

RESEARCH ON SPOKEN LANGUAGE PROCESSING
Progress Report No. 21 (1996-1997)
Indiana University

**Lexical Competition and Reduction in Speech:
A Preliminary Report¹**

Richard Wright

*Speech Research Laboratory
Department of Psychology
Indiana University
Bloomington, Indiana 47405*

¹ This work supported by NIH-NIDCD Training Grant DC0012 to Indiana University Bloomington. The author is grateful to Stefan Frisch, Ann Bradlow, and Gina Torretta for helpful comments during preparation of this manuscript.

Lexical Competition and Reduction in Speech: A Preliminary Report

Abstract. Lindblom (1990) among others has proposed that talkers accommodate listeners' communicative needs by controlling the degree of reduction (hyper- and hypo-articulation) in different contextual conditions, thereby maintaining sufficient intelligibility of words across a variety of contexts. Lindblom's proposal predicts that lexical factors that affect intelligibility of a word will affect the hypo- and hyper-articulation of words. Based on factors in lexical competition such as usage frequency and similarity-neighborhood density, previous research has characterized words as "easy" or "hard" to identify. This study examines the degree of centralization of vowels (a well known feature of reduction or hypo-articulation) in 34 "easy" and 34 "hard" monosyllabic (CVC) words of equal familiarity spoken in isolation by 10 talkers. Measurements of the first two vowel formants (F1, F2) were made at the point of maximal displacement in the vowel (excluding the initial and final 50 ms of the vowel). Centralization is measured by calculating the Euclidean distance from the center of a talker's F1-F2 vowel space. Three results emerge: 1) overall "easy" words were significantly more centralized than "hard" words, 2) peripheral vowels, such as /i/, /a/, /u/, showed the greatest effect, and 3) there was considerable between talker variability in the magnitude of the difference between vowels in "easy" and "hard" words. The results are interesting because they demonstrate that the talker takes into account a wider variety of sources of possible noise and information than previously thought. These results have implications for both diachronic and synchronic processes that involve reduction in linguistics. They have further implications for applications in speech recognition that model variation in spoken language.

Introduction

Understanding sources of variability in spoken language is one of the most important challenges that face speech researchers today. It is an issue that spans several fields such as linguistic phonetics with its interest in language specific vs. universal traits of spoken language; historical linguistics and sociolinguistics, which study patterns of sound change; and engineering with its interest in improving automatic speech recognition systems. Traditionally, variability had been treated as noise to be controlled or ignored in studying spoken language; however, there has recently been an increasing interest in exploring lawful variability in spoken language. Variability is a pervasive characteristic of spoken language that is introduced at nearly every level of an utterance. It goes well beyond the frequently noted physiologically based factors such as differences in larynx structure and vocal tract length. For example, in a study that examined a large number of talkers and utterances using the TIMIT speech database, Byrd (1994) found that sex and to a lesser degree dialect differences resulted in between-talker differences in the degree of reduction along a number of dimensions including speech rate, stop release, flapping, and quantity of central vowels. In addition to inter-talker differences which might be viewed as talker or group specific constants, there are many forces that act on spoken language that may change the way a word is pronounced from one utterance to the next by the same talker. Reduced and clear speech processes represent a significant source of within talker variability, much of which has been attributed to talkers varying their pronunciation to accommodate the communicative needs of the listener (e.g., Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon, 1997; Bolinger, 1963; Lieberman, 1963; Lindblom, 1990). In these studies talkers have been shown to produce more reduced speech when contextual information within the utterance or in the environment can aid the listener in recognizing what is said, and to produce more careful speech in when the talker is aware of conditions that may impede the listener's ability to understand what is said. When words are isolated from their spoken contexts, the speech produced under reducing conditions has a low intelligibility and the more careful speech has a high intelligibility. Although most studies have concentrated on contextual and environmental factors, relatively few have examined how word specific characteristics might affect production strategies. The

purpose of this study is to examine the degree to which factors in lexical competition that are known to affect intelligibility of individual words influence the carefulness which talkers produce words.

