Perception of pitch skip: Automatic or attentive?

SARA CATLIN
STONY BROOK UNIVERSITY
JANUARY 16, 2015

What is pitch skip?

- Initial perturbation of vowel pitch (F0)
- Cross-linguistic association
  - Voiced onset → low initial nucleus V pitch
  - Voiceless onset → high initial nucleus V pitch
- Duration varies across languages
    - English duration: ~100ms (Hyman 1978)
    - Yoruba duration: ~50ms (Hyman 1978)
    - Xhosa duration: halfway through the vowel (Jessen & Roux 2002)
- Happens in both tonal and non-tonal languages

Tonogenesis

- Voicing can influence development of tones
- Thai (Abramson & Erickson 1992)
  - Old Thai had 3 tonal categories
  - Each Old Thai tonal category split
    - Initial voiced consonant: lower tone
    - Initial voiceless consonant: higher tone
- Bantu languages, e.g. SiSwati, Xhosa (Jessen & Roux 2002)?
  - “Voiced” depressor consonants actually voiceless
  - Contrast characterized by speaker as voicing actually tone
    - [!óɓá] ‘chop’ vs. [!òɓá] ‘hit hard’

Production

- Tone
  - Some languages, e.g. Nupe, Ngizim and Ewe (Hyman & Schuh 1974) and Shanghainese (Chen 2011) have phonological restrictions for which tones can occur after which consonants
  - Tendency:
    - Voiceless C: require high tone / disallow low
    - Voiced C: require low tone / disallow high

Production

- Pitch skip in English
  - Kingston (2007), Hanson (2009)
    - Results of both show much individual variability
  - Kingston found no pitch skip effect for [t^h] vs [d] in 4 naïve subjects, but did in himself
  - Hanson found pitch skip effect for voicing in 10 subjects, no effect for aspiration
  - Compared to Hombert (1978): pitch skip effect for /t/ and /d/
- Similar pattern in languages with three-way VOT contrast (Chen 2011)
  - /t/, /d/ pitch skip effect
  - /t^h/ varies between languages and individuals

Production

- Pitch skip in languages with aspiration contrasts (Mandarin: Xu & Xu 2003, Cantonese: Wong 2001)
  - Initial vowel pitch high after unaspirated C
  - Initial vowel pitch low after aspirated C
  - Different from English pattern despite phonetic similarity
    - Not just a physical effect of aspiration

Perception

- English
  - Abramson & Lisker (1984)
    - Vowel onset F0 affected voicing judgement on ambiguous VOT values
      - High initial F0 → voiceless
      - Low initial F0 → voiced
  - Whalen et al. (1992)
    - Vowel onset AND steady-state F0 affected voicing judgement on ambiguous VOT values
      - High F0 → voiceless
      - Low F0 → voiced
    - Vowel onset and steady-state F0 affected voicing judgement RT even for non-ambiguous VOT values
      - But only for voiced in steady-state
      - Long VOTs: Faster with high initial F0 than low initial F0
      - Neg VOTs: Faster with low F0 than high F0

Perception

- Tone languages
  - Thai: Abramson & Erickson (1992)
    - Vowel onset F0 affected voicing judgement on ambiguous VOT values (only /b/ vs /p/, not /pʰ/)
    - VOT affected tone judgement on ambiguous pitch values (majority low: /b/; high: /b/; mid /pʰ/)
  - Cantonese: Francis et al (2006)
    - More likely to identify a syllable with an ambiguous VOT as aspirated when the initial F0 was higher
      - Not the same as Cantonese production!

Automatic or attentive?

What is the nature of the perceptual relationship between the voicing properties of onset consonants and the pitch of vowels in languages like English, where the effect is not phonological?

English speakers can use F0 to disambiguate voicing.

- Do they also use F0 to identify voicing when not ambiguous?
- Do they need to be paying attention to the stimuli, or can they detect the interaction of F0 & VOT as automatically as they do VOT & F0 alone?

