

The 0/0 problem in the Fuzzy-Logical Model of Perception

Jean-Luc Schwartz (schwartz@icp.inpg.fr)

Institut de la Communication Parlée,
CNRS UMR 5009, INPG-Université Stendhal

INPG, 46 Av. Félix Viallet, 38031 Grenoble Cedex 1, France

Suggested running title: The 0/0 problem

ABSTRACT

The “Fuzzy-Logical Model of Perception” (FLMP) has often been questioned for its presumed ability to fit any data, but no clear-cut evidence has been presented yet. This paper demonstrates the ability of the FLMP to fit random data in the “McGurk region”, that is in conditions involving conflicting stimuli. This is due to the so-called “0/0 problem”, consisting in the fact that any audio-visual response can be fitted by the FLMP if the audio and visual stimuli provide at least one null probability in each possible category. The consequence is a high instability of the Root Mean Square Error RMSE in the region of the best fit.

Suggested PACS Classification numbers

Main section: 43.71

Detailed classification: 43.71.An, 43.71.Ma

INTRODUCTION

Since the middle of the 70s, Massaro and his colleagues have extensively studied the Fuzzy-Logical Model of Perception (FLMP) (Massaro, 1987), firstly in the context of categorical perception and then, since the emergence of the “McGurk” paradigm (McGurk & MacDonald, 1976), in the context of audio-visual (AV) interactions and fusion in speech perception (Massaro, 1998). A systematic assessment campaign of quantitative models, comparing the FLMP with various other competitors, lead some researchers to suspect a “high flexibility” of the FLMP, enabling it to fit a too large set of data (e.g. Cutting et al., 1992; Grant & Seitz, 1998). However, the demonstration has never been really conclusive (Massaro & Cohen, 1993, 2000; Dunn, 2000). More recently, the debate switched towards tools for comparing models, in the “Bayesian Model Selection” (BMS) framework (Myung & Pitt, 1997; Massaro et al., 2001; Pitt et al., 2003).

This paper discusses a technical problem with the FLMP, particularly acute in the context of a major experimental paradigm for multisensory interactions, that is the perception of conflicting stimuli – including the most well-known situation provided by the McGurk effect. This may lead to question some of the model comparison benchmarks using fit as a basic model assessment tool, rather than BMS criteria available in the literature.

I. A TECHNICAL ANALYSIS OF THE FLMP IN THE MCGURK PARADIGM

A. FLMP and the 0 / 0 problem

In a speech perception task involving the categorization of auditory, visual and audio-visual stimuli, the FLMP may be defined as a Bayesian fusion model with independence between modalities, and the basic FLMP equation is:

$$P_{AV}(C_i) = P_A(C_i)P_V(C_i) / \sum_j P_A(C_j)P_V(C_j) \quad (1)$$

C_i and C_j are phonetic categories involved in the experiment, and P_A , P_V and P_{AV} are the model probability of responses respectively in the A, V and AV conditions (observed probabilities are in lower case and simulated probabilities in upper case throughout this paper). In most papers comparing models in the field of speech perception, the tool used to compare models is the “fit” estimated by the “root mean square error” RMSE, computed by taking the squared distances between observed and predicted probabilities of responses, averaging them over all categories C_i (total number n_C) and all experimental conditions E_j (total number n_E), and taking the square root of the result:

$$RMSE = \left[\left(\sum_{E_j, C_i} (P_{E_j}(C_i) - p_{E_j}(C_i))^2 \right) / (n_E n_C) \right]^{1/2} \quad (2)$$

In a typical McGurk situation with audio /b/ plus video /g/, the unimodal responses are almost incompatible, and hence all phonetic categories involved in the pattern of responses display at least one very low value, either in the A modality, or in the V modality, or in both. The consequence is that all terms $P_A(C_i)P_V(C_i)$ are likely to be close to zero for all involved categories. To take an example, consider what would happen in an extreme situation with two phonetic classes C_1 and C_2 , and a pair of A and V stimuli perfectly conflicting, that is with 100% of C_1 responses with the A stimulus, and 100% of C_2 responses with the V stimulus (see Table I). Then, it is easy to show that any response in the AV modality, with a probability of C_1 response equal to x and a probability of C_2 response equal to $(1-x)$ (x being any value between 0 and 1) can be fitted by the FLMP with an RMSE value exactly equal to 0. Indeed, suppose that the corresponding FLMP parameters for C_1 and C_2 are respectively set to $(1-\epsilon_A)$ and ϵ_A in the A modality, and to ϵ_V and $(1-\epsilon_V)$ in the V modality, with ϵ_A and ϵ_V two small values. Then the probability of the C_1 response in the AV condition is given, according to Eq. 1, by:

