

a. Specific Aims

As described in our competing application, our proposed research involves the contribution of visible information in face-to-face communication and how it is combined with auditory information in bimodal speech perception. The experimental research methodology, substantiated in previous psychological investigations (Massaro & Cohen, 1983; Massaro, 1987, 1989, 1998), is used. The research paradigm utilizes a strong-inference strategy of hypothesis testing, independent manipulations of multiple sources of information, and the testing of mathematical models of performance. Facial animation and synthetic speech will allow the auditory and visual signals to be manipulated directly, an experimental feature central to the study of psychophysics and perception. In addition, expanded factorial designs are used when possible to provide the most powerful test of quantitative models of perceptual recognition (e.g., Massaro & Cohen, 1990). This paradigm has been highly successful in discriminating among contrasting theories and models, and it is important to expand the approach to provide additional tests among extant theories, to evaluate a new set of procedures for model selection, to broaden the nature of the experimental tests, and to extend the prototypical studies to a series of new speech distinctions and a broader range of experimental conditions.

b. Studies and Results

The fuzzy logical model of perception (FLMP, Massaro, 1998) has been extremely successful at describing performance across a wide range of ecological domains as well as for a broad spectrum of individuals. Because the model predicts optimal or maximally efficient integration, an important issue is whether this is the case for most individuals. Two of our data bases and one provided by Grant and Seitz (1998) were evaluated to determine to what extent a significant quantitative improvement in predictive ability can be obtained if integration is assumed to be somewhat inefficient (Massaro & Cohen, 2000). For the most part, there were no significant signs of inefficient integration. The differences claimed by Grant and Seitz (1998) must be due to their measures of efficiency, which appear to be invalid and/or conflate information with integration efficiency.

One important question is whether the descriptive ability of the FLMP is theoretically informative or whether it simply reflects the model's ability to describe a wider range of possible outcomes. Previous contrasts of this model with others have been adjudicated on the basis of both a root mean square deviation (RMSD) for goodness-of-fit and an observed RMSD relative to a benchmark RMSD. Massaro et al. (in press) extended the model evaluation by another technique called Bayes factor. The FLMP maintains its significant descriptive advantage with this new criterion. In a series of simulations, the RMSD also accurately recovers the correct model under actual experimental conditions. When additional variability was added to the results, the models continued to be recoverable. In addition to its descriptive accuracy, RMSD should not be ignored in model testing because it can be justified theoretically and provides a direct and meaningful index of goodness-of-fit. We also make the case for the necessity of free parameters in model testing.

We assessed the relative importance of the acoustic properties of vowels versus consonants to recognition of words in fluent speech (Cole et al., submitted). Subjects listened to (a) unaltered utterances from the TIMIT speech corpus; (b) utterances in which all of the vowels were replaced by a substituting sound; or (c) utterances in which all of the consonants were replaced by a substituting sound. The results of six experiments show that recognition of words depends more on vowels than consonants. On average, about twice as many words are recognized when vowels

are present. An analysis of the phonetic structure of our stimuli showed that, from an information theoretic point of view, consonants provide more information about the words in our experiments than vowels, and thus cannot account for the observed (opposite) effect in our experiments. These experiments provide compelling evidence that acoustic information within vowel segments provides more information about words than information within consonant segments during perception of fluent speech. We hypothesize that acoustic cues in vowel segments (such as formant values and durations) provide more information about consonants than acoustic cues in consonants provide about vowels, and that the greater acoustic information provided by vowels about adjacent consonants leads to more accurate word recognition in our experiments. Furthermore, the outcome provides another illustration of how observers exploit several sources of information in sentence processing: in this case, acoustic information and a priori information are used together for recognition.

It is valuable to study analogous processes in reading and speech perception (Massaro & Jesse, in progress). Two of the main critical issues in visual word recognition have been 1) how letter processing mediates word recognition and 2) whether words are recognized directly from written print or are first translated into their pronunciation which then accesses the lexicon. The latter question has also been extended to ask how orthographic information maps into phonology and vice versa. We carried out two experiments in which we manipulated spelling-to-sound (feedforward) and sound-to-spelling (feedback) consistency in a 2 by 2 factorial design and test both German and English monosyllabic words in a perceptual identification task, the fragmentation task. In this task an incomplete fragmented stimulus is gradually completed by the participants until it is identified. Although significant effects were found with German test words, we did not observe these feedforward and feedback consistency effects when we test English monosyllabic words. A possible explanation for these results is that the English stimuli were better matched between groups on potentially influencing variables, such as subjective familiarity. The fragmentation task might also have not been sensitive enough to reflect the fast automatic phonological processes that are responsible for consistency effects. We also address certain problems with consistency measures, such as unit size, and show that the existence of a feedback consistency effect does not require interactive activation, but can be explained and predicted by a feedforward model, the FLMP.

