Massaro, D. W. (2017). Multiple influences in vocabulary acquisition: Parental input dominates. Proceedings of the 17th Annual Conference of the International Speech Communication Association (Interspeech 2016), pp. 878-882. ISSN 2308-457X. Available at www.isca-speech/archive/interspeech_2016/pdfs/0037.PDFMassaro.

The McGurk Effect:

Auditory Visual Speech Perception's Piltdown Man

Dominic W. Massaro [1]

[1] Department of Psychology
University of California, Santa Cruz, CA USA

Massaro@ucsc.edu

Abstract

I draw an analogy of the McGurk effect to an episode in natural science. Piltdown Man was claimed to be the fossilized remains of a previously unknown early human. It took roughly 4 decades of controversy to conclusively learn that Piltdown Man was as a hoax because the natural scientists focused on the fossil of Piltdown Man rather than searching for other paleoanthropological evidence. I argue that the slow progress in understanding the McGurk effect is analogous because behavioral and speech scientists have not broadened their scope of inquiry much beyond the original McGurk finding. I then review a few representative examples of misguided research and theory that has resulted from this type of narrow inquiry. These include a hindering of the development of theoretical models, the belief that there are qualitative differences among individuals in terms of how they process auditory-visual speech, that different language communities process auditory-visual speech differently, and that speech is somehow special. To provide an alternative to the Piltdown Man approach, the Fuzzy Logical Model of Perception (FLMP) is briefly described to serve as a more appropriate paradigm for research and theoretical inquiry. The limitations of various neural measures are described, and when these limitations are surmounted, there appears to be some neural evidence for the independence of processing auditory and visual speech at the initial stage of speech processing.

Index Terms: McGurk effect, FLMP, speech perception, bimodal speech perception, auditory-visual speech perception, models, individual differences, cross-linguistic differences

1. Introduction

To pursue my goal of situating the McGurk effect [1] and its subsequent role in speech science more generally, I would like to draw an analogy to an episode in natural science. As is well-known, the bone fragments of a so-called Piltdown Man [2] were claimed to be the fossilized remains of a previously unknown early human. Part of a human-like skull was putatively found in the Pleistocene gravel beds near Piltdown, East Sussex. When the "discovery" was announced, many credible scientists believed that a large modern brain necessarily preceded an omnivorous diet. A prolonged four decades of controversial debate was required before the Piltdown Man was conclusively demonstrated as a hoax.

My claim is the McGurk effect has been analogous to Piltdown controversy because scientists focused on the specific phenomenon. In their defense, geologists and related scientists could have justifiably argued that such new discoveries are rare and new ones are not easily obtained at will. No such excuse is warranted for the McGurk effect. Behavioral and speech scientists could have easily broadened the scope of inquiry well beyond the original McGurk finding that an auditory labial consonant paired with a visual velar consonant sometimes produces perception of an alveolar consonant. Even without facial animation to produce intermediate consonants, they had a large family of segments to study. Even four decades after its discovery, it is discouraging to observe how many studies handicap themselves by studying only the original McGurk stimuli and procedure rather than stretching outside this narrow paradigm.

2. Hindering Model Development

Probably the most demoralizing downside to focusing on the limited McGurk effect is that the data from this limited set of conditions *underdetermines* any valid explanation [3]. So investigators, who have a theoretical bent and only study the small number of conditions in the McGurk effect, can have a field day with their theoretical interpretation. However, their favorite explanation is as weak as any other because of the small set of independent measurements that have been obtained from this limited set of McGurk conditions. As an example, Magnotti and Beauchamp [4] propose a putatively new causal inference model of multisensory speech perception. Their proposal is in fact surprisingly similar to one that I offered to explain whether or not multiple sources of information are integrated to achieve perceptual recognition [5]. I proposed that multiplicative integration, as prescribed by the FLMP, will occur only if the sources of information are perceived as "the two inputs are perceived as belonging to the same perceptual event" [5, p. 77].

In the causal inference model [4], given auditory and visual speech events, the brain first computes the likelihood that the events came from a single or multiple talkers. If the auditory and visual events are interpreted to come from the same talker, they are then combined. If not, the categorization is based on only the auditory speech. One limitation of the model is that the likelihood of inferring whether the two inputs come from a single talker is unspecified for each pair of auditory and visual inputs. Thus, the model really cannot be legitimately tested because it must assume as much or more than it is predicting.

