INTRODUCTION

Among the many acoustic correlates of infant-directed speech (IDS) is an increase in speakers' overall pitch and slower speaking rate when compared to adult-directed speech (ADS) (Fernald 1992). These observations have been made in genetically unrelated languages such as English and Korean. While the effect of IDS serves the dual psychological and linguistic purposes of emotional regulation and phonetic enhancement (Fernald 1992, Werker et al. 2007, Narayan 2013), the fine-grained effects of IDS at the level of the segment are only recently being understood. While the sonorant aspects of speech, such as vowel quality, vowel length, and tonal range have been shown to be exaggerated in IDS, measures of consonantal contrast, such as voice onset time (VOT) in stops, are more nuanced (Baran 1977, Sundberg and Lacerda 1999). Recent studies of English and Korean have shown greater overlap in VOT between laryngeal types in IDS compared to ADS (Narayan et al. 2011, Narayan 2013). For the most part, the overlap in VOT between voiced and voiceless stops (in English) and plain, tense, and aspirated stops (in Korean) occurs as a result of shorter VOT in aspirated stops. The present study asks whether phonetic factors contribute to aspirated stops having shorter VOTs in IDS than in ADS. In particular we investigate whether increasing pitch affects the VOT of aspirated stops more than other laryngeal types.

Connections between VOT and F0

VOT and fundamental frequency (F0) are inextricably linked in the form of pitch perturbation, the natural phenomenon of depressed F0 immediately following voiced stops (Haggard et al. 1970) Speakers may enhance this natural effect (Kingston and Diehl 1994) which is thought to lead to the development of tone in languages which historically lose a laryngeal contrast (Hombert et al. 1979). While pitch perturbation is relatively well understood, very little research has examined the reverse effect of F0 on VOT. Aerodynamic models suggest that increasing F0 should be accompanied by relatively longer VOTs. In elevating the larynx, a gesture associated with raising pitch and voiceless stops (Ohala 1973), intraoral volume is decreased, thereby reducing airflow across the glottis, thus inhibiting vocal fold vibration (Stevens 1977) and extending the duration of the initiation of vibration or VOT.

McCrea and Morris (2005) investigated the effects of F0 level on VOT in male speakers and found that, contra aerodynamic hypotheses, voiceless stops produced with a high F0 had significantly shorter VOTs than productions made at low or mid F0s. McCrea and Morris suggest that glottal and vocal tract factors interact in a way different from what is predicted by aerodynamic models. The high degree of stiffness of the vocal cords, associated with raised pitch, allows them to return quickly from the abducted position for voicelessness in the initiation of voicing. The present study extends McCrea and Morris's study by examining the VOT behavior at varying F0s in female English speakers as well as female speakers of Korean, a language that contrasts three (rather than two) laryngeal states.

VOT in English and Korean

The most reliable acoustic correlate of laryngeal type or phonological voicing of plosives in English is VOT (Lisker and Abramson 1964; Klatt 1975). Countless studies have consistently shown that voiceless plosives (/p/, /t/, /k/) have VOT values longer than voiced plosives (/b/, /d/, /g/). Physiological factors contribute to variation of VOT within laryngeal type in English (and other languages) with velars having longer VOTs than bilabials and coronals (Cho and Ladefoged 1999). Speaking rate also contributes to VOT variation, with voiceless stops in English being shorter at fast rates (Kessinger and Blumstein 1997, Volatis and Miller 1992). While some studies have found speaker sex contributing to VOT variation in English, a rather large recent study of isolated CVs found no difference in the VOTs of males and females (Miller et al. 2008).

Unlike English, Korean contrasts three laryngeal types (plain, tense, and aspirated) that are often characterized by VOT (Abramson and Lisker 1964). Though there is considerable evidence that the role of VOT as an acoustic cue to laryngeal type is waning with the strengthening of tonal contrasts on stops (Silva 2006), the general pattern of VOT has remained relatively stable, with aspirated stops having the longest VOTs and tense stops having the shortest. Plain stops remain in between tense and aspirated with respect to VOT (Abramson and Lisker 1964, Choi 2002, Park 2010).

