A Measure of Auditory-Visual Integration Efficiency Based on Fechnerian Scaling

Hans Colonius[1], *Adele Diederich*[2]

[1]Department of Psychology, University of Oldenburg, Germany
[2]School of Humanities and Social Sciences, Jacobs University, Bremen, Germany
hans.colonius@uni-oldenburg.de, a.diederich@iu-bremen.de

Abstract

Auditory-visual integration efficiency (IE) is a presumed skill employed by subjects independently from their ability to extract information from auditory and visual speech inputs. Currently there are no established methods for determining a subject's IE. Here we present a novel measurement technique to address this issue without requiring explicit assumptions about the underlying audiovisual processing. It is based on a version of the theory of Fechnerian Scaling developed by Dzhafarov and Colonius (Dzhafarov, E.N. and Colonius, H. (2006)). Reconstructing distances among objects from their discriminability, Psychometrika, 72(2), 365-386), that permits the reconstruction of subjective distances among stimuli of arbitrary complexity from their pairwise discriminability.

Index Terms: auditory-visual integration, Fechnerian Scaling, regular minimality

1. Introduction

Over the last five decades, the superiority of auditory-visual (AV) speech recognition over hearing alone (A) has been documented for nonsense syllables, words, as well as sentences ([1]; see [2] for recent reviews). Similar effects have been found for visually presented letters and speech sounds (phoneme-grapheme pairings) ([3]). Although AV speech recognition is quite robust in general, individuals differ with respect to their performance in situations where audibility and/or visibility is low or near threshold. "Auditory-visual integration" refers to the process employed by the receiver to combine the information extracted from both acoustic and visual sources. Braida ([4]) emphasized the importance of distinguishing between the ability to extract auditory and visual cues, on the one hand, from an individual's capability of integrating these cues from different modalities, on the other. In particular, low performance of hearing-impaired listeners may not only be a consequence of a lower rate of extracting auditory cues but also of a reduced efficiency in integrating cues from both modalities. In numerous studies, Grant and colleagues ([5],[6],[7]) have investigated these issues and proposed a quantitative measure of integration efficiency (IE).

In this paper, we suggest an alternative measure, somewhat close in spirit to the work of Grant and coworkers, but based on a new approach to measuring dissimilarity developed by Dzhafarov and Colonius ([8]).

2. Measures of integration efficiency based on models of integration

Any process model of AV integration strives to predict bimodal performance from performance on separate visual and auditory inputs (like nonsense syllables, words, sentences). A common data format on which performance is being gauged is the identification probability matrix ("confusion matrix") with cell entry $f(i,j)$ denoting the frequency with which a stimulus i has been identified as stimulus j, as an estimator of the identification probability $p(i,j)$. There is a plethora of models on how information from several sources are combined ([9]) and, since AV integration fits into this framework, many models of AV integration have been proposed in the literature ([10], [6]). Since our primary aim is to introduce a new measure of integration efficiency which is basically model-independent, no in-depth analysis or discussion of these models is presented here. A prominent model is Massaro's fuzzy logical model of perception (FLMP) with an optimal integration rule equivalent to Bayes' theorem ([11], [12], [13]). Another is Braida's prelabeling model of integration (PRE) ([4]) presented below in some detail because one of its aspects is related to our approach.

2.1. Prelabeling model of integration

In the PRE model, each response R_j corresponds to a point in a D-dimensional Euclidean vector space of stimulus attributes (cue vectors) referred to as *prototypes*. Each presentation of a stimulus i generates a D-dimensional vector of cues X in the same space following a multivariate normal distribution with independent components, unit variance, and a given mean S_i not necessarily identical to the prototype corresponding to R_i. According to a decision rule of multidimensional signal detection theory ([14]), the subject responds R_j if and only if the (Euclidean) distance of X to the prototype of R_j is smaller than the distance to any other prototype. The prototype locations are assumed to reflect response bias effects, whereas the subject's sensitivity in discriminating stimulus i from stimulus j is given by the Euclidean distance between S_i and S_j, $d'(i,j)$. The model parameters, i.e., the components of vectors S_i and R_i, are estimated iteratively through nonmetric multidimensional scaling (MDS) by comparing observed and predicted confusion matrices. The decision space for the AV condition is assumed to be the Cartesian product of the space for the A condition and the space for the V condition. A subject's sensitivity in the AV condition then is related to the unimodal sensitivities by

