(Im)perceptible Incomplete Neutralization

Two experiments on flapping in American English

Aaron Braver
abraver@rutgers.edu

RUMMIT II
16 May, 2011

What I’ll be talking about

• Background
  – Incomplete neutralization
  – Some previous studies
• The perception experiments
  – Identification
  – ABX
• What does it mean?
  – Implications for incomplete neutralization

Background

Complete neutralization

• /X/ → [Z] / (Context A)
• /Y/ → [Z] / (Context A)

• The traditional picture of German final devoicing:
  – /wat/ ‘advice’ –> [ war ]
  – /wad/ ‘wheel’ –> [ war ]

Incomplete neutralization

• /X/ → [Z] / (Context A)
• /Y/ → [Z] / (Context A)

• The picture of German final devoicing from acoustic studies (Port and O’Dell 1985, though see fourakis and iverson 1986)
  – /wat/ ‘advice’ –> [ war ]
  – /wad/ ‘wheel’ –> [ war ]

Flapping

writer –> [ zairə ]

rider –> [ zairə ]

OR?

writer –> [ zairə ]

rider –> [ zairə ]
My production studies

- Acoustic study 1 (Braver 2010, 2011)
  - 13 speakers
  - Pre-/d/-vowels longer than pre-/t/-vowels (8.76ms)
- Acoustic study 2 (Braver 2010, 2011)
  - 12 speakers
  - Pre-/d/-vowels longer than pre-/t/-vowels (3.45ms)
- The neutralization is incomplete

(See also Fisher and Hirsh (1976), Fox and Tenboek (1977), Zue and Lafreniere (1979), Huff (1980), Herd et al. (2010). But, see (partially) contrary results in Juus (1942), Port (1976)).

The question

- Speakers produce a distinction between /d/-flaps and /t/-flaps…
- …but this distinction is tiny
- Can listeners perceive this distinction?

Herd et al. 2010

- Production task:
  - Pre-/d/-flap vowels were 6ms longer than pre-/t/-flap vowels

Herd et al. 2010

- Identification task
  - Listeners heard a word and were asked which of two words on the screen it was:
    - Hear: [lɪɾə] “liter”
    - See: “leader”

Motivation for the current experiments

- Mitigate frequency effects
  - Nonce words
- How hard is the task itself?
  - ID and ABX
  - Feedback
- Keep bias in mind during analysis

Herd et al. 2010

- How did they do?
  - Near chance
  - /d/-tokens were correctly identified more frequently than /t/-tokens
  - Lexical frequency effects
    - Low frequency /t/ words: 33% correct
    - High frequency /t/ words: 55% correct
- /d/-bias and frequency bias (Connine et al. 1993), rather than preceding vowel duration, help determine listeners’ responses

(For previous perception studies of incomplete neutralization in non-flapping contexts, see Port and O’Dell (1985), Warner et al. (2004)).
The perception experiments

Two tasks

- Identification
  - Did that word have a /d/ or a /t/ in the target location?

- ABX
  - Was word X the same as word A or word B?

The tokens

- Nonce words
  - First syllable: unstressed
    - Onsets: p/t/b/d
    - Nucleus: ə
  - Second (target) syllable
    - Onsets: p/t/k
    - Nuclei: i/ɛ/æ
    - Coda: d/t
  - “-ing” was added to each bisyllabic nonce word, putting the final /d/ or /t/ in a flapping environment

Some representative minimal pairs:

puhPEET-ing puhPEED-ing
puhKAET-ing puhKAAD-ing
puhTAET-ing puhTAAD-ing

The tokens

- Taken from speakers in the 2nd acoustic study
- 12 speakers produced each token in 2 tasks
  - “Wug” task (Berko 1958, Foulkes and Kernan 1984)
    - John learned how to buhKEED this week. He was _________ this whole week.
    - Speakers read the sentences, filling in the “-ing” form—e.g., “buhKEEDing”
  - Minimal pair reading task
    - John learned how to buhKEED this week. He was buhKEEDing this whole week.
    - John learned how to buhKEET this week. He was buhKEETing this whole week.
  - No significant differences across tasks

The tokens

- Tokens were selected from three speakers who had the biggest difference between pre-/d/ and pre-/t/ vowel duration, and who accurately produced a sufficient number of tokens
- Balanced for onset and vowel of target syllable, as well as /d/ vs. /t/

Identification
Identification

• Instructions and practice with real and nonce words
• 3 blocks, each from different speakers
• Each block had 36 tokens (half /d/, half /t/), randomized, repeated 3 times (=108 per block)
• Feedback on each trial
• Break between each block
• Order of blocks balanced (Latin Square) across 21 listeners

Identification trials

• Trial 1  t
• Trial 2  d
• Trial 3  t

How well did listeners do?

• Percent-correct
  – “Listners identified X from Y more than 50% of the time.”
  – But:
    • How biased were the speakers towards X or Y?
    • Herd et al.’s listeners were biased towards saying /d/, and as such /d/ has a high percent correct rate.
      – But does this mean that /d/ was more perceptible than /t/?

