MODELING THE PERCEPTION OF BIMODAL SPEECH

Dominic W. Massaro and Michael M. Cohen
Program in Experimental Psychology, University of California-Santa Cruz

ABSTRACT

How does visible speech contribute to speech perception? Extant theories and their methodological implementations are evaluated and the process of multimodal integration is discussed. Separate operations of the Fuzzy Logical Model of Perception (FLMP) are clarified and found to be consistent with empirical phenomena. An important property of the FLMP is that multiple representations can be held in parallel. We also discuss appropriate methods for model testing.

THEORIES OF BIMODAL SPEECH PERCEPTION

The occurrence of this symposium attests to the powerful impact that visible speech has been shown to have in face-to-face communication, and the recent interest scientists have shown in the process of multimodal integration. The natural integration of several sources of information from several modalities provides a new challenge for theoretical accounts of speech perception. Although it is potentially dangerous to interpret how extant theories are impacted by the positive role of visible speech, we see a negative impact for several of them.

One class of theory seems to be either contradicted or at least placed outside the domain of bimodal speech perception. Psychoacoustic accounts [1] of speech perception are grounded in the idea that speech is nothing more than a complex auditory signal, and its processing can be understood by the psychophysics of complex sounds, without any reference to language specific processes. The strict form of this view is no longer sufficient because speech perception is not a simple function of the auditory information. In addition to the convincing findings of the influence of higher-order linguistic context in speech perception, there is the overwhelming evidence on the important influence of visible speech from a speaker's (or even an animated character's) face. It turns out that the psychoacoustic account even fails in the arena of auditory speech perception [2].

Three other theories have survived or even basked in the findings of audible-visible speech perception. The Gibsonian theory [3] states that persons perceive the cause of the sensory input directly. In spoken language, the cause of audible-visible speech is the vocal-tract activity of the talker. Accordingly, it is reasoned that visible speech should influence speech perception because it also represents the vocal-tract activity of the talker. Furthermore, by this account, vocal tract activity can be picked up directly from touching the speaker's mouth [3] which was found to influence the perceiver's interpretation of the auditory speech presented at the same time. Fowler and Dekle [3] interpret their results as evidence against the FLMP because there would be no haptic information available in the prototype descriptions. Normal perceivers supposedly have not experienced directly haptic information, nor have they experienced acoustic and haptic information about speech occurring together. Accordingly, there would be experience that would allow the development of the appropriate prototype descriptions. However, it is only natural to relate experience along one modality to experience along others.

As speakers and perceivers of language, we can easily describe what global haptic differences would be between /ba/ and /da/, for example. It should not be surprising if perceivers are influenced by haptic information when haptic and acoustic information are presented jointly in a speech identification experiment. As noted by Brunswik and demonstrated in many experiments, perceivers have difficulty selectively attending to just a single dimension of the stimulus input. Independently of the intentions of the observer, he or she tends to integrate multiple sources of information. Given the compatibility of the results with the FLMP, the results do not unambiguously support the idea that it is the events of the articulatory tract that are perceived. If an uppercase letter is drawn on a person's back, it can be recognized even though the person has never experienced this event previously. Its accurate recognition does not mean that reading letters involves direct perception of the handwriting movements that produced the letter. Rosenblum and Fowler (1991) also state that the FLMP cannot predict a contribution of visual effort on perceived loudness. They state that the model does not have loudness prototypes. However, they interpret the use of prototypes in the model much too rigidly. As stated in several venues, "prototypes are generated for the task at hand" [4, p. 17]. Our experience as perceivers of speech in face to face communication includes the positive correlation between loudness and perceived vocal effort by the talker.

The Motor theory assumes that the perceiver uses the sensory input to best determine the set of articulatory gestures that produced this input [5, 6]. One consistent theme for this theory has been the lack of a one-to-one correspondence between the auditory information and a phonetic segment. The inadequate auditory input is assessed in terms of the articulation, and it is only natural that visible speech could contribute to this process. The motor theory has not been formulated, however, to account for the vast set of empirical findings on the integration of audible and visible speech. Traditionally, the motor theory assumes that listeners analyze the acoustic signal to generate hypotheses about the articulatory gestures that were responsible for it. The outcome of the hypothesis testing is derived from the listener's speech motor system. Although the motor theory is consistent with a contribution of visible speech, it has difficulty in accounting for the strong effect of higher-order linguistic context in speech perception [4]. That is, there is nothing in this theory, grounded in modularity, that would allow context to penetrate the the "innate vocal tract synthesizer."

