Does lexical frequency affect phonetic traces in speech errors? A replication study

Thomas Denby, Emily Cibelli, Gia Eapen, Erin Madigan, and Matt Goldrick

Replication is vital, but difficult
• Published experiments in psychological sciences may overestimate effect sizes
• Replicating speech production studies highly resource-intensive
• Manual measurement, Goldrick et al. (2011): 3000 person hours

High-powered replication of Goldrick et al. (2011) using automated phonetic analysis tools
• Tools from Goldrick et al. (2016) measure 97k stop consonant voice onset times (VOTs) elicited in tongue twister paradigm
• Replicate original results with mixed effects regression to better account for participant and item variance
• Pre-registered: https://osf.io/32bhw/
• Statistical power estimated with Monte Carlo simulations; number of participants set to maximize power (β > 0.85)

Speech error phonetics influenced by production target and lexicon
• Target productions leave phonetic “trace” in resulting speech error: [b] from pin → bin
• Error more [p]-like (longer VOT) than [b] from correct bin → bin production
• Lexical information modulates traces (e.g., larger traces for nonword vs. word outcome)

Target study: Lexical frequency modulates phonetic traces in errors
• Integration hypothesis: Lexicon includes abstract phonology, specification of allowable range of phonetic variation
• High frequency (HF) allows substantial variation, weakly constraining phonetics
• LF words specify narrow range of phonetic variation
• LF targets strongly indicate target properties should be present
• LF outcomes strongly indicate target properties should be absent

Traces larger for LF targets, HF outcomes (Goldrick et al., 2011)
But effect is very small—is it reliable?

Methods
• 35 participants quickly repeated twisters made up of rhyming pairs differing in initial stop voicing (e.g., bin → pin in bin)
• Frequency of target or error outcome manipulated to create 20 quadruplets (80 words)

Contrasting Pair

<table>
<thead>
<tr>
<th></th>
<th>Controlled Pair</th>
</tr>
</thead>
<tbody>
<tr>
<td>bill (freq: 54)</td>
<td>pill (freq: 13)</td>
</tr>
<tr>
<td>bin (freq: 5)</td>
<td>pin (freq: 13)</td>
</tr>
</tbody>
</table>

• 86k VOTs analyzed (9.6k excluded due to poor alignments/ outlier VOT values)
• Gamma mixture model, assuming two target categories (voiced, voiceless) fit to each participant’s VOTs
• Error: intended category ≠ category assigned by mixture

Results: Phonetic traces in errors

Outcomes

- Voiceless

- Voiced

- Original: 26% error [range: 13-50%]
- Replication: 11% error [range: 3.6-30%]

Results: Modulation of traces by lexical frequency

- Voiced, Target Frequency
- Voiced, Outcome Frequency
- Voiceless, Target Frequency
- Voiceless, Outcome Frequency

Large trace effect robust; Smaller interaction effects less robust

- 4 models, one for each target/outcome frequency combination
- Fixed effects: accuracy, frequency, and interaction
- Original/simple model: random intercepts only
- New maximal model: random intercepts + slopes (accuracy for items; all fixed effects for participants)
- Main effect of accuracy significant in all analyses.
- Crucial measure: accuracy:frequency interaction

Do LF targets and HF outcomes have larger effect on traces?

<table>
<thead>
<tr>
<th></th>
<th>Original paper</th>
<th>Simple model</th>
<th>Maximal model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced, target</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiced, outcome</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Voiceless, target</td>
<td>✔</td>
<td></td>
<td>n.s.</td>
</tr>
<tr>
<td>Voiceless, outcome</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next steps: Replicate and expand McMillan and Corley (2010)

- Relative to baseline non-switching context (/t/-/t/), VOTs more variable when segments differ in one feature (/t/-/d/) vs. two features (/t/-/k/)
- High-powered replication using automatic phonetic tools
- Extension 1: Do different baseline contexts (original: “X X X” vs. “they X X him”; Goldrick & Blumstein, 2006) change estimates of VOT variability?
- Extension 2: Compare original analysis methods to Goldrick, et al., (2011)