$$d'_{AV}(i,j) = \sqrt{d'_A(i,j)^2 + d'_V(i,j)^2}. \tag{1}$$

2.2. Model-based measures of integration efficiency

The improvement in *information transmission* (for definition see [15]) that occurs when auditory and visual information is available is often measured by the relative benefit measure (RB) defined as $(AV - A)/(100 - A)$, where AV is the percent information transmitted in the auditory-visual condition and A is the percent of information transmitted in the auditory alone reference condition ([1]). As pointed out in [7], if auditory and visual cues are highly redundant and A is low, AV cannot reach a value close to its maximum of 1 even if integration ability is perfect.

Several measures of integration efficiency (IE) that do not share these limitations have been proposed in the context of specific models of auditory-visual integration (see, e.g., [6]). Their common feature is that performance in the auditory-visual condition is compared with the model's predicted optimal performance in that condition given the performance in the unimodal conditions separately. In the PRE model, for example, a subject's IE is computed as the ratio between (average) observed recognition scores and the recognition scores predicted from Equation (1).

3. Fechnerian Scaling of discrete object sets (FSDOS)

The theory of Fechnerian Scaling (FS) deals with the computation of "subjective" distances among stimuli from their pairwise discrimination probabilities (see [16], [17], [18] for historical background, [19] for an overview, and [20] for a recent extension). Here we only present a sketch of a special case of FS, Fechnerian Scaling of Discrete Object Sets (FSDOS). For a detailed and more general presentation, we refer to [8] and [19]). 'Pairwise discrimination probabilities' are the probabilities with which the judgment "these two stimuli are different" is chosen over "these two stimuli are the same". As shown below, with additional assumptions the theory is also applicable to identification probabilities (confusion matrices) as they occur in speech recognition.

3.1. FSDOS procedure

Consider a set of stimuli (objects) $\{s_1, s_2, \ldots, s_N\}$, $N > 1$, presented two at a time to a subject whose task is to respond to each ordered pair (x, y) whether they are the same or different. Note that we distinguish (x, y) from the ordered pair (y, x) and that we can treat (x, x) as a pair rather than a single object, because in pairwise presentation the two stimuli generally belong to two distinct *observation areas*. This usually refers to spatial arrangement (say, one stimulus on the left, the other on the right) or temporal order (stimulus presented first/second). The probability

$$\psi(x, y) = \Pr[\text{subject judges } x \text{ and } y \text{ in } (x, y) \text{ to be different}] \quad (2)$$

constitutes the entry of a matrix with *row stimulus* x in the first observation area and *column stimulus* y in the second observation area. The most fundamental property of discrimination probabilities (introduced in [21]) and, in fact, the only requirement for the existence of a Fechnerian distance in FSDOS, is *regular minimality* defined as follows (in its simplest form):
For any $x \neq y$

$$\psi(x, x) < \min\{\psi(x, y), \psi(y, x)\}. \quad (3)$$

We define *psychometric increments* in the first and second observation areas as, respectively,

$$\phi^{(1)}(x, y) = \psi(x, y) - \psi(x, x), \quad (4)$$

$$\phi^{(2)}(x, y) = \psi(y, x) - \psi(x, x). \quad (5)$$

Due to regular minimality, these differences are positive. Consider now a chain of objects $s_i = x_1, x_2, \ldots, x_k = s_j$, leading from s_i to s_j, with $k \geq 2$. The *psychometric length of the first kind* for this chain is defined as

$$L^{(1)}(x_1, x_2, \ldots, x_k) = \sum_{m=1}^{k} \phi^{(1)}(x_m, x_{m+1}). \quad (6)$$

The set of different psychometric lengths across all possible chains of distinct elements connecting s_i to s_j being finite, it contains a minimum value $L^{(1)}_{\min}(s_i, s_j)$. This value is called the *oriented Fechnerian distance of the first kind from object s_i to object s_j*: $G_1(s_i, s_j) = L^{(1)}_{\min}(s_i, s_j)$. It can be shown to satisfy all properties of a metric, except for symmetry: in general, $G_1(s_i, s_j) \neq G_1(s_j, s_i)$. Any chain from s_i to s_j whose elements are distinct and whose length equals $G_1(s_i, s_j)$ is a *geodesic chain* from s_i to s_j.

