Ordering of functions according to multiple fuzzy criteria

Burgos-Madrigal, Andrea; Orihuela-Espina, Felipe; Reyes-García, Carlos Alberto

DOI:
10.1007/s00500-021-05719-z

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Burgos-Madrigal, A, Orihuela-Espina, F & Reyes-García, CA 2021, 'Ordering of functions according to multiple fuzzy criteria: application to denoising electroencephalography', *Soft Computing*, vol. 25, no. 13, pp. 8573-8593. https://doi.org/10.1007/s00500-021-05719-z

Soft Computing
https://doi.org/10.1007/s00500-021-05719-z

FUZZY SYSTEMS AND THEIR MATHEMATICS

Ordering of functions according to multiple fuzzy criteria: application to denoising electroencephalography

Burgos-Madrigal Andrea[1] · Orihuela-Espina Felipe[1] · Reyes-García Carlos Alberto[1]

Accepted: 25 February 2021

Abstract

We introduce a new relation of order over functions according to multiple fuzzy criteria. Proof of the complied properties for relations of partial orders is given. Convergent and divergent validity of the new membership functions is established. Tolerance to noise of the relation of order is evaluated by corrupting synthetic prototypes and observing changes in the retrieved ordering. The effect of weighting strategies is evaluated in terms of Jaccard and XOR indices. The performance of the ordering algorithm is quantified in terms of richness of the resulting Hasse diagram. Applicability is demonstrated in the context of de-noising electroencephalographic (EEG) signals exemplified over two datasets and evaluated by classification wrapping.

Keywords Fuzzy decision making · Fuzzy order relations · Membership functions · Multiple criteria evaluation · Electroencephalography

1 Introduction

Relations of order are characterized for complying with reflexivity and transitivity (preorders), and further defined in terms of compliance with symmetry (equivalence relations) or antisymmetry (partial orders) and trichotomy (total orders) (Klir and Yuan 1995). Relations of order are common in most fields of science and engineering as well as in our day life. They are the ones that permit us taking informed decisions, sort or rank collections, optimize processes and of course define sequences and algorithms. We use the natural (cardinality based) ordering of the common numerals sets inadvertently in our daily lives which is an example of a relation of (total) order for crisp scalar sets according to a single criterion. In different domains, relations may have yet to be defined or existing relations of order might not be adequate because assumptions are not well matched to those of the problems at hand. One of such domains is the definition of a relation of order for ranking functions according to multiple fuzzy criteria (Burgos-Madrigal 2018). This abstract problem matches many practical ones including that of isolating stimulus evoked activity from electroencephalographic (EEG) recordings, a problem that remains unsatisfactorily solved.

During the EEG measurement of a neurological process, a range of concurrent cognitive and artifactual subprocesses occur. These include the evoked activation of frequency bands as well as the baseline default activity plus unwanted artifacts like eye blinking or heart beats that contaminate the signals. The measured signals at the scalp carry information contributed by all these sources in a mixture which is often processed by using some blind source separation method as, for instance, independent component analysis (ICA). ICA yields a set of components—whether spatial or temporal functions—suitable to attempt the reconstruction of a cleaner signal by discarding noise-related components. However, discrepancies between the mathematical assumptions and the physiological reality means that those components are just approximations of the sources. A crisp allocation to signal-related or noise-related source sets is therefore inappropriate, and any subsequent filtering would either still carry unwanted noise or unnecessarily discard genuine signal information. Fuzzy sets, which allow membership of elements from binary to real, seem more appropriate in this scenario. Further, not all signal-related sources (and analogously noise-related

✉ Burgos-Madrigal Andrea
burgosmad@inaoep.mx

1 National Institute of Astrophysics Optics and Electronics, Luis Enrique Erro #1, Sta María Tonanzintla, San Andrés Cholula, Puebla 72840, Mexico

Published online: 18 March 2021

sources) share equal relevance for the neurophysiological phenomenon under scrutiny, so any decision to preserve or discard a component during reconstruction has to conciliate multiple sorting criteria of varying relevance. Such scenario from neuroimaging has motivated this research.