Magnotti and Beauchamp [4] tested only the nine syllables created by the factorial combination of auditory /ba/, /da/, and

/ga/ paired with visual /ba/, /da/, and /ga/. Not only are there an inadequate number of informative conditions, 3 of the 9 conditions have congruent inputs and therefore will be handicapped by ceiling effects. Another limitation of the experiment was that the participants were constrained to respond with just one of the 3 alternatives ba, da/tha, and ga.

The investigators focused on the McGurk condition and its counterpoint condition. The auditory /ba/ combined with a visual /ga/ produced 40% da-tha judgments. In contrast, the auditory /ga/ combined with a visual /ba/ produced just 2% da-tha judgments. They explained this difference in terms of differences in causal inference in perceiving the two syllables. To describe the results quantitatively, they assumed that the causal inference predicted a single talker 51% in the standard McGurk condition and only 3% in the counterpoint condition. These assumptions about causal inference predicted the two results, given the same bimodal representation of the auditory and visual inputs. The limitation in this explanation is that as much is being assumed as is being predicted. Similarly, it was necessary to assume that the causal inference of a single talker was near 1 for the congruent syllables and much below 1 for the incongruent syllables.

The authors [4] propose a model that assumes two qualitatively different processes based on the outcome of causal inference. The critical point is that on a given trial of the sources may or may not be combined. Thus, this assumption makes the general prediction

$$P(da|A_iV_j) = pf(a_i,v_j) + (1-p)f(a_i) \quad (1)$$

where $P(da|A_iV_j)$ is the probability of responding /da/ given stimulus A_iV_j, A_i is the ith auditory stimulus, V_j is the jth visual stimulus, p is the probability inferring a single talker, and f() is some arbitrary function, a_i is the representation of A_i for /da/, and analogously for the visual input. The manner in which the auditory and visual sources are combined can be any function f(). The important conclusion is that it is predicted that there are two qualitatively types of trials, which should be noticeable in the appropriate data analysis.

If the outcome is a single talker, then the percept results from some combination of the auditory and visual inputs. If the outcome is multiple talkers, then the auditory event is perceived, with no influence from the visual input. This predicted bifurcation of behaviors depending on the outcome of causal inference could also have been tested using reaction times [see 6], because the time required for these two different operations should differ.

This formalization in Equation 1 is valuable because it makes transparent how little is being predicted relative to what is being assumed. According to [4], the value of p for each stimulus combination has not been rationalized independently and therefore must be arbitrarily chosen. Thus to describe any results, the model must assume as much or more than what is being predicted. Furthermore the model, although based on very different assumptions, is formally similarly to the class of models that have been deemed inadequate when tested against individual results from an expanded factorial design [5, pp. 49-67].

3. Universal Across Individuals

As a general principle, any behavioral question that handicaps itself to the McGurk effect will not advance the field to say the least. This is because the sparse data underdetermine any possible valid explanation. So for example questions have been raised about individual differences in the McGurk effect. Studying only congruent and incongruent McGurk conditions, however, cannot address why various individual differences have been observed. The question whether the differences are due to differences in perceiving auditory speech, differences in perceiving visual speech, or differences in the integration of the auditory and visual speech, or some combination of these differences cannot be answered.

Nath and Beauchamp [7] claim to find differential activity in the STS depending on how susceptible a perceiver is to the McGurk effect. However, they based their taxonomy of individuals on only McGurk pairs which clouds any understanding of the reason or reasons for the individual differences. For example, the results do not necessarily mirror integration ability but will necessarily depend on lipreading skill and detecting AV incongruity [8]. Nath and Beauchamp [7] categorize their 14 participants as either “perceivers” or “non-perceivers” of the McGurk effect, claiming to find two populations of subjects.

Does claiming that there are two groups of perceivers help us understand individual differences in the processing of audible and visible speech? Speech scientists are well aware of individual differences in a perceiver’s ability to understand audible speech. However, these differences are understood as quantitative rather than qualitative and can be due to hearing differences or the differential use of top-down constraints. Guided by parsimony, it is probably premature to describe visible speech perception or bimodal speech perception as qualitatively different across individuals.