The oriented Fechnerian distances $G_2(s_i, s_j)$ of the second kind (in the second observation area) and the corresponding geodesic chains are computed analogously, using the chained sums of psychometric increments $\phi^{(2)}$ instead of $\phi^{(1)}$. It should be noted that the oriented distances are not computed across the two observation areas but rather within the observation areas: $G_1(s_i, s_j)$ is the (oriented) distance between s_i and s_j in the first observation area, $G_2(s_i, s_j)$ is the (oriented) distance between s_i and s_j in the second observation area. The asymmetry of the oriented Fechnerian distances lacks operational meaning: while $\psi(x, y) \neq \psi(y, x)$ may be due to stimulus x (or y) belonging to two different observation areas in the two cases and may therefore be perceived differently, the inequality $G_1(s_i, s_j) \neq G_1(s_j, s_i)$ prevents one from interpreting either of them as a reasonable measure of perceptual dissimilarity between s_i and s_j. Therefore, we add the two oriented distances to obtain the (symmetrical) overall *Fechnerian distance* G:

$$G(s_i, s_j) = G_1(s_i, s_j) + G_1(s_j, s_i). \quad (7)$$

The validity of this definition derives from the fact (see [8]) that

$$G_1(s_i, s_j) + G_1(s_j, s_i) = G_2(s_i, s_j) + G_2(s_j, s_i), \quad (8)$$

that is, the Fechnerian distance G does not depend on the observation area in which these objects are taken.

An equivalent way of defining $G(s_i, s_j)$ is to consider all *closed loops* $x_1, x_2, ..., x_n, x_1$ ($n \geq 2$) containing two given objects s_i, s_j: $G(s_i, s_j)$ is the shortest of the psychometric lengths computed for all such loops. Note that the psychometric length of a loop depends on the direction in which it is traversed: generally,

$$L^{(1)}(x_1, x_2, ..., x_n, x_1) \neq L^{(1)}(x_1, x_n, ..., x_2, x_1)$$
$$L^{(2)}(x_1, x_2, ..., x_n, x_1) \neq L^{(2)}(x_1, x_n, ..., x_2, x_1) . \quad (9)$$

The equality in (7) tells us, however, that

$$L^{(1)}(x_1, x_2, ..., x_n, x_1) = L^{(2)}(x_1, x_n, ..., x_2, x_1), \quad (10)$$

that is, any closed loop in the first observation area has the same length as the same closed loop traversed in the opposite direction in the second observation area. In particular, if $x_1, x_2, ..., x_n, x_1$ is a geodesic (i.e., shortest) loop containing the objects s_i, s_j in the first observation area (obviously, the concatenation of the geodesic chains connecting s_i to s_j and s_j to s_i), then the same loop is a geodesic loop in the second observation area, if traversed in the opposite direction, $x_1, x_n, ..., x_2, x_1$.

3.2. A toy example

We illustrate the FSDOS procedure with a simple example (see [8] for a more detailed analysis). Consider the matrix of hypothetical discrimination probabilities from a 3-element stimulus set {A,B,C} in the left part of Table 1. Regular minimality is satisfied since the smallest value in each row is also the smallest value in each column. The Fechnerian distances are listed in the right part of Table 1. The geodesic loops are presented in Table 2. The geodesic loops containing A and B also contain C, i.e., the sum of the psychometric lengths of the chains AB and BA is larger than the sum of the psychometric lengths ACB and BA. Note that there is no monotonic relationship between the Fechnerian distances and the discrimination probabilities. However, empirically there is often a strong correlation between Fechnerian distances and the probabilities (see [17] for an analysis of the "probability-distance hypothesis").