Fuzzy multicriteria analysis (FMCA) is a fuzzy ordering algorithm that generates a partial order to allow indifference and incomparability in the decision process (Van de Walle et al. 1995; Bruggemann et al. 2011; Annoni et al. 2008). In FMCA, the elements to be compared are real scalars expressing an evaluation of some criteria, and the relation of order is based on the containment of one element in another. Here, *we extend FMCA to work over real functions*. The functions are evaluated to high/low membership to a set of identified subprocesses or ordering criteria. This is then applied to EEG analysis whereby ICA separated components are ranked according to their contribution to a neurological process. In the case of EEG, the ICA components (the functions to be ordered) are evaluated to be members of the subprocesses contributing to the EEG, e.g., different brain rhythms, or other systemic sources.

The new relation of order proposed establishes the relation of each function (i.e., ICA component in our application) to a model of a studied process. A process in mathematics is just a collection of variables, just another word for sets, and here it represents the collection of ordering criteria. Here, the model of the process is simply a weighted aggregation of subprocesses (each individual ordering criterion). Depending on the known information about the fuzzy set, that is, a subprocess, a different membership function to the set is chosen to evaluate the membership of the function, again, in our exemplary application, the ICA components. Moreover, every criterion (subprocess) may have different relevance to the main process. On Van de Walle et al. (1995) the weighting strategy is an additive model that consists of an arbitrary weight for each criteria. Here, fuzzy modifiers (Zadeh 1972) are used as weighting strategies instead.

In summary, our goal is to extend FMCA to order real functions. We modified FMCA by substituting the "default" real scalar-based relation of order underpinning FMCA, with a brand new relation of order that support functions as inputs. Further, this new relation of functions can operate according to multiple fuzzy criteria. The main contribution is the newly proposed relation of order for real functions itself with proof for the ordering properties. Additional contributions include three new membership functions for fuzzy functions sets, a new use of existing fuzzy modifiers as a weighting strategy for aggregating the process model and a new score, richness, proposed as an indicator of spread of a tree.

The usefulness of the approach is exemplified to the problem of denoising EEG recordings.

2 Related work

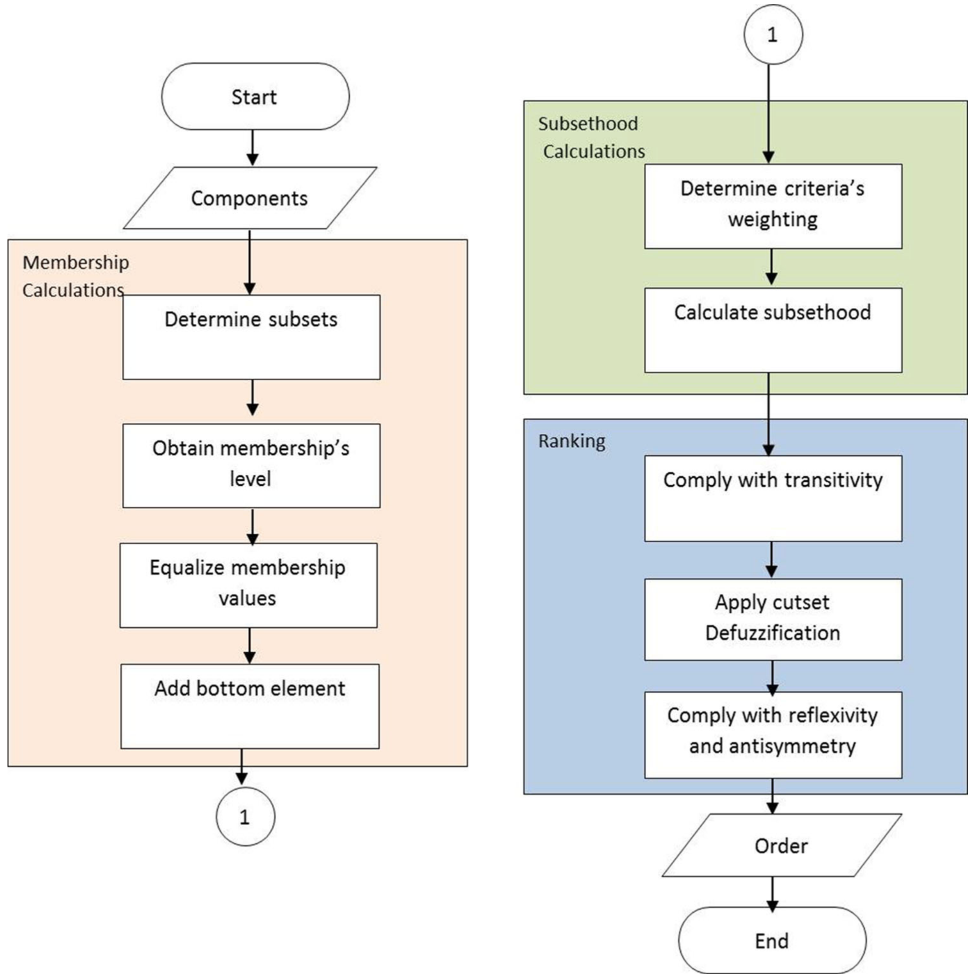
In this research, fuzziness is presented in three stages:

