Data Wareehouse for Speech Perception and Model Testing

Dominic W. Massaro and Michael M. Cohen

University of California, Santa Cruz

Santa Cruz, CA. 95064 U.S.A.

Send correspondence to:

Dominic W. Massaro

Department of Psychology

University of California

Santa Cruz, CA. 95064 USA

Work: 831-459-2330

Fax: 831-459-3519

E-mail: massaro@fuzzy.ucsc.edu

URL: http://mambo.ucsc.edu/psl/dwm/

Submitted to Behavioral Research Methods, Instruments, & Computers for the Special Issue on Archiving

ABSTRACT

Theories of speech perception, like most theories, have tended to be qualitative rather than quantitative. Progress in the field requires sufficiently detailed experiments and results to allow the development and testing of quantitative models. To meet this goal, we are establishing a data warehouse to provide a comprehensive but user-friendly database from speech perception experiments. The experiments involve several independent variables that are orthogonally varied in factorial or expanded factorial designs. In addition to the database, we will provide an easy-to-use application that will allow investigators to formulate and test different models in a variety of testing procedures. The goal of the data warehouse is to initiate the accumulation and dissemination of experimental results and formal model testing that will advance our understanding of speech perception and its relationship to other forms of pattern recognition.

Most extant theories of speech perception are stated in verbal rather than quantitative form. Although no one can deny that a qualitative fact is more informative than a quantitative one, qualitative theories are not usually sufficiently precise to be distinguished from one another. Another limitation of qualitative theories is that very different theories, when formalized, can be found to make very similar predictions. Formalization of theoretical ideas is an important step in the development of a scientific discipline. In addition, formalization also leads to insights about the nature of the experimental data required to test and distinguish among theoretical alternatives. Given that some quantitative refinement of the theories is usually necessary to create a chance for falsification and strong inference (Platt, 1964; Popper, 1959), an essential strategy for research is to quantify and test a family of specific models that represent the extant theories and also other reasonable alternatives (Massaro, 1987, 1998b).

Our goal is to provide a comprehensive but user-friendly warehouse of data from speech perception experiments in order to make the results available to the scientific community and to stimulate model development and testing. The database will include a broad range of experimental conditions, results from several languages, performance at several developmental and life-span periods, individuals who are hard of hearing, and conditions that include several modalities (auditory speech, visual speech, and gesture) and contextual variables (phonological, lexical, and sentential constraints. This warehouse will provide results that must be accounted for by any viable theory of speech perception and should make possible both qualitative and quantitative tests of models of speech perception.

In addition to the data warehouse, we will provide an easy-to-use application that will allow investigators to formulate and test different models against a broad range of results. This goal should also facilitate the development of new quantitative models of speech perception and their tests against robust and meaningful experimental results. This enterprise should advance the field significantly and help achieve a better understanding of how persons so easily understand one another by using many different sources of information present during the communicative exchange.

Research has also shown that model testing requires a fairly elaborate set of experimental conditions, with several independent variables, precise control and manipulation of the experimental stimuli, and a large number of observations from each participant (Massaro, 1998a, 1998b). Experiments with just a few conditions underdetermine a unique explanation. Analyses and tests of individual participant's results are necessary because data averaged across subjects might distort the actual individual outcomes. For example, averaging a set of different nonlinear outcomes will tend to make the average results more linear (Massaro & Cohen, 1993).

Notwithstanding the caveat of averaging results across individuals, some of the seminal databases are average data for a specific group (such as those who are hard of hearing) because these data have proven to be informative, and should lead to new research in which data from individual participants can be analyzed.

To further substantiate the model testing, Bayesian selection techniques as well as RMSD goodness-of-fit criteria are used in the evaluation of extant models.

We now discuss one domain of inquiry that will be in the data warehouse in order to illustrate the logic of the experimental paradigm and the model development and testing.

Auditory/Visual Speech Perception

Some of the research in the data warehouse 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 utilizes a strong-inference strategy of hypothesis testing, independent manipulations of multiple sources of information, and the testing of mathematical models against the results of individual participants. Synthetic speech allows 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 to provide the most powerful test of quantitative models of perceptual recognition (Massaro, 1998). Expanded factorial designs are used to study how auditory speech and visual speech are processed alone and in combination, and under different degrees of ambiguity. Experiments have clarified the classic McGurk effect, assessed the contribution of segment frequency in the language, and the psychophysical properties of the auditory and visual speech. Experiments have also contrasted the influence of visible speech with the influence from written text. The results can be used to address how these two sources of information are integrated with auditory speech.

The Data Warehouse

Table 1 gives a summary description of the planned databases to be included in the resource sharing application hosted by the Perceptual Science Laboratory. Each set of results will be linked to an electronic version of the original experimental report so that the details will be readily available. As can be seen in Table 1, the proposed database will include a wide variety of experimental results from several different laboratories and time periods. The database will include a broad range of experimental conditions, results from several languages, performance at several developmental and life-span periods, individuals who are hard of hearing, and conditions that include several modalities (auditory speech, visual speech, and gesture) and contextual variables (phonological, lexical, and sentential constraints). We will format the databases so that they can be easily accessed and analyzed using common tools such as Excel, SPSS, and MatLab. As an example of the nature of the data, we describe a typical experiment involving the independent variation of auditory and visual speech in an expanded factorial design.

