Massaro, D.W. (1998). Perceiving talking faces: Insights into Auditory Attention. In Proceedings of 16th International Congress on Acoustics, and 135th Acoustical Society of America, 3, (pp. 1573-1574).

Perceiving Talking Faces: Insights into Auditory Attention

Dominic W. Massaro

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

Abstract: Perceivers naturally integrate auditory from the voice and visual information from the face in the perception of speech and emotion. Although both modalities contribute to perception, instructions and intention can modulate the impact of these modalities. Although there is some influence of the to-be-ignored modality, perceivers can attenuate its influence so that some degree of control is possible. The fuzzy logical model of perception (FLMP) provides a good account of performance under different instruction conditions. Intention can be accounted for in terms of information with no qualitative changes in information processing. Within the framework of the FLMP, the fundamental pattern recognition algorithm is changeless. Bearing out this assumption, intention can enhance or attenuate the information contribution of sources that are available, but intention does not seem capable of enforcing another type of pattern recognition algorithm.

The study of auditory attention is about as old as the study of audition itself. We have had many experiences in which we feel capable of attending to one auditory input and ignoring others. For example, we follow one melodic line in a Baroque score without being confused by the other line. Or we comprehend one conversational dialog while ignoring several others. Our children tell us that they are listening even though they are engaged in some other activity. One question is to what extent we can succeed in enhancing the processing of an auditory input when only that input is presented relative to the case in which another input is simultaneously processed.

METHOD

We have addressed the question of auditory attention in our bimodal speech perception task. Observers identify auditory, visual, or bimodal speech syllables under varying instruction conditions. Here we report an extension of this research to assess how easily people could filter out one of the sources of emotion information. Two types of instructions were given to direct the intentional set or goal of the participant. Participants were instructed to watch the face and to listen to the word and to identify the emotion as happy or angry. For the bimodal instructions, they were instructed to make their judgment on the basis of both modalities. For the auditory instructions, they were instructed to "make the judgment on the basis of what you hear the voice to be expressing." Of course, it was necessary to warn the participants that sometimes only the face would be presented and, therefore, they would have to make their judgments on this basis. Is it possible to completely focus on one modality and shut out any influence from the other?

The same task was carried out in auditory/visual speech to allow a direct comparison of instructions across the speech and emotion domains. One of the themes of our framework is that there are analogous processes across a broad range of domains. Thus, we expect to find similar results for emotion and speech. The speech stimuli were chosen from the audible and visible synthetic speech. We created a good /ba/ and a good /da/ and an ambiguous syllable between these alternatives. These were used in a 3 by 3 expanded factorial design, exactly analogous to the emotion conditions. Each participant was tested under both instruction conditions in either the speech or emotion task. There were 24 participants in speech experiment and 26 in the emotion study.

RESULTS

Instructions had a significant impact on performance in the emotion task. With the bimodal instructions ("use the information from both the face and the voice"), the face had a much larger influence in the bimodal stimulus condition than did the voice. The voice had very little influence when the face was either happy or angry, and had a relatively small influence when the face was neutral. With the auditory instructions ("make the judgment on the basis of what you heard the voice to be expressing"), the voice had a large influence across all three levels of facial emotion. The auditory instructions also produced somewhat more extreme judgments than the bimodal instructions in the unimodal auditory condition. In contrast, performance on the unimodal face did not differ as a function of the bimodal and auditory instructions.

Similar results were found in the speech perception task. Replicating previous research, the two independent variables influenced performance in the expected direction with both types of instructions. The influence of one variable was larger to the extent that the other variable was ambiguous. For example, the face had a larger influence when the auditory speech was at the neutral level. With the auditory instructions ("make the judgment on the basis of what you heard the voice to be saying"), the voice had a larger influence than with the bimodal instructions. In summary, both variables influence emotion and speech judgments regardless of instructions. Instructions do modulate the degree of influence, however. In order to get a proper assessment of this interaction, a process model is necessary. Not surprisingly, we use the framework of the FLMP. We evaluate the influence of instructions and any performance differences between emotion and speech tasks within the context of the FLMP, and the information versus information processing distinction. For the estimation of the free parameters corresponding to the evaluation of the auditory and visual sources, the FLMP found different parameter values in the different instruction conditions. These different parameter values were sufficient to capture the results. In terms of the distinction between information and information processing, intention can thus be accounted for in terms of information with no qualitative changes in information processing. Why would the evaluation of the sources change with instructions? Subjects might simply learn more about the attended dimension over the course of the experiment and thus show a larger influence of that dimension at evaluation. Thus changes in instructions change the cue value of the information but do not change the information process of combining the cue values to arrive at an overall judgment (4).

DISCUSSION

We found that instructions and intentional set can modulate the influence of one of the modalities in bimodal perception. Thus, attentional set seems to play a role in processing. Other evidence for a contribution of attentional set comes from a creative study, Driver (1) displaced a talking head away from a noisy auditory message. This displacement facilitated recognition of the to-be-attended message relative to presenting the video at its actual location. Participants were evidently able to view the talking head to perceptually locate the relevant message away from the location of the distracting message, and thus facilitate their processing of the to-be-attended message. Given that instructions can modulate the influence of a modality, we might also expect that the influence of a cue might be modified in other ways. Gordon, Eberhardt, and Rueckl (2) showed that contribution of auditory speech cues could be modulated by a secondary task, which involved the arithmetic processing of numbers or the perceptual evaluation of lines. The voice onset time (VOT) and the fundamental (F0) onset frequency of consonant-vowel syllables were varied. The secondary task attenuated, but did not eliminate, the influence of F0 onset frequency on the identification of voicing. This result probably reflects the output of the evaluation process in the FLMP because the competing task probably degrades the information value of a particular cue. However, it appears that the secondary task does not disrupt the integration process.

As far as we can tell, instructions and intention modulate the impact of a given source of information but cannot preclude its influence completely. Although the influence of the to-be-ignored modality is significantly decreased, there is still a substantial influence. Thus, people are not able to completely filter out the influence of a to-be-ignored modality. On the other hand, they can attenuate its influence so that some degree of control is possible. It remains to be seen to what extent increasing a subject's practice in the task or the use of some particular intentional strategy can lead to successful use of just a single source of information (3). Future research using expanded factorial designs (4) should be able to provide more definitive understanding of the role of intention in pattern recognition.

REFERENCES

1. Driver, J., *Nature*, **381**, 66-68 (1996).
2. Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G., *Cognitive Psychology*, **25**, 1-42 (1993).
3. Massaro, D. W., *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1987.pp. 75-82.
4. Massaro, D. W., *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press, 1998.