ON THE PERCEPTION OF INCOMPLETE NEUTRALIZATION

Cynthia Kilpatrick[†], *Ryan Shosted*[‡], *Amalia Arvaniti*[†]

[†]University of California, San Diego [‡]University of Illinois, Urbana-Champaign
kilpatrick@ling.ucsd.edu, rshosted@uiuc.edu, amalia@ling.ucsd.edu

ABSTRACT

The perception of American English epenthetic and underlying stops (as in *prin[t]ce~prints*) was examined in a forced-choice identification experiment that controlled for word frequency and familiarity, closure duration and presence of burst. The results showed that listeners are largely unable to distinguish minimal pairs on the basis of differences in closure duration and the presence or absence of burst; word frequency and familiarity had little effect on the results. Generally, listeners had more difficulty with stimuli with strong [t]s (long closure, burst) than with stimuli with weak [t]s, which they tended to categorize as "nce" words. Overall the results suggest that [ns]~[nts] is close to complete neutralization in favor of [nts].

Keywords: epenthesis, incomplete neutralization, perceptual sensitivity, American English.

1. INTRODUCTION

Although epenthetic consonants in American English, such as the [t] heard in words like *prince* or *mince*, are well known ([2], [8]), few phonetic studies have examined them ([1], [3], [4], [14]), especially from a perceptual angle [6], [10]. In addition, while production studies, like [1] and [4], find that epenthetic [t]s have shorter closure duration and are less likely to have a burst than underlying [t]s—suggesting that this is a case of incomplete neutralization—the perceptual results indicate that listeners cannot distinguish minimal pairs, such as *tense* and *tents* [6]. Since the perceptual studies are somewhat limited (e.g. [6] used only three subjects for the perceptual experiment), or do not focus on the distinction between underlying and epenthetic stops as such (e.g. [10]), the present study aims to examine in more detail the perception of epenthetic [t] in American English. The results are interpreted in light of our recent production findings which show small yet observable differences between epenthetic and underlying [t], as well as effects of word-familiarity and position of the [n(t)s] sequence in the word [3]. These differences suggest that the [nts]~[ns] alternation may be gradually reaching a state of complete neutralization: although epenthetic [t]s are still different from underlying [t]s, the differences are minimal, especially in frequent words and in word-final position. (Note that in other varieties of English, neutralization may be achieved by eliding [t]; elision was rare in the variety examined here, Southern California English [3].)

Given that the differences between epenthetic and underlying [t] are small yet existent, as [3] shows, it was expected that in an identification task listeners would have difficulty distinguishing [n(t)s] minimal pairs, but that identification would be above chance level (though not as high as for "regular" minimal pairs). It was also expected that frequent and familiar words would be more difficult to distinguish than infrequent and unfamiliar ones, because the former show greater similarity in the production of [t]. Finally, it was expected that the presence of a burst and longer closure duration would favor "nts" responses because they make the presence of [t] more robust.

2. METHODS

2.1. Materials

Three word pairs, *prince, prints, quince, quints* and *mince, mints*, were chosen from among the small number of relevant monosyllabic minimal pairs. Monosyllables were chosen because they are the most numerous in the English lexicon among words with [n(t)s], and thus allowed us to control for frequency, familiarity and neighborhood density, which were obtained from the WU Speech and Hearing Lab Neighborhood Database (at http://128.252.27.56/neighborhood/Home.asp) and shown in Table 1. *Prince* and *prints* are relatively frequent and familiar, *quince* and *quints* are comparatively infrequent and unfamiliar, while *mince* is infrequent but familiar, and *mints* is both frequent (relative to *mints*) and familiar. Thus, it was expected that the use of *mince~mints* would allow us to disentangle frequency from familiarity effects. In addition, the words were chosen because

of their low lexical density: this meant that the (anticipated) difficulty of the task would not be aggravated by high density [13], and that the frequency results would not be confounded by density effects. The familiarity ratings were confirmed by means of questionnaires distributed to sixteen UCSD undergraduates who rated minimal pairs involving [n(t)s] for familiarity on a 7-point scale.

Table 1: Frequency, familiarity and neighborhood density data for the words used as stimuli.