Table 1: *A toy example of discrimination probabilities in a 3-element set.*

ψ	A	B	C	G	A	B	C
A	0.1	0.8	0.6	A	0	1.3	1
B	0.8	0.1	0.9	B	1.3	0	0.9
C	1.0	0.6	0.5	C	1	0.9	0

Table 2: *The geodesic loops where, by convention, each loop begins and ends with the row object.*

	A	B	C
A	A	ACBA	ACA
B	BACB	B	BCB
C	CAC	CBC	C

4. A measure of integration efficiency based on Fechnerian Scaling

The frequency $f(s_i, s_j)$ with which a stimulus s_i has been identified as s_j is taken as estimate for the identification probability $p(s_i, s_j) = \Pr[\text{response } "s_j" \,|\, \text{stimulus } s_i \text{ presented}]$. FSDOS can be applied to this paradigm under the additional assumption that $p(s_i, s_j)$ can be interpreted as $1 - \psi(s_i, s_j)$. Regular minimality then means that each stimulus s_i has a most frequent response "s_i" and any other stimulus s_j evokes response "s_i" less frequently than does s_i.

Let us now assume that Fechnerian distances G_A, G_V, and G_{AV} have been computed from the identification probabilities (confusion matrices) in the auditory, the visual, and the auditory-visual condition, respectively. For each pair of stimuli, s_i, s_j, the corresponding Fechnerian distances $G_A(i,j)$, $G_V(i,j)$, and $G_{AV}(i,j)$ are a measure of the subjective distance between the two stimuli under auditory, visual, and auditory-visual presentation, respectively. Note that, a-priori, these three distances are unrelated to each other since they are defined on different stimulus sets. On the other hand, if $G_{AV}(i,j)$ is larger than $G_A(i,j)$ or $G_V(i,j)$, this suggests that adding information about the two stimuli from the other modality (A or V) increases the subjective distance between s_i and s_j. *This increase in subjective distance from the unimodal presentation to the bimodal presentation is proposed to be an indicator of multisensory integration efficiency.*

In order to arrive at an overall index of integration efficiency, we suggest dividing by the sum of the unimodal distances and averaging across all stimulus pairs. This leads to the following index of *Fechnerian Scaling-based multisensory integration efficiency* :

$$\text{FS-IE} = \binom{N}{2}^{-1} \sum_{i>j} \frac{G_{VA}(i,j)}{G_A(i,j) + G_V(i,j)}. \qquad (11)$$

with N = the number of stimuli in each modality. Note that $G_{AV}(i,j) = G_A(i,j) + G_V(i,j)$ for all i,j $(i \neq j)$ yields an FS-IE value of 1. Ratios greater than one in (11) are not excluded a-priori (in which case one might speak of *super integration efficiency*). Alternative IE measures are certainly feasible.

The Fechnerian Scaling (FS) based approach to integration efficiency presented here and the prelabeling model (PRE) share the idea of converting the information contained in the confusion matrices into a spatial interpretation of the subjective distances between the stimuli. An important difference is that the FS based approach neither requires explicit assumptions about the space (e.g., Euclidean) and its dimensionality nor any parameter estimation.

Note that for the PRE model an index analogous to (11) could be computed by replacing the Fechnerian distances by the corresponding d' values. This index may be more informative than the recognition scores (percent correct recognition) commonly used (e.g., [7]).

At an intuitive level, it may be instructive to consider the effect of high redundancy of visual and auditory cues that compromised the relative benefit measure (RB) in Section 2.2. Combining such cues in the auditory-visual condition may not help in narrowing down the possible candidates in an identification task. Nevertheless, their combination may well be effective in increasing the subjective distance of a given audiovisual stimulus to some other audiovisual stimulus.