Reduction and Sources of Information

Although many studies have assumed reduction as a constant, Lindblom (1990) has proposed a more explicit model of the interaction between the forces that shape both reduced and clear speech. In his model, speech motor control is output-oriented and plastic. In this view, reduced speech and clear speech lie along a continuum of contextually determined variability. In a communicative context there is pressure on the system to maintain sufficient information in the signal for the listener to recover the intended message. As factors decrease the probability that a listener will be able to recover the message, in the talker's estimation, output constraints become more severe and the talker is forced to produce clearer speech (i.e., exaggerate contrast among lexical items), which Lindblom calls "hyper-speech." As output constraints become less severe, the system tends towards economy of effort becoming more system-oriented, resulting reduced speech ("hypo-speech"). In this view, speech perception involves discrimination among stored items—lexical access is a function of distinctiveness not invariance. Thus reduction and hyper-articulation will occur along dimensions that will increase or decrease the perceptual distance among lexical items. Reduction in speech can be measured in a number of ways; some of the better known characteristics are shortening of vowels, increased flapping, increased consonant and vowel deletion, and vowel centralization.

A commonly cited example of the talker modifying pronunciation to accommodate the hearer's needs is the Lombard effect (Lane & Tranel, 1971; Lane, Tranel, & Sisson, 1970; Lombard, 1911), in which the talker produces more careful and higher intensity speech in response to increased environmental noise. Similarly, in speech directed towards the hard of hearing there is less phonological reduction and an overall slower rate than is observed in conversational speech (Picheny, Durlach, & Braida, 1986; Uchanski, Choi, Braida, Reed, & Durlach, 1996). More careful productions have also been observed when talkers introduce unexpected or novel information into the discourse (Bolinger, 1963; Chafe, 1974; Hawkins & Warren, 1994). On the other hand, reduction occurs when the talker estimates that a listener will have little difficulty in identifying a word. For example, Lieberman (1963) found that words that are highly predictable from sentential context are more reduced than identical words produced in an equivalent sentence that does little to narrow the field of lexical candidates. When presented (without the rest of the sentential context) as stimuli to listeners, the words from the highly predictive contexts were less intelligible than the words from the less redundant contexts. Similar results have been found for a variety of predictive contexts (Bard & Anderson, 1983; Fowler & Housum, 1987; Hunnicutt, 1985).

Talkers are sensitive to a wide variety of sources of information not only from the auditory mode but also from the visual mode (and one would predict the haptic mode as well). For example, Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon (1997) found that a talker's performance reflects the listener's access to visual information from the talker's face. Visual information in the talker's face can aid the hearer enormously in recovering the spoken message: the equivalent of as much as an 18 dB gain by some estimations (Sumbly & Pollack, 1954). Anderson et al. (1997) found that tokens in spoken language when the talker estimated that listener had access to visual information were significantly degraded. In later intelligibility tests, the reduction of the auditory signal was offset only if subjects were given access to the accompanying video stimuli in addition to the audio stimuli. An interesting point that emerged from the study was that talkers maintained only a loose model of the information available to the listener; rather than tracking the listener's use of visual information from moment to moment, the talker adjusts the carefulness of speech in a more global fashion basing the model on talker internal conditions. That is, talkers did not pay attention to whether or not the listener was actually looking at his/her face, but rather to whether or not the listener had the ability to look.

Lexical Competition and Intelligibility

In addition to the message internal and environmental contextual factors that affect a word's intelligibility, there are lexical factors that may increase or decrease the probability of a talker identifying a word correctly. Word frequency is perhaps the best known lexical trait that may affect a

word's intelligibility. It is also a proposed factor in overall word shortening and acceleration of reduction processes (Balota, Boland, & Shields, 1989; Bybee, 1994; Zipf, 1935). However, overall word frequency alone has proven to be a rather poor predictor of intelligibility (Luce, 1986; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985) and has proven an unreliable predictor of reduction. Rather, word identification must be viewed in the context of lexical competition for the role of frequency in intelligibility to be clearly seen. In his dissertation, Luce (1986) studied patterns of auditory word confusion and found that a word's intelligibility is affected by two lexical factors: 1) *neighborhood density*: the number of phonologically similar words in the language, and 2) *relative frequency*: the frequency of the target word relative to its nearest phonological neighbors. In calculating nearest neighbors, Luce found a reasonably close match between his confusion matrices and neighborhood density determined using the *single phoneme substitution* method (Greenberg and Jenkins, 1964) in which all words that differ from the target word by a single phoneme are considered nearest neighbors. Luce proposed the Neighborhood Activation Model (NAM) in which the number of similar competitors that a word has and usage frequency have inhibitory and excitatory effects in lexical access and competition. In this model, while frequency will determine the probability of a word beating out its neighbors, a word with few neighbors is likely to be identified even if its usage frequency is low. Based on intelligibility properties, words from high density neighborhoods and with low relative frequency have been termed "hard" and those from low density similarity neighborhoods and with high relative frequency have been termed "easy" (Luce, 1986; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985).