	Frequency	Familiarity	Density
prince	33	7	2
prints	18	7	1
quince	2	4.4	1
quints	11	3.6	8
mince	1	6.3	7
mints	7	7	10

The materials were elicited from one female and two male naïve speakers of American English, all in their early 20s. The speakers produced three tokens each of the six words from a randomized Powerpoint presentation in which the test words were interspersed with an equal number of fillers. The test words were elicited in the carrier phrase *I will say ___ one more time*; the fillers were similar but more varied.

In order to prepare the stimuli, one token of each test word was selected from the recording of each speaker, typically from the second repetition, for a total of eighteen tokens. We derived stimuli from both words of each pair so as to control for the effects of potential differences that we did not manipulate. That is, we anticipated that if (as suggested by [4]) a pair of test words exhibit differences beyond closure duration and burst, and these differences are perceptually relevant, then listeners' responses would show a preference for the word from which the stimuli were derived.

For the preparation of stimuli, all traces of [t] closure and burst were first removed from the tokens using PRAAT. Next, 0-72 ms of silence were spliced in between [n] and [s], in 12 ms steps. For each adulterated token, two stimuli were prepared, one with and one without a [t] burst; the burst had been excised from one of the "nts" tokens and was used for all the stimuli in order to avoid any effects of burst quality on listeners' responses. This burst had an RMS of 0.002 Pa and was 5.4 ms in duration; thus it was as loud and less than one standard deviation longer then the mean burst in our production data [3].

2.2. Listeners

Twenty-six monolingual native speakers of American English, 18-23 years old, took part in the perception experiment. They were all naïve as to the purposes of the experiment, and reported no history of speech or hearing problems.

2.3. Procedures and measurements

The listeners performed a forced-choice identification task using SuperLab Pro 2.0.4: they saw a minimal pair on screen (e.g. *quints quince*) and selected the word they thought they heard by pressing a button on a keyboard.

The listeners heard the stimuli over headphones, while seated in the sound booth of the UCSD Phonetics Laboratory. First they read onscreen instructions at a self-selected pace. Then they pressed a button to initiate the experiment, which began with twelve practice stimuli derived from the data of one of the two male speakers (whose voice was not used in the rest of the experiment).

Each trial started with a silence of 1200 ms after which a red star (on white background) appeared in the center of the screen. Visible for 8 ms, the star was followed by an audible click. The words were then displayed simultaneously on the same line at the vertical center of the screen, one to the left and the other to the right of horizontal center; the left-right position of the "nts" and "nce" words was randomized. After the text had been displayed for 500 ms, an auditory stimulus was played. Stimuli were heard in random order, three times each, in blocks of ten. Each block contained six stimuli, interspersed with four fillers (none of which were words with [n(t)s] sequences). Breaks of 5 s (silence) occurred after every seven blocks, and a break of 10 s (silence) occurred in the middle of the experiment. The stimuli were blocked by speaker, and the order in which the two speakers were heard was counterbalanced across listeners. The experiment lasted approximately 40 minutes.

Responses and reaction times were recorded, unless the subject failed to respond within 1800 ms after the offset of a stimulus, at which point the next trial started. Several listeners reported difficulties with the task, but only two responded to less than 60% of the stimuli; the data of these two listeners were discarded.

After the data had been collected, it was discovered that due to a programming error, the listeners did not hear one stimulus at all (the burst-less stimulus for *quince* with 0 ms closure); in

addition, they heard a number of stimuli fewer than three times. For this reason, we decided to analyze only the responses to the first occurrence of each stimulus.

3. RESULTS

The reaction times (RTs) were statistically analyzed by means of repeated measures analysis of variance (ANOVAs) using stimulus origin, closure duration and burst as independent variables. The results showed that none of these factors affected RTs and there were no interactions. However, at 658 ms on average, RTs were relatively long for an experiment in which subjects were instructed to respond as quickly as possible to the stimuli (cf. [11]).