First, the process is modeled as a fuzzy set (Bellman and Zadeh 1970) conformed by subsets (referred to as criteria). Ranking fuzzy quantities has been explored using different approaches (Bortolan and Degani 1985) including Fuzzy sets type 1 (Wang and Kerre 2001), Fuzzy sets Type 2 (Wu and Mendel 2009; Figueroa-García et al. 2018), Hesitant Fuzzy sets (Wang et al. 2018a) and Intuitionistic Fuzzy sets (Garg and Kumar 2019; Xing et al. 2018; Kumar and Garg 2017). Wang and Kerre Wang and Kerre (2001) enumerate more than 35 fuzzy number ranking indices. These approaches usually treat a fuzzy set as an area to be compared or defuzzified, e.g., trapezoidal. A different strategy does not impose a curve to define the sets. Instead, FMCA generates a partial order Van de Walle et al. (1995) allowing indifference and incomparability in the decision process. FMCA was originally motivated by a problem of sorting cutting techniques in a nuclear reactor dismantling project. Later, FMCA was extended to other domain applications such as describing the structure of poverty (Annoni et al. 2008), ranking refrigerants (Bruggemann et al. 2011), for evaluating the performance of Internet of Things (IoT)-based supply chains (Wibowo and Grandhi 2018), structure for the periodic system of chemical elements (Leal and Restrepo 2019) or even the evaluation of predictive models (Früh et al. 2018). In these examples, the elements to be ranked are real numbers, indicating their respective evaluation to some criteria and the relation of order is then based on the containment of an element in another. While it is relatively easy to establish whether i precedes or proceeds j with $i, j \in \mathbb{R}$, it is not that clear how to decide whether $f(\bullet) \in \mathbb{R}^n$ precedes or succeeds $g(\bullet) \in \mathbb{R}^n$. Here, we propose a new relation whereby functions precedence is established by means of distance to the process model.

Second, a fuzzy set is characterized by a membership function which assigns to each element a grade of membership between zero and one (Zadeh 1965). Many common membership functions over numerical scalar domains have been thoroughly studied. Also, membership functions used to describe natural language, such as membership to fuzzy sets such as small or large, with shapes already defined can be found in literature. These, however, have been suggested to be inflexible for continuous variables describing problems of spatial variation, for instance, for soil (Zeng et al. 2017) or images (Chuang et al. 2006). Hence, in this research we define three new membership functions for which their applicability depends on the available knowledge about the fuzzy set.

Finally, when operating with multiple criteria, the final ordering ought to aggregate the marginalized orderings resulting from each individual criterion. This is often done by means of a weighting strategy. The weighting process based

Fig. 1 Flowchart illustrating the steps of Weighted Functions Fuzzy Multicriteria Analysis (WFMCA)



on the expert knowledge is usually vague and/or incomplete using linguistic information to describe it (Martinez et al. 2006). Using Linguistic hedges expressing the experts' opinions in qualitative decision making has been explored before (Chuang et al. 2006; Wang et al. 2018b). In this study, we made a comparison between the classical additive model used to describe importance with the linguistic hedges and the contrast intensification to automatically define the weighting instead of imposing a value.

3 Method

In this section, the new ordering algorithm named Weighted Functions Fuzzy Multicriteria Analysis (WFMCA) is detailed. The relation of order underpinning the new algorithm is based on the distance of the input functions to an aggregated weighted fuzzy model. The technique logical flow is illustrated in Fig. 1 .The sources recorded by EEG sensors are mixed, and ICA is applied to approximate to the original sources from the recorded signals. To do so, the sources are considered independent from each other. The resulting components are the input to our problem.

