Variability attenuates sensitivity to acoustic detail in cross-language speech production

Sean Martin*, Lisa Davidson*, Colin Wilson†

*Department of Linguistics, New York University, †Department of Cognitive Science, Johns Hopkins University

1. Phonetic detail in cross-language production

Speakers show systematic patterns of errors when processing non-native consonant clusters (e.g. misidentification, modification in production/ transcription, loanword adaptations):

- These patterns are frequently argued to arise from relatively abstract phonotactics such as sonority sequencing, syllable parsing, segmental phonotactics (e.g. Scholes, 1966; Hallé et al., 1998; Pitt, 1998; Dupoux et al., 1999; Moreton, 2002, Berent et al. 2007 et seq,)
- Davidson et al. (2012) provides evidence that both lower-level acoustic-phonetic properties of the stimulus and native language phonotactics taken together account for speaker behavior in non-native production.
- We hypothesize that sensitivity to fine acoustic detail will be attenuated if speakers are presented with stimuli containing a greater range of phonetic detail
  - **Experiment 1** (originally reported in Davidson et al., 2012): Subjects are given single-talker input with systematically manipulated acoustic-phonetic properties
  - **Experiment 2**: Replicates conditions of Experiment 1, but with the addition of multiple talkers producing the stimuli.

Question: Are the effects of low-level acoustic-phonetic variation modulated by the presence of additional sources of phonetic variability?

2. Methodology

- **Stimuli**: In each experiment, 24 English speaking participants heard and repeated critical items of the form [C1C2áCV] produced by native speaker(s) of Russian, along with fillers with initial schwa and medial schwa. Ex: **[ptake], [əptake], [pətake]; [zgamo], [əzgamo], [zəgamo]**

Cluster Type	C1 [-voice]	C1 [+voice]
SS	pt, tp, kp, kt	bd, db, gb, gd
SN	pn, tm, km, kn	bn, dm, gm, gn
FS		vd, vg, zb, zg
FN		vm, vn, zm, zn

- Subjects heard two (Exp 1) or three (Exp 2) repetitions of each stimulus item and repeated the stimulus aloud once into a head-mounted microphone connected to a Zoom H4n recorder.

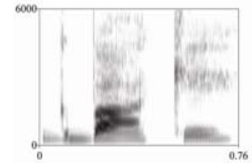
Acoustic Manipulations (based on Wilson & Davidson (2010)):

- **DUR**: duration of the acoustic transition (burst + aspiration) between stop and following consonant
  - **2 levels: 20ms, 50ms**
  - ***DUR longer → more epenthesis***
- **AMP**: amplitude of the acoustic transition of a stop (the burst) relative to the following consonant's amplitude
  - **2 levels: high and low**
  - ***AMP lower → more deletion & C1 change***
- **POV** (pre-obstruent voicing): interval of modal voicing *preceding* the onset of a voiced obstruent constriction
  - **2 levels: present vs. absent**
  - ***POV present → more prothesis***

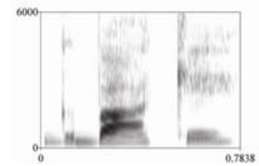
Speaker Variability:

- **Low-variability**: In Experiment 1, all stimuli were produced by a single talker
- **Higher-variability**: In Experiment 2, stimuli were produced by three talkers. Target (final) stimulus production was identical to Exp 1.
  - ***Higher variability → decreased effect of phonetic manipulations***

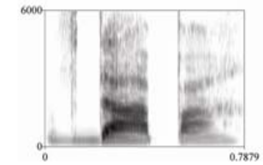
[gbaki]: 20ms DUR, hi AMP



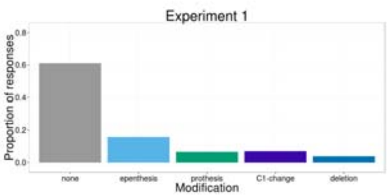
[gbaki]: 50ms DUR, hi AMP

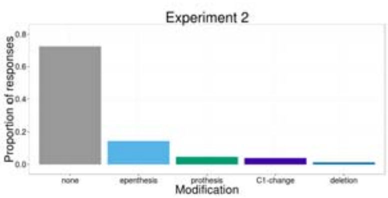


[vdapa]: POV present

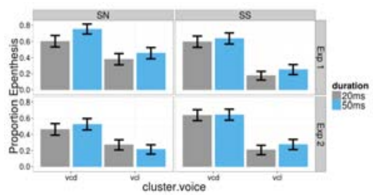


3. Results

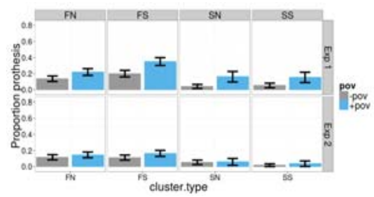




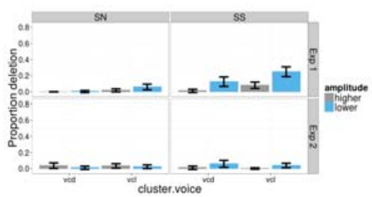
Experiments 1 and 2 showed similar overall patterns of errors. Figures to the right show the rates of specific errors broken down by experimental condition.



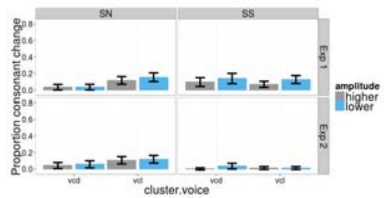
More epenthesis at longer duration in Exp 1 *(β = 0.5, p < 0.001);* non-significant in Exp 2



More prothesis when POV is present in Exp. 1 *(β = 1.1, p < 0.001)* **and Exp 2** *(β = 0.36, p < 0.02)*



More deletion with lower amplitude in Exp. 1 *(β = 1.8, p < 0.001);* non-significant in Exp 2



More C1 change with lower amplitude in Exp. 1 *(β = 0.6, p < 0.05);* non-significant in Exp 2

4. Discussion

Accounting for epenthesis

- Most common strategy for modifying SC stimuli, which contain an open transition.
- English speakers may interpret this as the cues to a vowel; the effect is enhanced for [+voice] clusters with voiced transitions
- No open transition in FC clusters results in lower proportion of epenthesis (9% for Exp 1, 8.2% for Exp 2)

Accounting for deletion & change

- With lower amplitude bursts, the information in release cues may be misperceived as a different stop, or as not being present at all

Accounting for prothesis

- Strong voicing preceded and lasted through obstruent constrictions in the stimuli, whereas English obstruents tend to be devoiced in initial contexts
- English speakers can interpret POV as a vowel

Effect of speaker variability

- Participants show a sensitivity to acoustic detail, but this can be modulated by the overall stimulus variability
  - o Experiment 1 showed significant effects of all three acoustic manipulations.
  - o Experiment 2 shows similar trends, but only the effect of the POV manipulation was statistically significant.
- The attenuation of the effects in Exp 2 suggests that varied stimuli lead speakers to better recognize which information corresponds to the relevant phonemes and which details indicate the range of variability in the phonetic implementation of those phonemes.
- Similarity of the numerical trends between Exp 1 and 2 suggests that sensitivity to acoustic detail is not entirely eliminated; the acoustic details of the last stimulus item still have some influence on subjects' productions.