**Perception of pitch skip: Automatic or attentive?**

Correlations between voicing and low pitch, and voicelessness and high pitch, are common cross-linguistically. In some languages, it has led to tonogenesis. For example, in some African tone languages, such as Nupe, Ngizim and Ewe, voiced obstruents have a tone-lowering effect, while voiceless obstruents have a tone-raising effect (Hyman & Schuh 1974). In SiSwati, the voicing contrast of clicks has been neutralized and replaced by a tonal contrast. In other tone languages, such as Mandarin, Cantonese, Thai and Yoruba, and even in non-tone languages such as French and English, the VOT of the onset consonant affects the pitch of the very beginning of the vowel, a phenomenon known as pitch skip (Hombert 1978; Wong 2001; Xu & Xu 2003). The fact that this occurs across unrelated languages which use pitch phonologically in very different ways suggests that there may be a universal tendency for voicing to pair with lower pitch, and voicelessness to pair with higher pitch. I propose that this relationship constitutes a natural relation, i.e. it is borne out of the physical effects of articulation, and is a universal tendency.

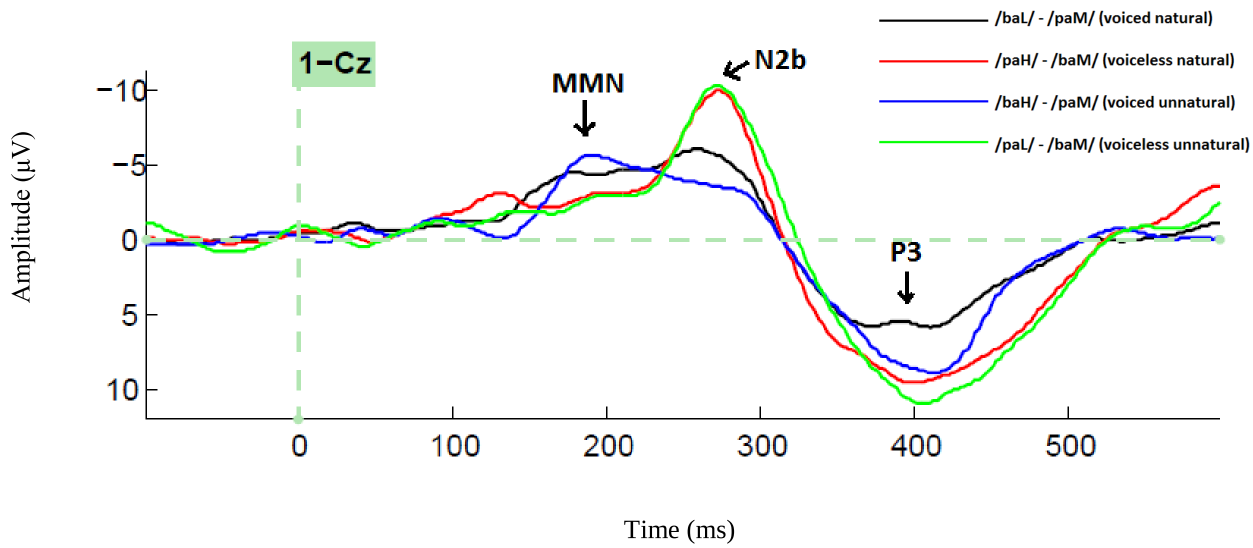
The central research question here is as follows: What is the nature of the perceptual relationship between the voicing properties of onset consonants and the pitch of vowels in English, where the effect is not phonological? In languages including English, it has been shown that this voicing-pitch association is active in both production (Hombert 1978, Kingston 2004 and Hanson 2009) and perception (Abramson & Lisker 1984, Whalen et al. 1992), and listeners can use pitch skip to help identify voicing of an obstruent with an ambiguous VOT. The present study further investigates the perceptual correlation of pitch and voicing, using electroencephalography in the form of event-related brain potentials (ERP), which offer more fine-grained details about the time-course in the perception of this correlation.

In this study, English-speaking listeners heard repetitions of mid-pitch [ba] standards in half the blocks, and mid-pitch [$p^h$a] standards in the other half, interspersed with high- and low-pitch [$p^h$a] and [ba] deviants (known as an “oddball” condition), where the voicing of the deviant was always a mismatch to the standard. They also had a deviant-identification task (the click of a mouse button). Mismatch Negativity (MMN) components are early ERP effects typically elicited at 100-200 ms in oddball conditions and correlated with automatic pre-attentive processing. MMN amplitude has been shown to correlate positively with the perceptual salience of the deviant (Näätänen 2007). Significant interaction of pitch and voicing was predicted for MMNs and deviant-identification reaction time, with greater ERP amplitude and longer RTs for the “natural” (more perceptually salient) deviants than for the “unnatural” (less perceptually salient) deviants.

No behavioral effect was found. There were significant differences between voiced and voiceless deviants in the two earlier ERP components, MMN and N2b, but there was no interaction of voicing and pitch for these two components. An interaction of voicing and pitch was seen in the P3 component (at 300-400 ms) with a significantly smaller effect for the low-pitch voiced (“natural”) deviant than any of the other deviants. These results suggest that processing of the pitch-voicing relationship in English involves attention, unlike simple voicing or pitch mismatch.

I suggest that the interaction of pitch and voicing for the P300, a component inversely correlated with stimulus probability (Polich 2007), is due to the general high probability of low-pitch [ba] in English speakers’ language experience outside the experiment, and may be masked in the voiceless deviants (where high-pitch [$p^h$a] is also highly probable in English) because of the much later voicing onset in [$p^h$a] than in [ba]. This also suggests that although pitch skip may be produced automatically for physiological reasons, it does not seem to be perceived automatically, and English speakers do not perceive the pitch/voicing relationship until/unless they are paying attention to the stimuli.

**Fig. 1**. ERP difference waves for midpoint of scalp
(Deviant minus standard, at electrode Cz)




**Selected References**


Abramson, A. S. and Lisker, L. (1984). Relative power of cues: F0 shift vs. voice timing. *Haskin Laboratories Status Report on Speech Research, 77/78*, 121-128.

Hanson, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J. Acoust. Soc. Am. 125*(1), 425-441.

Hombert, J. M. (1978). Consonant types, vowel quality, and tone. In V. Fromkin (Ed.), *Tone: A Linguistic Survey* (77-112). New York: Academic Press.

Hyman, L. M. and Schuh, R. G. (1974). Universals of tone rules: Evidence from West Africa. *Linguistic Inquiry, 5*(1), 81-115.

Kingston, J. (2004). Segmental influences on F0: Controlled or automatic? In C. Gussenhoven & T. Riad, (Eds.), *Tones and Tunes*, vol. 2, (171-210). Mouton de Gruyter.

Näätänen, R. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*, 2544-2590.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neuropsychology, 118*, 2128-2148.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1992). F0 gives voicing information even with unambiguous VOTs. *Haskin Laboratories Status Report on Speech Research, 111/112*, 11-22.

Wong, K.M. (2001). Fundamental frequency as an acoustic correlate of initial stop consonant aspiration in Cantonese. (Bachelor's thesis). The University of Hong Kong, Hong Kong.

Xu, C.X. & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association, 33*, 165-181.