METHODS

Speakers and recording

The present study replicated the methods of McCrea and Morris (2005) to the extent possible. Female speakers whose first language was either Canadian English (n=10) or Korean (n=10) were recruited via advertisement at the University of Toronto. Speakers were recorded individually in the sound-attenuated booth in the Phonetics Laboratory. In order to establish speakers' pitch range, each speaker was asked to phonate "aah" ([ɑ]) from their normal speaking pitch up to their highest pitch without straining and down to their lowest pitch without introducing glottal fry. Highest and lowest pitches were measured online using musical tuner applet. The highest and lowest pitches produced out of three trials were considered the extremes of the speaker's vocal range. Three pitch height conditions were established at 10%, 30%, and 50% of each speaker's vocal range. The raw Hz values of the three conditions were transformed, again using an online applet, to western music notes for playback. Speakers practiced producing [ɑ] at each frequency level while the experimenter listened for inconsistencies. Speakers attempted to imitate the pitch using a chromatic tuner for visual feedback of their accuracy. Speech materials for English were identical to those used in McCrea and Morris (2005) and consisted of a monosyllable word followed by a two-syllable related word in a carrier phrase. For example "tea and a teapot." The materials list for English reflected three places of articulation and two laryngeal states followed by the high-front vowel [i]. Materials for Korean speakers (written in Hangul script) were similar in that they reflected word-initial stop consonants followed by [i] at three places of articulation, but three laryngeal states. Speakers then practiced the speech materials (target words in carrier phrases) under normal pitch conditions at 180 BPM, according to a visual metronome. The instructed speech rate was identical to that used in McCrea and Morris (2005) who asked speakers to speak at three syllables per second, which was thought to mimic the pace of normal speech. Recording commenced after speakers' rehearsal of the materials. Speakers recorded the materials three times at each of their individualized target pitch heights (low, medium, high). Each recording commenced with the experimenter first presenting the target pitch height auditorily, which was then matched by speakers' production. The ordering of the conditions for recording was counterbalanced across speakers.

Measurements

In order to avoid phrase-initial strengthening effects, only target CVs occurring in second position (i.e., not phrase initial) were analyzed. Recordings were analyzed using spectrographic and waveform displays in Praat (Boersma 2001). VOT of word initial stops was considered the duration between the onset of the initial noise burst following release of the consonant and the onset of the quasi-periodic vocalizing of the following vowel. Post-consonant vowel duration was also measured along with F0 measures (PSOLA implemented in Praat) at vowel onset (analysis window centered in the first quartile) and midpoint.

RESULTS

Figure 1 plots VOTs for each laryngeal type and individualized pitch in English and Korean. Aspirated and Plain stops in Korean have very similar VOTs, with averages in the 50-100ms range across all three Pitch types. This result is consistent with recent studies (Choi 2002, Park 2010) showing a high degree of variability in Plain stops which have VOTs similar to aspirated stops. In Korean, within the Aspirated series, stops produced at a high pitch (average = 69ms) were significantly shorter than those produced at both medium (78ms, $p < 0.05$) and low pitches (91ms, $p < 0.005$). Within the Plain series, high pitch stops had a shorter VOT (55ms) than medium pitch stops (66ms, $p < 0.005$) and low pitch stops (71ms, $p < 0.001$). Within the Tense series there was no difference in VOT between Pitch types.

The English data showed a similar pattern with voiceless aspirated stops showing the most effect of Pitch. Within the Voiceless series, VOT for high pitch stops (70ms) was not significantly different from medium pitch stops (75ms), but significantly different from low pitch stops (80ms, $p < 0.01$). The VOT of voiceless stops at medium pitch was not different from low pitch. The VOT of voiced stops did not vary with Pitch type, averaging 24ms.