Perception: Event-Related Brain Potentials (ERPs)

- No previous ERP studies on voicing-pitch relationship
- Mismatch Negativity (MMN): Early preattentive perception in an oddball standard/deviant task
  - Effects have been found for VOT and F0 (Näätänen 2007)
  - The more perceptually distinct is the deviant is from the standard, the greater the MMN amplitude (Näätänen 2007)
  - Durvasala et al. (2008): Greater MMN for deviant [d] to standard [t] than deviant [d] to standard [t]
    - Conclusion: [d] laryngeally unspecified

Perception: ERP

- N2b (Sussman et al. 2004)
  - Elicited by deviants in oddball paradigm
  - Only elicited attentively (unlike MMN)
- P300 (Linden 2005, Polich 2007)
  - Elicited by deviants in oddball paradigm
  - Amplitude affected by perceptual distinctness of deviant from standard; probability; and surprise (related to probability).

The present study

- Subjects: English speakers
- Voiced/low and voiceless/high posited as natural correlations
  - Physiological source
  - Universal tendency
- Hypothesis: Natural deviants will be perceived as more distinct from standards than unnatural deviants
  - i.e. /paH/ sounds less like /ba/ than /paL/
- Is perception of natural correlation early in processing, even preattentive?
  - MMN & N2b = early; MMN = preattentive

The present study

Predicted Results:

- ERP
  - Larger amplitudes for natural deviants, indicating greater perceptual distinctiveness from standards than unnatural deviants
- Behavioral
  - Shorter reaction times for identification of natural deviants, also indicating greater perceptual distinctiveness from standards than unnatural deviants

The present study

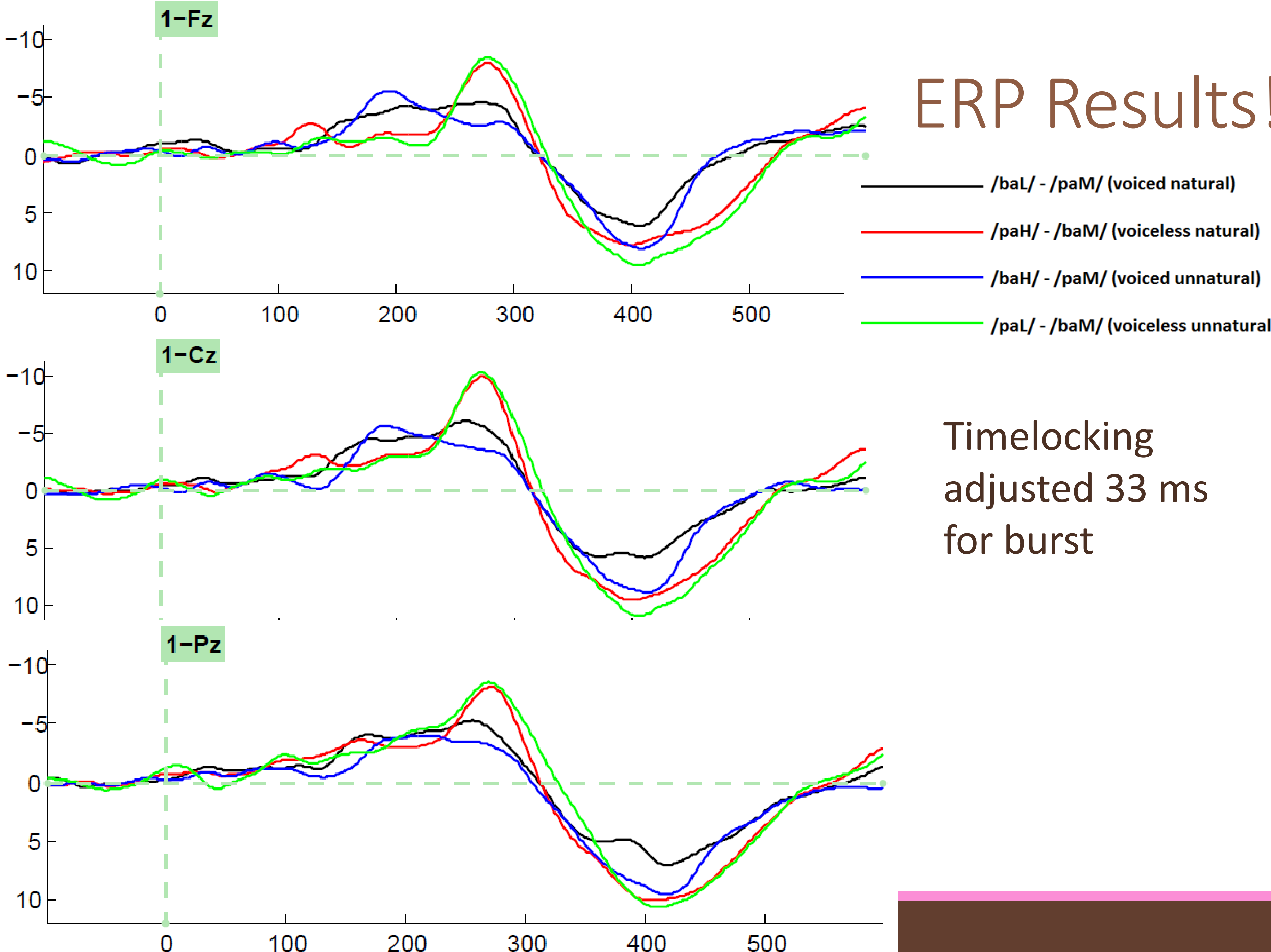
- Oddball paradigm with ERP and behavioral component
- 12 subjects, 2 dropped due to coding error
- While listening to stimuli, subjects wear an electrode cap to record EEG
- Subjects click the mouse whenever they hear a deviant
- 85/15 standard/deviant ratio
- Gap sizes between deviants varied consistently between blocks
- 4 versions, same 4 blocks arranged differently
- Stimuli each 700ms, ISI 600 ms

