Perception of Incompletely Neutralized /d/ and /t/ Flaps in AmE

Aaron Braver :: Rutgers University

Introduction

- **Incomplete Neutralization (IN)**
  - /X/ → [Z^X] / (Context A)
    /Y/ → [Z^Y] / (Context A)
  - Final devoicing is a commonly cited case of *IN*.
- **AmE Flapping** is a potential case of *IN*:
  - /t/ and /d/ become [ɾ] in certain prosodic configurations (Kahn 1980), but distinctions remain on the surface.
  - Most notably, **vowels before /d/-flaps are longer than vowels before /t/-flaps** (Braver 2011, Herd et al. 2010) (though see Port (1976)).
  - This distinction is generally less than 10ms.

Background, Questions, and Motivation

- **Perception studies of IN show mixed results**:
  - Port and O'Dell (1985), Warner et al. (2004): listeners can perceive the difference between *IN* segments.
  - Herd et al. (2010) present an identification task showing that listeners cannot correctly categorize /d/-flaps and /t/-flaps in actual AmE words.
- **Three questions:**
  1. Can AmE listeners *categorize* /d/-flaps and /t/-flaps?
  2. Can they *distinguish* /d/-flaps from /t/-flaps?
  3. Why do (some) speakers produce this distinction?
- Previous studies addressing these questions leave a number of issues open.
  1. They generally rely on actual words of a language, potentially introducing frequency bias on perceptual categorization.
  2. Even though listeners have a general bias towards /d/ (Herd et al. 2010), measures of performance do not take this into account.
  3. Most studies have relied solely on identification tasks (as opposed to discrimination tasks).
- Contributions of this study:
  - Shows that **AmE listeners can neither discriminate nor properly categorize /d/-flaps from /t/-flaps on the basis of category**.
  - Addresses concerns of frequency effects through the use of nonce word stimuli.
  - Addresses issues of bias through the use of d′ as a measure of performance.

Stimuli

- Tokens were trisyllabic nonce words taken from a related production experiment (Braver 2011).
  - 12 speakers produced each token in two tasks (no significant differences were found across tasks):
  - σ 1: onsets {p,t,b,d}, nuclei: ə
    σ 2: onsets {p,t,k}, nuclei: {i, ɛ, æ}, codas: {d,t}
    σ 3: '-ing' (places d/t in flapping environment)
- Some representative minimal pairs:

puhPEET-ing	puhPEED-ing
tuhKAT-ing	tuhKAD-ing
duhTEHT-ing	duhTEHD-ing

- Tokens were selected from three speakers, based on the following criteria:
  - Largest difference between pre-/d/ and pre-/t/ vowel duration.
  - Accurate production of a sufficient number of tokens.
  - Balanced for onset and vowel of target syllable, as well as /d/ vs. /t/.

Methods, Part I

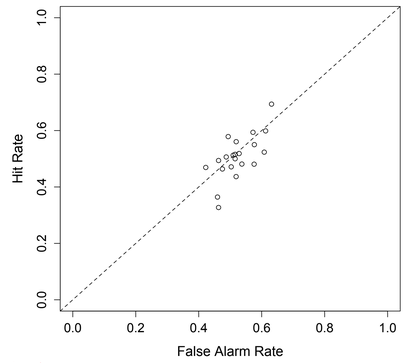
- 42 undergraduates participated in two tasks (21 per task).
- Feedback was given on each trial.
- Three blocks, each from a different speaker.
  - Block order was balanced (Latin Square) across all listeners.
- **Identification Task:**
  - Listeners heard a single token, and were asked whether the sound before the '-ing' was a /d/ or a /t/.
- **ABX Task:**
  - Listeners heard three stimuli per trial (A, B, then, X), and were asked to decide whether X was the same as A or the same as B.
  - A relatively long ISI (500ms) was used between sounds B and X with the goal of inducing a categorical, rather than purely auditory mode of perception.

