Epenthesis into nonnative consonant clusters: phonetic factors eclipse gradient phonotactics

Colin Wilson (colin@cogsci.jhu.edu) Johns Hopkins University

Lisa Davidson (lisa.davidson@nyu.edu) New York University

LabPhon 2016 Cornell University July 15, 2016

Introduction

Errors on nonnative sound structures are found in a wide range of natural behaviors and experimental tasks

Individual errors (and error patterns) could originate from many types of cognitive representation / processing that are shaped by the native language

- Ex. [d̥bif] classified as having *two syllables*
  - Misperception of voiced open transition as a reduced vowel?
  - Phonotactically-motivated epenthesis repair?
  - Extra-syllabic / appendix representation (d[bif]$_\sigma$)?
- Ex. [d̥baki] produced with an epenthetic vocoid [dᵊbaki]
  - Misperception? Phonotactic repair? Gestural mistiming?

Outline

- Epenthesis asymmetry in repetition of nonnative consonant clusters:
  more vocoid epenthesis into clusters beginning with voiced stops

  Ex. [d˚baki] → [dᵊbaki] > [t˚paki] → [tᵊpaki]

- Three hypotheses about the origin of the asymmetry

  - Releases of voiced stops are misperceived more often than releases of voiceless stops, due to greater acoustic-phonetic similarity to schwa

  - Clusters with voiced stops are less phonotactically well-formed than clusters with voiceless stops, and are actively repaired more often

  - **Transfer of native phonetic gestures and timing relations gives rise to a transitional vocoid more often for voiced stop-initial clusters**

- Evidence from an *orthographic transcription* task supports the misproduction hypothesis: asymmetry arises downstream of perception and phonology

Nonnative cluster repetition (Davidson et al. 2016)

American English speakers (N= 12) produced forms beginning with nonnative consonant clusters, modeled by a Russian talker, in an immediate repetition task

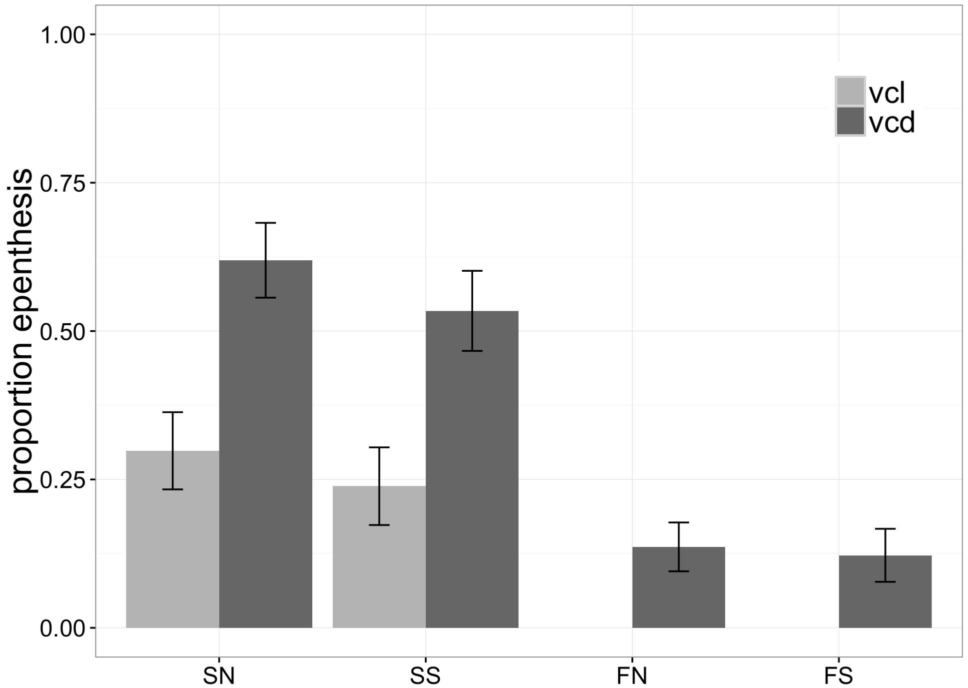
Stimuli were presented with computer speakers in a quiet classroom
Ex. **t˚páki** (Russian production 🔊) … **t˚páki** (replay 🔊) … [now you say it]

	S vcl	**S vcd**
SN (stop-nasal)	pn tm km kn	bn dm gm gn
SS (stop-stop)	pt tp kp kt	bd db gb gd

$\times$ manipulation of **release duration**, intensity, voicing profile $\sim$ 630 critical items
$+$ forms beginning with nonnative FN (vm vn zm zn) and FS (vd vg zb zg) clusters
$+$ matched fillers with intervening or initial schwas (ex. **təpáki**, **əgnátu**)
$\sim$ 100 counterbalanced critical trials (and $\sim$ 30 filler trials) per participant