We have accumulated a large database in which we can test the idea that there are qualitative differences in perceiving unimodal and bimodal speech [5]. We have posed this problem more generally in terms of whether the Fuzzy Logical Model of Perception (FLMP) is a universal description of the integration of auditory and visible speech across individuals, or whether there are qualitative differences in the integration of the two sources of information.

One method is to compare the goodness of fit of the FLMP across different subjects. As described in Massaro [5, pp. 135-139], determining a valid measure of goodness of fit is somewhat involved. A valid method is to compare the goodness of fit of the FLMP to the goodness of fit of another model, in this case the single channel model in which only a single source of information is used on any given trial. The single channel model can be considered a non-integration model so it can be considered as a description of Nath and Beauchamp’s [7] “non-perceivers”.

Figure 1 gives the RMSD value for the fit of the FLMP on the ordinate as a function on the RMSD value for the single channel model on the abscissa for each of the 82 subjects in the database. As can be seen in the figure, the advantage of the FLMP is fairly consistent across all subjects. Most importantly there is no significant gap separating two groups of subjects, putatively one group that follows the FLMP in terms of the integration of auditory and visual speech and another group does that does not follow this specific pattern.

Another important point deserves emphasis here: variability. Each of the 82 subjects was tested on a 5 by 5 expanded factorial design, with 24 observations for each of the 35 conditions for each subject [5. p. 18]. Even given this unusually large number of observations per condition, there will necessarily be significant noise or variability in the results. The investigator should not be seduced into thinking

that some observed differences are meaningful, even if they are statistically significant. This issue and various approaches to deal with it are discussed in detail in [5, Chapter 10].

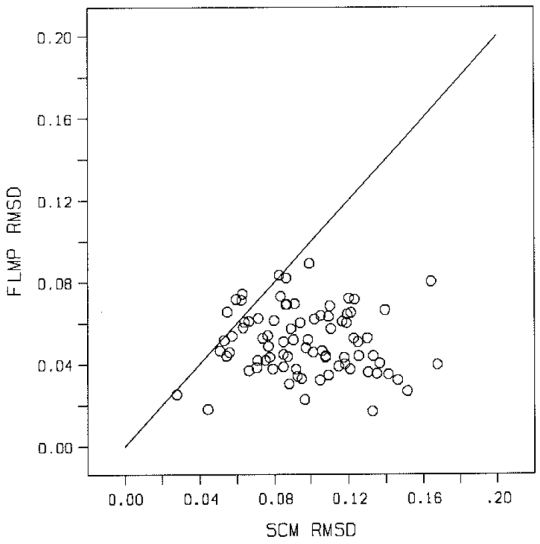

Figure 1*. The RSMD value for the fit of the FLMP on the ordinate as a function on the RMSD value for the single channel model on the abscissa for each of the 82 subjects in the database [5].*

4. Universal Across Languages

Another example of how one can be misled by focusing narrowly on the phenomenon of interest, we can look at the issue of whether there are similar processes that operate across different languages. By handicapping themselves to the phenomenon itself, the McGurk effect, investigators asked whether it occurs in other languages such as Japanese. They found very little influence of visible speech in the narrowly defined McGurk effect [9] and [10]. By limiting themselves to testing this simple "illusion" rather than being concerned with speech perception more generally, they concluded that for whatever reasons, Japanese perceivers are not as influenced by visible speech as perceivers of English. The investigators even speculated that perhaps Japanese are not as influenced by visible speech because their culture considers direct eye contact as inconsiderate or even rude. This was a very dangerous conclusion, given that languages have been proven over and over again to follow relatively universal principles, with similar processes or functions across all languages.

By broadening our telescope, we can observe that the oral deaf community in Japan flourishes in the presence of visible speech in the same manner that it does in English and other languages. Our research systematically analyzed the properties of segments that occur in Japanese versus those that occur in English. This analysis revealed very convincingly that the narrow McGurk effect would not be expected to occur in Japanese in the same manner that it occurs in English. In Japanese, the auditory labial phoneme rests in a zone of isolation since Japanese does have velar and interdental syllables. But, of course, this does not mean that perceivers are not influenced by visible speech in Japanese, as substantiated by several studies from our laboratory showing the commonalities in bimodal speech perception in Japanese and English, as well as a variety of other languages [11], [12], and [13]. The FLMP was shown to give adequate accounts of speech perception in English, Japanese, Spanish, Dutch, and Mandarin Chinese.