Results, Part I

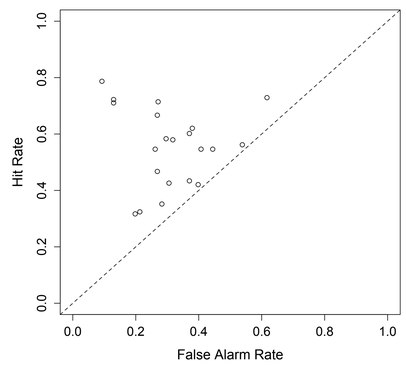
- **Identification Task:**
  - Listeners' d′ scores were not significantly different from 0 (mean d′ : -0.04, Wilcoxon test: $V = 76$, *n.s.*).
  - Listeners said "it's a /d/" just as often when they had heard a /d/ as when they had heard a /t/.

Results, Part 1 (Continued)

- **Hits vs. False Alarms for the Identification Task:**



- **ABX Task:**
  - Listeners' d′ scores were significantly different from 0 (mean d′ : 1.24, Wilcoxon test: $V = 231, p < 0.001$).
  - Listeners said "X is like A" more often when X was actually like A than when X was actually like B.
- **Hits vs. False Alarms for the ABX Task:**



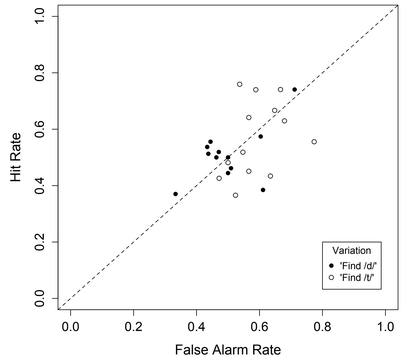
- **Why did listeners do better in the ABX task than the Identification Task?**
  - Listeners anecdotally reported using cues unrelated to the /t/∼/d/ distinction (e.g., intonation contour).

Methods, Part II

- To test whether listeners discriminated based on irrelevant acoustic differences between A and B in the ABX task, an **AB Task** (a.k.a 2AFC) was run.
  - Allows for comparisons (like the ABX task), but no two tokens are the same on a given trial (like the ID task).
- 21 listeners heard two tokens (from the same set as the previous tasks) per trial.
  - Half of the trials had /d/ first and half had /t/ first.

Results, Part II

- In the **AB task**, listeners' d′ scores were not significantly different from 0 (mean d′ : -0.02, Wilcoxon test $V = 148$, *n.s.*).
- **Hits vs. False Alarms for the AB Task:**



Discussion and Conclusions

- The low d′ scores in the identification task suggest that listeners were unable to categorize /d/-flaps and /t/-flaps.
- While listeners were able to distinguish /d/-tokens from /t/-tokens in the ABX task, they were unable to do this in the AB task where they heard only two sounds per trial, which are never identical.
  - Speakers were unable to use the 'unrelated cues' strategy in the AB task, suggesting that listeners cannot distinguish /d/-flaps from /t/-flaps on the basis of cues relevant to the underlying voicing contrast.
- These results hold in both an identification and a discrimination task, and when frequency effects are mitigated through the use of nonce words.
- If listeners are neither able to distinguish nor categorize /d/-flaps and /t/-flaps, speakers who maintain this distinction must be doing so for reasons other than listeners' benefit.

Selected References

A. Braver 2011. *Incomplete Neutralization in American English Flapping: A Production Study*. In PLC Proceedings 34 http://repository.upenn.edu/pwpl/vol17/iss1/5/

R. Port and M. O'Dell 1985. *Neutralization and Syllable-Final Voicing in German*. JPhon 13:455–471

W. Herd, A. Jongman, and J. Sereno 2010. *An Acoustic and Perceptual Analysis of /t/ and /d/ flaps in American English*. JPhon 38:504–516

Acknowledgements

Thanks are due to Shigeto Kawahara, Bruce Tesar, Kristen Syrett, and the Spring 2010 Rutgers Phonetics Seminar.

abraver@rutgers.edu :: http://www.aaronbraver.com

Perception of Incompletely Neutralized /d/ and /t/ Flaps in American English

Aaron Braver*, Rutgers University
abraver@rutgers.edu :: http://www.aaronbraver.com
NELS 42 :: 11–13 November, 2011