Voicing effect on epenthesis in repetition

Repetition in the classroom (N=12) [Davidson et al. 2016, *SLR*]



Mixed-effects logistic regression analysis of responses to cluster-initial stimuli

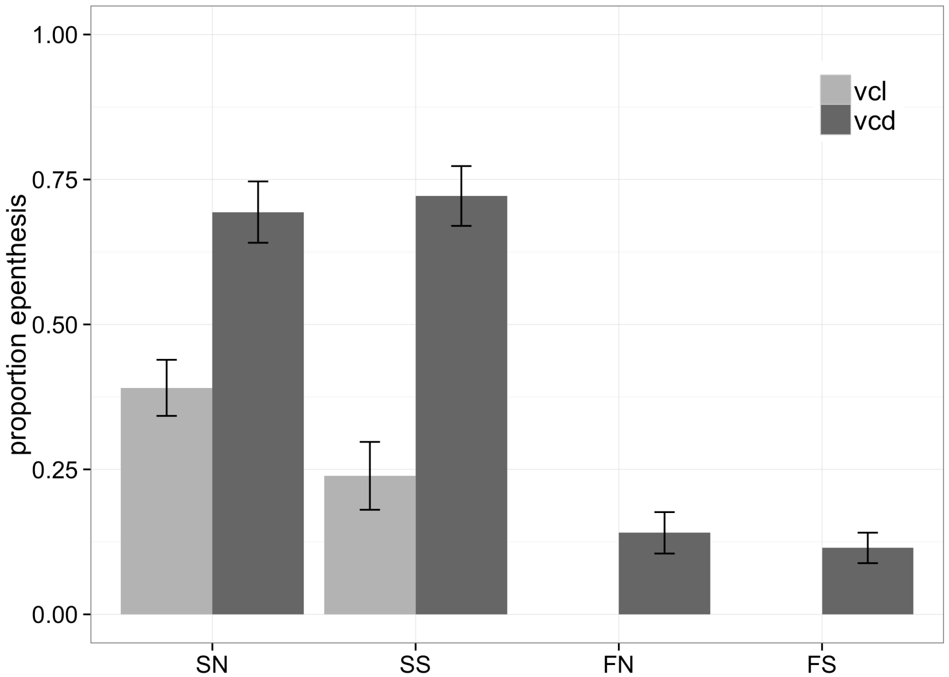
- **More epenthesis after voiced vs. voiceless stops** $\beta = 1.72, p < .01$
- More epenthesis after stops than fricatives $\beta = 2.00, p < .01$
- More epenthesis with longer stop releases $\beta = 0.14, p < .05$

Marginal anti-sonority-sequencing effect (small rise > plateau/fall, $p = 0.07$)

Accurate performance on filler items (epenthesis: 93%, prothesis: 88%)

Additional evidence for the voicing effect

Repetition of the same stimuli in the soundbooth (N=24) [Wilson et al. 2013, *JML*]



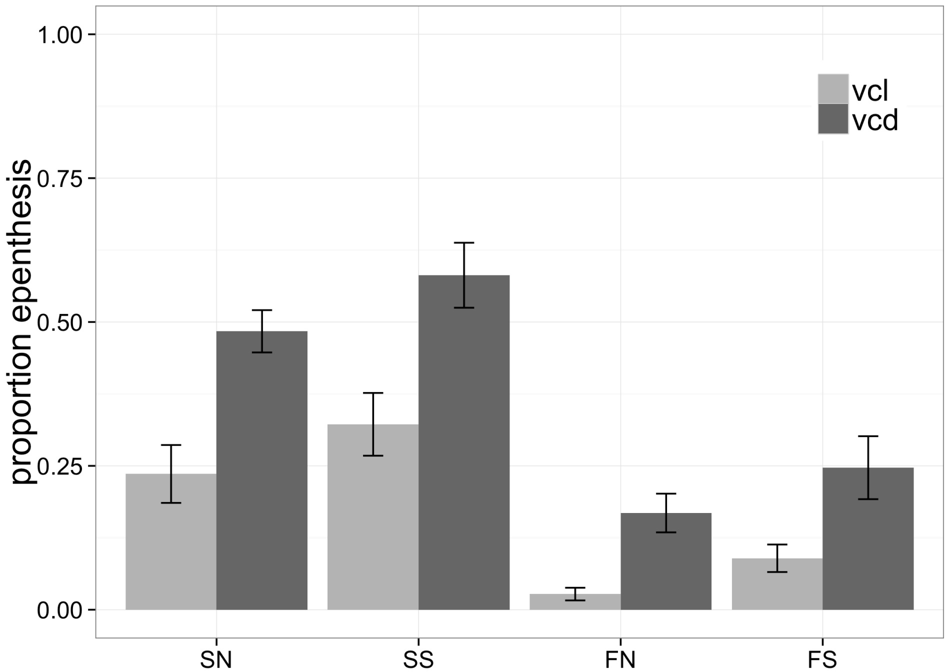
Mixed-effects logistic regression analysis of responses to stop-initial clusters

