AN ACOUSTIC & PERCEPTUAL STUDY OF UNDERSHOOT IN CLEAR AND CITATION-FORM SPEECH

Seung-Jae Moon

Department of Linguistics
University of Texas at Austin

ABSTRACT

This study investigated vowel reduction (the so-called undershoot phenomenon) in "Clear" (CS) vs. "Citation-Form" (CF) speech. Undershoot was readily observable for all 5 speakers. Furthermore, the results suggested that CS is not merely a louder version of normal speech, but it involves an active reorganization of phonetic gestures. A perception test showed that, in general, CS is more intelligible than CF under identical S/N conditions.

1. INTRODUCTION

In this study we investigated the acoustic characteristics of "clear speech" which was defined in terms of an explicit instruction to subjects to "overarticulate".

The following questions were addressed: Is clear speech merely a louder version of citation-form speech? Or does it also involve an active reorganization of speech gestures? If so, what is the perceptual significance of that reorganization?

The point of departure for the present experiments is unresolved issues of vowel reduction and the so-called undershoot phenomenon. The strong version of duration-dependent undershoot [1] makes vowel duration the only determinant of undershoot. On the other hand, there are findings in the literature[2] that are at variance with that model.

2. ACOUSTIC EXPERIMENT

2.1 Procedure

To induce duration-dependent undershoot, the following test words were used: wheel, wheeling, Wealingham, will, willing Willingham, well, welling Wellingby, wail, wailing and Wailingby. The following three criteria were considered in selecting the test words. First, get a maximum locus-to-target distance. Second, the vowels under analysis must have equal stress. Finally, the duration of the vowels of interest should vary systematically over a considerable range. The first condition was imposed because the larger a given formant movement, the greater the possibility that it will serve as a sensitive indication of articulatory undershoot. It was met by selecting front vowels in a labio-velar context. The second and third criteria were met by using so-called word-length effect.

In addition to the /w/-vowel-/l/ contexts, the same front vowels were also measured in an /h/-vowel-/d/ context. Those measurements were used to provide null-context target values.

For citation-form speech, subjects received no other instructions than to keep their effort and tempo constant and at comfortable levels. For clear-speech, they were explicitly instructed to overarticulate, that is to read the words as clearly as they could. To maintain this performance, during the recording of clear speech, at unpredictable moments, the subject was interrupted through the intercom by the experimenter who would pretend that the token just pronounced had not been understood, and would ask for a repetition. A total of 5 speakers were recorded and measured.

2.2. Results

There is a clear duration dependent undershoot effect: As vowels get shorter, the formant measurements are shifted further and further away from their null-context values and closer and closer to their position in [w].

A closer examination of the raw data indicates that the undershoot effects are vowel-specific. In general, tense vowels are more resistant to undershoot than lax vowels.

Also, the degree of undershoot is talker-specific. Each individual talker exhibits his own pattern of undershoot.

Undershoot is also style-specific: When the clear speech measurements are compared with the data from the other conditions in an F_2-F_1 vowel space diagram, it becomes clear that, for all speakers and all conditions, it is closer to the formant patterns of the null-context vowels. Another way of expressing that observation is to say that clear speech is more peripheral in the vowel space than citation-form speech. It seems as if the vowel space is a flexible object which speakers can adaptively expand or contract according to situational needs.

These findings refute the strong version of the undershoot model: Information on vowel duration alone is not sufficient to predict formant undershoot. Also these results suggest that clear speech is not merely citation-form speech spoken louder and more slowly. Clear speech transforms also involve an active reorganization of phonetic gestures.

A decaying exponential model was fitted to the data from 2 speakers to obtain a more systematical and economical description. The results indicate that the claims made above (vowel-specific, style-specific and talker-specific undershoot pattern) shall be weakened to some extent. It was shown that the dependence of degree of undershoot on identity of vowels and speaking styles is not as strong as it looked based on the raw data. For at

least one speaker, undershoot effects are fairly uniform for all vowels and all speaking styles, provided that appropriate target values are selected for styles and vowels. This modeling also shows that speakers differ in terms of the coefficients used to describe degree of undershoot. However, this fact does not suggest that the speakers need to control these variations directly. These variations are likely to be the results of the articulatory movements themselves and of the non-linear acoustic mapping of the articulatory gestures, not the results of active control over those constraints.