1 Introduction

(1) ***Complete* neutralization**: two underlyingly different segments become the same in the phonetic output in some context, effectively neutralizing the contrast.
   a. /X/ → [Z] / (Context A)
   b. /Y/ → [Z] / (Context A)
   c. Ex: the 'traditional' picture of German final devoicing:
      /ʁɑt/ 'advice' → [ʁɑt]
      /ʁɑd/ 'wheel' → [ʁɑt]

(2) ***Incomplete* neutralization**: two underlyingly different segments become *nearly* identical in the phonetic output—unlike complete neutralization, some small trace of the underlying distinction remains on the surface:
   a. /X/ → [Z^X] / (Context A)
   b. /Y/ → [Z^Y] / (Context A)
   c. Ex: the picture of German final devoicing from from acoustic studies (e.g., Port and O'Dell (1985))[1]:
      /ʁɑt/ 'advice' → [ʁɑt]
      /ʁɑd/ 'wheel' → [ʁɑːt]

(3) Final devoicing is the most commonly cited case of incomplete neutralization, with evidence from German (as above), Catalan (Dinnsen and Charles-Luce 1984), Polish (Slowiaczek and Dinnsen 1985, Slowiaczek and Szymanska 1989), Russian (Dmitrieva 2005), and Dutch (Warner et al. (2004), though see Warner et al. (2006) for caveats).

(4) American English Flapping as incomplete neutralization:
   a. In certain prosodic contexts, /d,t/ → [ɾ] (Kahn 1980)
   b. Previous studies show a difference between /d/-flaps and /t/-flaps (Herd et al. 2010, Fisher and Hirsh 1976, Fox and Terbeek 1977, Zue and Laferriere 1979, Huff 1980; but see (partially) contrary results in Joos 1942, Port 1976).

(5) My previous production studies (Braver 2010, 2011):
   a. Acoustic Study 1 (13 speakers)

*Thanks are due to Shigeto Kawahara, Bruce Tesar, Kristen Syrett, Wendell Kimper, Ryan Denzer-King, Sara O'Neill, and the Spring 2010 Rutgers Phonetics Seminar.

[1]Though, see Fourakis and Iverson (1984)

i. Pre-/d/ vowels longer than pre-/t/ vowels (by 8.76ms, on average)

b. Acoustic Study 2 (12 speakers)

i. Pre-/d/ vowels longer than pre-/t/ vowels (by 3.45ms, on average)

2 Background, Questions, and Motivation

(6) The differences between /d/-flaps and /t/-flaps found in previous studies are quite small

(7) Questions:

a. Can AmE listeners categorize /d/-flaps and /t/-flaps?

b. Can they distinguish /d/-flaps from /t/-flaps?

c. Why do (some) speakers produce this distinction?

(8) Previous perception studies of incomplete neutralization show mixed results:

a. Port and O'Dell (1985), Warner et al. (2004): listeners can perceive the difference between incompletely neutralized segments (in German and Dutch final devoicing)

b. Herd et al. (2010) present an identification task, showing that listeners cannot correctly categorize /d/-flaps and /t/-flaps in actual words of American English

i. Performance was near chance, though /d/ tokens were correctly identified more frequently than /t/ tokens

ii. Lexical frequency effects: low frequency /t/ words were correctly identified 33% of the time, while high frequency /t/ words were correctly identified 55% of the time

(9) These previous studies leave a number of issues open:

a. They generally rely on actual words of a language, potentially introducing frequency bias on perceptual categorization

b. Even though listeners have a general bias towards /d/ (Herd et al. 2010), measures of performance do not take this into account

c. Most studies have relied solely on identification tasks (as opposed to discrimination tasks)

(10) This study addresses these issues:

a. Frequency effects are mitigated through the use of nonce word stimuli

b. Bias is taken into account through the use of d′ as a measure of performance

c. The study involves both identification and discrimination tasks

3 Stimuli

(11) Token schema:

a. First syllable: unstressed

i. Onsets: p/t/b/d

ii. Nucleus: ə

b. Second ('target') syllable: stressed
   i. Onsets: p/t/k
   ii. Nuclei: i/ɛ/æ
   iii. Coda: d/t

c. '-ing' was added to each bisyllabic nonce word, putting the final /d/ or /t/ in a flapping environment