- **More epenthesis after voiced stops**
  $\beta = 2.26, p < 0.01$
- More epenthesis with longer releases
  $\beta = 0.44, p < 0.05$

Also more epenthesis for stop-initial clusters (> 47%) than for fricative-initial clusters (< 13%), accurate performance on filler items (> 87%)

Additional evidence for the voicing effect

Repetition of unmanipulated 'natural' Russian recordings [Davidson 2010, *JPhon*]



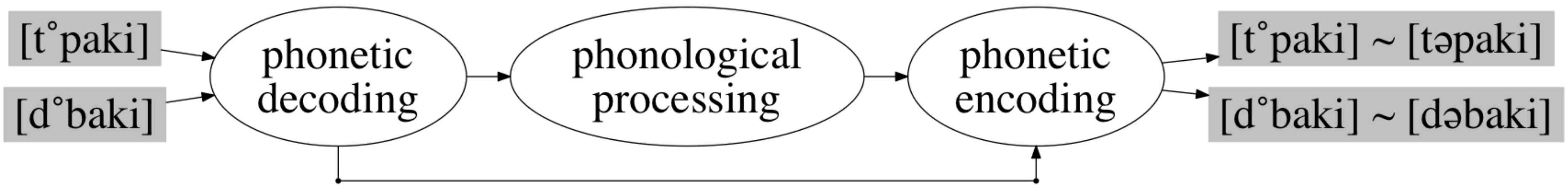
Mixed-effects logistic regression analysis of responses to stop-initial clusters

- **More epenthesis after voiced stops**
  $\beta = 1.44, p < 0.01$
- More epenthesis into SS than SN
  $\beta = -0.66, p < 0.05$

See Wilson & Davidson (2013, *NELS*) for discussion of release durations and other acoustic properties of these stimuli and associated effects on production responses

Where does the voicing asymmetry arise?

- Errors such as [bdazo] → [b^{ə}dazo] are ambiguous w.r.t. where they originate in the stream of processing from acoustic input to articulatory output
- Acoustic differences between epenthetic ([b^{ə}dazo]) and intended ([bədazo]) schwas (Davidson 2006, 2010) suggest that 'epenthesis' is gestural mistiming, but could arise from perceptual/phonological epenthesis + phonetic imitation



(similar cognitive architectures proposed by Ellis & Young 1988; Patterson & Shewell 1987; Ramus et al. 2010; Coltheart et al., 2001; Goldrick & Rapp 2007)

Hypothesis 1: similarity-based misperception

- English participants could epenthesize more often after voiced stops because Russian voiced open transitions (as in [d˚báki]) are
  - more acoustically similar to typical English schwas, and therefore
  - more likely to be misperceived as containing schwas

(cf. Russian voiceless open transitions, as in [t˚páki]) [Wilson & Davidson, AMP 2014]

- Related to theories of repair in phonetically-based phonology and perceptual approaches to loanword adaptation / cross-language speech processing

(see Flemming 1995/2002, 2004; Steriade 1997, 2001, 2009; Hermes 1998; Peperkamp & Dupoux 2003; Peperkamp 2007; Wilson 2006; Mielke 2008; Davidson 2010; Escudero et al. 2012; White 2014)

Acoustic measures of open transitions, schwas

Parallel measurements extracted from 20 ms window centered on

- open transitions of Russian clusters in the stimuli
- schwas produced by English participants for CəCáCV filler items
  (accurate responses only)