A preliminary study of the effect of neighborhood density on voice onset time (VOT), one of the main indicators of stop consonant voicing, was conducted by Goldinger and Summers (1989). In their study, they had talkers read minimal pairs of CVC words in which the word-initial stop consonant was either a voiced or a voiceless stop. The pairs were chosen so that both were from sparse neighborhoods or both were from dense neighborhoods. Each talker read each pair of words four times. Overall there was a greater difference in VOT between voiced-voiceless pairs from dense neighborhoods than in voiced-voiceless pairs from sparse neighborhoods. That is, the voicing contrast that is cued in part by VOT was more exaggerated in minimal pairs from dense neighborhoods than in minimal pairs from sparse neighborhoods. Moreover, across repetitions the difference in VOT between the dense pairs increased dramatically while the VOT difference between the sparse pairs increased slightly. The study was flawed because the method of presenting the words in minimal pairs attracts the talkers attention to the contrast being studied and generally results in an exaggeration of the contrast. Nevertheless, the fact that the sparse and dense neighborhoods predicted differences in talkers' behavior over repetitions is a preliminary indication that neighborhood density is a factor in variability of spoken language. Interestingly, the fact that all the words in the study showed the effect indicates that neighborhood density appears to have a global word level effect, at least on monosyllabic words, because the majority of lexical competitors would not have been confusable on the first phoneme. That is, although some of the words may have had neighbors that were close because they differed only in the initial phoneme, the single phoneme substitution method that the authors employed to calculate neighborhood density implies that roughly two thirds of the neighbors are based on the second or third phonemes in the words. The apparent whole word effect implies that talkers have only a loose sense of sources of neighborhood density rather than making fine adjustments. This last point is similar to the observation in Anderson et al. that talkers maintain only a crude estimation of what visual information listeners may be using.

The current study was designed with these preliminary VOT findings in mind. Because of the interaction between neighborhood density and relative frequency in the intelligibility studies mentioned above, it was decided that the first place to look for an effect was in "easy" vs. "hard" words. This choice was further driven by the existence of a publicly available prerecorded database of "easy" and "hard" tokens. One advantage of using a database is that a relatively large number of talkers can be studied, ten in this case. A second advantage is that the words of interest are well studied in independent research; tokens in the database have been measured for intelligibility and have been pre-coded for lexical characteristics. There are disadvantages to working with databases. The biggest one is lack of control the experimenter has over the recording conditions and over the choice of recorded material. In this case it would have been ideal to be able to have access to more detailed demographic information about the talkers and to be able to make follow-up recordings. Despite these shortcomings, this database and others like it provide researchers with rich tools for

testing hypotheses about spoken language. It is predicted in this study that factors in lexical competition should affect the degree of reduction and in production just as contextual factors do. That is, “easy” words should show a greater degree of reduction than “hard” words. Because differences in vowel centralization have been shown to differentiate clear from casual speech (e.g., Byrd, 1994; Lindblom, 1990; Lindblom & Moon, 1994) and because increased vowel dispersion is an established correlate of intelligibility (Bond & Moore, 1994; Bradlow, Torretta, & Pisoni, 1996; Picheny, Durlach, & Braida, 1985), “hard” words are predicted to have vowels that are more dispersed (i.e., less centralized) than “easy” words. Increased dispersion is characterized by an expanded vowel space or by overall increased acoustic distances between vowels from different categories.