Then, our solution exploits knowledge about the domain at hand. Such knowledge comes in the form of criteria (subsets) that are qualitatively or quantitatively known to be more or less (fuzzy) relevant to the studied phenomenon (universal set). The subsets are defined depending on the application for evaluating the input functions. The creation of a subset is based on the information known from the process being studied, i.e., the EEG measures include noise from other sources known as artifacts such as ocular artifacts, i.e., eye blink (Vigário 1997), artifacts related to cardiac activities, i.e., heart rate (Niazy et al. 2005). Also, the EEG signals present different frequency ranges (Moctezuma et al. 2019). During the membership calculations, we determine the membership of the function to the identified subsets. Next, every subprocess present has different relevance to the main process, and we studied different strategies of weighting and calculate the inclusion grades (also known as Subsethood (Young 1996)) to the weighted criteria. Finally, the ranking procedure ensures compliance with the properties that define an order resulting an order of the components.

3.1 Ordering of functions over multiple fuzzy criteria

Let $\mathcal{F}$ be the set of real functions, and $Y = \{y_j | y_j \in \mathcal{F}\}$ a set of arbitrary real functions to be sorted according to some set of criteria $C = \{c_1, c_2, ..., c_i\}$. The problem of ordering functions according to C can be stated as follows: generate a relation of order r constrained by C resulting in a partial or total indexing $J_r(Y)$ of the elements $y_j \in Y$.

If the sorting criteria C is known to encode a generating process for Y, without loss of generality, the problem can be restated as follows. Let $X = \{x_1, x_2, ..., x_j\}$ be a set of measured signals, e.g., EEG channels time courses, and P be a hidden generating process of interest, e.g., a certain stimulus-evoked cognitive activity, to which certain latent variables $C = \{c_1, c_2, ..., c_i\}$ are known to contribute $P \simeq f(C)$. In EEG signals, these correspond to subprocesses present during the main process as the activity in a certain frequency band or a systemic contribution. $C = \{c_1, c_2, ..., c_i\}$ is regarded as a set of sorting criteria to be considered for the process P. For a certain domain, they may be given or defined through the literature. Also, let $Y = \{y_j : X(C) \rightarrow \mathcal{F}\}$ be a set of functions y_j de-mixed by A from the measured signals x_l: $Y = AX$, e.g., the components found by ICA from the EEG recordings. Note that a blind source separation cannot guarantee a one-to-one relation among retrieved components and sources under mismatched assumptions. The problem is *to generate a relation of order r_P over Y, such that the resulting indexing $J_r(Y)$ of the elements $y_j \in Y$ reflects the relevance of each y_j for P according to a given appreciation of the importance of c_i for P, $W_P(C;\mu)$*, where μ represents the membership level and by definition, since they are variables, the sorting criteria are taken by sets.

3.2 Membership functions

The membership of y_j to a fuzzy set criteria c_i is denoted as:

$$\mu_{c_i}(y_j) \in [0, 1] \tag{1}$$

where $\mu_{c_i}(y_j)$ is the evaluation of function y_j in the criteria set c_i. Traditional membership functions in fuzzy set theory have a signature $\mathbb{R} \rightarrow [0, 1] \subseteq \mathbb{R}$, e.g., triangular (Klir and Yuan 1996). For functions, we need a signature $\mathcal{F} \rightarrow [0, 1] \subseteq \mathbb{R}$. For digitally sampled functions, this can be reduced to $\mathbb{R}^n \rightarrow [0, 1] \subseteq \mathbb{R}$ with n being a number of control points, e.g., the number of samples at which the function has been observed.

Individual criteria c_i may be described in different manners depending on the evidence or information known from the criteria set, and hence, μ_{c_i} will take different forms. Three scenarios have been identified, and a novel membership functions for each one is proposed.

3.2.1 Knowledge based

The description of fuzzy set c_i is given in terms of an explicit generative model, e.g., $c_i = g(Z) + \varepsilon_{c_i}$ with Z being some factor for which the element $y_j(Z) = g(\hat{Z}) + h(Z) + \varepsilon$ with $\hat{Z} = [a, b] \subseteq Z$. $g(\hat{Z})$ is the contribution of c_i into y_j and it is proportional to the membership of y_j to c_i, and $h(Z)$ is the contribution of any other source; this separation is, of course, not known. The membership level is defined in terms of the ratio between the area under the curve contributed by the fuzzy set and total area under the curve of the element as indicated in Eq. 2:

$$\mu^K_{c_i}(y) = \frac{\int_a^b g(Z)dZ}{\int_{-\infty}^{\infty} y_j(Z)dZ} \tag{2}$$