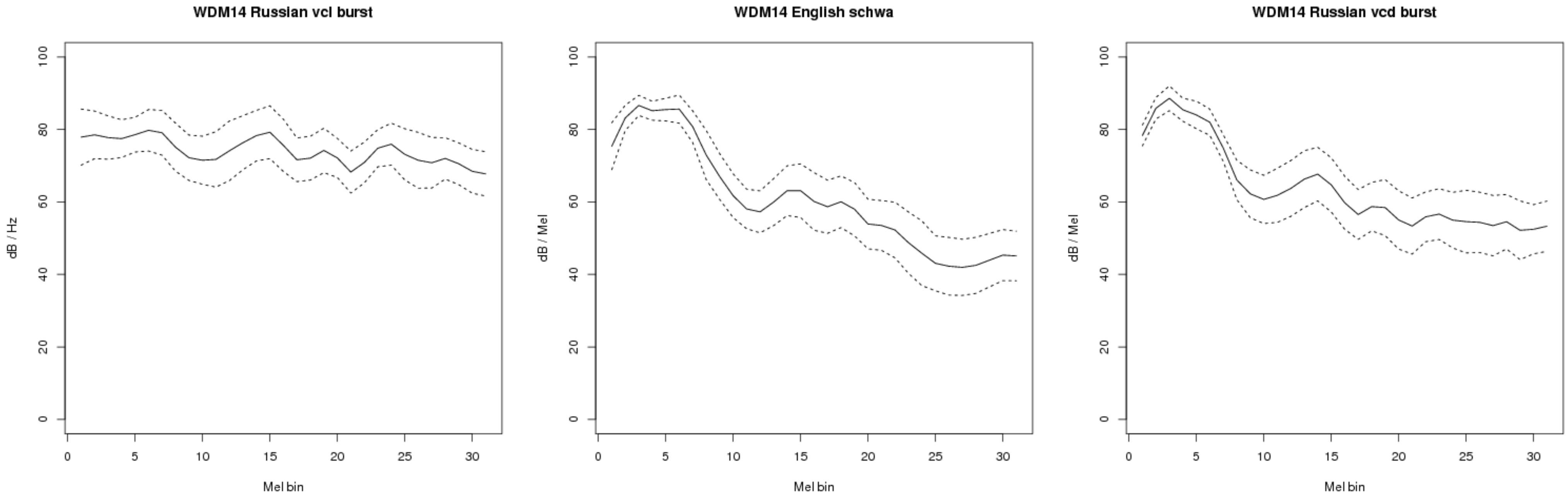
Measures

- Mel spectrum (100 Hz - 10 kHz, 31 bands)
  (each spectrum normalized by dividing DFT values by total power before conversion to dB / Mel scale)
- Peak frequency (Mel)
- RMS amplitude relativized to following consonant closure (dB)
- F1 - F3 (Hz) [used only for validation of schwas]

Spectral comparison of open transitions, schwas

[stimuli and responses from Wilson et al. 2014]

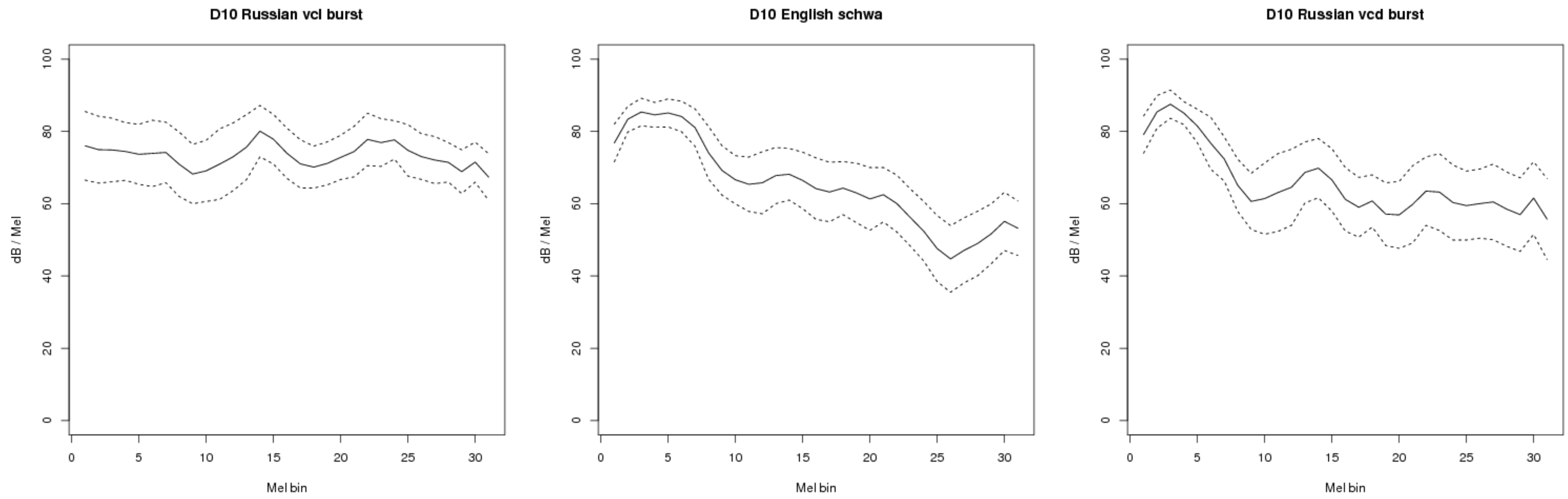
- Spectral comparisons support the possibility of asymmetric misperception



Key: bin5 (400 Hz), bin10 (1000 Hz), bin15 (2000 Hz), bin20 (3500 Hz), bin25 (5750 Hz), bin30 (9300 Hz)