5. Illustrative example

Computation of FS-IE is illustrated with a simple example ($N = 3$), the reduced confusion matrices for consonants reported in [22] (see also [23], pp. 71–74). The original confusion matrices with 8 consonants were simplified by combining stimuli and responses across voicing and manner categories. This produced matrices similar to those that might have been obtained in hypothetical identification experiments in which stimuli and responses were restricted to /b/, /d/, and /g/. It is easily checked that regular minimality is satisfied in all matrices.
Table 3 lists all 3 confusion matrices (auditory, visual, auditory-visual) together with their corresponding Fechnerian distances G_A, G_V, and G_{VA}. Computation of FS-IE is straightforward. Abbreviating

$$R(i,j) \equiv G_{VA}(i,j)/[G_A(i,j) + G_V(i,j)],$$

Table 3: *Example:* $\psi_{VA} = 1 - p_{VA}$ *(p =identification probabilities) at top of cell and* G_{VA} *(Fechnerian distances) at bottom of cell, for auditory, visual, and auditory-visual presentation (rows* $\equiv$ *stimuli, columns* $\equiv$ *responses).*

$\psi_{VA} = 1 - p_{VA}$ / G_{VA}	"b"	"d"	"g"
b–	0.437 0.000	0.717 0.450	0.846 0.589
d–	0.700 0.450	0.530 0.000	0.757 0.350
g–	0.746 0.589	0.689 0.350	0.566 0.000
–b	0.022 0.000	0.983 1.805	0.996 1.527
–d	0.990 1.805	0.146 0.000	0.871 0.864
–g	0.989 1.527	0.575 0.864	0.436 0.000
bb	0.007 0.000	0.996 1.860	0.998 1.704
dd	0.997 1.860	0.126 0.000	0.876 1.203
gg	0.991 1.704	0.731 1.203	0.278 0.000

we get

$$\begin{aligned}\text{FS-IE} &= \binom{3}{2}^{-1} [R(b,d) + R(b,g) + R(d,g)] \\ &= \frac{1}{3}\left(\frac{1.86}{0.450+1.805} + \frac{1.704}{0.589+1.527} + \frac{1.203}{0.350+0.864}\right) \\ &= \frac{1}{3}(0.8248 + 0.8053 + 0.9909) \\ &= 0.8737. \end{aligned} \quad (12)$$

This value happens to be very close to the correct identification score (87.1 %) predicted by the PRE model (see [22]).

6. Discussion and Conclusions

In this work, we have suggested a novel way of measuring multisensory integration efficiency. Unlike previous approaches, it is not based on a specific processing model but on a principled method of calculating subjective distances from discrimination probabilities, i.e., Fechnerian Scaling of Discrete Object Sets (FSDOS) ([8]).

Application of FSDOS to confusion matrices requires the additional assumption that the identification probabilities can be interpreted in terms of discrimination probabilities: we put $p(s_i, s_j) = 1 - \psi(s_i, s_j)$, but there may be alternatives.

Regular minimality, the property that each stimulus s_i has a most frequent response "s_i" and any other stimulus s_j evokes response "s_i" less frequently than does s_i, then is a necessary and sufficient condition for the existence of the Fechnerian distances. In the given context, we consider the Fechnerian distances to be determined uniquely (for a deeper discussion of the uniqueness aspect, see [24]). Note, however, that the FS-IE measure is invariant with respect to multiplication of the three distances by a common positive number.

Calculation of Fechnerian distances from the confusion matrices in the auditory, visual, and auditory-visual condition results in three corresponding metric spaces. None of these has to have a Euclidean structure, nor any other specific spatial property, except for the metric axioms. Beyond defining a measure of integration efficiency as done here, a direction for future research could be a systematic investigation of how these three spaces are related to each other.

7. Acknowledgements

This work was supported in part by a grant from Deutsche Forschungsgemeinschaft (Sonderforschungsbereich/Transregio 31 "Active Listening", Project B4) to Oldenburg University and by NSF grant SES 0620446 and AFOSR grant FA9550-06-1-0288 to Purdue University.

8. References

[1] Sumby, W. H. and Pollack, I., "Visual contributions to speech intelligibility in noise", J. Acoust. Soc. Amer., 26:212–215, 1954.