(12) Sample minimal pairs:

puhPEET-ing ~ puhPEED-ing

tuhKAT-ing ~ tuhKAD-ing

duhTAT-ing ~ duhTAD-ing

(13) Tokens were taken from speakers in a previous acoustic study (Braver 2011). 12 speakers produced each token in 2 tasks:

a. 'Wug' task (Berko 1958, Fourakis and Iverson 1984)
   i. John learned how to buhKEED this week. He was _____ this whole week.
   ii. Speakers read the sentences, filling in the '-ing' form—e.g., 'buhKEED-ing'

b. Minimal pair reading task
   i. John learned how to buhKEED this week. He was buhKEED-ing this whole week.
   ii. John learned how to buhKEET this week. He was buhKEET-ing this whole week.

c. No significant differences across tasks

(14) 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. Tokens were balanced for onset and vowel of target syllable, as well as for /d/ vs. /t/.

4 Methods, Part I

(15) 42 undergraduates participated in two tasks (21 per task).

(16) Each task was comprised of instructions and practice, followed by three blocks (each with tokens from a different speaker), with block order balanced (Latin Square) across all listeners. Feedback was given on each trial in both tasks.

4.1 Identification Task

(17) On each trial, listeners heard a single token, and were asked whether the sound immediately preceding the '-ing' was a /d/ or a /t/

(18) For example:

a. Listeners hear 'buhKEED-ing', and should respond '/d/'

b. Listeners hear 'buhKEET-ing', and should respond '/t/'

(19) Each block consisted of 36 trials (half /d/, half /t/), randomized, repeated 3 times (=108 trials per block)

4.2 ABX Task

(20) On each trial, listeners heard three stimuli (A, then B, then X), and were asked to determine whether the third (X) was the same as A or as B

(21) For example:

a. Listeners hear 'buhKEED-ing buhKEET-ing buhKEED-ing' and should respond 'A'

b. Listeners hear 'buhKEED-ing buhKEET-ing buhKEET-ing' and should respond 'B'

(22) The B–X ISI (500ms) was longer than the A–B ISI (250ms), in order to induce a categorical, rather than auditory mode of perception (in the sense of Gerrits and Schouten (2004))

a. Goal: get at a categorical distinction while using a task that's easier than identification.

b. Category labeling takes place after 100–200ms

c. Discrimination performance reaches a maximum between 500–1000ms

(23) Each block consisted of 72 trials (18 each of d-t-t, d-t-d, t-d-d, t-d-t), randomized

5 Results, Part I

5.1 d′

(24) d′ is a measure of sensitivity that takes bias into account. It can be thought of through a military analogy:

		What the radar operator says	
		"Missile"	"No Missile"
What's actually happening	Missile	Hit	Miss
	No Missile	False alarm	Correct rejection

(25) Crucially, d′ takes both the hit rate and the false alarm rate into account

(26) H = hits / (hits + misses)
F = false alarms / (false alarms + correct rejections)
For most simple cases[2], d′ = z(H) - z(F)