Typical Experiment: Varying the Ambiguity of the Speech Modalities

An informative manipulation in speech perception research is to systematically vary the ambiguity of each of the source of information in terms of how much it resembles each syllable.

Synthetic speech (or at least a systematic modification of natural speech) is necessary to implement this manipulation. In several experiments on bimodal speech perception, we used synthetic speech to cross five levels of audible speech varying between /ba/ and /da/ with five levels of visible speech varying between the same alternatives. We also included the unimodal test stimuli to implement the expanded factorial design, as shown in Figure 1.

Prototypical Method. The properties of the auditory stimulus were varied to give an auditory continuum between the syllables /ba/ and /da/. In analogous fashion, properties of our animated face were varied to give a continuum between visual /ba/ and /da/. Five levels of audible speech varying between /ba/ and /da/ were crossed with five levels of visible speech varying between the same alternatives. In addition, the audible and visible speech also were presented alone for a total of 25 + 5 + 5 = 35 independent stimulus conditions. Six random sequences were determined by sampling the 35 conditions without replacement giving six different blocks of 35 trials. An experimental session consisted of these 6 blocks preceded by 6 practice trials and with a short break between sessions. There were 4 sessions of testing for a total of 840 test trials (35 x 6 x 4). Thus there were 24 observations at each of the 35 unique experimental conditions. Participants were instructed to listen and to watch the speaker, and to identify the syllable as /ba/ or /da/. This experimental design was used with 82 participants and their results have served as a database for testing models of pattern recognition (Massaro, 1998b).

Typical Results. A critical feature of our database will be to archive individual participant results, Average results across individuals can distort the underlying pattern given by each individual (Massaro & Cohen, 1993; Massaro, 1998). We call these results typical because they are highly representative of many different experiments of this type. The mean observed proportion of /da/ identifications was computed for each of the 82 participants for the 35 unimodal and bimodal conditions. Figure 2 gives the results for a single participant who can be considered typical of the others in this task.

The points in Figure 2 give the observed proportion of /da/ responses for the auditory alone, the bimodal, and the visual alone conditions as a function of the five levels of the synthetic auditory and visual speech varying between /ba/ and /da/. Notice that the columns of points are spread unevenly along the x-axis. The reason is that they are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. This spacing reflects relative influence of adjacent levels of the auditory condition.

The unimodal auditory curve (indicated by the solid circles) shows that the auditory speech had a large influence on the judgments. More generally, the degree of influence of this modality when presented alone would be indicated by the steepness of the response function. The unimodal visual condition is plotted at .5 (which is considered to be completely neutral) on the auditory scale. The influence of the visual speech when presented alone is indexed by the vertical spread among the five levels of the visual condition.

The other points give performance for the bimodal conditions. This graphical analysis shows that both the auditory and the visual sources of information had a strong impact on the identification judgments. The likelihood of a /da/ identification increased as the auditory speech changed from /ba/ to /da/, and analogously for the visible speech. The curves across changes in the auditory variable are relatively steep and also spread out from on another with changes in the visual variable. By these criteria, both sources had a large influence in the bimodal conditions.

Finally, the auditory and visual effects were not additive in the bimodal condition, as demonstrated by a significant auditory-visual interaction. The interaction is indexed by the change in the spread among the curves across changes in the auditory variable. This vertical spread between the curves is about four times greater in the middle than at the end of the auditory continuum. It means that the influence of one source of information is greatest when the other source is neutral or ambiguous. We now address how the two sources of information are used in perception.

Evaluation of How Two Sources are Used.

Of course, an important question is how the two sources of information are used in perceptual recognition. An analysis of several results informs this question. Figure 3 gives the results for another participant in the task. Three points are circled in the figure to highlight the conditions in which the second level of auditory information is paired with the fifth (/da/) level of visual information. When presented alone, P(/da/ | A_2) is about .25 whereas P(/da/| V_5) is about .8. When these two stimuli occur together, P(/da/| A_2 V_5) is about .6. This subset of results is consistent with just about any theoretical explanation; for example, one in which only a single source of information is used on a given trial. Similarly, a simple averaging of the audible and visible speech predicts this outcome.

Other observations, however, allow us to reject these alternatives. Figure 4 gives the results for yet another participant in the task. Three points are circled in the figure to highlight the conditions in which the second level of auditory information is paired with the second level of visual information. Recall that in this forced-choice task, P(/ba/) is equal to one minus P(/da/). When presented alone, P(/ba/ | A_3) and P(/ba/| V_1) are both about .75. When these two stimuli occur together, P(/ba/| A_3 V_1) is about 9. This so-called super-additive result (the bimodal is more extreme than either unimodal response proportion) does not seem to be easily explained by either the use of a single modality or a simple averaging of the two sources. In order to evaluate theoretical alternatives, however, formal models must be proposed and tested against all of the results, not just selected conditions. The database will facilitate the formalization of competing models, which can be systematically tested against the results. We now turn to some important considerations in model development and testing.