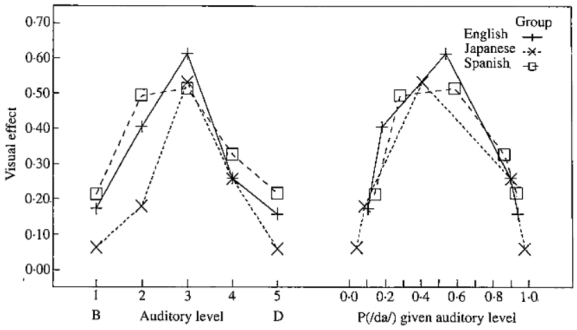

Figure 2. *The quantitative influence of visible speech across the five-step auditory continuum (left panel) and across the probability of a /da/ response (right panel) for native speakers of English, Japanese, and Spanish perceiving their native languages.*

Using results from our 5 by 5 expanded factorial design, it is possible to compute the influence of visible speech. Figure 2 gives the quantitative influence of visible speech across the five-step auditory continuum (left panel) and across the probability of a /da/ response (right panel) for native speakers of English, Japanese, and Spanish perceiving their native languages [5, p 151]. As can be seen in figure, the pattern of influence is very similar across the three languages. To further substantiate this conclusion, Chen and Massaro [13] also provide a detailed analysis and critique of a variety of studies that have claimed that the integration of audible and visible speech varies across different languages.

5. Speech is Special

Some investigators still believe that speech is special and also believe in categorical perception and motor theory. As I have said too many times, the goal of understanding language is categorization but this doesn't mean perception is categorical [6]. Perceivers easily rate the continuous degree to which one speech category has occurred versus other categories [14].

Similarly, the motor theory of speech perception might have outlived its usefulness [15]. Recently, we [16] offered a novel test of motor theory's assumption that motor processes are necessarily recruited for speech perception. We analyzed data of over 1000 children from the MacArthur-Bates Communicative Development Inventories [17] to measure individual children's understanding and production of vocabulary. If motor theory is correct, there should be a direct correspondence between the acquisition of receptive and productive vocabulary. Consistent with previous analyses of this database, the children comprehended many more words than they produced. The ease of articulation of the words acquired by individual children was measured based on the consonant segments in the word. Contrary to motor theory, ease of articulation significantly influenced production of vocabulary but much less so for comprehension.

The analyses of the vocabulary of individual children also revealed important differences between a child's productive and receptive vocabulary. As an example, one child understood 121 words and produced only 17. This child understood words with all of the 22 possible initial phonemes

but produced words with only 7 different phonemes. This child also understood words that began with several consonant clusters but only produced /vr/ in the iconic *vroom* sound. It is clear that the child was able to recognize words with various consonants even though none of the words produced contained them. The results were taken to falsify motor theory and to support the pattern recognition account of speech perception.

Notwithstanding empirical and theoretical evidence, controversies over issues such as speech is special will not be resolved sometime soon. Thus, we must be vigilant against impediments like confirmation bias [18] and safeguard appropriate inquiry [19].

6. Fuzzy Logical Model of Perception (FLMP)

There is a positive approach that can overcome the limitations of focusing on the McGurk effect. Early on [20], I advocated using factorial designs to independently manipulate multiple acoustic cues that might be functional in speech perception. The current paradigm at that time was to vary just a single cue, which weakens the ecological validity of the study [6]. This new paradigm delivered a much richer data set that sanctioned quantitative model building and testing [20]. In collaboration with Gregg Oden [21], we employed the Fuzzy Logical Model of Perception (FLMP) to account for the integration of multiple auditory cues in speech perception [21].

Three processes involved illustrated in Figure 3 include evaluation, integration, and decision. The evaluation process transforms these sources of information into psychological values indicating degrees of supports for various alternatives, which are then integrated to give an overall degree of support for each vocabulary alternative. The decision operation maps the outputs of this integration into some appropriate alternative.

The assumptions central to the model are: (a) each source of information is evaluated to determine the continuous degree to which that source specifies various alternatives; (b) the sources of information are evaluated independently of one another; (c) the sources are integrated to provide an overall continuous degree of support for each alternative; and (d) perceptual identification and interpretation follows the relative degree of support among the various alternatives. Given multiple sources of information, it is useful to have a common metric representing the degree of match of each feature. To serve this purpose, fuzzy-truth values [6] are used because they provide a natural representation of the degree of match.