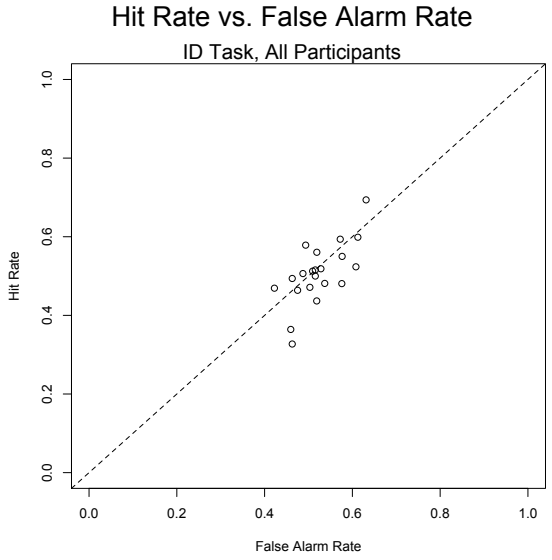
5.2 Identification Task

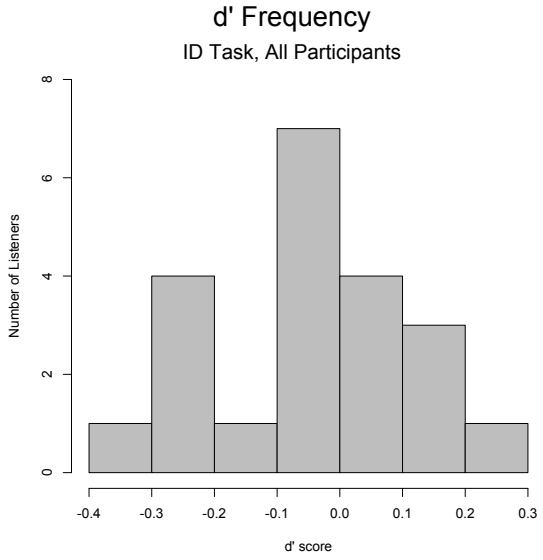
(27) Results from the Identification Task:

[2] d′ was computed this way for the Identification Task. d′ for the ABX task was computed with the R `PsyPhy` package. See Macmillan and Creelman (2005).

a. d′ is not significantly different from 0 overall (mean d' : -0.04, Wilcoxon test: $V = 76, n.s.$)

b. Listeners said "it's a /d/" just as often when they had heard a /d/ as when they had heard a /t/.





5.3 ABX Task

(28) Results from the ABX Task:

a. d′ is significantly different from 0 overall (mean d' : 1.24, Wilcoxon test: $V = 231, p < 0.001$.)

b. Listeners said "A is like X" more often when X was actually like A than when X was actually like B.

(29) Listeners anecdotally reported using cues unrelated to the /t/∼/d/ distinction (such as the intonation contour of individual tokens) in making their decisions

6 Methods, Part II

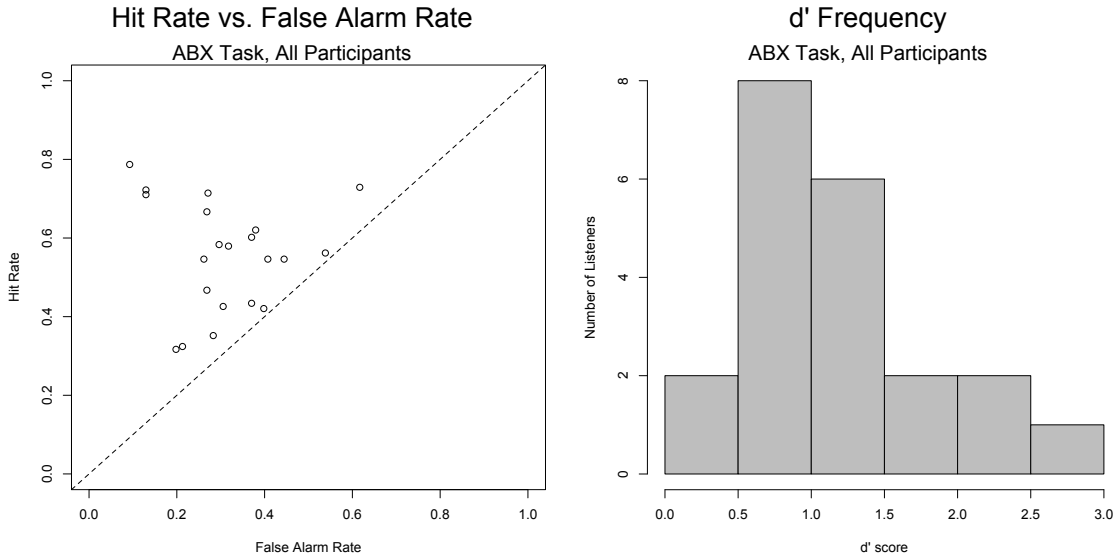
(30) To test whether listeners discriminated based on irrelevant acoustic differences between A and B in the ABX task, an AB Task (a.k.a. 2AFC) was run

a. This task allows listeners to make comparisons (as in the ABX task), but no two tokens are the same on a given trial (like the ID task).

b. Listeners cannot use irrelevant acoustic differences of single tokens to make their decisions in this task

(31) 21 undergraduates participated in the AB task.