Testing a Model's Predictions

One of the reasons that model testing is relatively rare in the speech perception field is that there are not easily accessible and usable techniques that the non-expert can use. We propose to remedy this situation by providing a user-friendly application that will allow investigators to test existing models against results in the library, formalize new models to test against these data, record a new model's description of existing results, or add new results and models to the library. We will also allow different techniques for model testing, as well as providing a variety of measures of goodness of fit. We briefly consider each of these aspects of the data warehouse.

There will be a simple user-friendly software interface to describe the model, and to implement its test against a particular database. We will ask students, colleagues, and others to use the interface and to give usability feedback to improve its functionality and friendliness. A current version of the model testing program available on the web is at http://mambo.ucsc.edu/psl/stepit.html, and with sample results at http://mambo.ucsc.edu/psl/Training/.

A new model or a new database can be submitted to the library in a straightforward manner. We will provide detailed formatting instructions for the submission of new models and databases. Investigators who submit results or models will be acknowledged for their contributions.

Different techniques for model testing will also be provided. One available technique that is possible is a parameter-free test. In this technique, the results of a subset of the experimental conditions are used to predict the other conditions (Braida, 1991; Massaro, 1998b). Although this technique cannot be expected to optimize the goodness-of-fit of a model (Massaro, 1998), it is a legitimate technique as long as the resulting goodness-of-fit is appropriately evaluated. A benchmark goodness-of-fit provides such an evaluation metric (Massaro, 1998b, chapter 10).

In most cases, however, parameter-free tests are not possible. Because of the many sources of variability inherent in experimental testing, we cannot expect a model's predictions of behavior to be very accurate without first taking into account what results are being predicted. As an example, we cannot know exactly how often a given person will identify one of the visible speech syllables as a particular alternative. Individual participants give similar but not identical results for the same experiment. We can know that one syllable might be more likely to be identified as /ba/ but we cannot predict ahead of time the actual probability of a /ba/ response by an individual participant. This uncertainty would preclude the quantitative test of models if we were not able to determine (estimate) the values of free parameters.

When applied to empirical data, most computational or quantitative descriptions have a set of free parameters. A free parameter in a model is a variable whose values cannot be exactly predicted in advance. We do not know what these values are, and we must use the observed results given to find them. The actual performance of the participant is used to set the value of this variable. This process is called parameter estimation. In parameter estimation, actual observations of behavior are used to estimate the values of the free parameters of the model being tested. Because we want to give every model its best shot, the goal is to find the values of the parameters that maximize how accurately the model is able to account for the results. The optimal parameter values can be found with an iterative search algorithm to find those parameter values that minimize the differences between the predicted and observed results. The parameters and parameter space must be specified for the search. In our model fitting technique, we usually estimate the free parameters based on all of the conditions, not just a subset. We have rationalized why this approach is more optimal (see also Schwartz, 2003).

RMSD Measure of Goodness-of-Fit

A factor that is often used to maximize the goodness-of-fit is the root mean squared deviation (RMSD) between the predicted and observed values. The best fit is one that gives the minimal RMSD. The RMSD is computed by a) squaring the difference between each predicted and observed value, b) summing across all conditions c) taking the mean, and d) taking the square root of this mean. (Squaring the differences makes all differences positive and also magnifies large deviations compared to small ones.) The RMSD can be thought of as a standard deviation of the differences between the predicted and observed values. The RMSD increases as these differences increase. In general, the smaller the RMSD value, the better the fit of the model.

The quantitative predictions of a model are determined by using any mimimization routine such as the program STEPIT (Massaro, 1998b). In STEPIT, a model is represented to the program in terms of a set of prediction equations and a set of unknown parameters. By iteratively adjusting the parameters of the model, the program maximizes the accuracy of the predictions by minimizing the RMSD. The outcome is a set of parameter values which, when put into the model, come closest to predicting the observed results. The RMSD is used to evaluate the goodness-of-fit of a model both in absolute terms and in comparison to other models.

An important consideration is whether a set of results is actually possible to distinguish between the predictions of two or more different models. One way to assess whether your data set is valid in discriminating two models is to 1) fit the models to the results, 2) cross-fit each model to simulated data generated from the predictions of the other model, 3) and find whether the two models are equally good at fitting their own simulated (predictive) data and equally poor at fitting the simulated (predictive) data from the other model (see Massaro & Friedman, 1990). If this is the case, then the data are discriminating. If a model can predict another model's predictions for a particular set of results, then the data set is not sufficiently detailed to differentiate between the models. If a model consistently predicts the predictions of other models across a broad range of results, then that model appears to be nonfalsifiable and probably not scientifically worthy (Massaro, 1988).

