Inhibition of Processing Due to Reduction of the American English Flap

Benjamin V. Tucker and Natasha Warner

University of Arizona, Dept of Linguistics
P O Box 210028, Tucson AZ 85721-0028 USA
E-mail: bvtucker@ualberta.ca, warner@u.arizona.edu

ABSTRACT

The speech we encounter in daily life casual conversation often contains impoverished or reduced acoustic information, in comparison to careful speech, and yet listeners can understand such speech with ease. This study explores differences in processing between reduced/ conversational speech and unreduced/careful speech. In a cross-modal identity priming experiment, listeners heard reduced vs. careful pronunciations of real words and then saw visual stimuli and decided whether the visual stimulus was a real word. This experiment investigates processing differences between reduced and unreduced speech using the American English flapped /d/ and word-medial /g/. American English listeners are shown to process unreduced (clear) targets more quickly than reduced targets.

1. INTRODUCTION

In spontaneous conversational exchanges, speech is acoustically highly variable. Segments, syllables or words are often reduced by being deleted and/or changed into a different sound [1]. For example, in the sentence: "We were supposed to see it yesterday." a careful reading of the sentence might be transcribed as [wi wɚ səpoʊzd tʰə si ɪt̚ jɛstɚdeɪ]. However in a spontaneous conversation recorded during this study, that same sentence was pronounced as [wiɚ sɸʊsə si: jɛʃe]. In the casual pronunciation, segments, syllables, and whole words were deleted (e.g. *it*) as well as changed (e.g. /sterd/ in 'yesterday' became [ʃ]). Despite this acoustic variability, listeners perceive conversational, spontaneous speech with ease.

While listeners probably hear such conversational speech more often in real life than any more careful type of speech, reduced, casual speech has been relatively unstudied [2, 3]. The vast majority of speech perception research has focused on unreduced or "laboratory" speech [2]. Recently, investigation of spontaneous speech has become the focus of a small but growing body of literature [2, 3, 4, 5, 6, 7, 8, 9].

This paper investigates the processing of reduced and unreduced flaps and word medial /g/, as described in Warner and Tucker [11]. The current paper uses the cross-modal identity priming task, which requires listeners to access the lexicon, providing a metaphoric window into processing of reduced and unreduced speech.

2. EXPERIMENT

This study tests the effect of reduction on listeners' recognition of words in a cross-modal identity priming task [10]. Listeners heard items, pronounced either in reduced or careful speech, and then were presented with a visual stimulus (a written word on a computer monitor). In some cases (priming trials), the visual word was the same word as had been presented auditorily. Listeners were asked to perform lexical decision on the visual stimulus. Listeners are expected to recognize the visual stimulus more quickly (to show priming) if they have just heard the same word presented auditorily than if they have just heard an unrelated word. The central question is whether listeners show greater facilitation in recognizing a visual word when the auditory prime is pronounced with reduced vs. unreduced speech.

There are two possible outcomes: (1) Due to the decreased clarity of acoustic cues, listeners might show a smaller effect for reduced than unreduced primes. Having heard the word "puddle" might facilitate recognition of the written word "puddle" less if the pronunciation one hears contains a reduced /d/ than if it contains a clear /d/. (2) Reduced speech targets are more frequently heard by listeners in daily life, so priming might be greater for reduced items than for unreduced items. That is, listeners might even be faster to recognize visually presented "puddle" after hearing a reduced

pronunciation of the same word than after hearing a clear pronunciation of it, because the reduced pronunciation is the more typical one.

2.1. METHOD

2.1.1. Participants

Thirty University of Arizona students either received extra credit or were paid for participation. All were native speakers of American English, with no known hearing problems.

2.1.2. Materials

The targets were 30 word medial /d/ (flap) items and 30 word medial /g/ items, such as "pu**dd**le" and "ba**gg**y." Items were produced in both reduced and unreduced forms (Fig. 1-2). Word medial /d/ items were selected because a phonological flapping rule applies [11, 12], while the word medial /g/ items do not undergo a phonological alternation. They do, however, reduce in acoustically similar ways to word medial /d/ items.