Figure 3 also illustrates how learning is conceptualized within the model by specifying exactly how the feature values used at evaluation change with experience. Learning in the FLMP is described by the following algorithm [5]. The initial feature value representing the support for an alternative is initially set to .5 (.5 is neutral in fuzzy logic). Given some speech with feedback, the prototypes would be updated appropriately.

This model has been successful in a wide variety of situations involving bimodal auditory-visual perception. It captures the outcomes of bimodal speech perception across the lifespan from age 3 to 83 [22]. The developmental trajectory we measured and modeled [23] has been recently shown to be consistent with domain-general statistical learning [24], which adds support for the ability of the FLMP to capture how audible and visible speech perception develop during language learning.

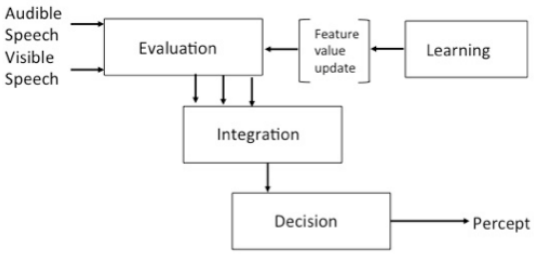

Figure 3. *Schematic representation of the three processes involved in understanding or producing a speech utterance. The three processes are shown to proceed left to right in time to illustrate their necessarily successive but overlapping processing. These processes make use of prototypes stored in long-term memory. The two sources of information are audible and visible speech. The evaluation process transforms these sources of information into psychological values. These sources are then integrated to give an overall degree of support, for each alternative. The decision operation maps the outputs of integration into some favored alternative. Learning from a speech event provides updates of the appropriate feature values.*

The model has been equally successful in describing the use of visible and audible speech in adults with hearing aids and with cochlear implants [25]. The FLMP has also given adequate descriptions of integrating a variety of different sources of language information [26] and [27]. For example, the FLMP gave a good account of how young children are capable of integrating auditory, visual, and gestural information in determining a word's referent [28]. Supporting the principle of a universal algorithm for processing multiple sources of information, the FLMP has been shown to describe a variety of non-speech domains. As an example, the FLMP described how size, height in the picture plane, occlusion, and motion parallax are used together to perceive relative depth [29]. A sampling of other domains includes emotion from the face and the voice [30]; visual perception of faces [31]; letter and word recognition in reading [32], and implicit and explicit memory [33].

Given that language perceivers so easily integrate multiple sources of information in language processing, it seems only natural to determine whether language perceivers are capable of learning new sources of information and actively using them during speech perception [34]. Utilizing new sources of information would be particularly valuable for perceivers that are limited in the sources of information they have available, such as the deaf and hard of hearing. We know that this population benefits immensely from watching the face during language presentations, and we asked whether they could utilize other visual information. We used Cued Speech as proof of concept to research this question. It consists of hand gestures while speaking to provide the perceiver with information that disambiguates the ambiguity of linguistic cues seen on the face. Analogously, we designed iGlasses, an automated wearable computer, to supplement face-to-face speech with added visible information.

We tested whether people can combine or integrate information from the face and information from newly learned cues in an optimal manner [35]. Subjects first learned the visual cues and then were tested just the face, just the visual cues, or both together. Performance was much better with both cues than with either one alone. Similar to the description

of previous results with audible and visible speech, the present results were well described by the FLMP. We also found, however, that the prolonged periods were required to learn the cues and to use them automatically. This led me to propose that learning must occur early in life, and early experience with written text could enable learning to read naturally [36].

Thus, we and other investigators [37] have found the FLMP to be a universal principle of perceptual cognitive performance that accurately simulates human pattern recognition [4] and [5]. People are influenced by multiple sources of information in a diverse set of situations. In most every case, these sources of information are ambiguous and any particular source alone does not usually specify completely the appropriate interpretation. These cues and constraints are graded (not categorical), suggesting further that they must be combined to give a more reliable understanding of the input. Evidence to date indicates that this combination process is highly efficient or optimal, as described by a Bayesian-like process [38] and [39].