Equation 2 can be interpreted as the amount of knowledge of the function y_j explained by criterion c_i. To be useful, $g(Z)$ in Eq. 2 has to be operationalized. In particular, here it is calculated for EEG frequency bands. For these, the $g(Z)$ is related to the Fourier transform of the y_j component. Let $FFT(y[t])$ be the discrete approximation of the Fourier transform of $y(t)$ found by the fast Fourier transform (FFT) algorithm (Smith et al. 1997). The total of energy across the spectrum $\mathcal{E}_T(y_j)$ (from 0.1 to 30 [Hz] for practical matters in EEG signals) can be approximated according to Eq. 3:

$$\mathcal{E}_T(y_j) = \sum_{k=0.1}^{k=30} \mathcal{E}_k(FFT(y[t])) \tag{3}$$

with $g(Z) \simeq g[k] = \mathcal{E}_k(FFT(y[t]))$, and $\mathcal{E}_k(FFT(y[t]))$ the estimated energy at frequency k [Hz]. Similarly, the energy in a given band of interest, e.g., Delta ($\delta \simeq [0.5, 3.5]$ [Hz] Knyazev (2013), is calculated as per Eq. 4:

$$\mathcal{E}_\delta(y_j) = \sum_{k=0.5}^{k=3.5} \mathcal{E}_k(FFT(y[t])) \tag{4}$$

The knowledge-based membership function for a given frequency band (e.g., Delta) is defined using Eq. 5:

$$\mu^K_{c_i:\delta}(y_j) = \mu_\delta(y_j) = \frac{\mathcal{E}_\delta(y_j)}{\mathcal{E}_T(y_j)} \tag{5}$$

Analogous expressions can be established for other frequency bands α (8-13 [Hz]), β (13-30 [Hz]) and θ (3.5-8 [Hz]). A more theoretical discussion will see Eqs. 3, 4 expressed as integrals following the analogous Fourier transform, but here we stick to the digitized FFT.

3.2.2 Prototype based

The description of fuzzy set c_i is given in terms of a collection of n prototypes of elements of the set $(s_n, \mu_{c_i}(s_n))$. Membership functions express the similarity of the elements to known examples of the set, and since the elements are functions, it seems appropriate to use the maximum across lags of the cross-correlation function as a measure of similarity. These prototypes (exemplary signals) are members of the fuzzy set, and the membership of other elements of the fuzzy set, e.g., y_j, is obtained by computing central tendency of the similarity to prototypes normalized by the membership value of the prototypes according to Eq. 6:

$$\mu^P_{c_i}(y_j) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\mu_{c_i}(s_k)}\max_{\tau}\left(\frac{E\left[(s_k[t]-\eta_{s_k})(y_j[t+\tau]-\eta_{y_j})\right]}{\sigma_{s_k}\sigma_{y_j}}\right) \tag{6}$$

where η_{s_n} and σ_{s_n} are the mean[1] and standard deviation of the synthetic prototype s^n_i, which are constant over time due to stationarity; and similarly for y_t, respectively, n is the number of prototypes available, τ the displacement or lag, and $E[\cdot]$ indicates the expected value. Equation 6 encodes the notion of the proximity of function y_j to previously seen examples (prototypes) of criterion c_i. If phase is important, then the zero lagged cross-correlation coefficient may be used instead in Eq. 6.

For instance, if we have a set of blinking prototypes with full membership to the blink criterion (i.e., $\mu_{c_i}(s_n) = 1$), then to calculate the membership of component y_j to the blink criterion, Eq. 7 applies

$$\mu^P_{c_i:Blink}(y_j) = \mu_{Blink}(y_j) = \frac{1}{n}\sum_{k=1}^{n}\max_{\tau}\left(\frac{E\left[(s_k[t]-\eta_{s_k})(y_j[t+\tau]-\eta_{y_j})\right]}{\sigma_{s_k}\sigma_{y_j}}\right) \tag{7}$$

That is, Eq. 7 expresses whether the y_j component shares some similarity to any prototype of blinking that we may have in our repository.