Figure 1: Waveform and spectrogram of the word "puddle" in an unreduced form. Intensity is superimposed on the spectrogram.

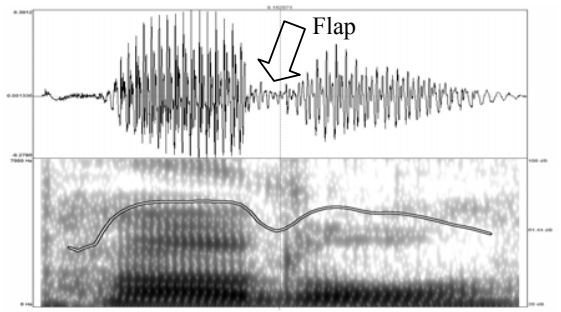
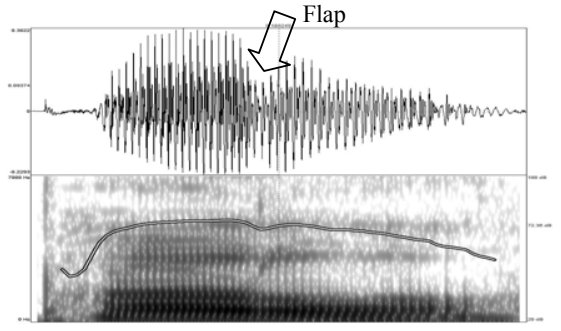


Figure 2: Waveform and spectrogram of the word "puddle" in a reduced form. Intensity is superimposed on the spectrogram.



In Fig. 1, the unreduced flap has a burst, a clear stop closure, and a large drop in intensity. In Fig. 2 the reduced flap has no clear burst or closure boundaries, and only a small dip in intensity. Importantly, the formants continue through the flap, causing it to resemble an approximant.

Sixty real word items were recorded as auditory controls (these items were produced in an unreduced manner). These items were selected by matching word frequency [13] to the frequency of one of the target items. These control items then were used as the auditory control for the visual target.

160 filler and 9 practice items were constructed in addition to the targets. 140 fillers contained pseudo-word visual stimuli matched with real-word auditory fillers. The visual pseudo-word for forty of these fillers contained partial phonological overlap with the auditory filler.

2.1.3. Procedure

Target and filler items were recorded in a sound attenuated booth by the second author, who is a native speaker of American English. Multiple repetitions of each target, including a variety of levels of reduction, were recorded. Tokens were selected based on visual inspection of spectrograms and comparison to data from a previous production study [11]. A small number of the fillers were also produced in a reduced form.

Each listener saw 30 /d/ target visual items and 30 /g/ target visual items. For each listener, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in reduced pronunciation, 10 /d/ and 10 /g/ items were paired with an auditory prime consisting of the same word in unreduced pronunciation, and 10 /d/ and 10 /g/ items were paired with an unrelated word as the auditory control stimulus. Items were counterbalanced, creating three lists. All fillers (160) appeared in each list. Lists were pseudo-randomized for presentation and were presented in two orders, creating six total lists.

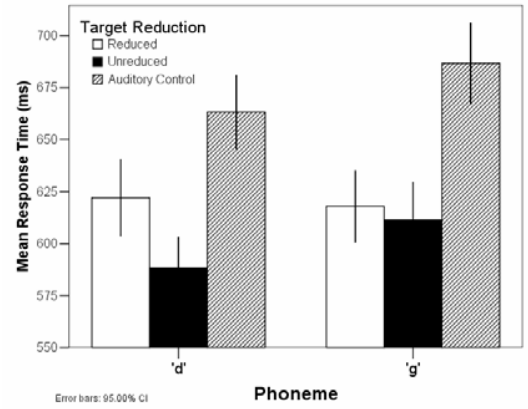
Listeners were tested individually in sound-attenuated booths. The auditory stimuli were presented over headphones and the visual targets appeared in lower case 36 point Arial font. The visual target appeared at the offset of the auditory stimulus (0 ms interval). Participants had 2 seconds to respond to visual targets. Listeners were instructed to respond to the visual probe using a button box as quickly and as accurately as

possible. The appropriate buttons were labeled "Yes" and "No", where a "Yes" response meant that the visual probe was a real word of English. Each listener first did a short practice test, then participated in one of the six stimulus lists, and then responded to a short comprehension test. Listeners were informed at the beginning of the procedure that there would be a test at the end of the experiment to discourage them from focusing exclusively on the visual stimuli. In the comprehension test, listeners were asked if they had heard each of 10 words in the main part of the experiment. Listeners scored an average of 78% correct on this test.