Given the delicate nature of testing among quantitative modes, we have explored alternative methods of model testing (Massaro, 1998, Chapter 10). The first involves the match between the goodness-of-fit of a model and a benchmark measure that indexes what the goodness of fit should be if indeed the model was correct. Because of sampling variability, we cannot expect a model to give a perfect description of the results, and the benchmark provides an absolute index for the observed RMSD. Second we have used a model selection procedure suggested by Myung and Pitt (1997; 1998; Massaro et al., 2001), and by Schwartz (2003), which we now describe.

Model Selection using Bayes Factor

This Bayes factor method of model selection seeks to handicap models to the extent they can predict a large range of outcomes with changes in their parameter values. One model might predict a large range of outcomes with changes in parameter values, whereas another model might predict only a small range of outcomes across changes in its parameter values. If these two models give an equally good description of an observed data set, then the second more constrained model should be preferred based on parsimony. This observation leads to the idea of handicapping models based on their flexibility in predicting a large range of outcomes.

The Bayes factor adjusts a model's goodness-of-fit index by the model's ability to describe a large range of different data configurations. One model capable of fitting a broader range of data configurations than another is not necessarily the better model. We desire a model to predict only a constrained set of data outcomes because if any configuration of data can be predicted, it is not falsifiable (Massaro, 1998). The Bayes factor handicaps a model to the extent that it can predict a broad range of data configurations other than the observed data, by simply assuming different parameter values. According to the assumptions underlying Bayes factor, a better model is one that predicts only data close to the data actually observed, regardless of the parameter values.

The data warehouse will therefore have several methods of selecting among models. It should be noted that as in all things, however, there is no holy grail of model evaluation for scientific inquiry. We offer several techniques for model testing to allow the researcher to provide converging evidence for the selection of one model over another. As an example, both RMSD and the Bayes factor can be used as independent metrics of model selection. Inconsistent outcomes should provide a strong caveat for the validity of selecting one model over another in the same way that conflicting sources of information create an ambiguous speech event for the perceiver.

Statistical Tests of the Models

The goodness-of-fit measures from different models can be evaluated using analysis of variance. When model fits are carried out on each participant's results individually, the goodness-of-fit measure can be used as the dependent variable and the different models as the independent variable.

Protection of Participants' Privacy

All of the previous publications of these databases safeguarded the privacy of participants, and this safeguarding will be maintained in the sharing of the databases.

References

Agelfors, E. (1996). A comparison between patients using cochlear implants and hearing aids. Part I: Results on speech tests. Royal Institute of Technology, Speech, Music and Hearing. *KTH Quarterly Progress and Status Report (TMH-APSR 1), 63-75.*

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

Campbell, R., Zihl, J., Massaro, D.W., Munhall, K. & Cohen, M.M. (1997). Speechreading in the Akinetopsic Patient LM, *Brain. 120*, 1793-1803.

Cathiard, M.-A., Schwartz, J.-L., & Abry, C. (2001). Asking a naïve question about the McGurk effect: Why does audio [b] give more [d] percepts with visual [g] than with visual [d]? In D. W. Massaro, J. Light, & K. Geraci (Eds.) *Proceedings of International Conference on Auditory-Visual Speech Processing* (pp.138-142). Santa Cruz, CA: Perceptual Science Laboratory.

Chen, T. H., & Massaro, D. W. (submitted). The Relationship between Unimodal and Bimodal Speech Perception of Mandarin Speakers. *Perception & Psychophysics*, submitted.

Derr, M. A., & Massaro, D. W. (1980). The contribution of vowel duration, Fo contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics 27*, 51-59.

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.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. Journal of the Acoustical Society of America, 104, 2438-2450

Grant, K. W., Walden, B. E., Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103*, 2677-2690.

Massaro, D.W., *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1987.

Massaro D.W., (1989). Multiple book review of *Speech perception by ear and eye: a paradigm for psychological inquiry*, by D.W. Massaro. *Behavioral and Brain Sciences, 12*, 741-794

Massaro, D. W. (1989). Testing between the TRACE model and the Fuzzy Logical Model of speech perception. *Cognitive Psychology, 21*, 398-421.

Massaro, D. W., (1984) Time's role for information, processing, and normalization. *The Annals of the New York Academy of Sciences.* New York: Science Press, Vol. 423, 372-384.

Massaro, D. W. (1988). Some Criticisms of Connectionist Models of Human Performance. *Journal of Memory and Language, 27*, 213-234.

Massaro, D. W. (1994). "Bimodal speech perception across the lifespan," in D.J. Lewkowicz and R. Lickliter (Eds.) *The Development of Intersensory Perception: Comparative Perspectives* (371-399). Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D. W. (1996). Integration of multiple sources of information in language processing. In T. Inui, and J.L. McClelland (Eds.) *Attention and Performance XVI: Information Integration in Perception and Communication*, Cambridge, MA: The MIT Press, 397-432.