$$\varepsilon_V(1-\varepsilon_A)/[\varepsilon_V(1-\varepsilon_A)+\varepsilon_A(1-\varepsilon_V)] \cong \varepsilon_V / (\varepsilon_V + \varepsilon_A) \quad (3)$$

Hence, x may be perfectly fitted by choosing an $(\varepsilon_A, \varepsilon_V)$ pair such that:

$$\varepsilon_V / [\varepsilon_V + \varepsilon_A] = x \Leftrightarrow \varepsilon_V = \varepsilon_A x / (1-x) \quad (4)$$

Then, setting ε_A and ε_V at an arbitrarily low value, provided that they respect Eq. (4), allows a perfect fit (RMSE equal to 0) to the pattern of experimental data in Table I, whatever x .

Table I

Exactly the same can be done with a 3-classes situation more similar to the McGurk effect (both for fusions and combinations). This corresponds to a configuration such as audio [b] (perceived as [b]) and video [g] (perceived as [d] or [g]), for which any response pattern can be perfectly fitted by the FLMP, with an RMSE value equal to 0.

This is due to a simple and well-known mathematical fact: 0/0 is an arbitrary value, or, to state this more precisely, $\lim(x/y)$ when $x \rightarrow 0$ and $y \rightarrow 0$, if it exists, may be any real value.

B. Application to real McGurk data

Of course, it could be argued that the previous section was not about real data, which seldom produce perfect zero values. However, the McGurk paradigm typically leads to quite similar situations, since it deals with conflicting A and V stimuli. In a study of the McGurk effect in French (Cathiard et al., 2001), the pattern of responses to [b_A], [d_V], [g_V], [b_Ad_V] and [b_Ag_V] for 126 French subjects, provided in Table IIa, surprisingly showed that there were less [d] and more [b] responses to [b_Ad_V] than to [b_Ag_V]. This pattern, coherent with many other published data,

seems difficult to understand on the classical view that “[b_A] is similar to [d_A] and [g_V] is similar to [d_V]”. Indeed, replacing [g_V] by [d_V] should obviously not result in a decrease of the [d] score in this reasoning. However, the FLMP performed very well on these data, with an RMSE of 0.0062⁽¹⁾. But actually, with the same A and V responses, any pattern of AV response can be fitted as well by the FLMP. This is displayed in Table IIb, providing the FLMP fits to hypothetical patterns of AV responses with [b_{AdV}] and [b_{AgV}] both perceived as mostly [b] (Response 1), mostly [d] (Response 2), mostly [g] (Response 3), [b_{AdV}] perceived as [b] and [b_{AgV}] as [d] (Response 4) or the inverse, [b_{AdV}] as [d] and [b_{AgV}] as [b] (Response 5). All the fits are equally good, and as good as the fit of true data: in this McGurk context, FLMP is able to fit everything, even a random pattern of response.

While *fitting both unimodal and multimodal data* by the FLMP in Table IIa provides excellent results (with RMSE as low as 0.0062), *predicting audiovisual from auditory and visual data* provides a dramatic RMSE increase up to 0.116, hence 20 times more. This is due to the fact that the FLMP ability to fit any pattern in this region has a severe drawback: the fit is highly unstable, hence the difference between the “fitting all” and the “prediction” strategies. Therefore, very small variations (± 0.01) applied to each FLMP parameter around the best fit to the experimental data in Table IIa, may lead to dramatic changes from the almost perfect value RMSE = 0.0062 to values as high as 0.25. The difference between a “fitting all” and a “prediction” strategy has been discussed in detail by Massaro (1998, Ch. 10), and the “fitting all” technique, called “variable FLMP”, is sometimes discarded in benchmarks, because of its presumed overfitting ability (e.g. Grant et al., 1998), though Massaro considers it to be the only valid one⁽²⁾.

