task as Li (2008) and even based her word list on Li's word list. Also, Shih applied the same acoustic measures as Li (2008), and the spectral patterns support her transcriptions. What does this predict for the Sōngyuán portion of the Paidologos corpus?

Recall that before the phonological change described in the previous section, Sōngyuán Mandarin was like the Shāndōng, Nánjīng, and Tiānjīn regional dialects and also like the Shànghǎi- and Taiwan-accented varieties of L2 Mandarin in not having a contrast between retroflex and dental fricatives, although the quality of this single apical category in Sōngyuán Mandarin was retroflex rather than dental as in those other varieties. If the Sōngyuán variety still had the original two-sibilant system, with a contrast only between this single apical sound and the alveolopalatal, then we might expect children in Sōngyuán to show a developmental pattern that is like the one that Shih (2012) documented for children in Kaohsiung. That is, we might expect children to master the contrast between the alveolopalatal and the native-dialect apical sound (which is [ʂ] rather than [s] in Sōngyuán) relatively early, through transmission from their parents, but we would not expect them to master the other apical sound (which is [s] rather than [ʂ] in Sōngyuán) until they enter primary school and begin to be taught the standard pronunciations for words in the dental class as they learn to read (in textbooks using first Pinyin alone and then using hànzì logographs with Pinyin annotations).

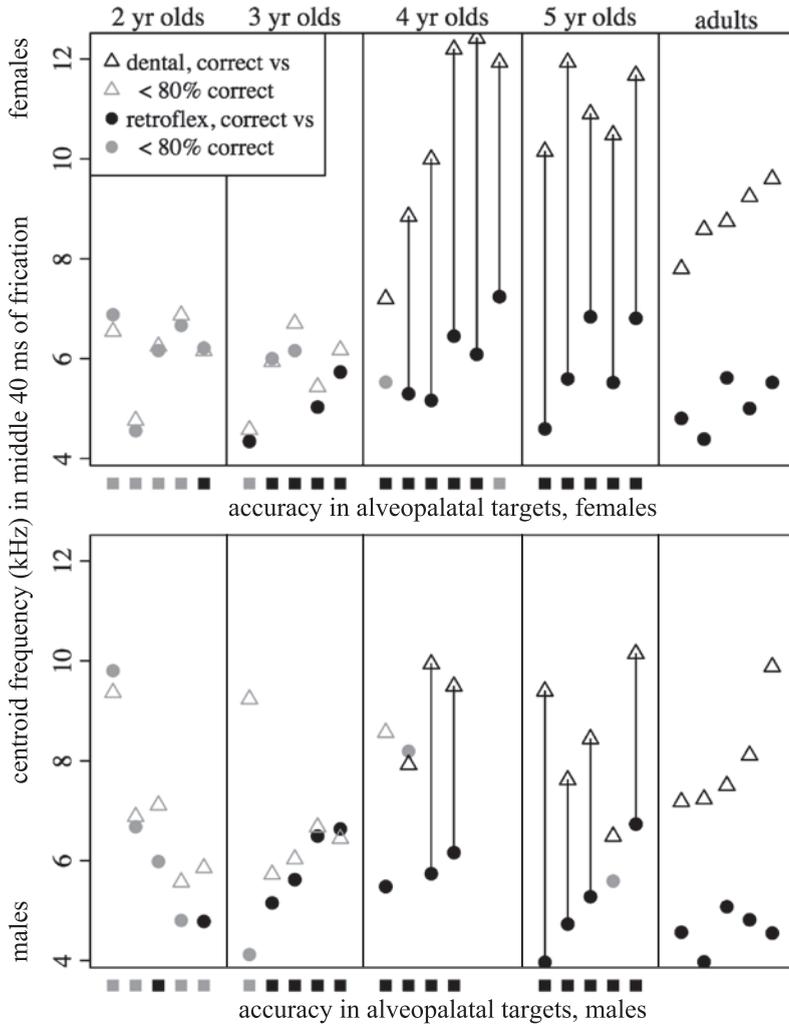
However, this is not what Li (2008) observed. Instead, she describes the Sōngyuán children as showing mastery of the borrowed apical [s] as well as of the originally native apical [ʂ] by the time they are 4, which is, if anything, even younger than the age at which Zhu and colleagues documented mastery of these

two fricatives by the children in Běijīng, where there is a robust contrast between retroflex and dental.