Method

Recording Materials

Tokens in the study were all monosyllabic CVC words drawn from a prerecorded database (see Torretta, 1995, for a detailed description). All words were of equally high familiarity, being between 6.8 and 7.0 on the 1-7 point Hoosier Mental Lexicon scale (Nusbaum, Pisoni, & Davis, 1984), but varied crucially in their similarity neighborhood density and in relative usage frequency. One set of words, termed “easy,” came from sparse similarity neighborhoods and had usage frequencies that were high relative to their neighbors. A second set of words, termed “hard,” came from dense similarity neighborhoods and had usage frequencies that were low relative to their neighbors. The words were chosen to provide a balanced segmental context for the vowel; consonantal contexts that could result in vowel coloring were avoided while maintaining similar contexts in easy and hard words. For example, postvocalic /r/ and /l/ were avoided altogether for both types of words, and nasal codas, when unavoidable, were balanced in both sets. The full set of tokens is listed in Appendix 1. Overall, there are 34 “easy” and 34 “hard” words spoken by 10 talkers (5 male and 5 female) of American English resulting in 680 tokens total. The sound files in the database are digital recordings of monosyllabic words presented singly in pseudo-random order on a CRT monitor and read once in isolation. The talkers were instructed to say each word at a “medium” rate. The utterances were antialias filtered and digitized directly to disk at 22050 Hz (see Torretta, 1995, for a detailed description).

Measurement

As the file names in the database included lexical neighborhood information, the files were renamed and randomized prior to measurement. The first and second formants (F1 and F2) of each vowel were measured at the point of maximal displacement on all 680 tokens. The initial and final 50 ms of the vowel were excluded to minimize the effect of flanking consonants on the measurement. The point of maximal displacement occurs when F1 and F2 are the most characteristic for that particular vowel. For example, for the vowel /i/ it is the point where F1 is lowest and F2 is highest and for /a/ it is the point where F1 is highest and F2 is lowest. Where F1 and F2 were not in agreement, F1 was taken as the point of reference and F2 was measured at that point. This measure is equivalent to what has typically been described in the literature as the “steady state” but takes into account the dynamic character of vowels which often results in the absence of a clear steady state. For diphthongs the measure was taken in the primary portion of the vowel (e.g., for /aj/ the measure was for the /a/ portion). Formant values were measured from a twelfth order LPC with a 25 ms window overlaid on a simultaneous 512 point FFT. A wideband spectrogram was used to locate the measurement point and for reference during the formant measures. The formant values were converted into the Bark scale (an auditory transform) using the formula given in (1) (Zwicker & Terhardt, 1980) where Z is bark and f is frequency in Hertz.

$$Z = \frac{26.81f}{1960+f} - 0.53 \quad (1)$$

The degree of dispersion was measured using a technique applied in Bradlow, Torretta, & Pisoni’s (1996) study of talker intelligibility: the Euclidean distance from the center of a talker’s F1

by F2 vowel space. Using this measure, Bradlow et al. found that one of the best correlates of talker intelligibility was the degree of vowel dispersion. Differences in dispersion were submitted to an analysis of variance with dispersion as the dependent variable and lexical category (easy/hard), vowel category, and talker as the independent variables.

Results and Discussion

Overall, The hypothesis was borne out; there was a reliable effect of lexical category (“easy” vs. “hard”) on dispersion (the Euclidean distance from the center of the vowel space). With an alpha level of .01 the analysis revealed a significant main effect of lexical category, $F(1,480) = 130.92, p < .0001$. There was also a significant interaction between lexical category and vowel type, $F(9,480) = 15.22, p < .0001$. Figure 1 shows the overall dispersion in Bark (vertical axis) in the height of the bars for “easy” vs. “hard” words collapsing across talker and vowel type. There is a clear difference in the degree of dispersion with the vowels from “hard” words being more dispersed on average than the vowels from “easy” words.

Insert Figure 1 about here

This difference in dispersion represents an overall expansion of the vowel space for hard words. Figure 2 is an F1 by F2 plot of the mean values for each vowel: vowels from “hard” words are plotted using darker slightly larger symbols and vowels from “easy” words are plotted with the lighter symbols. This plot illustrates two characteristics of the data: the overall expansion of the “hard” vowel space, and the tendency for certain vowels to show greater expansion than others.