(32) All tokens were from the same set as the Identification and ABX tasks



(33) The task consisted of instructions and practice, followed by three blocks (each with tokens from a different speaker), with block order balanced (Latin Square) across all listeners. Feedback was given on each trial.

(34) On each trial, listeners heard a two tokens—members of a minimal pair. Half of the listeners were asked whether the /d/ member came first or second. The other half were asked whether the /t/ member came first or second.

(35) For example, in the 'find /d/' variation:

a. Listeners hear 'buhKEED-ing buhKEET-ing', and should respond 'the /d/ member came first'

b. Listeners hear 'buhKEET-ing buhKEED-ing', and should respond 'the /d/ member came second'

(36) Each block consisted of 36 trials (half /d/, half /t/), randomized.

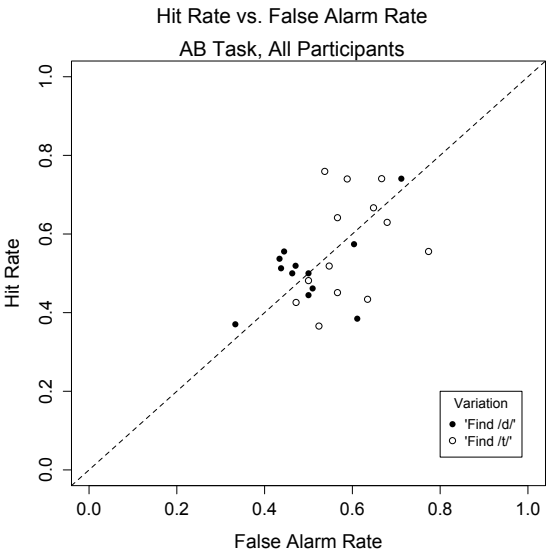
7 Results, Part II

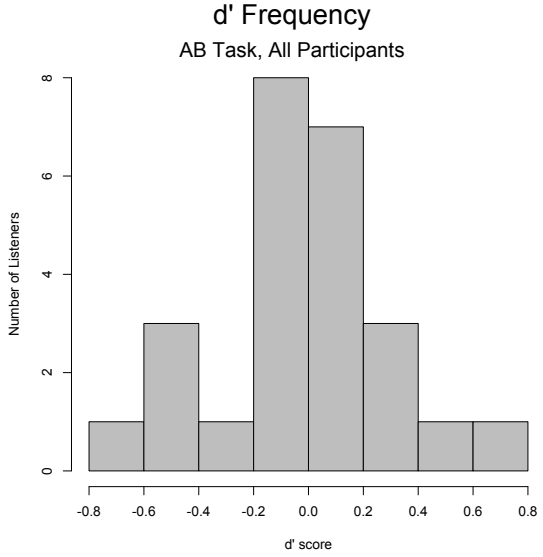
(37) Results from the AB task

a. d′ is not significantly different from 0 overall (mean d' : -0.02, Wilcoxon test: $V = 148, n.s.$)

b. In the 'find /d/' variation, listeners said "/d/ came first" just as often when they had heard a /d/ first as when they had heard a /d/ second.

c. In the 'find /t/' variation, listeners said "/t/ came first" just as often when they had heard a /t/ first as when they had heard a /t/ second.





8 Discussion and Conclusions

(38) The low d′ scores in the identification task suggest that listeners were unable to categorize /d/-flaps and /t/-flaps

(39) While listeners were able to distinguish /d/-tokens from /t/-tokens in the ABX task, they were unable to do so in the AB task

a. Explanation: Listeners used the 'unrelated cues' strategy in the ABX task, comparing acoustic cues unrelated to the underlying voicing distinction to determine which tokens were identical

b. In the AB task, listeners were unable to use the 'unrelated cues' strategy, since on any given trial, no two tokens were identical.

c. This suggests that listeners cannot distinguish /d/-flaps from /t/-flaps on the basis of cues relevant to the underlying voicing contrast

(40) These results hold in both an identification task and a discrimination task,, where frequency effects are mitigated through the use of nonce words.