2.2. RESULTS

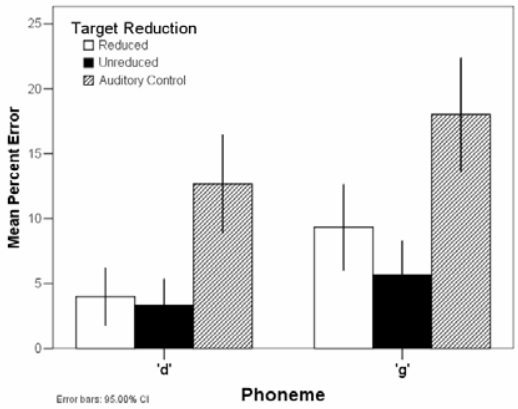
Response latencies were measured from the onset of the visual probe. Responses slower than 1100ms or faster than 350ms (less than 5% of total responses) were treated as errors and removed from the analysis. Six items were removed because response accuracy was less than 30%. Two listeners scored below 85% correct and were also excluded. Mean response times are shown in Figure 3.

Figure 3: Mean response time for cross-modal identity priming for /d/ and /g/ targets and reduced, unreduced and control primes.



An analysis of variance (ANOVA) of the response time data was performed for listeners (*F1*) and items (*F2*). The analysis (Target Reduction: Reduced, Unreduced, Auditory Control; Phoneme: /d/ and /g/) revealed an identity priming effect: listeners responded to targets presented with an identity prime more quickly than to targets presented with an unrelated auditory control stimulus ($F1(2,48)=38.292$, $p<0.005$; $F2(2,116)=33.402$, $p<0.001$). Furthermore, planned pairwise comparisons showed that listeners were primed significantly more for unreduced (clearly articulated) /d/ items ($F1(1,24)=13.843$, $p<0.005$; $F2(1,29)=6.073$, $p<0.05$) than for reduced /d/ items. While the comparison of the /g/ items trended in the same direction, the comparison of reduced vs. unreduced was not significant ($F1<1$; $F2<1$). A significant main effect of phoneme was found in the by-subjects analysis ($F1(1,24)=11.397$, $p<0.005$), but this effect was not significant in by-items ($F2(1,58)=2.153$, $p>0.05$). The interaction of phoneme and reduction was not significant ($F1(2,48)=1.235$, $p>0.25$; $F2(2,116)=1.000$, $p>0.05$). The mean percent error is shown in Figure 4 below.

Figure 4: Mean percent error of cross-modal identity priming of /d/ and /g/ targets and reduced, unreduced and control primes.



An ANOVA of the percent error data was performed for listeners (*F1*) and items (*F2*). Listeners made fewer mistakes to visual targets presented with identity primes than to those presented with unrelated auditory control stimuli ($F1(2,48)=13.444$, $p<0.001$; $F2(2,116)=18.558$, $p<0.001$). Planned pairwise comparisons showed a non-significant trend toward greater priming for unreduced /d/ items than for reduced /d/ items ($F1(1,24)=1.775$, $p>0.05$; $F2<1$). The comparison of the /g/ items showed a smaller trend in the same direction ($F1<1$; $F2(1,29)=2.233$, $p>0.05$). The main effect of phoneme was also significant ($F1(1,24)=6.892$, $p<0.05$, $F2(1,58)=4.378$, $p<0.05$). The interaction of phoneme and reduction was not significant ($F1(2,48)=1.235$, $p>0.05$; $F2<1$).