Massaro, D. W. (1998a). Illusions and Issues in Bimodal Speech Perception. Proceedings of Auditory Visual Speech Perception ’98. (pp. 21-26). Terrigal-Sydney Australia, December, 1998.

Massaro, D.W. (1998b). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle*. Cambridge, MA: MIT Press.

Massaro, D.W. (1999). Speechreading: Illusion or window into pattern recognition. *Trends in Cognitive Sciences*, 3, 310-317.

Massaro, D. W., & Cohen, M. M. (1983). Consonant/vowel ratio: An improbable cue in speech. *Perception & Psychophysics, 33*, 501-505.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception &Psychophysics, 34,* 338-348.

Massaro, D. W., & Cohen, M. M., (1991) Integration versus Interactive Activation: the joint influence of stimulus and context in perception. *Cognitive Psychology, 23*, 558-614.

Massaro, D.W., & Cohen, M.M. (1993). Perceiving Asynchronous Bimodal Speech in Consonant-Vowel and Vowel Syllables, *Speech Communication, 13*, 127-134.

Massaro, D.W., & Cohen, M.M. (1993). The Paradigm and the Fuzzy Logical Model of Perception are Alive and Well. *Journal of Experimental Psychology: General, 122*, 115-124.

Massaro, D.W. and Cohen, M.M. (1995). Perceiving talking faces, *Current Directions in Psychological Science*, *4,* 104-109.

Massaro, D.W., & Cohen, M.M. (1999). Speech Perception in Perceivers with Hearing Loss: Synergy of Multiple Modalities. *Journal of Speech, Language, and Hearing Research*, 42, 21-41.

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.; Campbell, C. S.; Rodriguez, T. (2001). Bayes factor of model selection validates FLMP. *Psychonomic Bulletin & Review, 8*, 1-17.

Massaro, D.W., Cohen, M.M., & Gesi, A.T., (1993) Long-term Training, Transfer, and Retention in Learning to Lipread. *Perception and Psychophysics*, *53*, 549-562.

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

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

Massaro, D.W., Cohen, M.M., & Smeele, P.M.T. (1996). Perception of Asynchronous and Conflicting Visual and Auditory Speech, *Journal of the Acoustical Society of America, 100*, 1777-1786

Massaro, D.W., Cohen, M.M., & Thompson, L.A. (1988). Visible language in speech perception: Lipreading and reading, *Visible Language, 22*, 8-31.

Massaro, D. W., Cohen, M. M., & Tseng, C. Y., (1985) The Evaluation and Integration of Pitch Height and Pitch Contour in Lexical Tone Perception in Mandarin Chinese. *Journal of Chinese Linguistics, 13*, 267-289.

Massaro, D.W., and Ferguson, E.L. (1993). Cognitive Style and Perception: The Relationship between Category Width and Speech Perception, Categorization and Discrimination, *American Journal of Psychology, 103*, 25-49.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information, *Psychological Review, 97*, 225-252.

Massaro , D.W. & Oden, G. (1980). Evaluation and Integration of Acoustic Features in Speech Preparation. *Journal of the Acoustical Society of America, 67*, 996-1013.

Massaro, D.W., & Oden, G.C. (1995). Independence of Lexical Context and Phonological Information in Speech Perception. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*, 1053-1064.

Massaro, D. W., Tseng, C., & Cohen, M. M. (1983). Vowel and lexical tone perception in Mandarin Chinese: Psycholinguistic and psychoacoustic contributions. *Quantitative Linguistics, 19,* 76-102.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4,* 79-95.

Myung, I. J., & Pitt, M. A. (1998). Issues in selecting mathematical models of cognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 327-355). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1037-1052.

Platt, J. R. (1964). Strong inference. *Science, 146*, 347-353.

Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.

Port, R.F., & Dalby, J. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics, 32*, 141-152.

Schwartz, J.-L. (2003). Why the FLMP should not be appliced to McGurk data: Or how to better compare models in the Bayesian framework. *Auditory Visual Speech Perception '03*. Saint Jorioz, France.

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

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

Tseng, C. Y., Massaro, D. W., & Cohen, M. M., (1986) Lexical Tone Perception in Mandarin Chinese: Evaluation and Integration of Acoustic Features. In Henry S.R. Kao and Rumjahn Hoosain, (Eds.) *Linguistics, Psychology, and the Chinese Language* (pp. 91-104). Centre of Asian Studies, University of Hong Kong.

Walden, B., Montgomery, A., Prosek, R. A., & Hawkins, D. B. (1990). Visual biasing of normal and impaired auditory speech perception. *Journal of Speech and Hearing Research, 33*, 163-173.

Table 1. Description of the planned databases to be included in the resource sharing application hosted by the Perceptual Science Laboratory.