3. PERCEPTUAL EXPERIMENT

What is the perceptual significance of the observed acoustic changes? It is reasonable to assume that, when people speak more clearly, they do so to communicate better and to make their speech more intelligible to the listener. We must ask then, are the clear speech tokens indeed more intelligible than the citation-form speech tokens? To address this question, the following perception experiment was carried out.

3.1. Procedure

The samples for the listening test were chosen from a subset of the words analyzed acoustically. A single representative token was selected for each combination of speaker, vowel, word-length and speaking style. From the various repetitions of the test items, the exemplar which showed a median "acoustic distance" to the null-context was taken as the representative token. For the present purposes, acoustic distance is defined as the Euclidean distance between two points in a three-dimensional formant space calibrated in Mel units.

The representative words were mixed with five different levels of low-frequency weighted Gaussian noise which had a spectral shape of -6dB/oct. One of the noticeable differences between citation-form speech and clear speech was their different intensities. For all five speakers, clear speech was approximately 3-5 dB more intense than citation-form speech. Since our aim was to undertake an intelligibility test based solely on acoustic characteristics other than amplitude, the differences in loudness was normalized by using a special computer program written by Jerry Lame.

Each stimulus was led by a 150ms segment of speech-free noise and was also followed by an interval of noise adjusted so that the duration of the whole stimulus would be the same within a given speaker. There were 120 stimuli per speaker.

These stimuli were presented to normal hearing subjects for identification through headphones. At least 24 responses were collected for each stimulus.

The responses were processed for each speaker and the percentage of correct identification was calculated for each step of S/N ratio.

3.2. Results

In general, the tense vowels show a strong clear-speech advantage while the lax vowels do not. This pattern is consistent for all 5 speakers.

However, let us now consider an alternative measure of S/N ratio. It is the same as before for citation-form speech but does not involve normalizing clear speech. It leaves intensity differences between clear and citation-form speech as they were on the original tape recordings.

When the second definition of S/N ratio is applied to the present data, all test words, when spoken clearly, tend to be more intelligible. With only marginal exceptions that observation is true for all speakers and for all test words.

It can be speculated that the reason for the perceptual advantage of clear speech is multi-dimensional. First, clear speech words tended to be 3-5 dB more intense than citation forms. Second, the formant patterns of the clear speech vowels were found to be closer to their null-context values. And also clear-speech is longer in duration than citation-form speech. In other words, speakers used various strategies to keep undershoot effects down in clear speech.

The intelligibility tests indicate that, in the case of tense vowels, these formant pattern adaptations and systematic duration changes are likely to be responsible for the improved identification scores of clear speech.

Although the style-dependent formant changes and duration changes in lax vowels were entirely analogous to those for the tense vowels, they were not sufficient to make the clear variants more intelligible.

4. CONCLUSION

It has been shown that clear speech is a speech act which involves active reorganization of acoustic patterns and the underlying articulatory gestures, and that it has clear perceptual advantages.

Everyday informal experience suggests that "clear speech" is invoked by a speaker to meet certain communicative and situational demands. And that speakers change and modify their speech according to the needs of their listeners.

The present results indicate that speakers are quite capable of doing so in an experimental situation. They show an ability to successfully adapt to varying demands for explicit signal information.

5. REFERENCES

[1] LINDBLOM, B. (1963), "Spectrographic study of vowel reduction", JASA 35: 1773-1781

[2] GAY, T. (1978), "Effect of speaking rate on vowel formant movements", JASA 63 : 223-230

6. ACKNOWLEDGEMENTS

This research was supported by a grant from the Advanced Research Program of the Texas Board of Coordination and grant No. BNS-9011894 from the NSF.