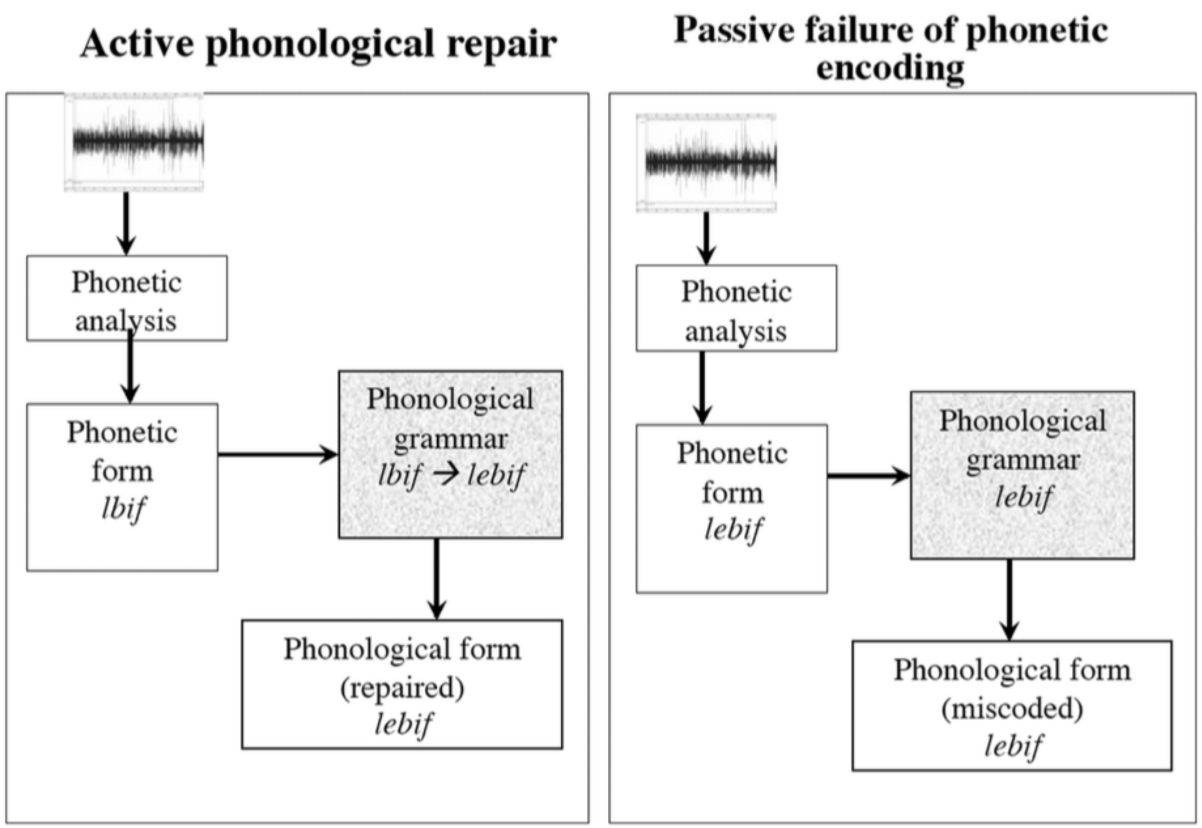
Spectral comparison of open transitions, schwas

(stimuli and responses from Davidson 2010)



Key: bin5 (400 Hz), bin10 (1000 Hz), bin15 (2000 Hz), bin20 (3500 Hz), bin25 (5750 Hz), bin30 (9300 Hz)

Hypothesis 2: Active phonological repair



(figure from Berent et al. 2012, *Lang & Speech*; see also Berent et al. 2007, et seq.)

Difference in phonotactic well-formedness

- Under the active phonological repair hypothesis, the difference between voiced and voiceless clusters arises at the level of phonological processing

- The voicing asymmetry could plausibly arise by 'projection' (e.g., Daland et al. 2011) from the statistical pattern of English word-initial / onset clusters
    - Some attested obstruent-nasal and obstruent-obstruent onsets begin with a voiceless obstruent: [sm sn sp st sk] (marginally [sf ʃm ʃn ʃp ʃt])
    - No attested obstruent-nasal or obstruent-obstruent onsets begin with a voiced obstruent (*[zm zn …])
    - No attested onsets have voiced obstruents non-initially ([sf] vs. *[zv])

- Maxent phonotactic learner (Hayes & Wilson 2008) induces strong constraints *[-son,+voice]X and *X[-son, +voice] from English onset lexicon

Hypothesis 3: Transfer of native gestures

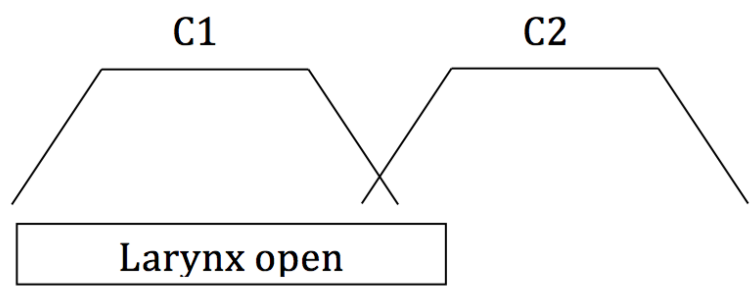
Many instances of 'epenthesis' do not involve insertion of a vocalic gesture, but failure to *tightly coordinate* the constriction gestures of the consonants
(Davidson 2003, et seq.; related proposals for native phonology by Gafos 2002; Hall 2006, 2011; Ridouane 2008; Gouskova & Hall 2009; Fougeron & Ridouane 2011)