Figure 9 shows Li's evidence for this description, which is based on half of the 80 children in the Sōngyuán portion of the Paidologos corpus. The figure combines the results of her transcription analysis with some of the results of the same kind of spectral analysis that we presented for adult speakers in Figure 7. Specifically, the figure plots the mean centroid values for the retroflex versus dental fricative targets for each of the 21 girls and 19 boys (as well as for the 10 adults that Li analyzed), using the same convention as in Figure 7 of circles for the retroflex target versus triangles for the dental target. Also, the black versus gray is the same convention as in Figure 4 for indicating whether a child was transcribed as producing an acceptable pronunciation in at least 80% of the tokens of that fricative target. The same black versus gray in the row of square plotting symbols below each panel of the figure encodes the transcribed accuracy of the children's productions of the alveopalatal target. As the figure shows, all but two of the 3-year-olds are producing the alveopalatal target accurately, but not one of them is producing the dental target accurately. Moreover, nearly all of the children in the two youngest groups have low mean centroid values for their dental targets, in accord with the transcribed alveopalatal or retroflex substitutions. In the older two groups, by contrast, the dental targets tend to have much higher centroid values, and all but four of these children are transcribed as having mastered the retroflex-dental contrast.

We applied the same kind of mixed-effects logistic regression model that we had applied to adult productions in Section 6 to the productions by the twenty 4- and 5-year-old children reported in Figure 9. As in the model applied to the values for the 10 adults in the Paidologos corpus (whose mean centroid values are shown in the rightmost panels of Figure 9), the model for the children's productions resulted in a false convergence warning because there was a complete lack of overlap between the two categories for 6 of these children. Moreover, prediction accuracy for 12 of these older children was at least as good as the prediction accuracies for the younger Sōngyuán adults (plotted in the middle two bars of Figure 8) from the model applied to the values shown in Figure 7 for the apparent-time study reported in Li (2005).

While this pattern of acquisition is very different from the pattern reported for the Kaohsiung children, it is virtually identical to the pattern reported for the Běijīng children, except for the even earlier mastery of the retroflex fricative in the Sōngyuán children compared to the Běijīng children. Thus, the evidence from the Paidologos corpus suggests that the system being transmitted to the Sōngyuán children is a new system with a robust contrast between the original native retroflex fricative and an imported standard dental fricative, and not the older



**Fig. 9:** Mean centroid value measured from a spectrum estimated over a 40 ms window in the center of the fricative for the target dental and the target retroflex fricatives elicited from each of the child and adult talkers in the Paidologos corpus that were measured by Li (2008). Values are grouped by gender (females in the upper panel, males in the lower), with individuals within each gender group ordered by age for children and by mean centroid in the dental for adults. Black versus gray plotting symbols indicate the transcribed accuracy of the targets for the children, with the row of square plotting symbols below each panel indicating accuracy for their alveopalatal targets. Also, black lines connect the two data points for children who have mastered the retroflex-dental contrast.

Sōngyuán system with only the single apical sound in contrast with the alveolo-palatal or some ‘mixed’ system that is intermediate between the two-sibilant system and the robust three-sibilant system.

## 8 Conclusions and caveats

Pulling together the results presented in Sections 7 and 5, then, we note the following differences that are in accord with our predictions about differing alignments between the timeline for development and the timeline for each type of phonological change. The Sōngyuán children show a developmental progression toward accurate retroflex and dental fricatives that is similar to the progression documented for children in Běijīng, and that differs markedly from the progression documented for the Mandarin productions of children in Kaohsiung. The distribution of their centroid values also looks much more like the distribution in the younger Sōngyuán speakers than it looks like the distribution in older men in Li (2005). This suggests an abrupt discontinuity in transmission patterns in the aftermath of a phoneme split due to dialect borrowing. By contrast, the Korean children show VOT values for the aspirated stops and  $F_0$  difference values for the lax versus aspirated stops that are more like the men’s. This suggests a continuity with previous generations such that the system that is being transmitted to them is one that is slightly less advanced than the system that these children will create once they are old enough to participate in the stage of incrementation appropriate for their cohort.