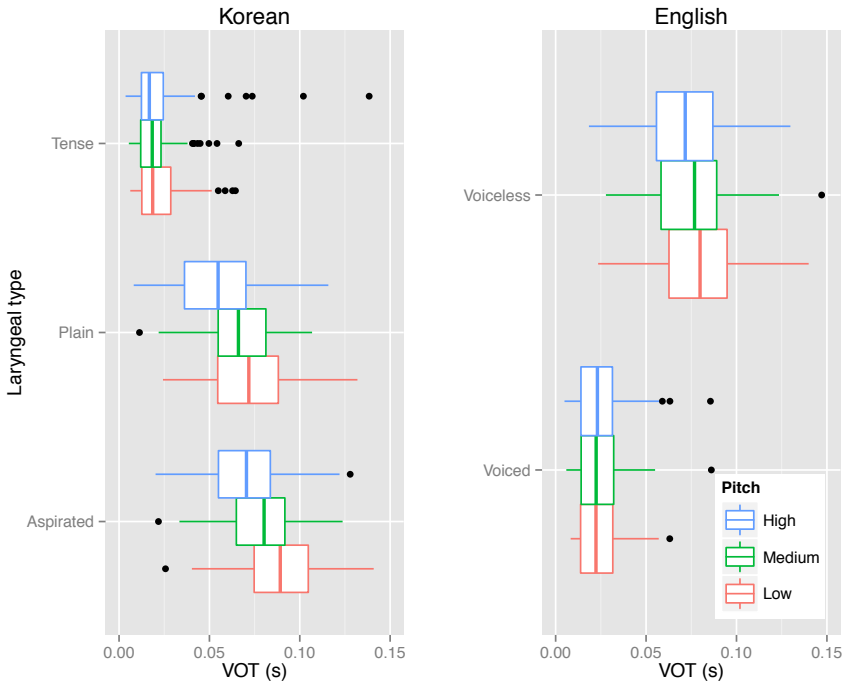

FIGURE 1. VOT according to Laryngeal and Pitch type in Korean and English. Pitch types reflect individualized pitch targets according each speakers pitch range.

The overlap of Aspirated and Plain VOTs suggested that an additional phonetic cue must differentiate the phonological categories. Figure 2 plots post-stop F0, which captures the natural pitch perturbation associated with different laryngeal states. Long-lag or aspirated stops are typically followed by higher F0 than following shorter lag stops. Figure 2 shows that, while there is considerable VOT overlap between Aspirated and Plain stops in Korean, F0 immediately following the release of the stop differentiates the categories within each Pitch type.

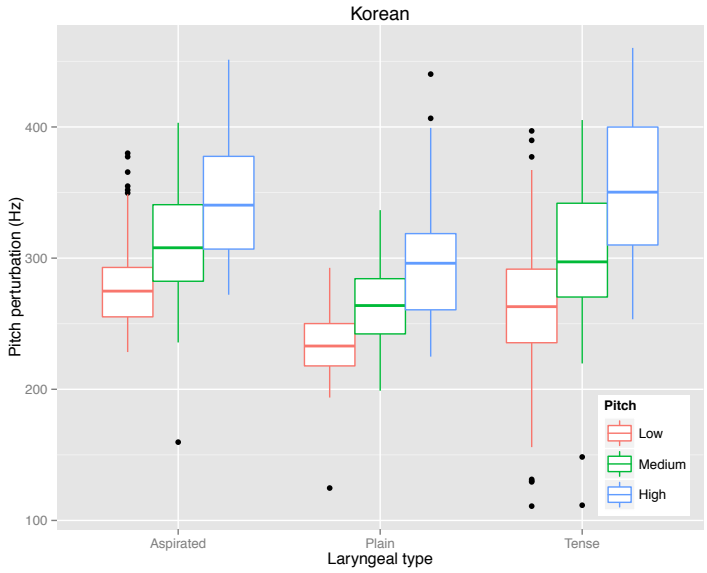

FIGURE 2. Pitch perturbation, or post-stop F0, at three Pitch heights. Aspirated and Plain stops, which have overlapping distributions along VOT, are differentiated when Pitch perturbation is considered. The differentiation remains stable across varying Pitch heights.

Pitch height as used in the descriptive statistics is an artificial designation, the implementation of which varies considerably between subjects. Figure 3 presents the VOT data as a function of continuous F0 across all speakers in the Korean and English groups.