3.2.3 Distribution based

The description of the fuzzy set c_i is given in terms of a known or assumed model distribution, $c_i \sim \mathcal{D}(\mathbf{p})$ with $\mathbf{p}$ being the parameters of the distribution. The membership of element y_j to the fuzzy set c_i is then given by the probability of the observation (the function) to be generated by the model (the fuzzy set), that is, the likelihood function in Eq. 8:

$$\mu^D_{c_i}(y_j) = P(y_j|c_i) \simeq P(y_j|\mathcal{D}(\mathbf{p})) \tag{8}$$

where $P(\cdot)$ is the probability of an event. Discrete Markov processes can be used to estimate these probabilities for functions (Sucar 2015), and for which a discretization of the function domain into states is convenient. Continuous density variants of Markov processes exist, but their relevance here is beyond this research. The estimation of the probability is affected by the number of states considered and the number of samples of the function available. The higher the number of states, the lower values of probabilities to any given initial state as well as the lower the probability of transition among any two given states. The higher the number of samples, i.e., length of sample observations, it is less likely to be generated by a nominal theoretical model. Having the initial and transition probabilities, the states in which the function exists will have a probability associated. These probabilities will soon become close to 0, and so a normalization step is considered. The higher probability a signal can present is given by Eq. 9:

$$P_{\max}(y_j|\mathcal{D}(\mathbf{p})) = \max(\pi_i)(\max(a_i))^{t-1}) \tag{9}$$

where $t-1$ is the number of samples considered discarding the initial probability, π is the initial probability vector, and a_i is a probability element from the transition probability matrix A.

A Box-Cox transformation Box and Cox (1964) with $\omega = 0$ (logarithm) alleviates the low probabilities (Eq. 10).

$$\mu^D_{c_i}(y_j) \simeq \tilde{P}(y_j|\mathcal{D}(\mathbf{p})) = BoxCox\left(\frac{P(y_j|\mathcal{D}(\mathbf{p}))}{P_{\max}(y_j|\mathcal{D}(\mathbf{p}))};\omega\right) : \omega = 0 \tag{10}$$

The distribution-based membership function in Eq. 10 expresses the likelihood of observing y_j should the true model be described by criterion c_i. For our domain, when the criterion is an artifact, as it is the case of heart rate, we may opt for Eq. 11 to facilitate interpretation, the higher the probability, the less the component contributes to the cognitive process of interest:

$$\mu_{HR}(y_j) = 1 - \mu^D_{c_i}(y_j) = 1 - \frac{|\tilde{y}_j(\lambda)|}{\max(|\tilde{y}(\lambda)|)} \tag{11}$$

The decision to complement the μ (e.g., $1-\mu$) is arbitrary as we could later use negative weightings.

3.3 Weighted functions fuzzy multicriteria analysis

The new relation of order and the supporting ordering algorithm named Weighted Fuzzy Multicriteria Analysis applied to functions are an extension to FMCA from Van de Walle et al. (1995) where objects to be ordered are functions and

[1] To avoid the ambiguity on the term μ denoting membership and the usual μ denoting the mean, here the mean is denoted as η

the criteria considered are weighted using modifiers (Zadeh 1972). The use of modifiers relaxes the need to impose a quantitative weighting and instead also permits expressing the weights qualitatively (with the modifier being responsible to quantify the qualitative appreciation), to an extent more natural to humans. Also, equalization enhances differentiation among the elements to be ordered. Further, a bottom element is added as a reference ensuring an order is always generated. Finally, the relation of order is based on the weighted membership of each element to the subprocesses model formed from the studied process. We now describe these changes in detail and later provide the new algorithm in full in Algorithm 1.

3.3.1 Equalization

If the membership values for the *components* obtained for a given criteria present a highly skewed distribution, this may hinder the distinction between the elements. To alleviate this, an equalization of membership values for the *components* in a criteria is applied as per Eq. 12:

$$\tilde{\mu}_{c_i} = \frac{\mu_{c_i}(y_j) - \min(\mu_{c_i})}{\max(\mu_{c_i}) - \min(\mu_{c_i} + k_{eq})} \tag{12}$$

A constant k_{eq} is added to avoid degeneracy by a denominator equal to 0. Here, k_{eq} was empirically chosen to 0.05, and the influence of the choice of the constant value is not further discussed here. Notwithstanding, our internal testing suggest that if $k_{eq} \gg 0.05$, then $\bar{\mu}_{c_i}$ concentrates in low membership levels, whereas if $k_{eq} \ll 0.05$, then the weighting strategy (below) loses its capacity to affect the highest membership values (which may be an issue if not all criteria are equally important).

Note that if the opening premise does not hold (i.e., the distribution of memberships is not skewed), the equalization is harmless.