Before we draw any more sweeping generalizations from these suggested interpretations, however, some other potential explanations and differences also need to be considered. First, Labov’s model of incrementation is based primarily on shifts in the pronunciation of vowels, which are among the first sounds that infants are able to control. Although Tagliamonte and D’Arcy (2009) have argued for the model as applicable to proportional frequencies of discretely differing variants for language changes more generally, we know of no prior examples of applications of the incrementation model to the kinds of continuous acoustic evidence for changes in consonant sounds that we examine here. So we need to be cautious in applying the model to interpret the real-time and apparent-time data presented in Section 4 as evidence of a Neogrammarian sound change by incrementation and then contrasting that interpretation to our interpretation of the patterns in the apparent-time data presented in Section 6, which we take as evidence instead of an abrupt phonological change by borrowing.

This need for caution is compounded when we then contrast the developmental trajectories described in Sections 5 and 7, because the two types of

consonant contrast that we compare are also at different points along the scale of motor difficulty. Stop constrictions are produced by simple ballistic gestures and are among the first consonants to be produced by infants in canonical babbling at 6–8 months. Control of fundamental frequency is also mastered at a very young age, so that Korean-learning children could, in theory, begin to differentiate lax stops from the other two types by the pitch pattern before they master the more challenging property that distinguishes the Korean lax and aspirated stops, which is the control of temporal coordination between oral and laryngeal gestures to produce two subtly different long-lag VOT categories. The Korean phonation type contrast is thus more motorically challenging than vowel contrasts, but less motorically challenging than the control of lingual postures to produce voiceless sibilants with contrasting spectral shapes. Where 13 out of 20 three-year-olds in the Seoul Korean sample are transcribed as having at least 80% accuracy for the phonation type for both the lax and aspirated stops, not one of the 3-year-old children in the Sōngyuán Mandarin sample is transcribed as having more than even 50% accuracy for [s], and it is only in the 4-year-olds that mastery of this sibilant contrast reaches the levels observed for the phonation type contrast in the Korean 3-year-olds.

This difference in age at which the children master the relevant lexical contrast interacts with another apparent difference for which we have only qualitative observations to date. The specific model of incrementation that Labov (2001) presents assumes that until children are about 5 years of age, their pronunciations will reflect an older stage of a sound change in progress. This is because the source of early transmission is the speech patterns of their primary caretakers, and children do not begin to participate in the social negotiations that drive the next stage of incrementation until their social networks expand beyond the immediate family circle, a point which Labov sets at about 5 years for the sake of developing the model. It is not clear how we could adapt this aspect of the model to test it against the Sōngyuán data, given that even the youngest children in the Sōngyuán sample were recorded at school. In Sōngyuán, many women work, and it seems that their children are placed in government-funded nursery schools, where they can be cared for all day, at age 2 or 3 years. By contrast, many of the 2-year-olds in the Seoul sample were recorded in their homes, because most children in Seoul are not placed in (private) nursery schools until 3 years of age at the earliest, and they tend to spend less time each day at school. We have recordings of primary caretakers interacting with very young children in both communities, and are currently in the process of analyzing those recordings to quantify production patterns in child-directed speech. However, we unfortunately have no recorded productions by the children's nursery school teachers in either speech community. So we will not be able to compare the production patterns in the two

potentially different adult models for the children at home and at school until we can return to do more fieldwork.