The present study

Stimuli produced by female native English speaker, modified in Praat (two base tokens, manipulated for pitch)

	Std	VOT	Pitch	Dev	VOT	Pitch
unnatural deviant (UN)	paM	83 ms	230 Hz	baH	-22 ms	260 Hz
natural deviant (N)	paM	83 ms	230 Hz	baL	-22 ms	200 Hz
unnatural deviant (UN)	baM	-22 ms	230 Hz	paL	83 ms	200 Hz
natural deviant (N)	baM	-22 ms	230 Hz	paH	83 ms	260 Hz

ERP Results!



Timelocking adjusted 33 ms for burst

ERP Results, in stats! **=p<0.05; *=p<0.1

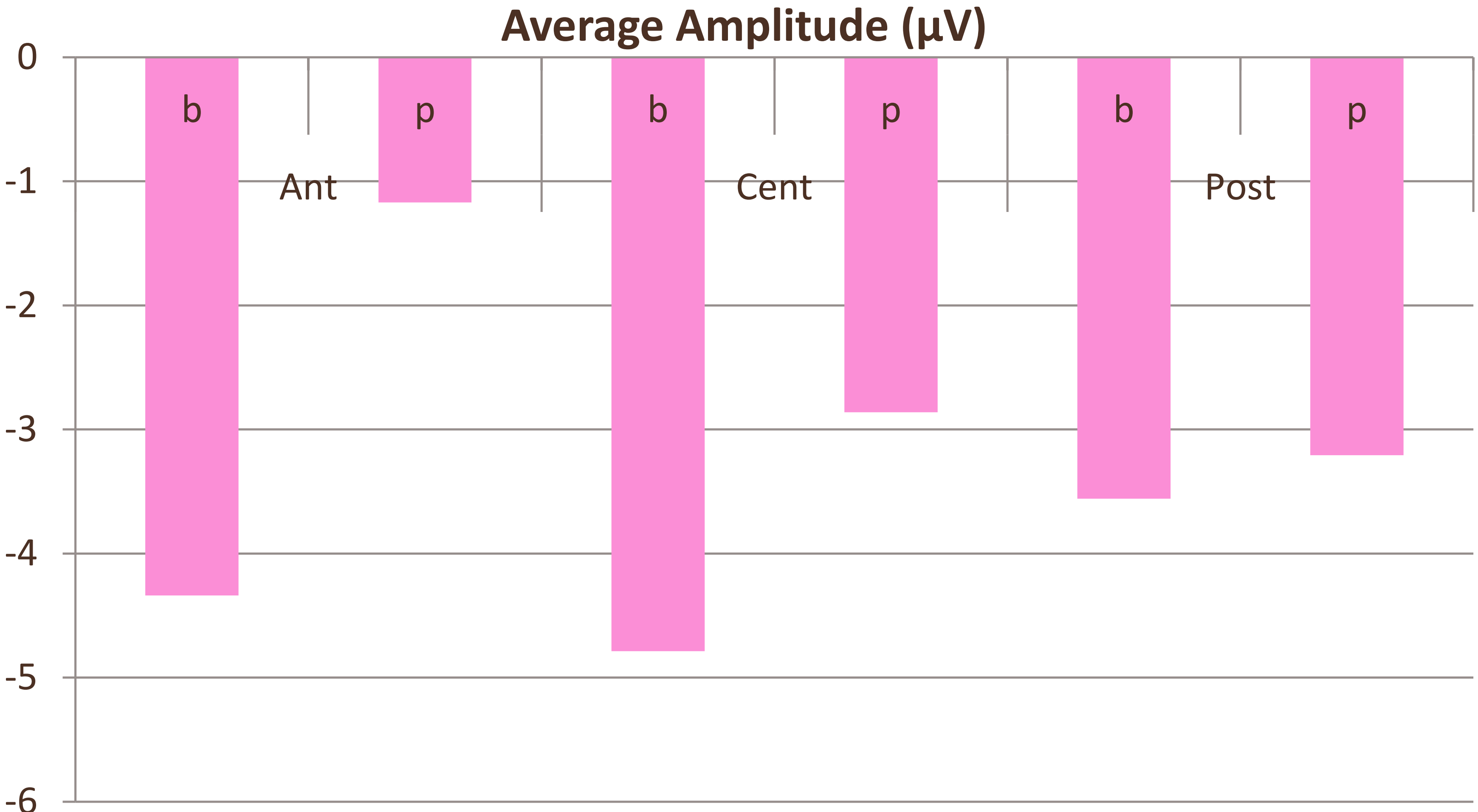
	df	MMN	N2b	P3
voicing	1, 9	F=6.114; **p=0.035****	F=7.884; **p=0.020****	F=2.677; p=0.136
voicing x anteriority	2, 18	F=8.315; **p=0.003****	F=2.137; p=0.147	F=0.398; p=0.678
pitch	1, 9	F=0.315; p=0.588	F=2.036; p=0.187	F=1.278; p=0.288
pitch x anteriority	2, 18	F=1.923; p=0.175	F=1.923; p=0.175	F=2.668; **p=0.097***
voicing x pitch	1, 9	F=0.111; p=0.746	F=0.148; p=0.709	F=4.156; **p=0.072***
voicing x pitch x anteriority	2, 18	F=0.085; p=0.919	F=0.203; p=0.818	F=0.375; p=0.692

There's more to the MMN…

	Frontal	Central	Parietal
voicing	F=11.008; **p=0.009****	F=5.001; **p=0.052***	F=0.278; p=0.611