Reference	Experiment	Participants, Responses, Conditions	Description
Oden & Massaro, 1978	Place and VOT Cues	16 Ss, 4 responses, 35 conditions	Acoustic Cues to Place and Voicing
Massaro & Oden, 1980, Exp 1	Place and VOT Cues	11 Ss, 4 responses, 49 conditions	Acoustic Cues to Place and Voicing
Massaro & Oden, 1980, Exp 2	Place, VOT, and Aspiration Cues	8 Ss, 4 responses, 84 conditions	Acoustic Cues to Place and Voicing
Derr & Massaro, 1980	Duration Cues to Final Voicing	7 Ss, 2 responses, 16 Conditions	Duration Cues to Voicing
Derr & Massaro, 1980	Duration Cues to Final Voicing	10 Ss, Rating responses, 16 Conditions	Duration Cues to Voicing
Port & Dalby, 1982, exp. 1, Massaro & Cohen, 1983	Duration Cues to Voicing	16 Ss, 2 responses, 45 Conditions	Duration Cues to Voicing
Port & Dalby, 1982, exp. 2, Massaro & Cohen, 1983	Duration Cues to Voicing	10 Ss, 2 responses, 45 Conditions	Duration Cues to Voicing
Port & Dalby, 1982,	Sentence Tempo Cue	12 Ss, 2 responses, 64	Duration Cues to

Massaro, 1984	to Voicing	Conditions	Voicing
Erber, 1972; Massaro & Cohen, 1999	Hearing Level in Children	3 different hearing populations, 8 responses	Auditory and Visual Cues in Perception
Dowell et al., 1982, Massaro & Cohen, 1999	Cochlear Implant	1 patient, 12 responses, 3 conditions	Auditory and Visual Cues in Perception
Agelfors, 1996, Massaro & Cohen, 1999	Hearing Aids vs. Cochlear Implants	12 HA and 8 CI Ss, 16 responses, 3 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1983	Tone and Vowel Perception	6 Ss, 4 responses, 49 conditions	Auditory Cues to Vowel and Tone
Massaro et al., 1985	Auditory Cues to Tone Perception	6 Ss, 2 responses, 49 conditions	Auditory Cues to Tone
Walden et al., 1990, Massaro & Cohen, 1999	Hearing Level in Older Adults	2 groups, 15 Ss each, 3 responses, 42 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1993	Cross-linguistic influences	3 groups, 62 Ss, 2 responses, 35 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1993	Cross-linguistic influences	2 groups, 26 Ss, 8 or 6 responses, 35 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1995, exp 1	Cross-linguistic influences	1 group, 20 Ss, 2 responses, 35	Auditory and Visual Cues in Perception

		conditions	
Massaro et al., 1995, exp 2	Cross-linguistic influences	1 group, 10 Ss, 8 responses, 35 conditions	Auditory and Visual Cues in Perception
Massaro, 1987, Chapter 8, pp. 224-234	Developmental and Life-Span Differences	6 groups, 1-16 Ss each; 17 conditions, 2 responses	Auditory and Visual Cues in Perception
Massaro, 1994	Life Span Differences	2 groups, 13 Ss each, 35 conditions, 8 responses	Auditory and Visual Cues in Perception
Massaro & Cohen, 1995	Auditory and Visual Cues in Perception	10 Ss, 12 responses, 24 conditions	Auditory and Visual Cues in Perception
Massaro & Cohen, 1996	Inverted vs. Upright Face	20 Ss, 12 responses, 44 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1996, Exp. 1	Timing of Auditory and Visual Speech	10 Ss, 5 responses, 7 SOAs, 24 conditions	Auditory and Visual Cues in Perception
Massaro et al., 1996, Exp. 2	Timing of Auditory and Visual Speech	18 Ss, 5 responses, 7 SOAs, 24 conditions	Auditory and Visual Cues in Perception
Campbell et al., 1997	Visual Movement in Speechreading	Patient LM, 2 Controls	Auditory and Visual Cues in Perception
Massaro, 1998b	Prosopagnosia and Speech Perception	Patient HJA, 5 responses, 35 conditions	Auditory and Visual Cues in Perception
Pitt, 1995; Massaro & Oden, 1995	Lexical Influences	12 Ss, 2 responses, 12 conditions	Acoustic Cues and Lexical context

Grant & Seitz, 1998, Massaro & Cohen, 2000	Efficiency of Auditory-Visual Processing	40 Ss, 18 response alternatives,	Auditory and Visual Cues in Perception
Chen & Massaro, 2003, exp. 1	Cross-linguistic Influences	2 groups, 7 Ss each, 2 response alternatives, 35 conditions	Auditory and Visual Cues in Perception
Chen & Massaro, 2003, exp. 1	Cross-linguistic Influences	1 group, 7 Ss, 8 response alternatives, 35 conditions	Auditory and Visual Cues in Perception
Sekiyama, 1997; Chen & Massaro, 2003	Cross-linguistic Influences	14 Ss, 4 groups, 2 response alternatives	Auditory and Visual Cues in Perception
Cathiard et al., 2001; Massaro, 2003	McGurk Effect	126 Ss and 63 Ss, 11 responses	Auditory and Visual Cues in Perception

*The number of responses refers to the number of unique responses coded for the model tests.