Insert Figure 2 about here

When the data is analyzed by vowel the differences between vowel types becomes clear: the vowels / i, Q, a, ç, u / (point vowels) show the greatest difference between “easy” and “hard words whereas the remainder of the vowels are only slightly expanded or not expanded at all. For some of the vowels / I, E, o,

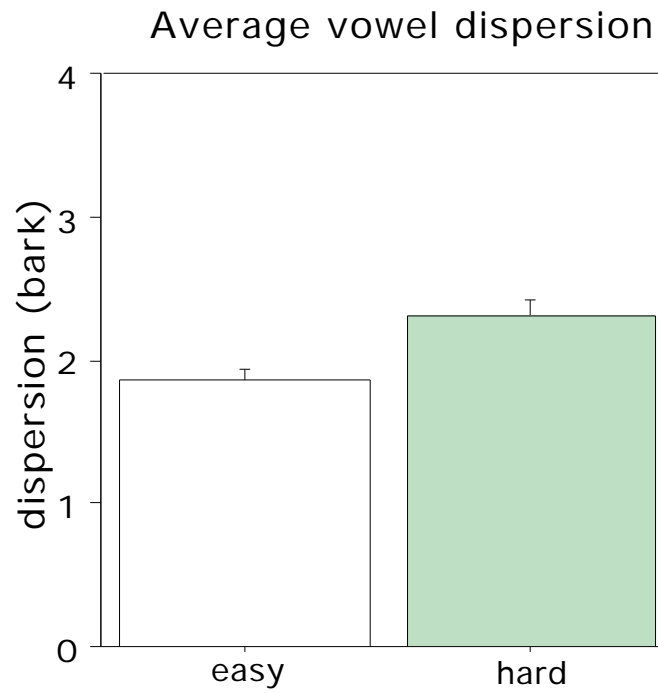


Figure 1. Vowel dispersion for “easy” and “hard” words averaged across talker and vowel, error bars indicate 95% confidence intervals.