- Typical MMN results: Concentrated at frontal electrodes

MMN amplitudes

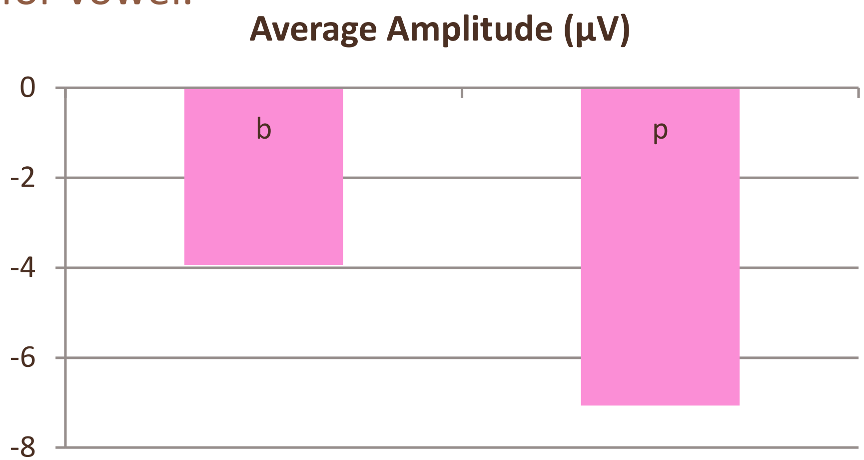


MMN

- Effect of voicing: larger for /ba/
- Replication of Durvasala et al. (2008)
  - Larger MMN effect for voiced English consonants
  - Supporting evidence for laryngeal underspecification of /b, d, g/

N2b

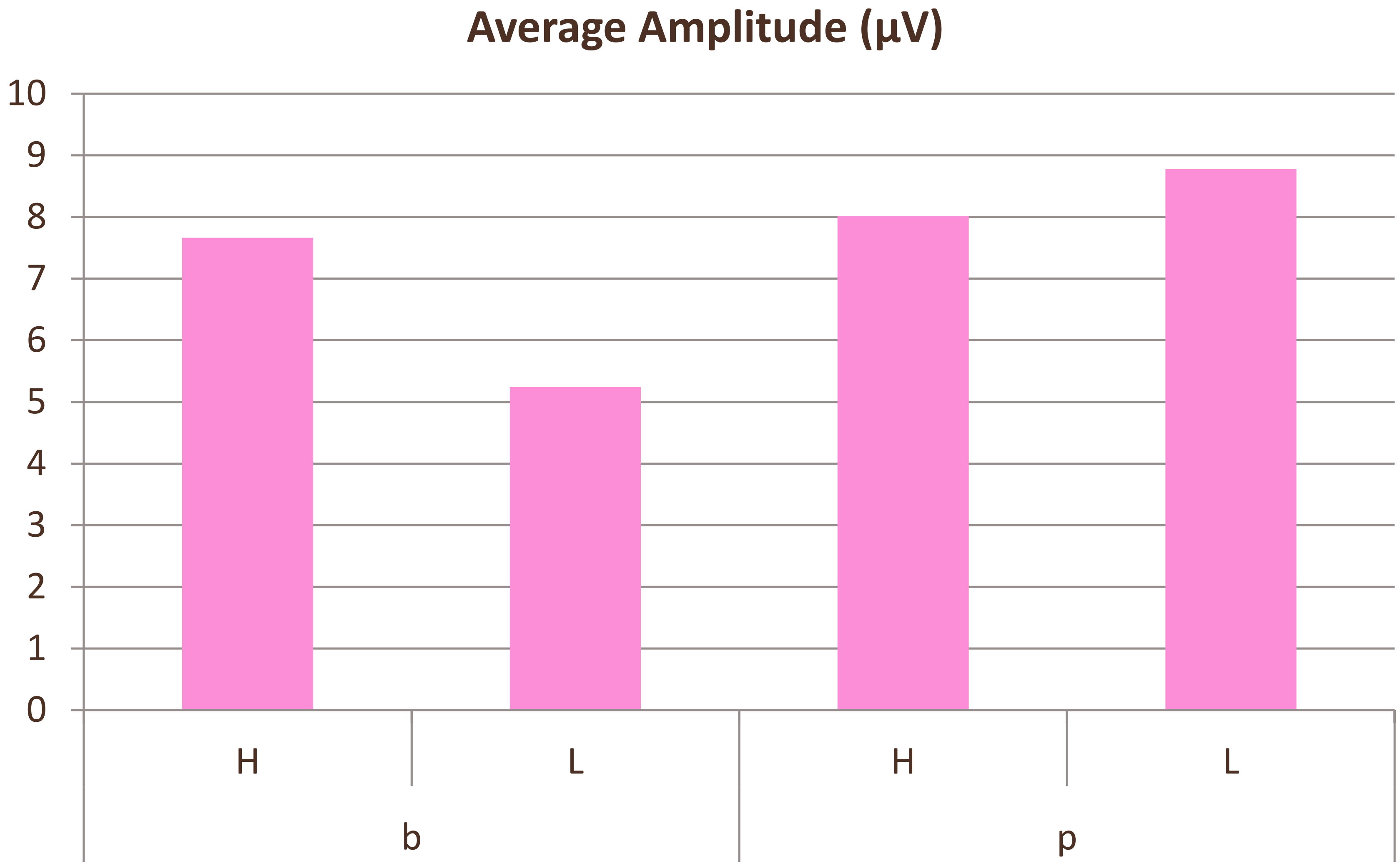
- Effect of voicing: larger for /pa/
- N2b (unlike MMN) indicates focal attention
- After initial burst, listeners wait for vowel onset: 3 ms for /ba/, 83 ms for /pa/
  - i.e. must wait for the vowel after /pa/
  - Perhaps when vowel does not immediately follow burst, listeners focus attention on stimuli while waiting for vowel.