List of Figures

Figure 1. Expansion of a typical factorial design to include auditory and visual conditions presented alone. The five levels along the auditory and visible continua represent auditory and visible speech syllables varying in equal physical steps between /ba/ and /da/.

Figure 2. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 9.

Figure 3. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 41. The lines are drawn through the observed points. The three large-circled points A_2V_5 give two unimodal conditions and the corresponding bimodal condition. The relationship among the three points can be explained by the use of a single modality, a weighted averaging of the two sources, or a multiplicative integration of the two sources.

Figure 4. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 25. The lines are

drawn through the observed points. The three large-circled points A_3V_1 give two unimodal conditions and the corresponding bimodal condition. The relationship among the three points *cannot* be explained by the use of a single modality or a weighted averaging of the two sources, but can be described by a multiplicative integration of the two sources.

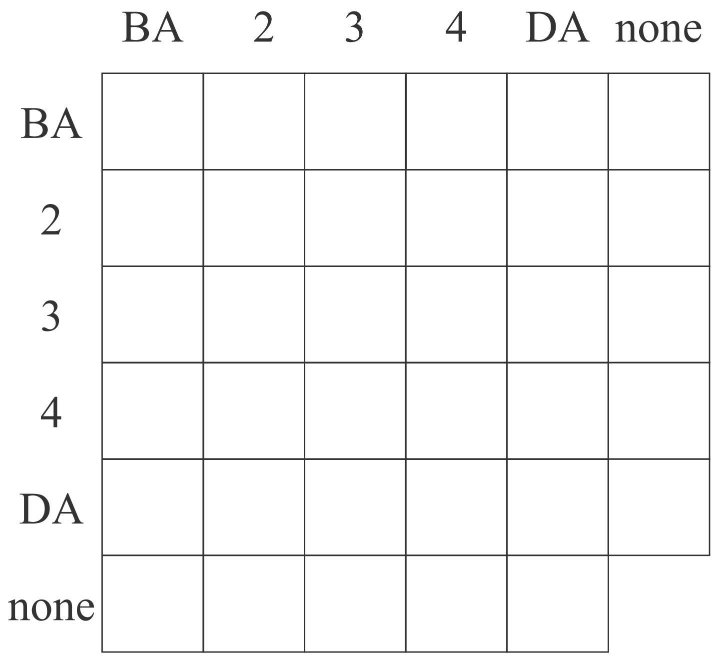

Figure 1. Expansion of a typical factorial design to include auditory and visual conditions presented alone. The five levels along the auditory and visible continua represent auditory and visible speech syllables varying in equal physical steps between /ba/ and /da/.

Figure 2. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 9.