Remez and his colleagues [7] use the perception of sine-wave speech to argue for a view of speech perception very similar to our own.. They assume that the distal objects of perception are phonetic objects. We agree, but would replace phonetic with linguistic so as not to limit ourselves to a particular type of object, or to preclude higher-order recognition at the word or sentence level without recognition at the phonetic level. People might easily perceive a word without being aware of the phonetic segments that make it up. Remez et al assume that there is an unlimited set of cues that can be used to perceive a message. These cues have no prior grouping relationship to one another: the meaningfulness of the input binds them together. Neither the Gestalt laws of organization nor a schema-based grouping [8] can account for the perceptual grouping of these cues. Finally, somehow the appropriate sensory convergence takes place without reference to prototypes or standards in memory.

The major difference between the Remez et al. view and our view probably has to do with the role of prototypes or standards in memory. We have shown that perception occurs in the framework of one's native language [9, 10]. The same speech signal has very different consequences for speakers of different languages. It is difficult to comprehend how this could occur without a central role of memory.

INTEGRATING AUDIBLE AND VISIBLE SPEECH

More generally, many of us have grappled with the appropriate metaphor for audible-visible speech perception. A simple metaphor comes from the use of visible information in speech recognition by machine [11]. The auditory information is the workhorse of the machine, and the visible information is used in a relatively posthoc manner to decide among the best alternatives determined on the basis of the auditory information. At a psychological level, this model is similar to but differs from an auditory dominance model in which the perception is controlled by the auditory input unless it is ambiguous [9]. An ambiguous auditory input forces the system to use the visible information.

Other metaphors build on the idea of combination or integration. Somehow the visible and auditory information is combined, integrated, or joined together. The formalization of this operation is hotly debated. The two inputs are said to be fused [4], morphed, or converged [6] Fusion and morphing imply some type of blending, whereas convergence appears to be a brain metaphor describing how the different brain systems processing inputs from separate modalities converge for further processing in another brain area. The nature of this blending thus becomes a focal point for investigation and theorizing. We believe that these metaphors must be refined to specify the mathematical manner in which the different modalities are combined.

Of these major theories of speech perception, only the FLMP has provided a formal quantitative description of *how* the auditory and visual sources are processed together to determine perceptual recognition. The FLMP is well-qualified for describing the integration of audible and visible speech because it is centered around the theme of the influence of multiple sources of information. In addition, addressing the nature of the integration process cannot be adequately addressed independently of the dynamics of bimodal speech perception, and it is only the FLMP that takes a stand on the time course of audible-visible speech perception. As shown in Figure 1, the model consists of three operations: feature evaluation, feature integration, and decision. The sensory systems transduce the physical event and make available various sources of information called features. These continuously-valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness-of-match of the stimulus information with the relevant prototype descriptions.

During feature evaluation, the features of the stimulus are evaluated in terms of prototype descriptions of perceptual units of the perceiver's language. For each feature and for each prototype, feature evaluation provides information about the degree to which the feature in the signal matches the corresponding feature value of the prototype. Some investigators have argued against this early analysis of the input relative to knowledge in memory. However, it is well-documented that speech perception occurs in real time and we see no justification for some type of temporal delay in the contact of sensory input to information in memory. As an example, we have evidence that speech readers begin accumulating information of a /ba/ even before stop closure [11].