An interim conclusion from this comparison of the youngest speakers of two different languages with two different phonological changes in progress, then, is that we see a very large qualitative difference between the two cases, which we attribute to the effects of acquiring a system that is undergoing a gradual change from below versus a very abrupt change from above. However, we need considerably more detailed ethnographic information on language socialization patterns in early childhood in the two speech communities before we can be confident of our interpretations of these two cases. And we also need more case studies of developmental patterns in other communities where there are changes affecting consonant contrasts like the ones studied here before we can extend the interpretations to be a more confident general hypothesis about differing alignments to the timelines for incremental regular sound change versus potentially abrupt borrowing.

**Acknowledgments:** Portions of the work reported here were supported by NSF Grant BCS0739206 to Mary Beckman, by an Ohio State University Department of Linguistics Targeted Investment in Excellence award to Fangfang Li and Eun Jong Kong, and by NIDCD grant 02932 to Jan Edwards. We thank Jeffrey J. Holliday for lending us the raw data from Holliday and Kong (2011, 2013) to build the third model reported in Table 2.

## References

- Ahn, Hyunkee. 1999. Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. Austin, TX: University of Texas dissertation.
- Cameron, Richard. 2010. Growing up and apart: Gender divergences in a Chicagoland elementary school. *Language Variation and Change* 22. 279–319.
- Chambers, J. K. 1992. Dialect acquisition. *Language* 68(4). 673–705.
- Chen, Ping. 1999. *Modern Chinese: History and sociolinguistics*. Cambridge: Cambridge University Press.
- Cheng, Chin-chuan. 1979. Language reform in China in the seventies. *Word* 30. 45–57.
- Cho, Taehong, Sun-Ah Jun, & Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30. 193–228.
- Clark, Eve, & Melissa Bowerman. 1986. In Joshua A. Fishman, A. Tabouret-Keller, M. Clyne, B. Krishnamurti, & M. Abdulaziz (eds.), *The Fergusonian impact: In Honor of Charles A. Ferguson on the occasion of his 65<sup>th</sup> birthday*. Vol. 1 *From phonology to syntax*, 51–68. Berlin: Mouton de Gruyter.
- Eckert, Penelope. 1997. Age as a sociolinguistic variable. In Florian Coulmas (ed.), *The handbook of sociolinguistics*, 151–167. Oxford, UK: Blackwell Publishers.