[2] Calvert, G., Spence, C., and Stein, B. E.(eds.), Handbook of multisensory processes, MIT Press, Cambridge, MA, 2004.

[3] Larsen, A., McIlhagga, W., Baert, J., Bundesen, C. "Seeing or hearing? Perceptual independence, modality confusions, and crossmodal congruity effects with focused and divided attention", Perc. Psychophys., 65(4):568–574, 2003.

[4] Braida, L. D., "Crossmodal integration in the identification of consonant segments", Q. J. Exp. Psychol., 43:647–677, 1991.

[5] Grant, K. W. and Seitz, P. F., "Measures of auditory-visual integration in nonsense syllables and sentences", J. Acoust. Soc. Amer., 104:2438–2450, 1998.

[6] Grant. K. W., "Measures of auditory-visual integration for speech understanding: A theoretical perspective (L)", J. Acoust. Soc. Amer., 112:30–33, 2002.

[7] Grant, K. W., Tufts, J. B., Greenberg, S., "Integration efficiency for speech perception within and across sensory modalities by normal-hearing and hearing-impaired individuals", J. Acoust. Soc. Amer., 121:1164–1176, 2007.

[8] Dzhafarov, E. N. and Colonius, H., "Reconstructing distances among objects from their discriminability", Psychometrika, 71:365–386, 2006.

[9] Massaro, D. W. and Friedeman, D., "Models of integration given multiple surces of information", Psychol. Rev. 97:225–252, 1990.

[10] Massaro D. W. and Cohen, M. M., "Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception", J. Acoust. Soc. Amer., 108:784–789, 2000.

[11] Oden, G. C. and Massaro D. W., "Integration of featural information in speech perception", Psychol. Rev. 85:172-191, 1978.

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

[13] de la Vaux, S. K. and Massaro, D. W., ”Audiovisual speech gating: examining information and information processing”, Cogn. Process. 5:106-112, 2004.

[14] Macmillan, N. A. and Creelman, C. D., Detection Theory: A User’s Guide, 2nd Ed., Lawrence Erlbaum Ass., Mahwah, NJ, 2005.

[15] Luce, R. D., ”Detection and recognition”. In R. D. Luce, R. R. Bush, E. Galanter, Handbook of mathematical psychology, John Wiley, New York, 1960, pp. 103–188.

[16] Dzhafarov, E. N., ”Fechnerian Psychophysics”. In N. J. Smelser and P.B. Baltes (Eds.), International Encyclopedia of the Social and Behavioral Sciences, Vol. 8, Pergamon Press, New York, pp. 5437–5440.

[17] Dzhafarov, E. N., ”Multidimensional Fechnerian Scaling: Probability-distance hypothesis”, J. Math. Psychol. 46:352–374.

[18] Dzhafarov, E. N. and Colonius, H., ”Fechnerian metrics in unidimensional and multidimensional stimulus spaces”, Psychonomic Bull. Rev. 6:239–268.

[19] Dzhafarov, E. N. and Colonius, H., ”Generalized Fechnerian Scaling”. In H. Colonius and E. N. Dzhafarov (Eds.), Measurement and Representation of Sensations, Lawrence Erlbaum Ass., Mahwah, NJ, 2006, pp. 47–88.

[20] Dzhafarov, E. N. and Colonius, H., ”Dissimilarity cumulation theory and subjective metrics”, J. Math. Psychol., 2007, doi:10.1016/j.jmp.2007.02.004.

[21] Dzhafarov, E. N.,”Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity”, J. Math. Psychol. 46:583–608.

[22] Braida, L. D., Sekiyama, K., Dix, A. K., ”Integration of audiovisually compatible and incompatible consonants in identification experiments”, Proceedings of AVSP98, Terrigal, NSW, Australia.

[23] Johnson, K., Acoustic and Auditory Phonetics, 2nd ed., Blackwell, Oxford, 2003.

[24] Dzhafarov, E. N. and Colonius, H., ”Psychophysics without physics: Extension of Fechnerina scaling from continuous to discrete and discrete-continuous stimulus spaces”, J. Math. Psychol. 49:125–141, 2005.