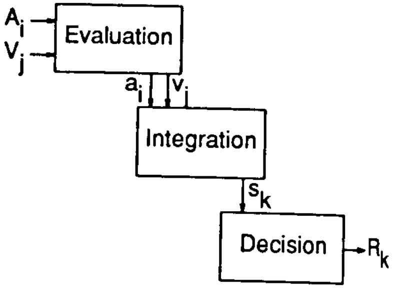

Figure 1. Schematic representation of the three processes involved in perceptual recognition. The three processes are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. These processes make use of about prototypes stored in long-term memory. The sources of information are represented by uppercase letters. Auditory information is represented by A_i *and visual information by* V_j*. The evaluation process transforms these sources of information into psychological values (indicated by lowercase letters* a_i *and* v_j*) These sources are then integrated to give an overall degree of support,* s_k*, for a given speech alternative* k*. The decision operation maps the outputs of integration into some response alternative,* R_k*. The response can take the form of a discrete decision or a rating of the degree to which the alternative is likely.*

During the second operation of the model, called feature integration, the features (actually the degrees of matches) corresponding to each prototype are combined (or conjoined in logical terms). The outcome of feature integration consists of the degree to which each prototype matches the stimulus. The third operation is decision. During this stage, the merit of each relevant prototype is evaluated relative to the sum of the merits of all relevant prototypes. This relative goodness-of-match gives the proportion of times the stimulus is identified as an instance of the prototype, or a rating judgment indicating the degree to which the stimulus matches the category. A strong prediction of the FLMP is that the contribution of one source of information to performance increases with the ambiguity of the other available sources of information.

Clarifying the FLMP

Some clarification of the FLMP is necessary because neither real processing nor predicted processing corresponds to a strict single channel discrete flow of information. The three processes shown in Figure 1 are offset to emphasize their temporal overlap. Evaluated information is passed continuously to integration while additional evaluation is taking place. Although it is logically the case that some evaluation must occur before integration can proceed, these two processes overlap in time. Similarly, integrated information is continuously made available to the decision process.

It is also necessary to emphasize that information transformed from one process to another does not obliterate the information from the earlier process. Thus, evaluation maintains its information even while simultaneously passing it forward to the integration process. This parallel storage of information does not negate the sequential process model in Figure 1. What is important to remember is that transfer of information from one process to another does not require that the information is lost from the earlier process. Integrating auditory and visual speech does not compromise or modify the information at the evaluation process. In the FLMP, the representation at one process continues to exist in unaltered form even after it has been "transformed" and transmitted to the following process. As an example, the abstract or amodal categorization of a speech signal does not replace its multimodal sensory representation. The simultaneous maintenance of several levels of information is central to the FLMP. We have shown that perceivers can report modality-specific information being maintained at feature evaluation after this same information has been combined at feature integration [11, 12].

More generally, information in the evaluation process maintains its integrity, and can be used independently of the output of integration and decision. Perceivers are not limited to only the output of integration and decision: they can also use information at the level of evaluation when appropriate. It is well-known, for example, that the relative time of arrival of audible and visible speech can greatly reduce the uncertainty about voicing [13, 14]. We know since the time of Hirsh's seminal studies [15] that perceivers are highly sensitive to temporal onset differences in the two modalities. It would not be surprising, therefore, if perceivers used this temporal asynchrony as a cue to voicing. Furthermore, the temporal asynchrony should be conceptualized as a derived cue that can be integrated with other audible and visible cues.

The FLMP predicts prototypical results of integration, as in the case in which a visual /da/ and an auditory /ba/ produces the percept /ða/. However, it is not inconsistent with a perceiver's ability to determine the temporal relationship between the auditory and visual input, as in the case when the temporal alignment of the lip movements and auditory tone pulses generated by vocal fold activity can be used as a cue to voicing. This latter phenomenon was used [13] to argue against the independence assumption of the FLMP—that the two sources of information are evaluated independently of one another. By independence, however, we simply mean that the representation of one cue at evaluation is not modified by another cue.

The degrees of support provided by the features from one modality for a given alternative are not modified by the information presented along other modalities. At the same time, the temporal relationship between two modalities might be used as an additional source of information. This comparison could therefore make available "higher-order" multimodal information indicating the temporal relationship between the audible and visible speech. This relative time of arrival could accordingly be used as a cue to voicing, which would be sent forward to the integration process. Comparisons across modalities could also provide information about the degree to which there was a phonetic discrepancy, and permit perceivers to make some other judgment such as rating the degree to which there was a discrepancy between the auditory and visual inputs [8]. The assumption of independence does not imply that there is no knowledge about what information is available from each modality, and when it is available.