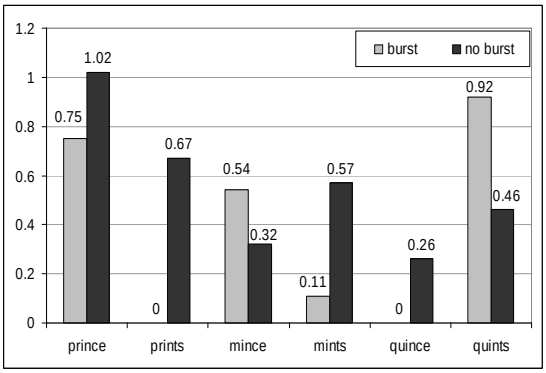
Table 2: Percentage of the *mince~mints*, *quince~quints* and *prince~prints* series identified as the "nce" word of each pair. Identification rates significantly different from chance are shaded. *B* stands for *burst*, *NB* for *no burst*.

	[t] closure in ms						
	0 ms	12 ms	24 ms	36 ms	48 ms	60 ms	72 ms
mince NB	58	67	67	54	54	46	46
mince B	54	58	50	67	63	54	33
mints NB	25	46	46	42	50	38	46
mints B	58	21	54	58	63	42	63
quince NB	79	67	83	67	79	58	71
quince B	*	67	67	75	50	71	63
quints NB	25	33	46	38	50	29	42
quints B	21	38	46	29	33	58	54
prince NB	79	71	46	71	46	46	42
prince B	67	63	50	50	67	63	38
prints NB	25	38	13	50	38	33	50
prints B	58	33	54	50	29	63	58

Table 2 presents the percentage of stimuli identified as the "nce" word in each minimal pair; responses to stimuli derived from the "nce" and "nts" word of each pair are presented separately. Binomial tests (one-sample tests for proportions) showed that the identification rates were rarely significantly different from chance (except for the shaded data points, for which $p<0.05$). As can be seen in Table 2 the stimuli for which responses were not random were mostly stimuli without bursts (8/11 cases) or with short closure duration (7/11 cases involved closures of 0 or 12 ms) or both (5/7 cases). In addition, the results show that there were no substantial differences between stimuli originating from one or the other word of each pair; e.g. stimuli originating from *prince* were not identified as *prince* more frequently than stimuli originating from *prints*. Rather, it was the length of closure and the presence or absence of burst that influenced listeners the most.

A similar picture emerges from measuring perceptual sensitivity (*d'*); *d'* was calculated separately for stimuli with burst and stimuli without; Fig. 1 presents *d'* values for stimuli with 0 ms vs. 72 ms of silence. The *d'* values were typically lower than 1.0, suggesting that subjects were not strongly sensitive to the distinction between stimuli with short and long closure durations [7]. Despite the overall low degree of perceptual sensitivity, the two most frequent words, *prince* and *prints* showed mostly high *d'* values that were higher for stimuli without bursts than for stimuli with bursts; for the other two pairs no such pattern emerged, and *d'* was lower overall (0.61 on average for *prints* and *prince* vs. 0.39 for the other two pairs together).

Figure 1: Perceptual sensitivity (*d'*) to all series of stimuli.



Response bias (*c*) for each stimulus series was also computed; *c* ranged from -0.2 to 0.8. Given that *c* may range from approximately -2.33 to +2.33, where 0 means no bias [7], we assume that the subjects were not predisposed to categorize the stimuli as "nts" or "nce". Thus, the response bias results support the validity of the *d'* measures.

4. DISCUSSION AND CONCLUSION

The long reaction times indicate that the listeners found the task difficult and this agrees with the reports of the subjects themselves. The mostly random responses and low *d'* values also show that even in the presence of strong evidence for or clear absence of a [t], listeners had difficulty distinguishing words containing an [ns] sequence from words containing [nts].

The identification rates were lower than those of some previous experiments testing incomplete

neutralizations, such as [9] and [12]. This may be due to the fact that in other cases of incomplete neutralization, such as German and Dutch devoicing, other cues are available to the speakers ([9], [12]), while [t] epenthesis does not appear to involve consistent cues beyond those associated with [t] ([3], contra [4]). This conclusion is supported by the present results: the words from which the stimuli were derived did not condition listeners' responses. Thus, the present data suggest that even if this case of neutralization is not yet complete, it is close to being so, as production results also show [3].