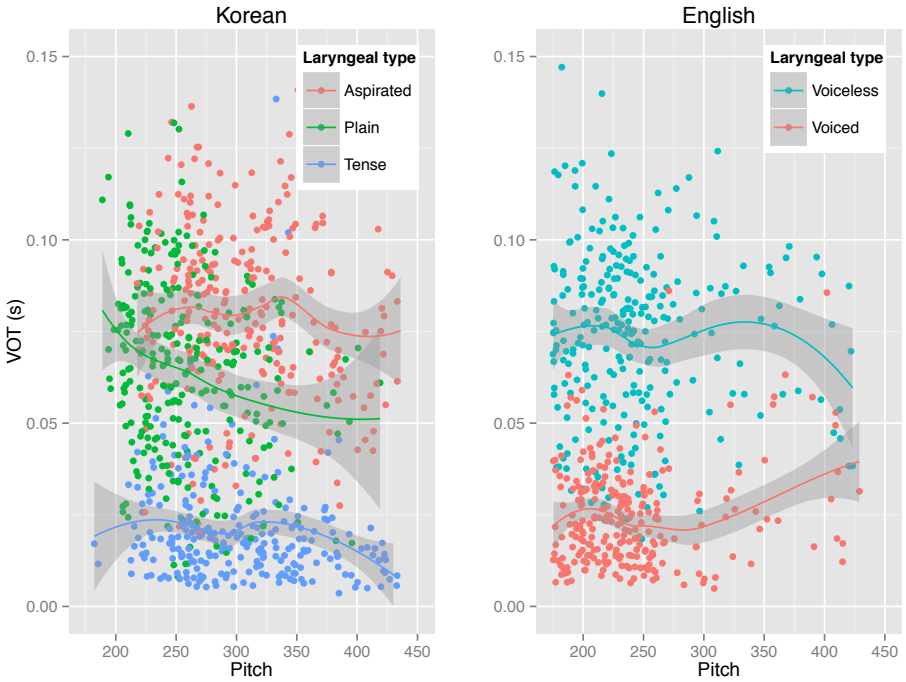

FIGURE 3. VOT x Pitch in Korean and English with smoothing done locally according by least-squares. Shaded portion around smooth line shows local confidence intervals.

The English data were submitted to logistic regression models using raw VOT values in order to determine whether stops produced at varying Pitch heights are better or worse at predicting voicing. Table 1 presents the results of models for each of the Pitch heights.

TABLE 1. Logistic regression models of Voicing in English at three Pitch heights with VOT as the predictor.

Pitch height	**VOT *β***	**Error**	***z***	**Deviance**
Low	5.82	0.93	6.23	74.26
Medium	4.96	0.79	6.25	84.33
High	3.96	0.55	7.14	113.60

While the *z* values show that VOT is significantly predicts voicing at all three Pitch heights, the parameter coefficients reveal that the effect of VOT on voicing is strongest in the Low pitch, with decreasing strength as Pitch height increases. The fit of the models, as indicated by the Deviance (*-2 log liklihood*) suggests a best fit, and strongest prediction in the Low pitch model, with increasing deviance as pitch increases.

DISCUSSION

The results of the present study confirm previous research showing a decrease in the duration of VOT with increasing pitch (McCrea and Morris 2005). Our results extend previous research by showing pitch effects on VOT in a language that has a three-way laryngeal contrast. Interestingly, the Plain/Aspirated phonological dimensions in Korean have collapsed along the phonetic VOT dimension, with pitch affecting both categories in a similar way. We might conclude that increasing pitch affects the long-lag VOT category in general, encompassing a variety of phonological distinctions such as voiced-voiceless (in English), and lenis-aspirated (in Korean).

This research was spurred by the observation that the distinction between long-lag and short-lag phonological categories in English and Korean infant-directed speech is abated relative to similar contrasts in adult-directed speech. Given that IDS is characterized by slower than normal speech rates, and slower speech rates are associated

with longer than normal VOTs in long-lag stops, the increased pitch of IDS is thought to directly contribute to the shortening of voiceless stops in English and aspirated stops in Korean. While the precise physiological mechanism underlying the connection between high pitch and short VOT in long-lag stops has yet to be fully understood, we might postulate that the variability in the glottal-oral timing exhibited in IDS can have consequences for the development of speech perception in infancy.