6.1. Neural Evidence for Evaluation Independence

It has been at least two decades since brain recordings held the promise of understanding the integration of auditory and visual speech [40]. It would be an easy answer to the question of how quickly the two sources merge into some combined representation in speech perception. Theoretical interpretation can be pigeon-holed into early and late merges. Motor theory or analogously gesture theory would assume that audible and visible speech would interact early in the processing chain. The FLMP, on the other hand, assumes an initial evaluation stage during which the two sources remain independent of one another [5] and [6].

Of course, the nature of brain measures must be sensitive to the processing questions being addressed. Functional magnetic resonance imaging (fMRI) studies are grounded in metabolic changes that are driven by neural activity but these changes only occur well after neural activity has occurred. Thus, the relatively slow time course of the blood-oxygen level dependent (BOLD) response cannot address the early time course of bimodal speech perception [e.g., 41].

Electrophysiology measures such as EEG and MEG, on the other hand, are more temporally tied to neural activity and therefore could inform us the time course of the activation of specific neural structures. Moreover, Non-invasive electrophysiology (EEG, MEG) offers more detail about the time course of visual interaction with auditory information. Unfortunately, these measures are lacking in spatial resolution of the neural structures. Electrocorticography ECoG can overcome these problems by combining high temporal precision with increased spatial resolution.

Rhone et al. [42] capitalized on the opportunity of testing awake, behaving humans who were undergoing chronic intracranial monitoring as part of pre-surgical evaluation for treatment of medically intractable epilepsy. These neurosurgical patients were implanted with multi-contact depth electrodes and subdural grid arrays that allowed for simultaneous recordings from primary, non-primary auditory and frontal cortex. Thus, measures using (ECoG) would have both high temporal resolution and high spatial resolution for the brain areas with electrodes.

The patients were presented with unimodal speech and nonspeech in both auditory and visual modalities. The results are fairly involved and I can only summarize them here. Although visual input activated primary auditory areas, there was "little speech-specific activation" [42, p. 294). Most importantly, the auditory and visual speech influenced activation in the superior temporal gyrus relative to non-speech. This influence was not observed in the auditory cortex at Heschl's Gyrus.

The study was also successful in locating the influence of visible speech at the Precentral gyrus. This activity occurred both before and after the auditory input and was much greater for speech than nonspeech. Although the authors bravely acknowledge the limited stimulus set that was used, they conclude that "we did not find strong auditory effects in primary motor cortex (PreC). Instead, only non-primary auditory areas on the STG were sensitive to both factors, with meaningful visual speech content showing distinct advantage (high gamma increase and beta suppression). This is consistent with an integration model in which visual and auditory information are transduced independently and combined at higher levels of processing…" [42, p. 299].

7. Discussion

Will the McGurk effect obtain the notoriety of Piltdown Man for similar reasons? With thousands of publications since its discovery, it has occupied many individuals in a variety of disciplines. Roughly two and a half centuries ago, Benjamin Franklin disclosed the value of visible speech during his diplomatic service in France when he was conversing in a non-native language. (He must have gained a fondness of lip rounding given its prominence in French). Serendipity led to the modern discovery of visible speech but its study was hindered by the illusion it could create (everybody loves an illusion). It is true that the technology required to animate visible speech was not easily available but the same was true for audible speech during its initial investigations. But on the other hand, a few investigation showed that even outline drawings of lip motions could influence speech perception.

My goal for this retrospective is to encourage the field to move beyond the McGurk effect and to study speech perception given multiple sources of potential information. We are daily observing the value of big data, and there is no reason why the same cannot be true for our field. I am looking forward to a data warehouse of results of audible and visible speech perception across a wide variety of speech segments, individuals, languages, and most importantly under an extended set of experimental conditions.

8. Acknowledgements

The author acknowledges the helpful discussions with Bill Rowe for many discussions and his suggestion of Piltdown Man as an example of a natural science dead end, and to Grace Shefcik, Mike Beauchamp, Tobias Anderson, and an anonymous reviewer for comments on the paper.

9. References

[1] H. McGurk, and J. MacDonald, (1976). Hearing lips and seeing voices. Nature, 264P, 746-748.