(41) If listeners are neither able to distinguish nor categorize /d/-flaps and /t/-flaps, speakers who maintain this distinction must be doing so for reasons other than listeners' benefit.

References

Berko, Jean (1958). *The Child's Learning of English Morphology*. Word 14:150–177.

Braver, Aaron (2010). *Incomplete Neutralization in American English Flapping*. Ms. Rutgers, The State University of New Jersey.

Braver, Aaron (2011). *Incomplete Neutralization in American English Flapping: A Production Study*. In *Proceedings of the 34th Annual Penn Linguistics Colloquium*, volume 17 of *University*

of Pennsylvania Working Papers in Linguistics. Penn Linguistics Club. `http://repository.upenn.edu/pwpl/vol17/iss1/5/`.

Dinnsen, Daniel and Charles-Luce, Jan (1984). *Phonological Neutralization, Phonetic Implementation and Individual Differences*. Journal of Phonetics 12:49–60.

Dmitrieva, Olga (2005). *Incomplete Neutralization in Russian Final Devoicing: Acoustic Evidence from Native Speakers and Second Language Learners*. Master's Thesis, University of Kansas, Lawrence, Kansas.

Fisher, William M. and Hirsh, Ira J. (1976). *Intervocalic Flapping in English*. In *Papers from the Twelfth Regional Meeting of the Chicago Linguistic Society*, pp. 183–198. Chicago Linguistic Society.

Fourakis, Marios and Iverson, Gregory (1984). *On the 'Incomplete Neutralization' of German Final Obstruents*. Phonetica 41:140–149.

Fox, Robert A. and Terbeek, Dale (1977). *Dental Flaps, Vowel Duration, and Rule Ordering in American English*. Journal of Phonetics 5:27–34.

Gerrits, Ellen and Schouten, M.E.H. (2004). *Categorical perception depends on the discrimination task*. Perception and Psychophysics 66(3):363–376.

Herd, Wendy; Jongman, Allard; and Sereno, Joan (2010). *An acoustic and perceptual analysis of /t/ and /d/ flaps in American English*. Journal of Phonetics 38:504–516.

Huff, Charles T. (1980). *Voicing and Flap Neutralization in New York City English*. Research in Phonetics 1:233–256.

Joos, Martin (1942). *A Phonological Dilemma in Canadian English*. Language 18(2):141–144.

Kahn, Daniel (1980). *Syllable-based Generalizations in English Phonology*. Garland, New York.

Macmillan, Neil A. and Creelman, C. Douglas (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates Inc., Mahwah, NJ, 2nd edition.

Port, Robert (1976). *The Influence of Speaking Tempo on the Duration of Stressed Vowel and Medial Stop in English Trochee Words*. Doctoral Dissertation, University of Connecticut.

Port, Robert and O'Dell, Michael (1985). *Neutralization and Syllable-Final Voicing in German*. Journal of Phonetics 13:455–471.

Slowiaczek, Louisa M. and Dinnsen, Daniel (1985). *On the Neutralizing Status of Polish Word-Final Devoicing*. Journal of Phonetics 13:325–341.

Slowiaczek, Louisa M. and Szymanska, Helena (1989). *Perception of Word-Final Devoicing in Polish*. Journal of Phonetics 17:205–212.

Warner, Natasha; Good, Erin; Jongman, Allard; and Sereno, Joan (2006). *Orthographic vs. Morphological Incomplete Neutralization Effects*. Journal of Phonetics 34(2):285–293.

Warner, Natasha; Jongman, Allard; Sereno, Joan; and Kemps, Rachèl (2004). *Incomplete Neutralization and other Sub-Phonemic Durational Differences in Production and Perception: Evidence from Dutch*. Journal of Phonetics 32:251–276.

Zue, Victor W. and Laferriere, Martha (1979). *Acoustic Study of Medial /t, d/ in American English*. Journal of the Acoustical Society of America 66:1039–1050.