As the infant assembles her working phonetic space over the first year of life (Narayan et al. 2010), the nature of the phonetic input becomes increasingly important evidence for phonetic categories. IDS is often though to provide infants with enhancing phonetic cues to speech sound categories in the domain of vowel quality and phonological duration (Werker et al. 2007). The diminished contrast between laryngeal categories evidenced by shortening VOTs in long-lag stops suggests that not all aspects of the input to infants is enhancing, and this may be purely an artifact of the physiology of increasing pitch. We might consider an abated long-lag/short-lag contrast when deciphering the highly variable infant speech perception results with VOT.

ACKNOWLEDGMENTS

The authors thank Derek Denis and Sae Jin Byeon for assistance. This work was sponsored by the Social Sciences and Humanities Research Council of Canada, Standard Research Grant #410092742.

REFERENCES

Baran, J., Zlatin Laufer, M., and Daniloff, R. (**1977**). Phonological contrastivity in conversation: a comparative study of voice onset time. *Journal of Phonetics*, 5:339–350.

Boersma, P. (**2001**). Praat, a system for doing phonetics by computer. *Glot International*.

Cho, T. and Ladefoged, P. (**1999**). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27:207–229.

Choi, H. (**2002**). Acoustic cues for the Korean stop contrast: Dialectal variation. *ZAS Papers in Linguistics*, 28:1–12.

Fernald, A. **(1992)**. Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In Barkwo, J., Cosmides, L., and Tooby, J., editors, *The Adapted Mind*, pages 391–428. Oxford University Press, New York.

Haggard, M., Ammbler, S., and Callow, M. (**1970**). Pitch as a voicing cue. *Journal of the Acoustic Society of America*, 47:613–617.

Hombert, J.-M., Ohala, J., and Ewan, W. (**1979**). Phonetic explanations for the development of tones. *Language*, 55:37–58.

Kessinger, R., & Blumstein, S. (**1998**). Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26, 117–128.

Kingston, J. and Diehl, R. (**1994**). Phonetic knowledge. *Language*, 70:419–454.

Klatt, D. (**1975**). "Voice onset time, frication and aspiration in word-initial clusters." *Journal of Speech and Hearing Research*, 18, 686-705.

Lisker, L. and Abramson, A. (**1964**). A cross-linguistic study of voicing in initial stops: Acoustical measurements. *Word*, 20:384–422.

McCrea, C.R. and Morris, R.J. (**2005**). The effects of fundamental frequency levels on voice onset time in normal adult male speakers. *Journal of Speech, Language, and Hearing Research*, 48(5):1013–1024.

Morris, R., McCrea, C., and Herring, K. (**2008**). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics*, 36:308–317.

Narayan, C., Werker, J., and Beddor, P. (**2010**). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13(3): 407–420.

Narayan, C., Yoon, T., and Denis, D. (**2011**). VOT and f0 in Korean infant-directed speech. *Canadian Acoustics*, 39(3).

Narayan, C. (**2013**). Developmental perspectives on phonological typology and sound change. In Yu, A., editor, *Origins of Sound Change: Approaches to Phonologization*, chapter 6. Oxford University Press.

Ohala, J. (**1973**). The physiology of tone. In Hyman, L., editor, *Consonant Types and Tone*, number 1 in Southern California Occasional Papers in Linguistics. Linguistics Program, University of Southern California.

Park, S. (**2010**). *The Laryngeal Contrast in Stops in Korean-English Bilingual Children: An Acoustic Study*. PhD thesis, University of Toronto.

Silva, D. (**2006**). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23:287–308.

Stevens, K. (**1977**). The physics of laryngeal behavior and larynx modes. *Phonetica*, 24:264–279.

Sundberg, U. and Lacerda, F. (**1999**). Voice onset time in speech to infants and adults. *Phonetica*, 56:186–199.

Volaitis, L., & Miller, J. (**1992**). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92, 723–735.

Werker, J., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (**2007**). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*, 103:147–162.