[2] https://en.wikipedia.org/wiki/Piltdown_Man

[3] D. W. Massaro (1998). Illusions and Issues in Bimodal Speech Perception. Proceedings of AuditoryVisual Speech Perception '98. (pp. 21-26). Terrigal-Sydney Australia, December, 1998.

[4] J. F. Magnotti and M. S. Beauchamp (2017). A Causal Inference Model Explains Perception of the McGurk Effect and Other Incongruent Audiovisual Speech. PLoS Comput Biol 13(2): e1005229. doi:10.1371/journal.pcbi.1005229

[5] D. W. Massaro (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, MA: MIT Press.

[6] D. W. Massaro (1987). Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Hillsdale, NJ: Erlbaum Associates.

[7] A. R Nath and M. S. Beauchamp (2012). A Neural Basis for Interindividual Differences in the McGurk Effect, a Multisensory Speech Illusion. NeuroImage, 59(1), 781–787. http://doi.org/10.1016/j.neuroimage.2011.07.024

[8] J. Strand, J., A. Cooperman, J. Rowe, and A. Simenstad (2014) Individual differences in susceptibility to the McGurk effect: Links with lipreading and detecting audiovisual incongruity. Journal of Speech, Language, & Hearing Research, 57, 2322-31. doi: 10.1044/2014_JSLHR-H-14-0059

[9] K. Sekiyama (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. Perception & Psychophysics, 59, 73-80.

[10] K. Sekiyama and Y. Tohkura (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. Journal of the Acoustical Society of America, 90, 1797-1805.

[11] D. W. Massaro M.M Cohen, A. Gesi, A., and R. Heredia, (1993). Bimodal Speech Perception: An Examination across Languages. Journal of Phonetics, 21, 445-478

[12] D. W. Massaro, M. M. Cohen, and P.M.T. Smeele (1995). Cross-linguistic Comparisons in the Integration of Visual and Auditory Speech. Memory and Cognition, 23, 113-131.

[13] T. H. Chen and D. W. Massaro (2004). Mandarin speech perception by ear and eye follows a universal principle. Perception,& Psychophysics, 66(5), 820-836.

[14] D. W. Massaro and M. M. Cohen (1983). Categorical or continuous speech perception: A new test. Speech Communication, 2, 15-35. D. W. Massaro and T. H. Chen, (2008). The Motor Theory of Speech Perception Revisited. Psychonomic Bulletin & Review (pp. 453-457), Vol. 15(2).

[15] D. W. Massaro and T. H. Chen (2008). The Motor Theory of Speech Perception Revisited. Psychonomic Bulletin & Review (pp. 453-457), Vol. 15(2).

[16] D. W. Massaro and B. Rowe (2015). Comprehension outscores production in language acquisition: Implications for Theories of Vocabulary Learning. Journal of Child Language Acquisition and Development – JCLAD, Vol: 3 Issue: 3 121-152, 2015, September ISSN: 2148-1997

[17] V. A. Marchman, , P. S. Dale, , J. S. Reznick, D. Thal, and E. Bates, (2007). MacArthur-Bates Communicative Development Inventories: User's Guide and Technical Manual. 2nd Ed.. Baltimore, MD: Brookes Publishing Company.

[18] https://en.wikipedia.org/wiki/Confirmation_bias

[19] D. W. Massaro (2012). A Quarter Century of Book Reviews in The American Journal of Psychology (pp. 499-500). DOI: 10.5406/amerjpsyc.125.4.0499. Stable URL: http://www.jstor.org/stable/10.5406/amerjpsyc.125.4.0499

[20] D. W. Massaro and M. M. Cohen (1976). The Contribution of Fundamental Frequency and Voice Onset Times to the /zi/-/si/ Distinction. Journal of the Acoustical Society of America, 60, 704-717.

[21] G. C. Oden, and D. W. Massaro (1978). Integration of Featural Information in Speech Perception. Psychological Review, 85, 172-191.

[22] D. W. Massaro, L.A. Thompson, B. Barron, and E. Laren (1986). Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology, 41, 93-113. PDF

[23] D. W. Massaro (1994). Bimodal speech perception across the lifespan. In D.J. Lewkowicz and R. Lickliter (Eds.), The Development of Intersensory Perception: Comparative Perspectives (pp. 371-399). Hillsdale, NJ: Lawrence Erlbaum.