Modality-Specific Representations

We have demonstrated that observers have access to modality-specific information at evaluation even after integration has occurred. This result is similar to the fact that observers can report the degree to which a syllable was presented even though they categorically labeled it as one syllable or another. A system is robust when it has multiple representations of the events in progress, and can draw on the different representations when necessary. In the Massaro and Ferguson [11] study, 20 subjects performed both a perceptual identification task and a same-different discrimination task. There were 3 levels (/ba/, neutral, /da/) of visual information and 2 levels (/ba/, /da/) of auditory. This design gives 6 unique syllables for identification, and there were 20 types of discrimination trials: 6 types of same trials, 6 types of trials with auditory different, 4 types of trials with visual different, and 4 types of trials with both auditory and visual different.

The predictions of the FLMP were derived for both tasks, and the observed results of both tasks were described with the same set of parameter values. For integration in the identification task, the degree of auditory support for the alternative /ba/ in a two-alternative forced choice task is a_i. The visual support for /ba/ is v_j, With just two alternatives /ba/ and /da/, if a visual feature supports /ba/ to degree v_j, then it supports alternative /da/ to degree $(1 - v_j)$, and similarly for the auditory feature. In this case, the overall support for alternative /ba/, S(/ba/), given audible and visible speech, is

$$S(/ba/) = a_i v_j \qquad (1)$$

The support for /da/, S(/da/) is equal to

$$S(/da/) = (1 - a_i)(1 - v_j) \qquad (2)$$

The predictions of a /ba/ judgment, P(/ba/), is equal to

$$P(/ba/) = \frac{a_i v_j}{a_i v_j + (1-a_i)(1-v_j)} \qquad (3)$$

Given the FLMP's prediction for the identification task, its prediction for a same-different task can also be derived. Faced with a same-different task, we assume that the observer evaluates the difference along both the auditory and visual modalities and responds different if a difference is perceived along either or both modalities. Thus, the task is basically a disjunction decision within the framework of fuzzy logic. The perceived difference, d_v, between two levels j and j+1 of the visual factor is given by

$$d_v = v_j - v_{j+1}. \qquad (4)$$

Analogously, the perceived difference d_a, between two levels i and i+1 of the auditory factor is

$$d_a = a_i - a_{i+1}. \qquad (5)$$

Given two bimodal speech syllables, the perceived difference, d_{va}, between them can be derived from the FLMP's assumption of a multiplicative conjunction rule, using DeMorgan's Law,

$$d_{va} = d_v + d_a - d_v d_a. \qquad (6)$$

It is also assumed that the participant computes the degree of sameness from the degree of difference, using the fuzzy logic definition of negation. In this case, the degree of sameness, s_{va}, is equal to

$$s_{va} = 1 - d_{va}. \qquad (7)$$

The participant is required to select a "same" or "different" response in the discrimination task. The actual same or different response is derived from the RGR. The probability of a different response, P(d), is thus equal to

$$P(d) = \frac{d_{va}}{d_{va} + s_{va}} = d_{va}, \qquad (8)$$

where d_{va} is given by Equation 3.

The predictions of the FLMP were fit to both the identification and discrimination tasks of each of 20 subjects. For each subject, all 26 points were fit with the same set of parameter values. The simultaneous prediction of identification and discrimination insures parameter identifiability, even when only the factorial conditions are tested. There 6 unique syllables in identification, and there were 14 types of different trials and 6 types of same trials. These 26 independent observations were predicted with just 5 free parameters, corresponding to the 3 levels of the visual factor and the 2 levels of the auditory factor. The FLMP gave a good description of the average results, with an RMSD of .0805.

An alternative model was formulated to test the idea that the auditory and visual sources are blended into a single representation, without separate access to the auditory and visual representations. The only representation that remains after a syllable is presented is the overall degree of support for the response alternatives. What is important for this model is the overall degree of support for /ba/ independently of what modalities contributed to that support. In this six-parameter model, it is possible to have two syllables made up of different auditory and visual information, but with the same degree of support for /ba/. For example, a visual /ba/ paired with an auditory /da/ might give a similar degree of overall /ba/ as a auditory /ba/ paired with an visual /da/. When formulated, this model gave a significantly ($p < .001$) larger RMSD of .1764. These model fits provide evidence that the auditory and visual sources of information are maintained independently of one another in memory, even after integration has occurred.