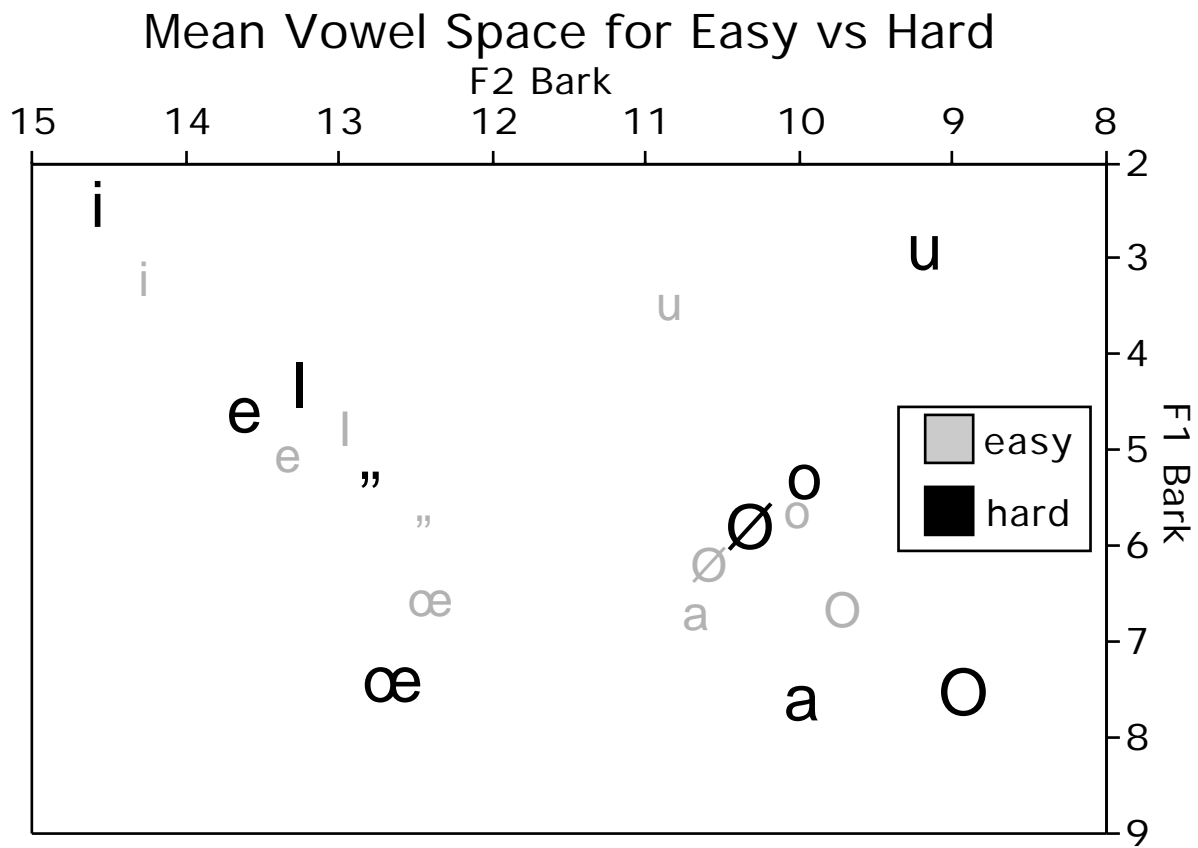


Figure 2. A traditional F1 by F2 vowel plot with F1 on the Vertical axis and F2 on the horizontal axis. Vowel category means are plotted using dark symbols for vowels from “hard” words and lighter symbols for vowels form “easy” words.

÷/ there is no reliable difference between the two conditions. Figure 3 illustrates the latter point. It is a plot of dispersion that splits “easy” vs. “hard” by vowel category. There is a marked difference in the height of the dispersion bars between /i, Q, a, ç, u/ (point vowels) and the remainder of the vowels.

 Insert Figure 3 about here

The difference in dispersion across conditions between the point vowels and the rest of the vowels is in one sense unsurprising. Given the physiological limitations in producing vowels, and given that vowels contrast with each other rather than with the Euclidean center of the space, moving the point vowels while leaving the other vowels fixed maximizes the acoustic distance between vowel of different categories. This stretching of the acoustic space should make the vowel contrasts more salient by increasing the perceptual distance between vowels of differing categories. This finding for English mirrors a crosslinguistic simulation conducted by Liljencrants & Lindblom (1972) that explored the relationship between the number of vowels in a system and the shape of vowel spaces. Their simulation, which took into account the physiological limitations of the vocal tract aimed to find the vowel spaces which maximized the distance between vowels in a two dimensional F1 by F2 space. Their distance formula, given in (2), calculates the repulsive force in a system of vowels using the inverse sum of the squared distances between vowels. With this measure, the closer the different items are the greater the repulsive force.

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} 1/r^2_{ij} \tag{2}$$

where r^2_{ij} is $\sqrt{(F1_i - F1_j)^2 + (F2_i - F2_j)^2}$

This measure of distance was applied to the set of monophthongs and resulted in a clear difference in repulsive force between the vowel systems from “easy” and “hard” words. Across talkers the “easy” system showed dramatically more repulsive force than the “hard” system, indicating a more expanded vowel space for “hard” words. In looking at individual talkers, this measure reveals striking individual differences. Figure 4 plots repulsive force on the vertical axis by talker (greater repulsive force indicates less overall distance between vowels).

 Insert Figure 4 about here

Although the “hard” vowel space showed less repulsive force for all the individuals, some of the individuals exhibited a much greater magnitude of difference. For example, the magnitude of talker M9’s difference is several times that of talker F1.

Conclusion

The data supports the hypothesis that vowels from “hard” words are more hyper-articulated than vowels from “easy” words. The expansion of the vowel space occurs in such a way that overall distances between vowels are maximized; only the point vowels, which can move without obscuring the vowel