There's more to the P300…

	/L/	/H/	/ba/	/pa/
voicing	F=5.826; **p=0.039****	F=0.067; p=0.802		
pitch			F=6.632; **p=0.030****	F=0.406; p=0.540

P300 amplitudes



P300

- Voicing x pitch interaction: Smaller effect for /baL/ only
  - Opposite from predicted direction of voicing x pitch interaction!
- No effect for pitch on voiceless deviants
  - /paL/, /paH/, /baH/ all larger than /baL/

P300

Why the asymmetry for /baH/?

- P300 factors:
  - perceptual distinctness of deviant from standard
  - probability of deviant occurring
  - surprise (related to probability)
- Using 85/15 ratio and controlling deviant-to-deviant gaps controlled for probability *in the experiment*
  - But what about probability of /baH/ vs /baL/ in English?

P300

- No literature on effects of frequency of phonological pattern in a language
- P300 effects *are* seen for lexical frequency
- If /baH/ is less frequent in subjects' language experience than /baL/, then maybe /baH/ is more surprising
  - Hence /baH/ > /baL/ in P300

P300

- If /baH/ is less frequent and more surprising than /baL/, then /paL/ and /paH/ should have the same relationship
- Both /pa/s have same P300 effect as /baH/
- Why? Possibly: voicing surprise factor
  - N2b indicates attention to /pa/, possibly in anticipation of vowel onset
  - After 83 ms of voicelessness, maybe any voicing is surprising/unexpected

P300

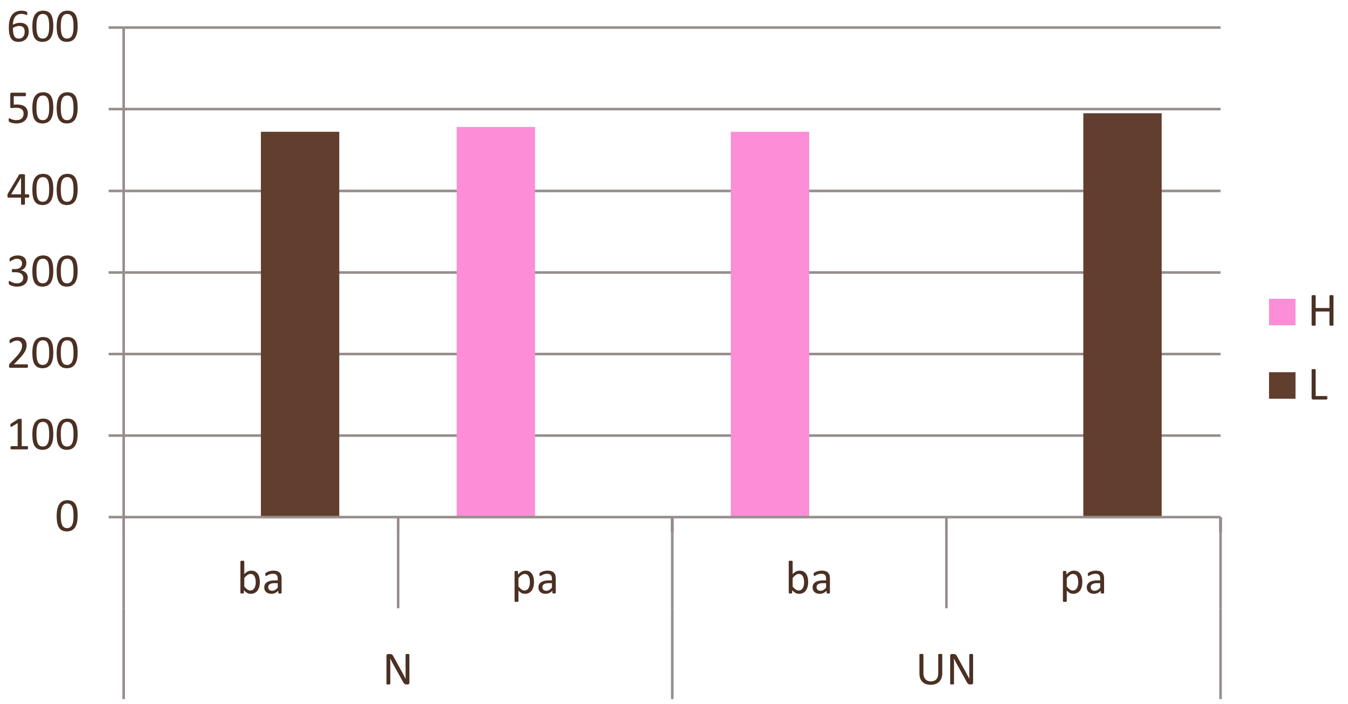
Precedent: Whalen et al. (1992)