METHODS FOR TESTING MODELS

Grant and Walden (this volume) test the FLMP and the prelabeling model of Braida [17] against confusion matrices of individual subjects. The Prelabeling model (PM) putatively outperformed the FLMP in terms of accounting for bimodal performance as a function of unimodal performance. We believe, however, that for several reasons the test of the prelabeling model against the FLMP has been inadequate and biased, and the results of the comparison incorrect. The limitations are a) the FLMP was fit with no free parameters whereas the fit of the PM allowed many parameters to vary, b) the PM, as used by its adherents, has been biased in favor of fitting bimodal data, and c) the fit of the FLMP and PM has wrongly assumed that the unimodal data are noise free estimates of performance. We now discuss these related issues in greater detail.

In our model tests, the free parameters are adjusted to maximize the overall goodness of fit between the entire data set and a model's predictions. In the PM theorists tests of the FLMP, the bimodal performance was predicted directly from the unimodal performance. The FLMP predicts that the probability of a response to a unimodal condition is equal to the truth value supporting that response. Thus, it seems reasonable to set the free parameters in the FLMP equal to the unimodal performance levels. For example, if a participant correctly identified a visual /ba/ 85% of the time, the the visual amount of /ba/-ness given that visual stimulus would be .85. This value would be used along with the other parameters derived in the same manner to predict the bimodal performance. This would be valid test of the FLMP, however, only if the unimodal identifications are noise free measures and have very high resolution. The first requirement is certainly wrong: behavioral scientists have yet to uncover a noise free measure of performance. The second requirement is also important when confusion matrices are generated. Many cells of the confusion matrix might be 0 or 1 simply because of a relatively small number of observations per condition. Both of these factors can lead to a poor fit of the FLMP when the. unimodal probabilities are used to predict the bimodal responses. To determine the truth of a theory, we believe that it is necessary to measure how well it accounts for the entire pattern of results, rather than how well some conditions predict others. In our model tests, the optimal parameter values are used to predict the entire data set. We do not reserve this technique for tests of the FLMP but allow each competing model to do its personal best.

Braida [17] fit Erber's [18] severely impaired (SI) and profoundly deaf (PD) confusion data. In the fit, the unimodal data was first fit using a KYST technique to find stimulus and response centers in a multidimensional space based on the unimodal data. In a second stage, these centers are further adjusted to improve the fit of observed and predicted data. This space and categories were then used to predict the bimodal performance. This fit was contrasted with the fit of the FLMP when the unimodal proportions were taken as estimates of the truth values. In this case, the argument is made that comparable methods are being used to fit the PM and the FLMP. However, in addition to having adjustable parameters, the PM's predictions of the unimodal results was not optimal. A multidimensional representation was derived that gave a good fit of the bimodal results at the expense of a poor description of the unimodal data. With the Erber SI data, for example, the coordinates used by Braida yielded RMSDs of .0522 for the visual condition, .0367 for the auditory condition, and .0366 for the bimodal condition. For the PD results, the RMSDs were .0651 visual, .0756 auditory and .0400 for the bimodal. However, when the fit to the unimodal data is maximized, the RMSDs for the SI data are .0255 visual, .0288 auditory, and .0443 bimodal. For the PD data set, the RMSDs are .0299 visual, .0343 auditory and .0435 bimodal. Thus, optimizing the fit of the unimodal data decreases the accuracy of the bimodal predictions. In contrast, the FLMP yielded RMSDs of .0385 and .0509 for the SI and PD bimodal data when the parameters were fixed by the unimodal data. When the parameters were estimated from all of the results, the RMSDs dropped to .0121 and .0114.