Table II

II. A BAYESIAN-MODEL-SELECTION SOLUTION TO THE 0/0 PROBLEM

The 0/0 problem may result in a difficulty in testing FLMP in light of McGurk data (e.g. Cathiard et al., 2001; Tiippana et al., 2004). A possible solution could be to discard McGurk data from model assessment studies, but this seems odd in light of the many facts discovered about audio-visual speech perception by studying this paradigm through both behavioral (e.g. Green, 1998) and, more recently, neurophysiological (e.g. Jones & Callan, 2003; Sekiyama et al., 2003) paradigms.

The solution recommended by Massaro (1998) is to use large datasets rather than restricted McGurk data to assess and compare models. However, the benchmark set he used in a number of studies (e.g. Massaro, 1998), crossing a synthetic five-level audio /ba-/da/ continuum with a synthetic video similar continuum (<http://mambo.ucsc.edu/ps1/8236/>), does contain typical conflicting configurations involving the 0/0 problem. Therefore, this problem actually interferes with the results of the performed benchmark studies. The question is to know how much it interferes, and the answer is not at all obvious. Hence, our interest in another tool for comparing models, namely Bayesian Model Selection.

A. The Bayesian framework for model assessment

The fit may be derived from the logarithm of the *maximum likelihood of a model*, considering a data set. If \mathbf{D} is a set of k data d_i , and M a model with parameters Θ , $L(\Theta|M)$ is the likelihood of parameter Θ for the model, considering the data:

$$L(\Theta|M) = p(\mathbf{D}|\Theta, M) \tag{5}$$

The θ parameters maximizing the likelihood of M are provided by⁽³⁾:

$$\boldsymbol{\theta} = \operatorname{argmax} L(\boldsymbol{\Theta}|M) \quad (6)$$

and it is possible to show that maximizing likelihood is not very different from minimizing RMSE, that is searching the best fit to the experimental data. However, comparing two models by comparing their best fits means that there is a first step of estimation of these best fits, and it must be acknowledged that the estimation process is not error-free. Therefore, the comparison must account for this error-prone process, which is done in BMS by computing the total likelihood of the model knowing the data. This results in integrating likelihood over all model parameter values:

$$p(\mathbf{D}|M) = \int p(\mathbf{D}, \boldsymbol{\Theta}|M) d\boldsymbol{\Theta} = \int p(\mathbf{D}|\boldsymbol{\Theta}, M) p(\boldsymbol{\Theta}|M) d\boldsymbol{\Theta} = \int L(\boldsymbol{\Theta}|M) p(\boldsymbol{\Theta}|M) d\boldsymbol{\Theta} \quad (7)$$

Taking the opposite of the logarithm of total likelihood leads to the so-called ‘‘Bayesian Model Selection’’ (BMS) criterion that should be minimized for model evaluation (MacKay, 1992, Pitt & Myung, 2002):

$$\text{BMS} = -\log \int L(\boldsymbol{\Theta}|M) p(\boldsymbol{\Theta}|M) d\boldsymbol{\Theta} \quad (8)$$

B. How BMS integrates fit and stability

The integral in Eq. (7) means that the total likelihood of a model knowing the data evaluates the volume of $\boldsymbol{\Theta}$ values providing an ‘‘acceptable’’ fit (not too far from the best one) relative to the whole volume of possible $\boldsymbol{\Theta}$ values. This relative volume decreases if the function $L(\boldsymbol{\Theta}|M)$ decreases too quickly around its maximum value $L(\boldsymbol{\theta}|M)$: this is what happens if the model is too sensitive, as is the FLMP around its best fit in the McGurk region.

The difference between best fit and global likelihood is illustrated in Fig. 1. This figure deals with a two-category audio-visual experiment with one audio condition, one visual condition, and

one audio-visual condition combining the audio and the visual stimuli. An FLMP simulation of this experiment needs two free parameters, that is the audio probability P_A and the visual probability P_V of the first category, the probabilities of the second category being $1-P_A$ and $1-P_V$. Three experimental configurations are considered in the figure.

In the first one, the data provide an ambiguous perception both in the A, V and AV modalities ($p_A=p_V=p_{AV}=0.5$). On Fig. 1a, the distribution $L((P_A, P_V)|FLMP) = p(\mathbf{D}|(P_A, P_V), FLMP)$ is a smooth curve peaking at the point (0.5, 0.5) in the (P_A, P_V) plane. This peak provides the maximum likelihood, which is high, since the data are compatible with the FLMP prediction. The smooth behavior of the likelihood function leads to a large value of the total integral below the surface, which is precisely the global FLMP likelihood for these data.

