Perceptual Test of a Phonological Rule

Linda Shockey

ABSTRACT

Synthetic speech is characteristically produced in a highly formal, maximally differentiated style. This experiment shows that the output of one phonological rule manifested frequently in conversational English (ð-assimilation) can be simulated easily by lengthening appropriate word-final consonants. Listeners accept the resulting long consonants as consonant + ð clusters. It is suggested that the inclusion of this and other casual speech rules will improve the naturalness and hence the "listenability" of synthetic speech.

It has been noted by Wolfram and Fasold (1974), Gimson (1962), Kohmoto (1965), Hubbell (1950), and Shockey (1974), among others, that in relaxed or conversational speech a word-initial [ð] will assimilate completely to a preceding nasal, fricative, or [l] and will cause the nasal, fricative, or [l] to lengthen as well. Examples from natural speech follow (taken from Shockey (1974):

S # Assimilations:

course they ['kɔrz·eɪ]
effects the [ɪfɛks·o]

Z # Assimilations:

broads that [bɔːdz·æʔ]
cause the [kɔːz·o]
years there [ɪz·ɛx]

N # Assimilations:

seen the [sin·ə]
on their [ɔn·ɛk]
line that [laɪn·æt]

The ð-assimilation will presumably only occur when the cluster in question receives a relatively low degree of stress. Therefore, if one were to construct a sentence of the sort "I said put it in this box, not in that one!," the assimilation would be less likely to occur. However, since virtually all English words beginning with [ð] carry very little semantic load, ð-assimilation can be expected to be very frequent. This experiment was
conducted to investigate whether the lengthening of the assimilating (word-final) consonant is sufficient to induce a cluster percept. If so, \( \delta \)-assimilation could easily be included in a speech synthesis strategy designed to output natural-sounding English. The specific condition examined here is ubiquitous in conversational speech: in the word "the", [\( \delta \)] assimilates to a preceding word-final consonant so that the distinction between definite and indefinite articles is preserved mainly in the duration of the final consonant. (It is also possible that another cue may be retained in the consonant transitions. Final alveolar consonants may be fronted and thus be more dental in cases where the interdental [\( \delta \)] has been assimilated. This possibility must be examined further).

To test the perceptual effect of final consonant lengthening, the following experiment was conducted: two utterances by one female speaker were digitized and stored in a computer-accessible file. These utterances were "miss a guy" ['mɪsəgə] and "warn a guy" ['wɜːnəgə]. Using digital splicing techniques, the duration of the frication in the first utterance was varied from 80 to 200 msec in 20 msec steps, and the duration of the low amplitude, low frequency portion of the acoustic signal that presumably corresponded articulatorily to the closure portion of [n] in the second, varied from 0 to 120 msec in 10 msec steps. The lengthening was done by holding the transitions into and out of the steady-state consonant in question and repeating a characteristic portion of the waveform judged to be the center of the consonant enough times to give the desired durations. It was judged that the nasal closure began when the waveform became smooth and lacked high-frequency components. This decision lies behind the inclusion of an [n] of 0 msec. The heavily nasalized transitions into and out of the closure were sufficient to give an impression of the nasal consonant.

Two tests were constructed. For each test, each stimulus was included four times in a randomized-order listening test with three seconds between stimuli. The resulting tests were presented over headphones to 30 undergraduate students at Ohio University. The students were asked to judge whether the middle word in the three-word sequence was "a" [\( \alpha \)] or "the" [\( \theta \)]. They were instructed that the signal had been degraded and, therefore, that their decision was to be based on which English article the stimulus reminded them of the most.

Results are depicted graphically in Figure 1. At the top we see results for the [\( \alpha \)] test (miss a guy). It demonstrates that when the [\( \alpha \)] assumes a duration of 130 msec, subjects cease to hear the sequence as containing the indefinite article and begin to hear it as containing the definite article.

Figure 1 (bottom) shows the same result at 120 msec for the lengthened nasal segment. We have thus been able to induce the impression of an \( s + \delta \) cluster or \( n + \delta \) cluster by increasing the length of the assimilating consonant.

The curves shown in Figure 1 represent data points for all subjects who responded systematically to the stimuli. Of 31 subjects, 5 responded randomly to both tests. Five additional subjects displayed random results for part 2 (nasal + \( \delta \) assimilation) while performing adequately on part 1. This means that there were 24 subjects for part 1 and 21 for part 2. It is difficult to
Figure 1: Number of "a" and "the" judgments as the nasal segment in "warn a guy" is lengthened; number of "a" and "the" judgments as the fricative segment in "miss a guy" is lengthened.
say why many people did not respond in a patterned fashion to these stimuli; perhaps the utterance is not long enough to induce a casual speech frame of perception for some. It seems unlikely that the frequency with which one hears these processes is a contributing factor, since both assimilations are extremely common in the dialect areas in which the subjects live (Shockey, 1974).

Practical applications are foreseeable for the results of this experiment:

1) Since [ə] is a difficult sound for learners of English to perfect, they can be taught quite early on to assimilate it to appropriate preceding consonants while also lengthening the consonant. This will not only make their English easier to articulate, but closer to standard conversational speech, since consonant + [ə] clusters are scarce in spoken English. Of course, nonassimilable [ə] must still be dealt with. (Current English-for-non-natives texts regard ə-assimilation as substandard. However, the author has observed it to be ubiquitous in the speech of American newscasters, actors, politicians, and others who depend on effective oral communication.)

2) In speech synthesis strategies, consonant + [ə] clusters can be approximated by lengthening the consonants that participate in this rule, which could make the resulting speech phonologically more natural, less stilted, and easier to listen to for extended periods (as is called for in reading machines for the blind). It is likely that other phonological properties of casual or connected speech are equally easy to simulate and should be included in a synthesis strategy that hopes to approximate natural speech output.

REFERENCES