2.3. DISCUSSION

The response time and percent error results indicate that priming occurred in both the reduced and unreduced conditions. Listeners responded more quickly and accurately to items primed by an unreduced pronunciation of the same word than to items primed by a reduced pronunciation. For example, an unreduced production of "puddle" produced greater priming than a reduced production of "puddle". However, this difference in strength of priming was very small, and was significant only in reaction times to the /d/ items. Still, the difference is consistently in the direction of greater priming by unreduced pronunciations in all data. The results provide some support for the first prediction: lesser acoustic clarity in reduced forms appears to hinder recognition of the word, as reflected by how well the auditory stimulus primes recognition of the same word when presented visually. The speech produced in normal, casual, fluent conversation, because it contains acoustically reduced sounds, makes it harder to recognize words than carefully pronounced unreduced speech does. Still, listeners recognize reduced speech with apparent ease, and do so quite often: reduced speech probably forms the majority of the speech most of us hear each day.

The greater priming by unreduced than reduced forms suggests that clarity of acoustic information plays a larger role in spoken word recognition than does the frequency of occurrence of a particular pronunciation of a word. This finding is in agreement with results by Ernestus et. al. [4]. However, the current results reflect processing of words (reduced or unreduced) presented in isolation. Listeners might expect isolated words to be pronounced quite carefully, without reduction. Future research will investigate the processing of reduced vs. unreduced forms in context.

The overall difference in results for /d/ vs. /g/ items (the phoneme effect) might be attributed to word frequency. Medial /d/ words are generally more frequent than medial /g/ words (mean frequency of /d/ target items = 53.225 and /g/ = 6.55, [13]). Because this is an overall effect of phoneme, it is not relevant for the primary question of how reduction affects speech processing.

3. CONCLUSION

These results suggest that in the processing of reduced speech, at least for words in isolation, acoustic information plays a greater role than frequency of occurrence of a particular style of speech. They also suggest that even though listeners can readily understand reduced speech, there is a processing cost to recognizing words produced with acoustic reduction.

ACKNOWLEDGEMENTS

We thank Karen Morian, Mary Dungan, Raiza Betters and Chris Crawford for their help.

4. REFERENCES

[1] Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.) Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium. Tokyo, Japan: The National International Institute for Japanese Language. pp. 29-54.

[2] Cutler, A. (1998). The recognition of spoken words with variable representation. In Proceedings of the ESCA Workshop on Sound Patterns of Spontaneous Speech (pp. 83-92). Aix-en-Provence: ESCA.

[3] Ernestus, M., Lahey, M., Verhees, F., Baayen, R.H. (2006). Lexical frequency and voice assimilation. J. Acoust. Soc. Am. 120, 1040-1051.

[4] Ernestus, M., Baayen, R.H., Schreuder, R. (2002). The recognition of reduced word forms. Brain and Language, 81, 162–173.

[5] Mitterer, H., Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. Journal of Phonetics, 34, 73-103

[6] Pluymaekers, M., Ernestus, M., Baayen, R.H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. The Journal of the Acoustical Society of America 118 (4), p. 2561-2569.

[7] Gaskell, M.G., Marslen-Wilson, D. (1996). Phonological variation and inference in lexical access. J. Exp. Psychology: Human Perc. and Perf., 22, 144–158.

[8] Gaskell, M.G., Marslen-Wilson, D. (1998). Mechanisms of phonological inference in speech production. J. Exp. Psychology: Human Perc. and Perf. 24, 380–396.

[9] Zwitserlood, P., Coenen, E. (2000). Consequences of assimilation for word recognition and lexical representation. Presented at the SWAP workshop, Nijmegen, 1999.

[10] Norris, D., Cutler, A., McQueen, J., Butterfield, S. (2006). Phonological and conceptual activation is speech comprehension. Cognitive Psychology, 53, 146-193.

[11] Warner, N., Tucker, B.V. (2007). Categorical and Gradient Variability in Intervocalic Stops. Talk given at the annual meeting of the Linguistic Society of America.

[12] Zue, V. Lafferriere, M. (1979). Acoustical study of medial /t/, /d/ in American English. J. Acoust. Soc. Am. 66, 1039–1050.

[13] Kučera, H., Francis, W. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.