In the second configuration, the A and V percepts are still ambiguous ($p_A=p_V=0.5$), but the AV percept is not ($p_{AV}=0.8$). This is of course contradictory with the FLMP prediction, hence the peak of the likelihood distribution in Fig. 1b is very low, and both best likelihood and global likelihood are small. Altogether, Fig. 1a and Fig. 1b illustrate the case of a strong FLMP prediction in which fit RMSE and global likelihood BMS provide the same kind of behavior (both good or both poor depending on the coherence of A, V and AV data).

The third configuration is typical of the McGurk effect, with conflicting A and V stimuli and any AV response (that is, the AV probability of the first category p_{AV} can be any value between 0 and 1). Fig. 1c shows that here, the FLMP maximum likelihood is quite high, but the likelihood distribution is not smooth at all. The reason is that to be able to fit everything in this case, the region around $(P_A=0, P_V=1)$ must be divided into an infinity of small sub-regions able to predict any value of p_{AV} . As a consequence, the integral below the likelihood curve is small. A high maximum likelihood (or a small RMSE) together with a small global likelihood: this is typically an over-fitting configuration, with no prediction ability at all. In such kinds of configurations, FLMP provides a better fit than almost any other model while global likelihood naturally

combines fit and stability into an integrated measure, making model assessment sounder. A detailed implementation of the so-called Laplace approximation of BMS together with its conditions of use is provided in <http://www.icp.inpg.fr>.

Figure 1

III. CONCLUSION

The 0/0 problem raises a difficulty in model fitting based on RMSE criteria for comparing FLMP with other models on conflicting stimuli. BMS appear as a better model comparison technique in this case, though RMSE may remain an interesting additional criterion enabling to assess the quality of the fit, apart from model comparison per se. We suggest that BMS could be of great interest in future model comparison studies in the AV speech perception domain.

Footnotes

1. RMSE in Cathiard et al. (2001) was slightly larger, because of the use of a threshold on the minimal acceptable probabilities, not used here, apart from the classical constraint that a probability is within [0-1].
2. Massaro (1998, Ch. 10) discusses in detail the relationship between “prediction” and what he calls “post-diction”, that is the “fitting all” procedure. The proposal he makes for connecting these is based on so-called “benchmark goodness of fit”. However, this technique, in which the data are varied according to the observed data statistics, suffers from exactly the same problem: in McGurk cases, any variation of the data can be perfectly fit by FLMP.
3. In the following, bold symbols deal with vectors or matrices, and all maximizations are computed on the model parameter set Θ .

References

- Cathiard, M.A., Schwartz, J.L., and Abry, C. (2001). "Asking a naive question to the McGurk effect : why does audio [b] give more [d] percepts with visual [g] than with visual [d] ?," Proceedings of AVSP'2001, pp. 138-142.
- Cutting, J.E, Brady, N.P., Bruno, N., and Moore, C. (1992). "Selectivity, scope, and simplicity of models: A lesson from fitting judgements of perceived depth," J. Experimental Psychology: General **121**, 364-381.
- Dunn, J.C. (2000). "Model complexity: the fit to random data reconsidered," Psychological Research **63**, 174-182.
- Grant, K.W., & Seitz, P.F. (1998). "Measures of auditory-visual integration in nonsense syllables and sentences," J. Acoust. Soc. Am. **104**, 2438-2450.
- Grant, K.W., Walden, B.E., and Seitz, P.F. (1998). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677-2690.
- Green, K.P. (1998). "The use of auditory and visual information during phonetic processing: implications for theories of speech perception," in *Hearing by Eye II*, edited by R. Campbell, B. Dodd and D. Burnham (Psychology Press, Hove, UK), pp. 3-25.
- Jones, J. A., and Callan, D. E. (2003). "Brain activation during an audiovisual speech perception task: An fMRI study of the McGurk effect," NeuroReport **14**, 1129-1133.
- MacKay, D.J.C. (1992). "Bayesian interpolation," Neural Computation **4**, 415-447.
- Massaro, D.W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. London: Laurence Erlbaum Associates.
- Massaro, D.W. (1998). *Perceiving Talking Faces*. Cambridge: MIT Press.