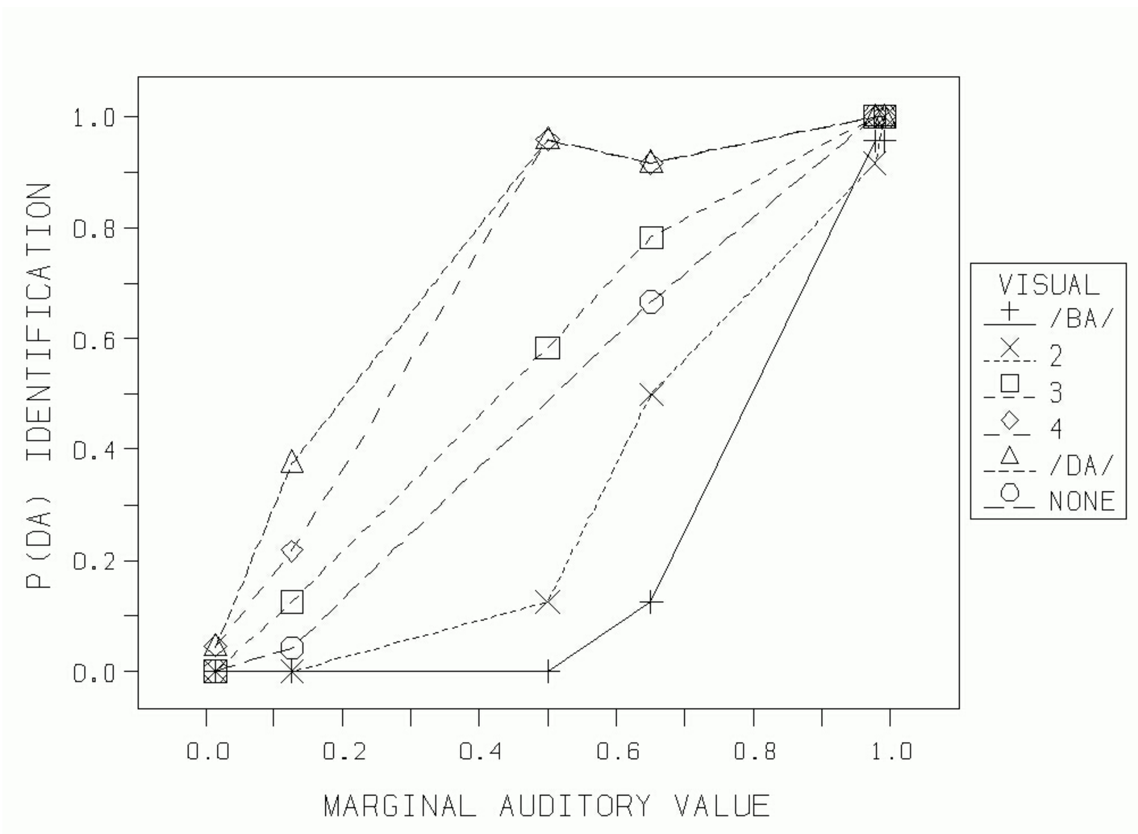

Figure 3. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 41. The lines are drawn through the observed points. The three large-circled points A_2V_5 give two unimodal conditions and the corresponding bimodal condition. The relationship among the three points can be explained by the use of a single modality, a weighted averaging of the two sources, or a multiplicative integration of the two sources.

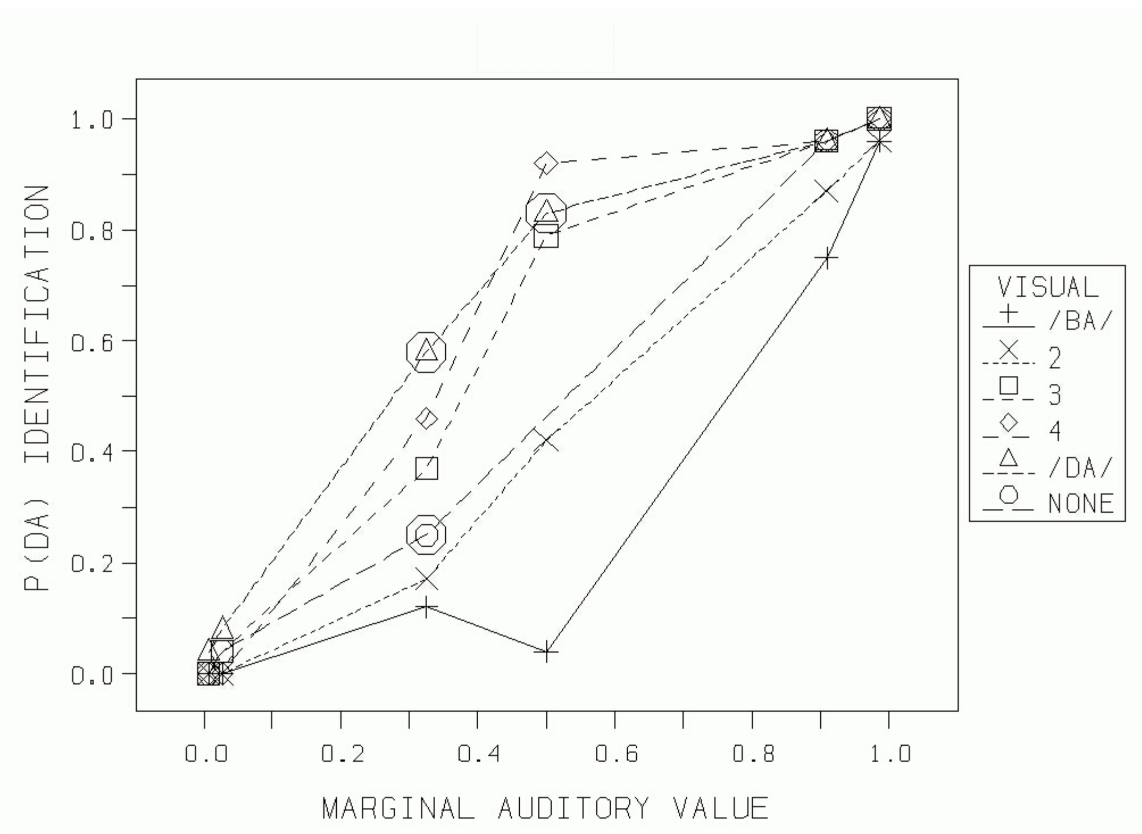

Figure 4. The points give the observed proportion of /da/ identifications in the unimodal and factorial auditory-visual conditions as a function of the five levels of synthetic auditory and visual speech varying between /ba/ and /da/. The columns of points are placed at a value corresponding the marginal probability of a /da/ judgment for each auditory level on the independent variable. The auditory alone conditions are given by the open circles. The unimodal visual condition is plotted at .5 (completely neutral) on the auditory scale. Results for participant 25. The lines are drawn through the observed points. The three large-circled points A_3V_1 give two unimodal conditions and the corresponding bimodal condition. The relationship among the three points *cannot* be explained by the use of a single modality or a weighted averaging of the two sources, but can be described by a multiplicative integration of the two sources.