
Perception of visible speech: influence of spatial quantization

Christopher S Campbell, Dominic W Massaro

Department of Psychology, University of California at Santa Cruz, Santa Cruz, CA 95064, USA

Received 15 October 1996, in revised form 17 April 1997

Abstract. Visible speech reading was studied to determine which features are functional and to test several models of pattern recognition. Nine test syllables differing in their initial consonant were presented in intact form or under various levels of spatial quantization. Performance decreased in increasing quantization but remained relatively good at moderate levels of degradation. Different models were tested against the confusion matrices. Six features were identified as functional in distinguishing among the nine consonant–vowel syllables. These features were used as sources of information in a fuzzy-logical model of perception and an additive model. The fuzzy-logical model provided a significantly better description of the confusion matrices, showing that speech reading is analogous to other domains of pattern recognition such as face recognition and facial-affect perception.

1 Introduction

The talker's face as well as his or her voice convey linguistic information in face-to-face communication. This contribution of visible speech has been shown in various speech-perception contexts; namely, with auditory speech that is conflicting (McGurk and McDonald 1976), ambiguous (Massaro and Cohen 1990), or degraded (Sumbly and Pollack 1954). In the absence of audible speech subjects with normal hearing have been shown to speech read reliably without any systematic training (Heider and Heider 1940; Massaro and Cohen 1995). Research has also demonstrated that subjects with normal hearing (Massaro et al 1993) and hearing-impaired subjects (Walden et al 1977) can be trained to discriminate nine classes of visible consonant phonemes (visemes) without the aid of audible speech. In addition, eight vowel-viseme categories can also be speech read reliably (Montgomery and Jackson 1983).

In further inquiry into visible-speech perception it has been asked what features of the face convey the perceptually functional information for speech reading. Not surprisingly, studies have shown that functional features reside in the lower half of the face. For example, experiments by Erber (1974) with $\frac{3}{4}$ -facial views indicate that jaw rotation and cheek movement are important. Functional features have also been shown to reside in the lips. Summerfield (1979) presented subjects with different lip representations in four conditions to determine their contribution to perception when the acoustic speech was distorted with interfering prose: (a) a control in which the entire face was shown; (b) lips isolated by painting with phosphorescent paint and filming in ultraviolet lighting; (c) a point-light display of the lips similar to that of Johansson (1974); (d) a moving-ring representation of the lips. In all conditions, the visual display was shown in synchronization with auditory speech. Correct responding was increased by 42.6% when the entire face was shown relative to hearing the distorted auditory speech alone. With only the painted lips, correct responding was improved by 31.3%. The third and fourth conditions, however, did not significantly improve performance. Summerfield concluded that untrained subjects use the functional features of the lips to perceive speech. It was speculated that these features may include lip occlusion, horizontal lip extension, and oral area. Benoit et al (1996) found that a view of the jawbone enhanced speech reading relative to just a view of the lips. Including the skin around the jaw improved speech reading even more.