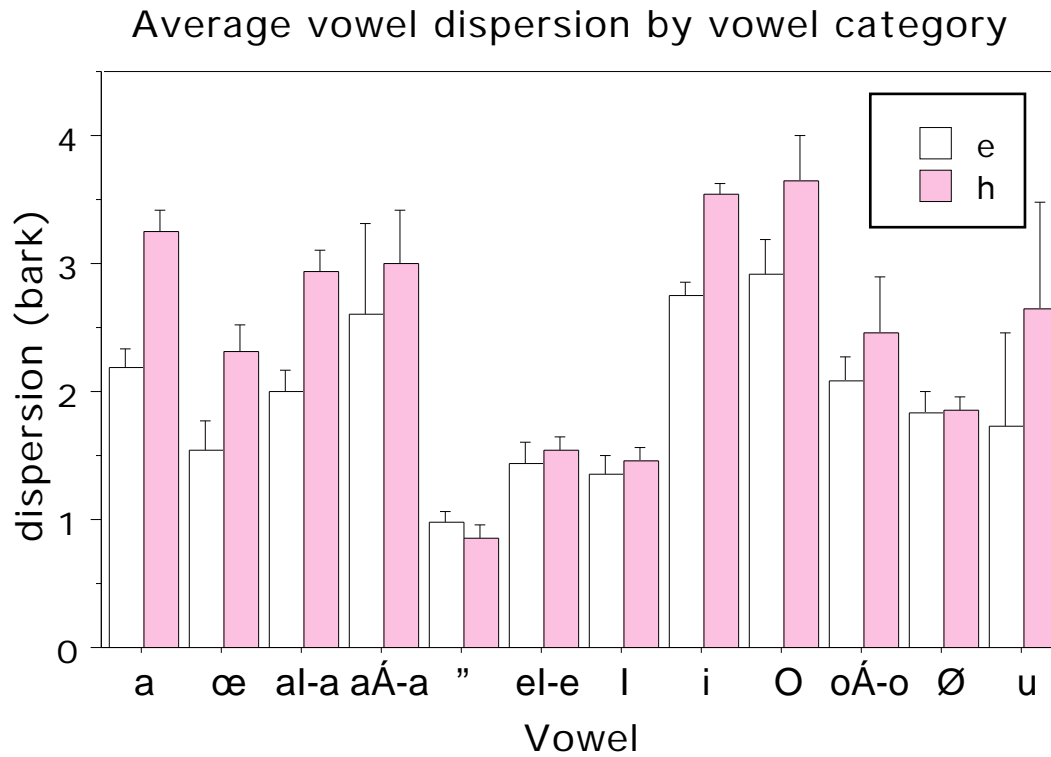


Figure 3. Vowel dispersion for “easy” and “hard” words by vowel type averaged across talker, error bars indicate 95% confidence intervals.

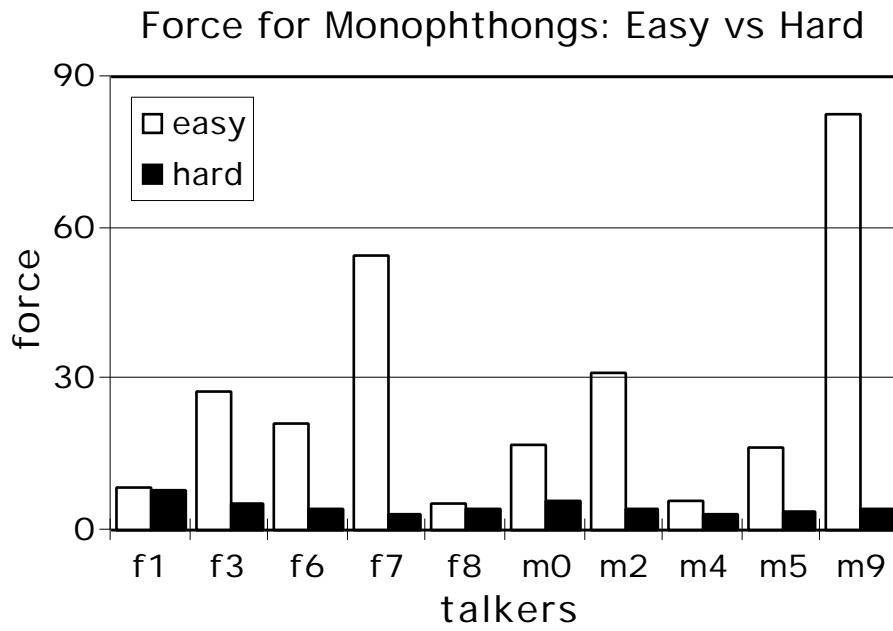


Figure 4. Repulsive force in vowel systems from “easy” and “hard words. Greater force indicates an overall more compact vowel system with less distance between vowels tokens from

contrasts, become more dispersed while the others remain relatively unchanged across conditions. This finding replicates previous studies' findings that talkers adjust the degree of hyper-articulation to compensate for factors that may impede the intelligibility of a message. This study is novel in that it finds compensatory hyper-articulation for lexical characteristics of individual words. This finding has implications for speech recognition research because it represents a potentially significant reduction in the amount of *random* variability which must be dealt with in an ad-hoc fashion. It also has implications for linguistics, sociolinguistics and historical linguistics in that processes that refer to reduction should take into account lexical properties of words.