- **Phonetic transfer**
  Native phonetic knowledge (gradiently) biases patterns of gestural overlap and timing in favor of those that are typical of the participant's language
  - Stop constrictions should not overlap word-/syllable- initially
  - Realization of 'voicing' should be as in English word-/syllable- initial stops

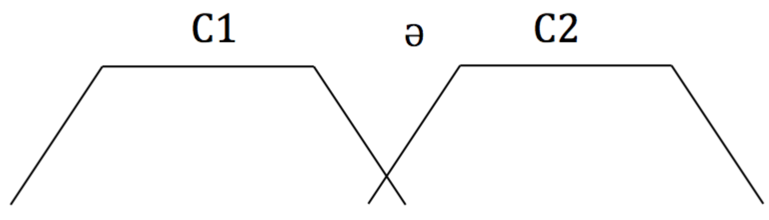
- **Phonetic imitation**
  Even when a vocalic gesture is inserted, participants attempt to match the acoustic-phonetic properties of the stimuli and may apply schwa devoicing
  (see Davidson 2006 on schwa 'deletion' as devoicing, and Chistovich et al. 1966; Marler & Mundinger 1977; Flege & Eefting 1988; Goldinger 1998; Nielsen 2011 on phonetic imitation)

Articulatory account of voicing asymmetry

Native word-initial voiceless stops have long aperiodic releases that inhibit transitional vocoids even when C1 and C2 constrictions have minimal overlap

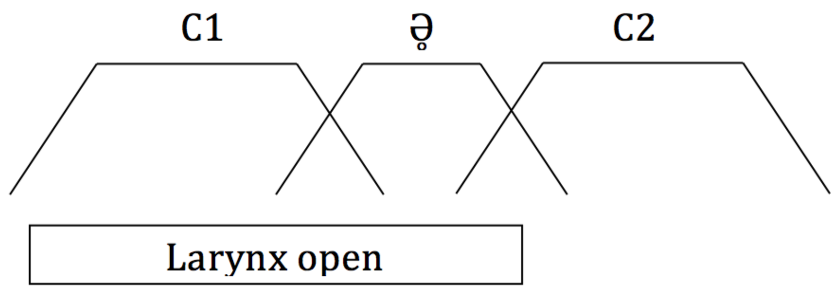


Native voiced stops have short aperiodic releases, transitional vocoids ("vocal releases") arise when C1 and C2 constrictions are minimally overlapping

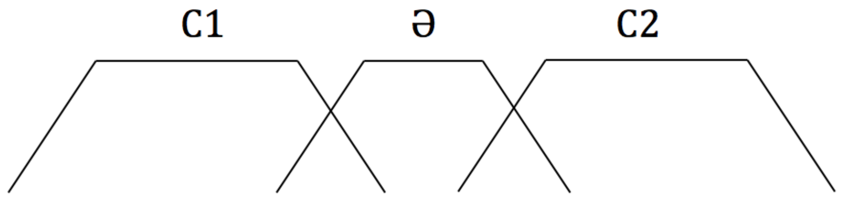


Articulatory account of voicing asymmetry

Native word-initial voiceless stops have long aperiodic releases that can sometimes devoice schwas inserted between C1 and C2



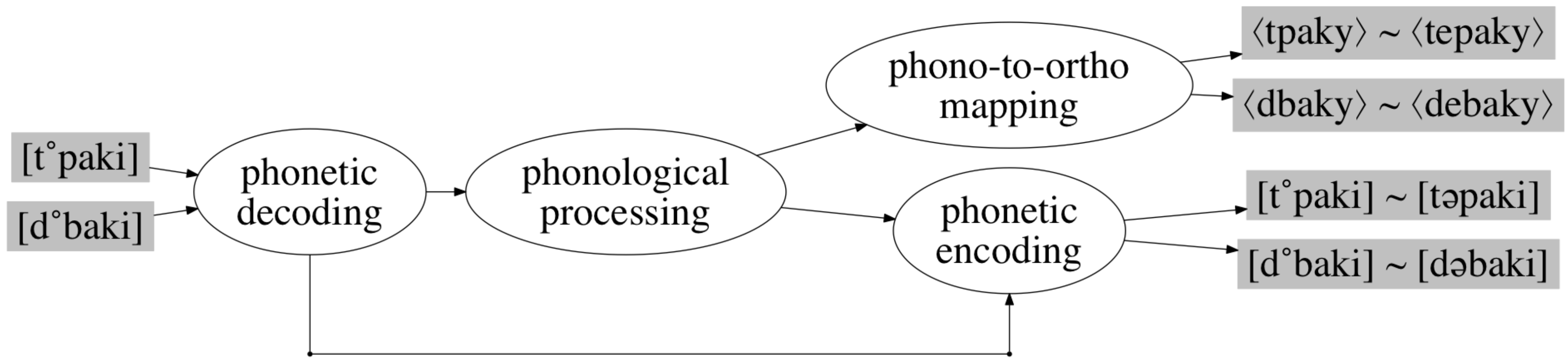
Native voiced stops have short aperiodic releases, therefore inserted vowels are produced with voicing and formant structure



