ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb12. Effects of variation on processing of word-medial consonants
Benjamin V. Tucker*, Kaitlin Mackie and Tatiana Kryuchkova

*Corresponding author's address: Linguistics, University of Alberta, 4-32 Assiniboia Hall, Edmonton, T6G 2E7, AB, Canada, bvtucker@ualberta.ca

The present study investigates processing of variation in word-medial stops and the role of this variation in phoneme recognition. Variation in word-medial stops has been shown to influence lexical access but it is unclear whether this variation also affects recognition at the phoneme level [Tucker, J. Phon., 39, 312-318, (2011)]. A phoneme monitoring task is used to investigate the role of production variation in the identification of word-medial stop consonants. Following Warner & Tucker [J. Acoust. Soc. Am., 130, 1606-1617, (2012)], the stimuli comprise English nonwords with target consonants /b, p, d, t, g, k/ in an intervocalic post-stress position produced with a range of production variation. This variation was grouped into three sets including groups at each extremes and a group in the middle of the range. The acoustic characteristics of the stops, which reflect this variation (e.g. consonant duration and intensity difference) were extracted and used to model the reaction times collected during the phoneme monitoring experiment. The reaction times were statistically modeled using a linear mixed-effects regression. Models of spoken word recognition and models of lexical storage are used to interpret the results and the contribution of the results to our understanding of these models is discussed.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

An understanding of the how variability in speech is understood is necessary for modeling the spoken word recognition process. The variability occurring in speech occurs in many types and forms. This variation can be found in many speaking genres, e.g., list reading, connected reading, monologue and conversation. The present paper focuses on the types of variation found in word-medial stops across all of genres of speech. More specifically, this study investigates the role of reduction based variation in the production of word-medial stops on the recognition of these sounds.

Work by Warner & Tucker (2011) investigated the variation found in the acoustic characteristics of word-medial stops. They found, among other things, that reduced (more approximated forms) were most likely in the more connected and natural speech but that reduced forms of the stops occurred in all three genres tested. Processing of the variation of word-medial stop was also investigated by Tucker (2011) and Tucker & Warner (2007). This research focused on word-medial /d/ and /g/ with two degrees of reduction. They found that the reduced, more approximated forms made lexical access more difficult. These studies along with many others investigating the processing of reductions in spontaneous speech have focused on lexical access as a way to investigate the role of variation in the speech signal and found a similar inhibition of access with reduced forms (e.g. Ernestus, Baayen, & Schreuder, 2002; Ernestus & Baayen 2007; Mitterer & Ernestus, 2007; Ven et al, 2011).

While much of this research has focused on lexical access it is unclear whether the difficulty falls in accessing the individual sounds or if they are processed as a whole. It is possible that the delay in accessing the reduced forms is due to difficulty in accessing the identity of the sound or the delay may be that the sound identified does not match the stored form of the lexical item in question. Further, if this delay occurs during natural speech, why do speakers produce the speech signal with so much variation? It has been proposed that these types of reductions may be speaker or listener oriented. This delay due to the variation in the signal may be a result of the speaker trying to find a balance between hypo- and hyper-articulated speech (Lindblom, 1990).

The present paper focuses on the recognition of word-medial stops when lexical information is not contributing to the identification of the word. It attempts to find out if the delay on the more reduced end of the variation spectrum is due to accessing the phoneme or if it is part of accessing lexical items. Further we hope to learn more about the way in which listeners identify variable speech sounds.

EXPERIMENT

To investigate these questions a phoneme monitoring experiment (Mehta & Cutler, 1988; Conine & Titone, 1996) was designed to assess how well participants identify reduced and unreduced word-medial stops in a speech sequence.

Method

Participants: Participants were undergraduate students at University of Alberta registered in an Introductory Linguistics class and received course credit for their participation. Data from 66 native English speaking participants between the ages of 16-27 is analyzed here (11 male and 45 female). Two participants responding to fewer than 50% of the items were excluded from the analysis.

Materials: Materials were designed such that items were non-word sequences where the phoneme of interest was embedded in a flapping environment, in other words, all stops followed a vowel and were preceded by an unstressed vowel. This environment allowed us to investigate the variance in the production of these stops as defined by the acoustic characteristics of the consonants (Warner & Tucker, 2011; Tucker, 2011) as predictors in the model. We tested all English stop sequences (/p, t, k, b, d, g/) in this environment. Fillers were also included which contained one of the stops of interest in either word onset or word final position. A total of 84 target items and 150 fillers were recorded by a male native speaker in a sound attenuated booth, who did not know the purpose of the experiment. Items were recorded at least 10 times each and the speaker was asked to modulate the speed and the “formality” with which the items were produced allowing us to get variation in the production of the stop. A research assistant manually selected the items looking for items that represented 3 approximated levels of variation: careful, mid, and reduced. Acoustic measures were later extracted from these items by the same research assistant. We extracted multiple measures based on the acoustic analysis and description in Warner & Tucker, (2011) and used the measures of consonant duration and intensity difference (the change in intensity measured by subtracting the minimum
intensity during the consonant from the average peak intensity of the surrounding vowels) as predictors in the model.