- Massaro, D.W., and Cohen, M.M. (1993). "The Paradigm and the Fuzzy Logical Model of Perception are alive and well," *J. Experimental Psychology: General* **122**, 115-124.
- Massaro, D.W., and Cohen, M.M. (2000). "Tests of auditory-visual integration efficiency within the framework of the fuzzy-logical model of perception," *J. Acoust. Soc. Am.* **108**, 784-789.
- Massaro, D.W., Cohen, M. M., Campbell, C.S., and Rodriguez, T. (2001). "Bayes factor of model selection validates FLMP," *Psychonomic Bulletin & Review* **8**, 1-17.
- McGurk, H., and MacDonald, J. (1976). "Hearing lips and seeing voices," *Nature* **264**, 746-748.
- Myung, I. J., and Pitt, M. A. (1997). "Applying Occam's razor in modeling cognition: A Bayesian approach," *Psychonomic Bulletin & Review* **4**, 79-95.
- Pitt, M.A., Kim, W., and Myung, I.J. (2003). "Flexibility versus generalizability in model selection," *Psychonomic Bulletin & Review* **10**, 29-44.
- Pitt, M.A., and Myung, I.J. (2002). "When a good fit can be bad," *Trends in Cognitive Science* **6**, 421-425.
- Sekiyama, K, Kanno, I, Miura, S, and Sugita, Y. (2003). "Auditory-visual speech perception examined by fMRI and PET," *Neuroscience Research* **47**,277- 287.
- Tiippana, K., Andersen, T.S., and Sams, M. (2004). "Visual attention modulates audiovisual speech perception," *European Journal of Cognitive Psychology* **16**, 457-472.

Tables

Table I – Fitting “perfect” two-category McGurk data with FLMP.

Fit is perfect for any value of x , provided that ϵ_A and ϵ_V follow Eq. (4), with arbitrary low values

	<i>C1 responses</i>		<i>C2 responses</i>	
	Data	FLMP	Data	FLMP
<i>A cond.</i>	1	$1-\epsilon_A$	0	ϵ_A
<i>V cond.</i>	0	ϵ_V	1	$1-\epsilon_V$
<i>AV cond.</i>	x	x	$1-x$	$1-x$

Table II – Fitting experimental McGurk data by the FLMP. (a) McGurk data for 126 French subjects obtained by Cathiard et al. (2001); (b) RMSE obtained when using the FLMP to fit the data in (a), or arbitrary AV Responses 1, 2, 3, 4, 5 with the same A and V unimodal data (see text).

(a)

<i>Responses</i>	[b]	[d]	[g]	other
[b _A]	0.98	0	0	0.02
[d _V]	0.005	0.88	0.06	0.055
[g _V]	0	0.125	0.845	0.03
[b _A d _V]	0.835	0.095	0	0.07
[b _A g _V]	0.68	0.23	0.02	0.07

(b)

		<i>Answer [b]</i>	<i>Answer [d]</i>	<i>Answer [g]</i>	<i>RMSE</i>
<i>True response</i>	[b _A d _V]	0.835	0.095	0	0.0062
	[b _A g _V]	0.68	0.23	0.02	
<i>Response 1</i>	[b _A d _V]	0.9	0.1	0	0.0049
	[b _A g _V]	0.9	0.1	0	
<i>Response 2</i>	[b _A d _V]	0.1	0.9	0	0.0053
	[b _A g _V]	0.1	0.9	0	
<i>Response 3</i>	[b _A d _V]	0.1	0	0.9	0.0061
	[b _A g _V]	0.1	0	0.9	
<i>Response 4</i>	[b _A d _V]	0.9	0.1	0	0.0082
	[b _A g _V]	0.1	0.9	0	
<i>Response 5</i>	[b _A d _V]	0.1	0.9	0	0.0047
	[b _A g _V]	0.9	0.1	0	

Figure captions

Figure 1 – Likelihood distributions in 2-category 3-condition configurations

(a) $p_A = p_V = p_{AV} = 0.5$

(b) $p_A = p_V = 0.5$ $p_{AV} = 0.8$

(c) $p_A \approx 0$ $p_V \approx 1$ p_{AV} arbitrary (any value between 0 and 1)