- Edwards, Jan, & Mary E. Beckman. 2008. Methodological questions in studying phonological acquisition. *Clinical Linguistics and Phonetics* 22(12). 937–956.
- Edwards, Jan, Mary E. Beckman, & Benjamin Munson. 2012. Cross-language differences in acquisition. Manuscript in review for publication in Melissa A. Redford (ed.), *Handbook of speech production*. Malden, MA: Blackwell.
- Ferguson, Charles A., & Carol B. Farwell. 1975. Words and sounds in early language acquisition. *Language* 51. 419–439.
- Foulkes, Paul, & Marilyn M. Vihman. In press. Language acquisition and phonological change. In Patrick Honeybone & Joseph Salmons (eds.), *The Oxford handbook of historical phonology*. Oxford: Oxford University Press.
- Gottschang, Thomas R., & Diana Lary. 2000. *Swallows and settlers: The great migration from North China to Manchuria*. Ann Arbor: Center for Chinese Studies, University of Michigan.
- Grammont, Maurice. 1902. Observations sur le langage des enfants. In D. Barbelenet & Paul Jean Marie Boyer (eds.), *Mélanges Linguistiques offerts à M. Antoine Meillet*, 61–82. Paris: C. Klincksieck.
- Guy, Gregory. 1980. Variation in the group and the individual: The case of final stop deletion. In William Labov (ed.), *Locating language in time and space*, 1–36. New York: Academic Press.
- Han, Mieko, & Raymond S. Weitzman. 1965. *Studies in the phonology of Asian languages III: Acoustic characteristics of Korean stop consonants*. Los Angeles: Acoustic Phonetics Research Laboratory, University of Southern California.
- Hardcastle, William J. 1973. Some observations on the tense-lax distinction in initial stops in Korean. *Journal of Phonetics* 1(3). 263–272.
- Hockett, Charles F. 1950. Age-grading and linguistic continuity. *Language* 26(4). 449–457.
- Hockett, Charles F. 1965. Sound change. *Language* 41(2). 185–204.
- Hoenigswald, Henry M. 1978. The annus mirabilis 1876 and posterity. *Transactions of the Philological Society* 76(1). 17–35.
- Holliday, Jeffrey J., & Eun Jong Kong. 2011. Dialect variation in the acoustic correlates of Korean stops. *Proceedings of the 17<sup>th</sup> International Congress of Phonetic Sciences*. 878–881.
- Holliday, Jeffrey J., & Eun Jong Kong. 2013. Cross-generational acoustic correlates of Jeju Korean stops. Manuscript under review, Indiana University and Korea Aerospace University.
- Jakobson, Roman. 1941. *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist and Wiksell.
- Jiang, Zemin. 2000. Law of the People's Republic of China on the Standard Spoken and Written Chinese Language (Order of the President No. 37). [http://www.gov.cn/english/laws/2005-09/19/content\\_64906.htm](http://www.gov.cn/english/laws/2005-09/19/content_64906.htm) (accessed 29 December 2012).
- Jun, Sun-Ah. 1998. The Accentual Phrase in the Korean prosodic hierarchy. *Phonology* 15(2). 189–226.
- Kang, Hyeon-Seok. 1997. Phonological variation in glides and diphthongs of Seoul Korean: Its synchrony and diachrony. Columbus, OH: The Ohio State University dissertation.
- Kang, Kyoung-Ho, & Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *Journal of the Acoustical Society of America* 124(6). 3909–3917.
- Kewley-Port, Diane, & Malcolm S. Preston. 1974. Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2. 195–210.
- Kim, Chin-Woo. 1965. On the autonomy of the tensivity feature in stop classification (with special reference to Korean stops). *Word* 21. 339–359.

- Kim, Mi-Ran Cho. 1994. Acoustic characteristics of Korean stops and perception of English stop consonants. Madison, WI: University of Wisconsin dissertation.
- Kim, Youngho. 1995. Acoustic characteristics of Korean coronal stops, affricates and fricatives. Arlington, TX: University of Texas at Arlington thesis.
- Kiparsky, Paul. 1968. Linguistic universals and linguistic change. In Emmon Bach & Robert T. Harms (eds.), *Universals in linguistic theory*, 171–202. New York: Holt, Rinehart, and Winston.
- Kiparsky, Paul. 2008. Universals constrain change; Change results in typological generalizations. In Jeff Good (ed.), *Linguistic universals and language change*, 23–53. Oxford: Oxford University Press.
- Kong, Eun Jong, Mary E. Beckman, & Jan Edwards. 2011. Why are Korean tense stops acquired so early: The role of acoustic properties. *Journal of Phonetics* 39(2). 196–211.
- Koops, Christian, Elizabeth Gentry, & Andrew Pantos. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics* 14(2). 93–101.
- Labov, William. 2001. *Principles of linguistic change, Vol. 2. Social factors*. Malden, MA: Blackwell.
- Labov, William. 2007. Transmission and diffusion. *Language* 83(2). 344–387.
- Lee, Ki-Moon, & S. Robert Ramsey. 2012. *History of the Korean Language*. Cambridge, UK: Cambridge University Press.
- Li, Fangfang. 2005. The production and perception of dental vs. retroflex sibilants in the Songyuan dialect of Northeastern Mandarin Chinese. Poster presented at NNAV 34, New York.
- Li, Fangfang. 2008. The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese. Columbus, OH: The Ohio State University dissertation.
- Li, Fangfang, Jan Edwards, & Mary E. Beckman. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics* 37. 111–124.
- Lisker, Leigh, & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422.
- Martin, Samuel E. 1951. Korean phonemics. *Language* 27(4). 519–533.
- Menn, Lise. 1975. Counter example to ‘fronting’ as a universal of child language. *Journal of Child Language* 2. 293–296.
- Oh, Eunjin. 2011. Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics* 39. 59–67.
- Reardon-Anderson, James. 2005. *Reluctant pioneers: China’s expansion northward, 1644–1937*. Stanford: Stanford University Press.
- Roberts, Julie Lee. 1994. Acquisition of variable rules: (-t,d) deletion and (ing) production in preschool children. Philadelphia, PA: University of Pennsylvania dissertation.
- Roberts, Julie. 2002. Child language variation. In J. K. Chambers, Peter Trudgill, & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*, 333–348. Malden, MA: Blackwell.
- Sankoff, Gillian. 2004. Adolescents, young adults and the critical period: Two case studies from Seven Up. In Carmen Fought (ed.), *Sociolinguistic variation: Critical reflections*, 121–139. Oxford: Oxford University Press.
- Shih, Ya-ting. 2012. Taiwanese-Guoyu bilingual children and adults’ production patterns of sibilants. Columbus, OH: The Ohio State University dissertation.