3.3.2 Lifting

A lower bound bottom $\perp$ is an element that by definition precedes every other element in an order (Santana and Santiago 2013). The bottom element itself is a theoretical entity; it is a reference that allows to have a smallest element in common. *Lifting* adds $\perp$ to an ordering (Gunter 1985). Specifically, the lifting operation redefines set Y as $Y \leftarrow \{Y, \perp\}$. $\perp$ is further used to calculate the depth and centrality of elements in a resulting order graph (Hasse diagram). Properties such as continuity or density of the set are not affected by lifting (Abramsky and Jung 1994).

3.3.3 Weighting strategy

In a trivial case, all criteria have the same relevance for the relation of order. This assumption can be relaxed by considering a weighting strategy encoding the relative importance of each criterion to the final order. Weighting modulates raw memberships. It can be used to adjust relevance of common criteria to different processes. A popular way of weighting memberships is by setting an arbitrary weight vector. Such arbitrariness is both a bless (easiness of application) and a curse (lack of rigor). Hence, other options have been explored in the literature in which the weight vector may be informed by an expert, e.g., Van de Walle et al. (1995); Chuang et al. (2006). In related work (Bruggemann et al. 2011; Van de Walle et al. 1995), the data are unweighted or simply weighted by using an additive model which needs an expert to give a quantitative answer to a qualitative question. In the context of decision making, several works (Cheng et al. 1999; Hameed 2017; Dalal and Zaveri 2014) have used linguistic terms to represent the opinion of an expert and measure the importance of relative weight or sets reporting thoughtfulness, flexibility and efficiency of the proposed method as fairer with subjective valuations of the decision-makers. Hence, we propose exploiting modifiers to quantify qualitative appreciations about the criteria relevance, alleviating the demand for an explicit weighting. We tested the following:

Concentration: Concentrating a fuzzy set c_i results in a small reduction of the membership of y_j to c_i for those y_j which originally have a large grade of membership to c_i and a large reduction of membership for those y_j with low membership to c_i (Zadeh 1972). Concentration obeys Eq. 13.

$$\mu_{con(c_i)}(y_j) = [\mu_{c_i}(y_j)]^{m_c} \tag{13}$$

where m_c is any real number bigger than 1.

Dilation The effect of dilation is the opposite of that of concentration (Zadeh 1972) spreading the grade of memberships according to Eq. 14:

$$\mu_{dil(c_i)}(y_j) = [\mu_{c_i}(y_j)]^{\frac{1}{m_d}} \tag{14}$$

where m_d is any real number bigger than 1.

Contrast intensification Intensification increases/diminishes the values of $\mu_{c_i}(y_j)$ above/below a threshold according to Eq. 15 Zadeh (1972):

$$\mu_{int(c_i)}(y_j) = \begin{cases} 2(\mu_{c_i}(y_j))^2 & if\, \mu_{c_i}(y_j) \in [0, th] \\ [1 - 2(1 - \mu_{c_i}(y_j))]^2 & \text{Otherwise} \end{cases} \tag{15}$$

where $th \in \mathbb{R}$ is a given threshold.

3.4 Relation of order

Definition 1 Let $C = \{c_1, ..., c_i\}$ be the set of sorting criteria. The fuzzy set y_j is *contained* in the fuzzy set y_l (or, equivalently, y_j is a subset of y_l, or y_j is smaller than or equal than to y_l) if and only if $\forall y \in C : \mu_{c_i}(y_j) \leq \mu_{c_i}(y_l)$ (Zadeh 1965).

Kosko (1986) contends that if this inequality holds for all but just a few c_i, one can still consider y_j to be a subset of y_l to some degree known as *Fuzzy subsethood* (SH).

Definition 2 The *transitive closure* SHT of a SH relation is the smallest relation which is transitive and of which SH is a subset (Kundu 2000).

Definition 3 Let $Y = y_1, y_2, ..., y_j$ be a set of elements and B be a fuzzy set $B = (Y, \mu_B(y_j))$. The set of elements y_j that belong to the fuzzy set B at least to the degree $\alpha \in [0, 1] \in \mathbb{R}$, i.e., $\mu_B(y_j) \geq \alpha$, denoted B_α is called α-cut or α-level set.

The α-cut is a popular defuzzification approach. For a given α, the α-cut can be applied to the transitive subsethood SHT yielding SHT_α. Details of how to calculate SHT and SHT_α can be found in Burgos-Madrigal (2018).