\(d’\) for the intelligent non-psychophysicist

\[
\begin{array}{c|c|c|c|c}
\text{What the operator says} & \text{What is happening} \\
\hline
\text{Missile} & \text{Hit} & \text{Miss (uh-oh)} \\
\text{No Missile} & \text{False alarm} & \text{Correct rejection} \\
\end{array}
\]

• Hit rate
  • When there’s a missile, how often do they say “missile!”?
• False alarm rate
  • When there’s no missile, how often do they say “missile!”?

Hit Rate vs. False Alarm Rate

(Says “missile!” when there’s a missile, but not when there isn’t one.)

Macmillan and Creelman (2005)
(Im)perceptible Incomplete Neutralization

5/16/11

Hit Rate vs. False Alarm Rate

d' for ID

What the listener says they heard:
- "That was a /d/"
- "That was a /t/"

What's actually happening:
- Hear /d/ → Hit
- Miss (uh oh)
- Hear /t/ → False alarm
- Correct rejection

ID results

- d' is not significantly different from 0 overall (mean d': -0.04, Wilcoxon test: V=76, n.s.)
- So—listeners said “it’s a d'” about the same number of times for /d/ as for /t/
- (If these were missiles, we’d be in trouble)
**What’s ABX?**

- Three stimuli per trial (A, then B, then X)
- Participants decide whether the third (X) was the same as A or as B

**ABX**

- Instructions and practice with real and nonce words
- 3 blocks, each from different speakers
- Each block had 72 trials (18 each of d-t-t, d-t-d, t-d-d, t-d-t), randomized
- Feedback on each trial
- Break between each block
- Order of blocks balanced (Latin Square) across 21 listeners

**ABX: ISI**

A (250ms) B (500ms) X

- Longer B-X ISI to induce categorical, rather than auditory mode of perception  
  (in the sense of Gerrits and Schouten 2004)
- Goal: get at a categorical distinction, using an easier task than identification
  - Category labeling takes place after 100-200ms
  - Discrimination performance reaches a maximum between 500–1000ms  
  (Gerrits and Schouten 2004)

**ABX trials**

- A (250ms) B (500ms) X
- **Trial 1** A (d t d)
- **Trial 2** B (t d d)
- **Trial 3** B (t d d)
- **Trial 4** A (t d t)

**d’ for ABX**

<table>
<thead>
<tr>
<th>What actually happening</th>
<th>What the listener says is happening</th>
</tr>
</thead>
<tbody>
<tr>
<td>X = A</td>
<td>Hit</td>
</tr>
<tr>
<td>X = B</td>
<td>False alarm</td>
</tr>
<tr>
<td>X = B</td>
<td>Correct rejection</td>
</tr>
</tbody>
</table>

What’s actually happening

What the listener says is happening

Aaron Braver (abraver@rutgers.edu)
(Im)perceptible Incomplete Neutralization

ABX results

• **d’ is significantly different from 0 overall** (mean d’: 1.24; Wilcoxon test: V=231, p < 0.001)

• So, listeners said “X is like A” more often when X was actually like A than when X was actually like B

• BUT…

ABX: Statistical vs. linguistic significance

• Listeners reported using cues like “the up and the down of each one” (intonation)

• Remember, X is literally the same as one of either A or B
  – Listeners were likely to have used any auditory differences between A and B, including intonation
  – Some of these cues may have nothing to do with an underlying voicing contrast

Testing the “any auditory differences” strategy hypothesis

• AB task
  – Which one had a /d/—word A or word B?

• Speakers can’t use the “any auditory differences” strategy, since there are only two sounds, and they are never the same

• (Forthcoming)

So what?

• Speakers might be able to discriminate between flapped /d/ and /t/, but don’t identify them—in ideal lab conditions, even in tokens taken from the minimal pair reading task

• These results generally support the findings of Herd et al. (2010) that listeners can’t identify flapped /d/ and /t/
  – This holds even in relatively easy tasks, when frequency effects are mitigated, and bias is taken into account
So what?

- Many speakers produce this distinction anyway, in both "wug" and minimal pair reading tasks (Braver 2010, 2011).
- Given the perception results, they probably do not do it for the benefit of hearers.

Conclusions

- American English speakers produce a small preceding vowel duration distinction between flapped /d/ and /t/.
- The ID experiment suggests that listeners can't tell them apart, even in an ideal situation.
- The ABX experiment showed moderate discriminability, but speakers reported using cues unrelated to the /d/-/t/ distinction.
- Speakers don't maintain the distinction for the benefit of listeners, since they can't perceive it.

Thanks!

Thanks are due to Shigeto Kawahara, Bruce Tesar, Kristen Syrett, Peter Staroverov-Vandana Bajaj, the members of the Spring 2011 Rutgers phonetics seminar, and the audience of RULing VI for their comments and insights on this project.

References