Distinguishing among the hypotheses

If voiced stop-initial clusters are more often represented with a true epenthetic schwa — as the result of (1) misperception and/or (2) active repair — *the voicing asymmetry should not be limited to repetition or other speech production tasks*

Misperception and active repair hypotheses predict that voiced stop-initial clusters should also be *spelled* more often with a vowel (e.g., ⟨e⟩)



(English spelling reflects (morpho)phonemic structure, not allophonic detail such as stop aspiration, vowel nasalization, final devoicing, …; Chomsky & Halle 1968)

Forced-choice transcription experiment

English listeners (N=90) heard the **same recordings** as in the earlier repetition studies and performed a **forced-choice transcription task** on MTurk
(similar tasks: Berent et al. 2007; Peperkamp et al. 2008; Dupoux et al 2011; Escudero & Williams 2011)

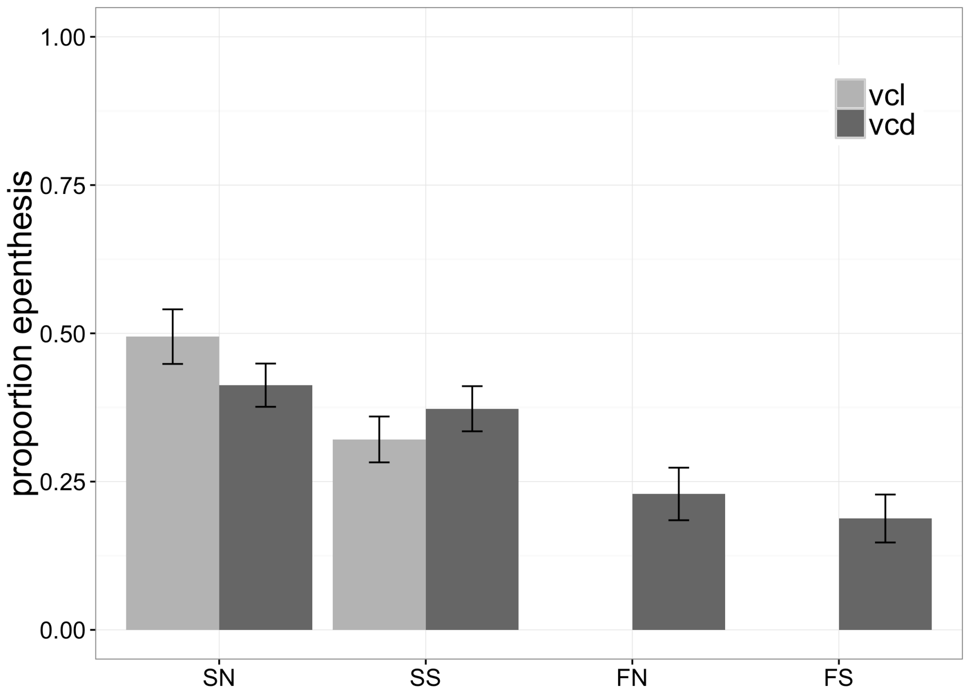
Ex. **t˚páki** (Russian production) … **t˚páki** (replay) … [now choose one]

etpaky	tpaky
tepaky	paky

Transcription choices, arranged randomly across participants:
cluster (C1C2), epenthesis (C1eC2), prothesis (eC1C2), other (C1 del / ftr change)

Large number of participants ensured coverage of the stimuli with a small number of trials (M=28) per run ~ 15 min including orthography instructions

Voicing effect on epenthesis in transcription?



Mixed-effects logistic regression analysis of responses to cluster-initial stimuli

- **No main effect of voicing on epenthesis** $\beta = 0.02, p > .9$
- More epenthesis after stops than fricatives $\beta = 1.41, p < .01$
- More epenthesis with longer stop releases $\beta = 0.45, p < .01$

Anti-sonority-sequencing effect on epenthesis (small rise > plateau/fall, $p < 0.01$)

Accurate performance on filler items (epenthesis: 89%, prothesis: 99%)

Interaction analyses

No main effect of stop voicing found in transcription, but there was a moderate interaction between voicing and C2 manner ($\beta = 0.88, p = .03$)

- Numerically *higher* rate of epenthesis after *voiceless* stops for SN stimuli
- Effect of voicing did not reach significance in post-hoc tests of SN and SS subsets (cf. release duration significant for both subsets)
- Is the voicing × C2-manner interaction reliable in transcription?