- Exp II: stimuli similar to present study, pitch manipulated throughout vowel (not just onset)
- RTs for voicing judgement were longer for /baH/ than /baL/, not different for /paL/ and /paH/
- RTs for voicing judgement were longer for /paL/ than /paH/ when pitch manipulated *only* in vowel onset
- *Pitch skip* (F0 change) seems to affect perception of voicing in long VOTs, but overall F0 of the vowel does not.
- Overall F0 of vowel *does* seem to affect perception of voicing in short/negative VOTs (as well as F0 change)

RT results

No significant differences!

- Voicing effect, pitch effect, voicing x pitch interaction: $F < 1$, $p > 0.1$



General conclusions

- MMN and N2b indicate lack of early processing of pitch-voicing correlation in English
- P300 indicates later processing of pitch-voicing correlation, but only in interaction with *voiced* deviants
- Reaction times indicate that perceptual use of pitch may not last very long, or at least do not affect conscious decisions

What might be next?

- Phonological vs phonetic voicing
  - What do we mean when we say "voicing"?
    - Pos/neg VOT, or any categorical VOT distinction?
    - Present study uses pos/neg VOT, would long/short VOT distinction lead to the same results?
- Similar study to present, but actually using *pitch skip* instead of steady-state vowel
  - Would we see the same asymmetry in P300?

What might be next?

Learnability questions

- Are /paH/ and /baL/ easier to learn than /paL/ and /baH/?
- Do the natural correlations make an easier system to learn?
- If given an artificial tone language (e.g. resembling Thai), would English (or French) speakers learn /paH/ and /baL/ faster/easier?

References

Abramson, A. S. and Erickson, D. M. (1992). Tone splits and voicing shifts in Thai: Phonetic plausibility. *Haskin Laboratories Status Report on Speech Research, 109/110*, 255-262.

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.

Durvasula, K., Hestvik, A., Bradley, C., and Bradley, E. (2008). ERP evidence for laryngeal underspecification in English: A mismatch negativity study. *Western Conference on Linguistics*, Davis, CA.

Grabe, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics, 26,* 129-143.

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.

Jessen, M. & Roux, J. C. (2002). Voice quality di4erences associated with stops and clicks in Xhosa. *Journal of Phonetics, 30,* 1-52.

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.

Kingston, J. and Diehl, R. L. (1994). Phonetic knowledge. *Language, 70*(3), 419-454.

References con’t

Linden, D. E. (2005). The P300: Where in the brain is it produced and what does it tell us? *The Neuroscientist, 11*(6), 563-576.

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.

Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190,* 128-140.

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.