It is expected that in follow-up studies, lexical factors will be shown to interact with other processes that promote reduction or hyper-articulation such as information redundancy, noise, the new vs. given status of an utterance, and the familiarity of a talker with a particular hearer. It should be the case that "hard" words will be proportionally more hyper-articulated in conditions where the hearer is expected to have greater difficulty recovering the utterance, and that "easy" words will be proportionally more reduced in conditions where the hearer is expected to have greater ease in recovering the utterance. The individual differences seen in this study are interesting because they indicate non-uniform behavior in the face of relatively uniform differences in factors that affect intelligibility. The individual differences might be thought to be due to differences in the individual's lexicon. Although this is a possibility, it seems unlikely considering the high familiarity of the items used. It may rather indicate differing sensitivity to the effects of lexical factors on intelligibility, different levels of willingness to take advantage of lexical factors, or differential responses to the demands of the recording setup. Further studies of individual differences in hyper-articulatory strategies and their correlations with intelligibility are needed to answer these questions.

References

- Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception & Psychophysics*, **39**(4), 580-592.
- Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language*, **28**, 14-36.
- Bard, E. G., & Anderson, A. H. (1983). The unintelligibility of speech to children. *Journal of Child Language*, **10**, 265-292.
- Bolinger, D. (1963). Length, vowel, juncture. *Linguistics*, **1**, 5-29.
- Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, **14** (4), 325-337.
- Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, **20**, 255-272.
- Bybee, J. L. (1994). A view of phonology from a cognitive perspective. *Cognitive Linguistics*, **5**(4), 285-305.
- Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, **15**, 39-54.
- Chafe, W. (1974). Language and consciousness. *Language*, **50**, 111-133.
- Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Memory and Language*, **26**, 489-504.
- Goldinger, S. D., & Summers, W. V. (1989). Lexical neighborhoods in speech production: A first report, *Research on Speech Perception Progress Report No. 15* (pp. 331-342). Bloomington, IN: Speech Research Laboratory, Indiana University.

- Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, **20**, 157-177.
- Hawkins, S., & Warren, P. (1994). Phonetic influences on the intelligibility of conversational speech. *Journal of Phonetics*, **22**, 493-511.
- Hunnicuttt, S. (1985). Intelligibility vs. redundancy—conditions of dependency. *Language and Speech*, **28**, 47-56.
- Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, **14**, 677-709.
- Lane, H., Tranel, B., & Sisson, C. (1970). Regulation of voice communication by sensory dynamics. *Journal of the Acoustical Society of America*, **47**, 618-624.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, **6**, 172-187.
- Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, **48**, 839-862.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling*, pp. 403-439. Dordrecht: Kluwer.
- Lindblom, B., & Moon, S. J. (1994). Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, **96**, 40-55.
- Lombard, E. (1911). Le signe de l'élévation de la voix. *Annales des maladies de l'oreille, du larynx, du nez et du pharynx*, **37**, 101-119.
- Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Unpublished Ph.D. dissertation, Indiana University.
- Nusbaum, H. C., Pisoni, D. B., & Davis, C. (1984). Sizing up the Hoosier mental lexicon, *Research on Speech Perception Progress Report No. 10*, . Bloomington, IN: Speech Research Laboratory, Indiana University.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, **28**, 96-103.
- Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, **29**, 434-446.
- Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition, and the structure of the lexicon. *Speech Communication*, **4**, 75-95.
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.
- Torretta, G. M. (1995). The “easy-hard” word multi-talker speech database: An initial report, *Research on Spoken Language Processing Progress Report No. 20* (pp. 321-333). Bloomington, IN: Speech Research Laboratory, Indiana University.

Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., & Durlach, N. I. (1996). Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate. *Journal of Speech and Hearing Research*, **39**, 494-509.

Zipf, G. K. (1935). *The Psycho-biology of Language*. Boston: Houghton Mifflin.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, **68**, 1523-1524.

Appendix 1: Words Used in Study“easy”

job
 watch
 shop
 gas
 jack
 path
 five
 wife
 vice
 mouth
 gave
 faith
 shape
 page
 chain
 death
 check
 leg
 peace
 deep
 teeth
 give
 thing
 ship
 thick
 wash
 both
 vote
 food
 young
 love
 judge
 hung
 rough

“hard”

wad
 knob
 cod
 pat
 hack
 hash
 rhyme
 white
 lice
 rout
 fade
 dame
 mace
 sane
 wade
 den
 wed
 pet
 bead
 teat
 weed
 kit
 hick
 kin
 mitt
 cot
 goat
 moat
 hoot
 hum
 pup
 mum
 bum
 bug