Further, our results show effects of burst and closure duration though not in the expected direction: it was anticipated that stimuli with bursts and stimuli with longer closure durations would be more readily identified as "nts" words [11], but in fact the opposite obtained: stimuli without bursts were perceptually better separated along the closure dimension (as *d'* values show) and more readily identified as "nce" words (as identification rates show); long closure durations and bursts, on the other hand, led to more random responses.

This unexpected result is consistent with the production data of [3]: since [t] is present in both "nce" and "nts" words, when listeners hear a [t] in an [n_s] context, they cannot reliably identify the word they hear. On the other hand, when acoustic evidence for [t] is absent or weak, the listeners opt for "nce" words, possibly because these lack orthographic "t." In this sense, the results show better perceptual separation of the stimuli when [t] is absent; to put it differently, they suggest that in this context listeners are sensitive to the absence, not the presence of [t].

Finally, it is worth considering why high word frequency, which had been expected to lead to more chance results, had the opposite effect. Hay et al. [5] offer the following explanation for similar data from the New Zealand merger-in-progress of the vowels in *near* and *square*: they suggest that although listeners categorize the vowels in both types of words as "the same" vowel, and hence report that identification is difficult, they still manage to disambiguate (to an extent) minimal pairs such as *bare* and *beer* because they have two clouds of exemplars they categorize as the "same" vowel at a more abstract level, one for *near* vowels and one for *square* vowels.

Our results support this line of reasoning: frequent minimal pairs, though often confused, were more perceptually distinct than infrequent pairs. That is, for items such as *quince* and *quints* the listeners most likely do not have strong representations, so they are less sure how to categorize the stimuli.

In conclusion, this experiment showed that listeners have only limited sensitivity to the differences between epenthetic and underlying [t] in the [n_s] context, and are primarily attuned to the absence, not the presence of [t] in this context. These results support the view that the [ns]~[nts] alternation is rapidly progressing from incomplete to complete neutralization.

5. ACKNOWLEDGEMENTS

Grant LIN681C from the UCSD General Campus Subcommittee on Research (COR) is gratefully acknowledged.

6. REFERENCES

[1] Ali, L., Daniloff, R., Hammarberg, R. 1979. Intrusive stops in nasal-fricative clusters: an aerodynamic and acoustic investigation. *Phonetica* 36, 85-97.

[2] Anderson, S. R. 1976. Nasal consonants and the internal structure of segments. *Language* 52, 326-344.

[3] Arvaniti, A., Shosted, R., Kilpatrick, C. Submitted. Incomplete neutralization? The case of epenthetic [t].

[4] Fourakis, M., Port, R. 1986. Stop epenthesis in English, *JPhon* 14, 197-221.

[5] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *JPhon* 34, 458-484.

[6] Lee, S. 1991. The duration and perception of English epenthetic and underlying stops. *JASA* 89, 1999.

[7] Macmillan, N. A., Creelman, C. D. 2005. *Detection Theory: A User's Guide*. 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates.

[8] Ohala, J. J. 1974. Experimental historical phonology. In: Anderson, J. M., Jones C. (eds), *Historical Linguistics II*. Amsterdam: North Holland, 353-389.

[9] Port, R., O'Dell, M. 1985. Neutralization of syllable-final voicing in German. *JPhon* 13, 455-471.

[10] Shinya, T. 2005. The perception of epenthetic stops in English: The effects of cluster type and silent interval duration. In: Flack, K., & Kawahara, S. (eds), *University of Massachusetts Occasional Papers in Linguistics 31: Papers in Phonetics/Laboratory Phonology*.

[11] Warner, N., Weber, A. 2001. Perception of epenthetic stops. *JPhon* 29, 53-87.

[12] Warner, N., Jongman, A., Sereno, J., Kemps, R. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *JPhon* 32, 251-276.

[13] Vitevitch, M., C. Luce, D. Pisoni, E. T. Auer. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language* 68, 306-311.

[14] Yoo, I. W., Blankenship, B. 2003. Duration of epenthetic [t] in polysyllabic American English words, *JIPA* 33, 153-164.