[24] L. M Getz, E. R Nordeen, S. C. Vrabic, and J. C. Toscano (2017). Modeling the Development of Audiovisual Cue Integration in Speech Perception. Brain Sciences 7(3):32 · March 2017. DOI: 10.3390/brainsci7030032

[25] D. W. Massaro and M. M. Cohen (1999). Speech perception in hearing-impaired perceivers: Synergy of multiple modalities. Journal of Speech, Language, and Hearing Research, 42, 21-41.

[26] R. J. Srinivasan, R. J and D. W. Massaro (2003). Perceiving prosody from the face and voice: Distinguishing statements from echoic questions in English. Language and Speech, 46, 1-22.

[27] D. W. Massaro (1998). Models for reading letters and words. In S. Sternberg,& D. Scarborough (Eds.), Invitation to Cognitive Science, 4 (pp. 301-364). Cambridge, MA: MIT Press.

[28] L. A. Thompson, and D. W. Massaro (1986). Evaluation and integration of speech and pointing gestures during referential understanding. Journal of Experimental Child Psychology, 42, 144-168.

[29] (1988). Ambiguity in perception and experimentation. Journal of Experimental Psychology: General, 117, 417-421.

[30] D. W. Massaro and P. B. Egan, (1996). Perceiving Affect from the Voice and the Face. Psychonomic Bulletin and Review, 3, 215-221.

[31] D. W. Massaro and G. Schwarzer (2001). Modeling face identification and processing in children and adults. Journal of Experimental Child Psychology 79, 139-161.

[32] D. W. Massaro and A. Jesse (2005). The Magic of Reading: Too Many Influences for Quick and Easy Explanations. In T. Trabasso, J. Sabatini, D.W. Massaro,& R.C. Calfee (Eds.), From orthography to pedagogy: Essays in honor of Richard L. Venezky (pp.37-61). Mahwah, NJ: Lawrence Erlbaum.

[33] Weldon, M.S.,& Massaro, D.W. (1996). Integration of Orthographic, Conceptual, and Episodic Information on Implicit and Explicit Tests. Canadian Journal of Experimental Psychology, 50, 72-85.

[34] Massaro, D.W., Carreira-Perpinan, M.A. & Merrill, D.J. (2009). iGlasses: An Automatic Wearable Speech Supplement in Face-to-Face Communication and Classroom Situations. In LaSasso, C., Leybaert, J. & Crain, K. (Eds.) Cued Speech and Cued Language Development of Deaf and Hard of Hearing Children (Chapter 20). Plural Publishing Inc.

[35] D. W. Massaro, M. M. Cohen, H. Meyer, T. Stribling, C. Sterling, and S. Vanderhyden, (2011) Integration of Facial and Newly Learned Visual Cues in Speech Perception. American Journal of Psychology, 124, 341-354.

[36] Massaro, D. W. (2012). Acquiring Literacy Naturally: Behavioral science and technology could empower preschool children to learn to read naturally without instruction, American Scientist, 100, 324-333.

[37] Movellan and J. L. McClelland, (2001). The Morton–Massaro law of information integration: Implications for models of perception. Psychological Review, 108, 113-148.

[38] D. W. Massaro (1992). Broadening the domain of the fuzzy logical model of perception. In H.L. Pick, Jr., P. Van den Broek,& D.C. Knill (Eds.), Cognition: Conceptual and Methodological Issues (pp.51-84). Washington, DC: American Psychological Association

[39] D. W. Massaro and D. G. Stork (1998). Sensory integration and speechreading by humans and machines. American Scientist, 86, 236-244.

[40] Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K.,…David, A. S. (1997). Activation of auditory cortex during silent lipreading. Science, 276(5312), 593–596. doi:10.1126/science.276.5312.593

[41] Bernstein, L. E., & Liebenthal, E. (2014). Neural pathways for visual speech perception. Frontiers in Neuroscience, 8. doi:10.3389/fnins.2014.00386

[42] A. E. Rhone, K. V. Nourski, H. Oya, H. Kawasaki, M. A. Howard III and B. McMurray (2016). Can you hear me yet? An intracranial investigation of speech and non-speech audiovisual interactions in human cortex, Language, Cognition and 1Neuroscience, 31:2, 284-302, DOI: 10.1080/23273798.2015.1101145J.1