Phonological bias, phonetic likelihood: a Bayesian model of non-native cluster production errors

Lisa Davidson*, Colin Wilson†, Sean Martin*

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

1. What explains non-native production?

Under a variety of natural and experimental conditions, non-native sounds and sequences are difficult and dispreferred relative to native sound structures:

- misidentification and poorer discrimination (e.g., Werker & Tees 1984; Pitt 1998; Dupoux et al. 1999; Moreton 2002; Berent et al. 2007; Davidson 2011)
- modifications in production and transcription (e.g., Flege 1987; Davidson 2006, 2010)
- loanword and L2 adaptations (e.g., Kang 2004; Davidson 2007; Zuraw 2007; Peperkamp et al. 2008; Boersma & Hamann 2008)
- lower acceptability ratings (e.g., Greenberg & Jenkins 1964; Scholes 1966; Albright 2009; Daland et al. 2011)

These dispreferences are often attributed to relatively abstract **phonotactics**:

- sonority sequencing, syllable parsing, segmental phonotactics (Berent et al. 2007-09, Kabak & Idsardi 2007, Dupoux et al. 1999; Moreton 2002; Hayes & Wilson 2008)

Question: What effects do fine-grained, non-contrastive phonetic properties of the stimulus have on non-native speech production?

- Previous work on production can be seen as focusing on the stimulus-independent *prior distribution,* p(z), over phonetic and phonological representations.
- We instead sought to study the *likelihood function*, p({x}|z), which assesses the 'match' between candidate representations (z) and richly detailed stimulus encodings ({x}).

2. Methodology

- **Stimuli**: 24 English speaking participants heard and repeated critical items of the form [C1C2ə́CV] produced by a native Russian speaker, and corresponding fillers with initial schwa and medial schwa. Ex: **[ptake], [əptake], [pətake]; [zgamo], [əzgamo], [zəgamo]**
- Each cluster appeared in 4 distinct stimulus items. Each participant heard and produced 288 items, as per breakdown in the table + 48 CəCX, 48 əCCX fillers

Cluster Type	C1 [-voice]	C1 [+voice]
SS (N=64)	pt, tp, kp, kt	bd, db, gb, gd
SN (N=64)	pn, tm, km, kn	bn, dm, gm, gn
FS (N=32)		vd, vg, zb, zg
FN (N=32)		vm, vn, zm, zn

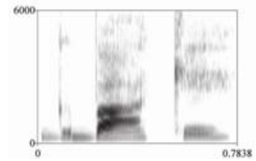
- Each stimulus item was played 2x and the participant repeated the stimulus aloud once into a head-mounted microphone connected to Zoom H4n digital recorder.

Acoustic manipulations (based on Wilson & Davidson (2010), showing that listeners are sensitive to variation in these areas)

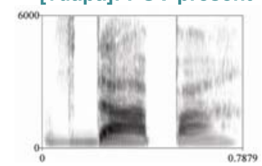
- DUR: duration of the acoustic transition (burst + aspiration) between stop and following consonant
  - **4 levels: 20ms, 30ms, 40ms, 50ms**
  - ***DUR longer → more epenthesis***
- AMP: amplitude of the acoustic transition of a stop (the burst) relative to the following consonant's amplitude
  - **2 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***

[gbaki]: 30ms DUR, hi AMP

[gbaki]: 50ms DUR, hi AMP



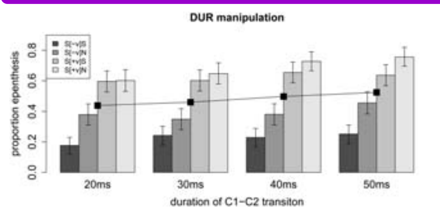
[vdapa]: POV present



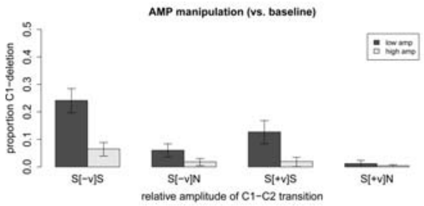
Summary of acoustic manipulations

S[-voice]C: DUR x AMP (8 conditions); **S[+voice]C**: DUR x AMP (8 conds.), DUR x POV (8 conds.); **FC**: POV (2 conditions)

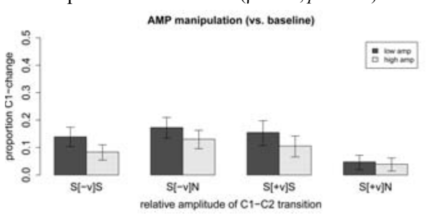
3. Results



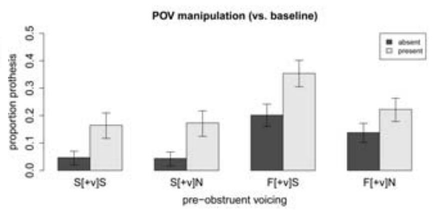
Less epenthesis at shortest duration (indicated by line from 20 to 50ms) (β = -.23, p <.01)
More epenth. for voiced clusters (β = .92, p < .001)
More epenth. before nasals (β = .33, p < .001)



More deletion with lower amplitude (β = .80, p < .001)
Less del. for voiced clusters (β = -.68, p < .05)
Less del. before nasals (β = -.94, p < .001)



More C1 change with lower amplitude (β = .22, p < .05)
Less C1 change for nasals (β = -.36, p < .001)



More prothesis when POV is present (β = .66, p < .01)
Less proth. for stop-initial clusters (β = -.68, p < .05)

4. Discussion

Accounting for the epenthesis modification

- Russian productions of SC clusters have an open transition which English speakers may interpret as containing a devoiced vowel and then reanalyzed as an underlying vowel: [ptake] → [pʰə̥take] → [pʰətake] (acoustic properties of stimuli are compatible with English schwa reduction (Davidson, 2006))
- There is no open transition in FC clusters, so proportion of epenthesis is much lower (9% of FC clusters)

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

- In natural Russian productions, strong voicing preceded and lasted through obstruent constrictions, whereas English obstruents tend to be devoiced (when not in post-voiced environments)
- English listeners can interpret POV as a vowel

Generative model of cross-language production

Perception: A detailed encoding ({x}) of the stimulus—containing quantitative information about duration, formants, voicing, etc.—is used together with the prior to determine the probability of phonetic/phonological representations (z).

$$p(z|\{x\}) \propto p(\{x\}|z)\, p(z)$$

Production: The speaker generates a production ({y}) according to language- and context- specific native knowledge of phonetic realization (possibly attempting to imitate low-level properties of the encoding).

$$p(\{y\}|z,\{x\}) \propto p(\{y\}|z)\, p_{(\{y\}|\{x\})}$$

The present results highlight the importance of stimulus properties, and their interpretation by the native system (as quantified by the likelihood function), in explaining patterns of modification in non-native cluster production.

- Effects of fine-grained stimulus encodings are well-known in perception but less studied in the domain of production.
- Phonotactic biases (gradient or absolute) are present in the prior, but may be less central than previously suggested.