- Shimizu, Katsumasa. 1989. A cross-language study of voicing contrasts of stops. *Studia Phonologica* 23. 1–12.
- Silva, David J. 2002. Consonant aspiration in Korean: A retrospective. In Sang Oak Lee & Gregory K. Iverson (eds.), *Pathways into Korean language and culture: Essays in honor of Young-Key Kim-Renaud*, 387–410. Seoul: Pagijong Press.
- Silva, David J. 2006. Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology* 23. 287–308.
- Stampe, David. 1979 [1973]. *A dissertation on natural phonology*. New York: Garland.
- Sūn, Wéizhāng, Yě Lù, & Lìjūn Lǐ. 1986. Jílín fāngyán fènqū lüèshuō [A synopsis of Jílín dialect regions]. *Fāngyán* [Dialect] 1. 39–45.
- Tagliamonte, Sali A., & Alexandra D’Arcy. 2009. Peaks beyond phonology. Adolescence, incrementation, and language change. *Language* 85(1). 58–108.
- Vihman, Marilyn May. 1980. Sound change and child language. In Elizabeth C. Traugott, Rebecca LaBrum, & Susan Shepherd (eds.), *Papers from the Fourth International Conference on Historical Linguistics, Stanford, March 26–30, 1979* (Current Issues in Linguistic Theory 14), 303–230. Amsterdam: John Benjamins.
- Wang, Jianjun. 2012. Curriculum reform in Mainland China, 1978–2008: Change, maintenance, and conflicts. *Chinese Education and Society* 45(1). 59–68.
- Warren, Paul, Jen Hay, & Byrnmor Thomas. 2007. The loci of sound change effects in recognition and perception. In Jennifer Cole & José Ignacio Hualde (eds.), *Laboratory Phonology 9*, 87–112. Berlin: Mouton de Gruyter.
- Weinreich, Uriel, William Labov, & Marvin I. Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for Historical Linguistics*, 97–195. Austin: University of Texas Press.
- Wong, Pusan. 2013. Perceptual evidence for protracted development in monosyllabic Mandarin lexical tone production in preschool children in Taiwan. *Journal of the Acoustical Society of America* 133(1). 434–443.
- Wright, Jonathan D. 2007. Laryngeal contrast in Seoul Korean. Philadelphia, PA: University of Pennsylvania dissertation.
- Yin, Bo, & Richard B. Baldauf, Jr. 1990. Language reform of spoken Chinese. *Journal of Multilingual and Multicultural Development* 11(4). 279–289.
- Zhang, Qíng. 2005. A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society* 34. 431–466.
- Zhu, Hua. 2002. *Phonological development in specific contexts: Studies of Chinese-Speaking children*. Clevedon: Multilingual Matters.
- Zhu, Hua, & Barbara Dodd. 2000. The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language* 27. 3–42.
- Zhu, Liyi. 2012. Retroflex and non-retroflex merger in Shanghai accented Mandarin. Seattle, WA: University of Washington M.A. thesis.