The gating paradigm was extended to audiovisual speech in two experiments examining the integration of partial information in bimodal speech perception (De la Vaux & Massaro, in progress). Auditory, visual and bimodal single syllable consonant vowel consonant (CVC) words were presented at one of eight stimulus durations. Performance improved systematically with increases in stimulus duration under all three presentation conditions. Performance in the bimodal conditions was better than either unimodal condition. The FLMP was found to provide a superior fit to the word identifications and confusion matrices than other models.

c. Significance

We have studied how a speaker's face provides valuable and effective information for the perception and understanding of language and its accompanying emotional content. We envision speech perception as an instance of a more general process of pattern recognition in which persons use multiple sources of information. Recognition is achieved via a variety of bottom-up and top-down sources of information. Our research addresses both empirical and theoretical issues. At the empirical level, experiments are carried out to determine how visible speech is combined with auditory speech for a broad range of individuals and across a wide variation of situational domains. At the theoretical level, the assumptions and predictions of several models are formalized, analyzed, contrasted, and tested. Our recent analysis of research from several different laboratories has shown that both children and adults with hearing loss benefit greatly

from having visible speech presented jointly with the necessarily degraded audible speech. These positive findings encourage the use of multimodal environments for persons with hearing loss.

d. Plans

One area of inquiry is aimed at clarifying the classic McGurk effect. Previous studies have failed to measure a sufficient number of experimental conditions and to account for other potential influences on performance. The proposed studies will assess the contribution of the frequency of the phonological segments in the language, the role of potential lexical alternatives created by the two sources of information, and the psychophysical properties of the auditory and visual speech. A series of experiments is proposed to study the uniqueness of visible speech by contrasting it with written text in terms of how it is integrated with auditory speech. Experiments are designed to test the hypothesis that other sources of information, in addition to visible speech, are capable of influencing the perceptual (dominantly auditory) experience of speech. In addition to these studies of the perception of segmental information, the proposed research includes the study of the perception of paralinguistic information, such as the auditory and visual cues to word emphasis or stress. Potential cues in both of these modalities will be manipulated in order to determine which ones are functionally valid and to assess how they are evaluated and integrated in speech perception.

e. Publications

- Massaro, D.W., Cohen, M. M., Campbell, C.S., & Rodriguez, T. (in press). Bayesian method of model selection validates FLMP. *Psychonomic Bulletin & Review*, in press. http://mambo.ucsc.edu/psl/dwm/model_selection.doc
- Massaro, D. W. "Speech Perception" (in press). W. Kintsch (Ed) *International Encyclopedia of the Social & Behavioral Sciences*, expected publication 2001
- Massaro, D. W. (2000) The horse race to language understanding: FLMP was first out of the gate, and has yet to be overtaken. *Behavioral and Brain Sciences*, 23, 338-339.
- Massaro, D.W., & Cole, R. (2000). From "Speech is special" to talking heads in language learning. In *Integrating Speech Technology in the (Language Learning and Assistive Interface*, University of Abertay Dundee, Dundee, Scotland, August 29-30, 153-161.
- Massaro, D.W., Cohen, M. M. (2000). Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *Journal of the Acoustical Society of America*, 108, 784-789.
- Massaro, D.W., Cohen, M. M., Beskow, J., & Cole, R. A. (2000). Developing and evaluating conversational agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.) *Embodied conversational agents*. Cambridge, MA: MIT Press.
- Massaro, D.W. (2000). Reply to Vroomen and de Gelder. *Trends in Cognitive Sciences*, 4, 38-39
- Massaro, D. W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3, 310-317.
- Cole, R.A., Massaro, D.W., Yan, Y., Mak, B. & Fenty, M. (submitted) The Role of Vowels versus Consonants to Word Recognition in Fluent Speech. *Journal of the Acoustical Society of America*.
- Massaro, D.W., & de la Vaux, S.K. (in progress) Audiovisual Speech Gating: Examining Information and Information Processing.
- Massaro, D.W., & Jesse, A. (in progress). Explorations of reading processes within the framework of the fuzzy-logical model of perception.