As stated earlier, fitting the FLMP on the basis of the unimodal judgments makes the necessarily inaccurate assumption that the unimodal observations are noise free. In order to illustrate the fallibility of this method, we carried out a simulation using some previous results. In this study, participants identified auditory, visual, and bimodal syllables in which the visual syllables were either /ba/, /va/, /ða/, or /da/ and the auditory syllables fell on a ten-step continuum between these same syllables. The FLMP was fit to each subjects results and provided an excellent description of performance, with a mean RMSD of .0198. In this case, the FLMP was fit to 46 individual subjects data by estimating the free parameters using *all* of the observed results from the expanded factorial design. These predictions of the FLMP were used to generate 10 simulated subjects from each of the original 46. Rather than taking the noise-free predictions of the FLMP, each predicted point was assumed to be a value from a binomial distribution with a variance based on the number of observations (16) in the actual experiment. This new set of points now corresponded to a simulated subject. The FLMP was now fit to the simulated subjects, using Grant and Walden's method of predicting the bimodal results with the unimodal observations. Using their method, the FLMP gave a very poor description of the bimodal results of the simulated FLMP subjects. The fit of the bimodal results derived from the FLMP predictions with added noise had a mean RMSD of .0478 for the 460 pseudo-subjects. This demonstration exposes the limitations of testing models by using unimodal performance to predict bimodal results.

MOTOR REPRESENTATIONS

Robert-Ribes, Schwartz, and Escudier (this volume) advocate an amodal motor representation to account for the integration of audible and visible speech. We believe that this account suffers from many of the same problems posed for motor theories of speech perception more generally, such as accounting for the influence of higher-order linguistic context. Furthermore, it is not obvious how this model can account for the cue value of the temporal arrival of the two sources of information—a result they use against the FLMP. Some type of representation is necessary to account for the joint influence of audible and visible speech but we see not compelling reason that this representation should be a motor one.

ACKNOWLEDGMENT

The research reported in this paper and the writing of the paper were supported, in part, by grants from the Public Health Service (PHS R01 NS 20314) and the National Science Foundation (BNS 8812728).

REFERENCES

[1] Diehl, R. L., & Kluender, K. R. (1987). On the categorization of speech sounds. In S. Harnad (Ed.) *Categorical perception* (pp. 226-253). Cambridge: Cambridge University Press.

[2] Massaro, D. W. (1987). Psychophysics versus specialized processes in speech perception: An alternative perspective. In M. E. H. Schouten (Ed.) *The psychophysics of speech perception* (pp. 46-65). Amsterdam: Marinus Nijhoff Publishers.

[3] Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 816-828.

[4] Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[5] Liberman, A, Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21,* 1-33.

[6] Mattingly. I. G., & Studdert-Kennedy, M. (Eds.) (1991). *Modularity and the motor theory of speech perception.* Hillsdale, NJ: Erlbaum.

[7] Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., & Lang, J. M. (1994). On the perceptual organization of speech. *Psychological Review, 101,* 129-156.

[8] Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT press.

[9] Massaro, D. W., Cohen, M. M., & Smeele, P.M.T. (1995). Cross-linguistic comparisons in the integration of visual and auditory speech *Memory & Cognition, 23,* 113-131.

[10] Massaro, D. W., Tsuzaki, M., Cohen, M. M., Gesi, A., & Heredia, R. (1993). Bimodal speech perception: An examination across languages. *Journal of Phonetics*, 21, 445-478.

[11] Petajan, E. (1985). Automatic lipreading to enhance speech recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*, 40-47.

[12] Massaro, D. W., & Ferguson, E. L. (1993). Cognitive Style and perception: The relationship between category width and speech perception, categorization, and discrimination. *American Journal of Psychology,* 106, 25-49.

[13] Breeuwer, M. & Plomp, R. (1985) Speechreading supplemented with formant-frequency information for voiced speech. *Journal of the Acoustical Society of America, 77,* 314-317.

[14] Bernstein, L. E. (1989). Independent or dependent feature evaluation: A question of stimulus characteristics. *Behavioral and Brain Sciences, 12,* 756-757.

[15] Hirsh, I.J. (1959). Auditory perception of temporal order. *Journal of the Acoustical Society of America, 31,* 759-567.

[16] Cohen, M. M., & Massaro, D. W. (1995). Perceiving visual and auditory information in consonant-vowel and vowel syllables. In C. Sorin, J. Mariani, H. Meloni, & J. Schoentgen (Eds.) *Levels in speech communication: Relations and interactions* (pp.25-38). Amsterdam: Elsevier.

[17] Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *Quarterly Journal of Experimental Psychology, 43A,* 647-677.

[18] Erber, N. P. (1972). Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *journal of Speech and Hearing Research, 15,* 423-422.