Procedure: Participants were seated in a sound attenuated booth and the experiment was presented using E-Prime (CITATION). Following instructions and a brief practice block, participants began the trial. In each trial a participant was presented with an orthographic character of the sound they should monitor, for example “t”. Participants saw this prompt for 1500ms before the onset of the auditory stimulus and during the entire auditory stimulus and until a response was recorded. Participants were given a total of 2000ms to respond to the stimulus. Response time was measured from the offset of the target consonant. Following the response the trial ended and an inter-trial interval was set at 1000ms.

Results & Discussion

A total of 4305 data points were gathered from the speakers after RTs shorter than 150ms and longer than 1500ms were excluded. A linear mixed-effects analysis was performed to identify the significance of the predictors. Two analyses were performed on the data. The first analysis was performed using Target Reduction as a predictor to verify that an effect was found based on the categorization of the research assistant. The results showed that the careful items elicited the fastest response times, with the reduced items eliciting the slowest response times and the mid items being in the middle (Figure 1).

![FIGURE 1. Plot of the initial model which illustrates the main effect of the factor Target Reduction.](image)

A second analysis was performed with Age, Gender, Education, Trial, Letter (the letter the participant was listening for), Word Duration, Consonant Duration, and Intensity Difference as predictors. Target Reduction is not included in this model because the predictors Consonant Duration and Intensity Difference should account acoustically for the effects found in the categorization in the Target Reduction factor. Random-effects of subject and item were included in the model and random slopes were also investigated. No effects of Age, Gender, Education or Trial were found and were excluded from the final model. Consonant Duration was found to strongly correlate with Word Duration (R = 0.43); to remove the colinearity induced by this relationship and to confirm that Consonant

![FIGURE 2. Plot of the final model showing the main effects of Duration, Residualized Consonant Duration, Intensity Difference, Letter.](image)
Duration was indeed contributing to the model, we used the residuals of a model predicting consonant duration from word duration as a new predictor: Residualized Consonant Duration. Random slopes of Duration and Residualized Consonant Duration were the only predictors found to significantly improve the fit of the model. No interactions were found in the modeling and thus only main-effects are reported. A summary of the main-effects are shown in Figure 2 and are summarized below.

Generally we find that the more reduced an item is the slower the response times. Thus we see for Word Duration that the longer a word is the shorter the response time. The predictors Residualized Consonant Duration and Intensity Difference also show a similar effect thus the longer the stop is or the greater the intensity difference the faster the response time, in other words the more carefully produced the target it is the faster the response time. Thus we see for Word Duration that the longer word is the shorter the response time. The more acoustic information the likely the listener is to respond faster. Duration also acts as a predictor of Reduction in this model in that items that are more reduced are shorter than items that are not.

We also find interesting effects from the factor Letter when the intercept is set to /k/. The /k/ and /p/ phoneme are not significantly different from each other and group together. The remaining stops, /t, d, b, g/, are significantly different from /k/. Thus the "voiced" consonants group together and the voiceless consonants do as well, noting that these items are in a flapping environment and /t/ should be produced as [ɾ].

CONCLUSION

The finding in the present paper is in accord with previous research in that we see that items on the reduced end of the variation spectrum are more difficult to process, specifically at the level of phoneme recognition, excluding lexical effects on the word. This indicates that the difficulty for listeners lies in the recognition of the sound [This presumes that we believe that there are phonemes that have to be mapped first]. It is possible that at the level of processing there is no need to access phonemes and that these patterns are mapped directly to a word level representation. However, since there is no word level representation in this experiment it would seem to indicate that what we are finding indicates that listeners have difficulty in identifying the individual sound and mapping it to the form of the sound that has been stored.

The current results fit well with an H&H interpretation (Lindblom, 1990). If we think about the data in terms of the fact that as speakers we want to produce speech in such a way that it is efficient and balances both ease of perception and ease of production it is likely that the production of more reduced variation occurs as part of a trade-off where speakers may reduce in one location while emphasizing others. This likely occurs in interaction with the greater context of spontaneous speech and more information is required about the contributions of the acoustic, syntactic and semantic context to truly understand how this difficult variation allows for efficient communication.

ACKNOWLEDGMENTS

We would like to thank Ryan Pudlubny and the research assistants who helped us run this experiment. This research was funded by the Social Sciences and Humanities Research Council of Canada.

REFERENCES