Most crucially, **strong interaction between task and stop voicing** when repetition and transcription were analyzed together ($\beta = 1.62, p < .01$)

- *Voicing effect on 'epenthesis' is larger (and consistently in the direction voiced > voiceless) for repetition in comparison to transcription*

Additional transcription experiments

1. Replication of MTurk experiment in the soundbooth (N=16, original lists combined to give more trials per participant)
   - More epenthesis for **stop-initial vs. fricative-initial clusters** ($\beta = 0.90, p < .01$) and for **longer stop releases** ($\beta = 0.50, p < .01$)
   - **No main effect of stop voicing**, moderate interaction of C2 manner and voicing suggesting *more* epenthesis for *voiceless* SN clusters ($p < .05$)

2. MTurk replication with additional filler trial types in which 'C1 deletion' (e.g., *paky*) and 'C1 change' (e.g., *faky*) were correct response options
   - Same pattern as in first MTurk experiment, again with interaction between voicing and C2-manner (but no significant main voicing effect)

3. MTurk replication with unmanipulated stimulus items from Davidson (2010)
   - Same effects of stop vs. fricative, stop release duration, and anti-sonority-sequencing, **no main effect of stop voicing or interaction**

Conclusions / Implications

- Some epenthesis errors observed in production do arise from schwa insertion by misperception and/or phonological repair (see also Shaw & Davidson 2011)
  - True schwa insertion is also reflected in choice of ⟨e⟩ transcription
- **Voicing asymmetry does not reflect a difference in the rate of schwa epenthesis, but arises in phonetic encoding**
  (compare similar analysis of native clusters in Berber: Ridouane 2008, et seq.; Georgian: Chitoran et al.)
  - Voiceless clusters, when represented faithfully by perception and phonology, have glottal spreading gestures that inhibit transition vocoids
  - Voiceless clusters, when repaired by schwa insertion, can be produced with schwa devoicing in the service of phonetic imitation
  - Voiced clusters develop transitional schwas if constriction overlap is minimal and reveal epenthetic schwas when they are repaired

Conclusions / Implications

- Phonetic encoding account depends upon fundamental properties of the stop voicing contrast in English and restricted phonetic imitation
  - Laryngeal spreading gesture of voiceless stops, not available to voiced stops, can inhibit voicing in open transitions and epenthetic schwas
  - Speakers attempt to match the phonetic profile of nonnative clusters but are (gradiently) constrained by transfer native phonetic patterns
- This account of the asymmetry requires that upstream processes represent nonnative stop-initial clusters intact at a fairly high rate (> 50%)
  - Nonnative clusters are not consistently mapped to native structures, even at later stages of speech perception & with high spectral similarity
  - Phonotactics does not ineluctably drive active repair of nonnative clusters, even those that are highly marked

Conclusions / Implications

- Measure of perceptual similarity relevant for phonetic decoding must selectively weight available acoustic-phonetic cues for schwa
  - Greater spectral similarity between voiced releases and schwa is apparently not sufficient to induce higher rates of misperception
  - Aperiodicity and duration appear to be more important for perceptual determination of English schwa presence (absent top-down influence)
- Lexical knowledge — and projected gradient phonotactics — were not strongly engaged by the present repetitition and transcription tasks
  - Stimuli were not very word like, tasks did not involve word learning (but: similarity to frequent lexical items such as *tomorrow*, *tomato*, *c'mon*, *c'mere* may underly voiceless > voiced epenthesis asymmetry found in transcription of SN clusters)
  - Effects of sonority sequencing were weak and inconsistent with projection
  - Nonlexical phono-ortho mapping sufficient for transcription choice

Summary

- Effect of stop voicing on 'epenthesis' rate is large (20% - 30%) and consistent across many experiments on repetition of nonnative clusters
- Voicing asymmetry in repetition could be due to misperception, active phonotactically-driven repair, or transfer of native gestural patterns
- No consistent voicing asymmetry was found in forced-choice transcription experiments with the same auditory stimuli
  - Maximally 10% voiced > voiceless for SS clusters
  - Opposite effect (voiceless > voiced) for SN clusters
- Combining results from multiple tasks can resolve the origin of errors in nonnative speech processing: voicing asymmetry arises 'upstream' of perceptual and phonological processing (after ortho-to-phono conversion)

Thank you!

Acknowledgments

- Experiment participants in the lab and cyberspace
- Alice Hall, Francesca Himelman, Johnny Mkitarian, Elizabeth George, Steven Foley for their assistance in coding the data
- Members of the JHU and NYU Phonetics/Phonology groups, and audiences at AMP 2014 and 2015, LabPhon 2014, ICPhS 2015, Harvard, Michigan
- National Science Foundation grants BCS-1052784 to Wilson, BCS-1052855 to Davidson