In this work, the criteria in C defined for evaluating the functions in Y are used to model the process P of interest. The model itself is the triplet $< C, \tilde{\mu}_\perp, W_p >$. The rationale of the proposed relation r_P of order in Eq. 16 is a comparative evaluation of how far (distance d) is each function to the model.

$$r_P = R_<(y_j, y_l) = \begin{cases} y_j < y_l & d(y_j, y_l; C, \tilde{\mu}_\perp, W_p) < 0 \\ y_j > y_l & d(y_j, y_l; C, \tilde{\mu}_\perp, W_p) > 0 \\ y_j = y_l & d(y_j, y_l; C, \tilde{\mu}_\perp, W_p) = 0 \\ y_j || y_l & d(y_j, y_l; C, \tilde{\mu}_\perp, W_p)! \end{cases} \tag{16}$$

where ! denotes that it takes on an indeterminate form. We define the model-based oriented distance among functions as per Eq. 17.

$$\begin{aligned} d(y_j, y_l; C, \tilde{\mu}_\perp, W_p) &= \frac{SHT_\alpha(y_j, y_l; C, \tilde{\mu}_\perp, W_p)}{SHT_\alpha(y_l, y_j; C, \tilde{\mu}_\perp, W_p)} \\ &\quad - \frac{SHT_\alpha(y_l, y_j; C, \tilde{\mu}_\perp, W_p)}{SHT_\alpha(y_j, y_l; C, \tilde{\mu}_\perp, W_p)} \end{aligned} \tag{17}$$

where SHT_α denotes the α-cut level of subsethood (Kosko 1986) with a transitive closure already applied which is the smallest fuzzy relation which is transitive (Xiu et al. 2012; Klir and Yuan 1995). Equation 17 is a subtraction of ratios and expresses a distance between two components as the relative closedness to a third object, in this case the model $< C, \tilde{\mu}_\perp, W_p >$. Distance d is signed or oriented; it is positive if the second operand is a closer candidate to the model $< C, \tilde{\mu}_\perp, W_p >$ than the first as dictated by the membership functions and weightings.

Cut sets SHT_α are preorders. In order to obtain a symmetric relation (partial ordering), a relation of equivalence can be defined by collecting together indistinguishable functions under cut α according to Eq. 18.

$$[y_j]_\alpha \geq [y_l]_\alpha \Leftrightarrow (y_j, y_l) \in SHT_\alpha \text{ while } (y_l, y_j) \notin SHT_\alpha \tag{18}$$

Finally, the sorting algorithm is presented in Algorithm 1.

4 Properties of the order relation

By definition, a relation of order has to comply with the properties of reflexivity, transitiveness and antisymmetry. For fuzzy relations, (Zadeh 1971) defined these properties accordingly. We here provide evidence of this being the case for our proposed relation of order.

1. **Reflexive**: For each class $[y_j]_\alpha$, we have that $[y_j]_\alpha \preceq [y_j]_\alpha$ by definition.
2. **Antisymmetric**: Let $[y_j]_\alpha$ be a membership level to one or more the criterion c_i denoted as Eq. 19

$$\forall y_j \in [y_j]_\alpha, \exists c_i \in C : \mu_{c_i}([y_j]_\alpha) > 0 \tag{19}$$

Let $[y_l]_\alpha$ be conformed only by noise and thus not a member of any criteria. This is denoted in Eq. 20:

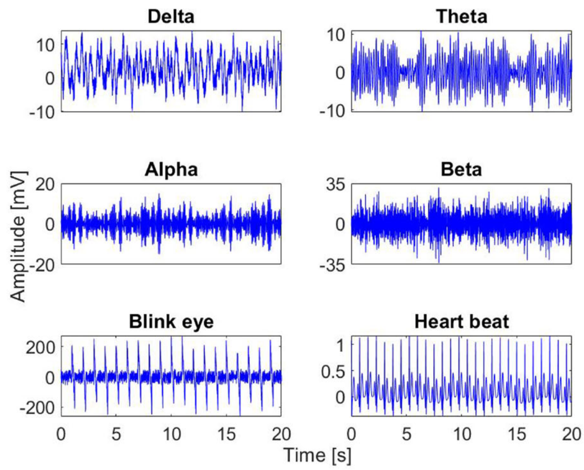
$$\forall y_l \in [y_l]_\alpha, \forall c_i \in C : \mu_{c_i}(y_l) = 0 \tag{20}$$

Therefore, $[y_j]_\alpha \succ [y_l]_\alpha \Longrightarrow [y_l]_\alpha \prec [y_j]_\alpha$
3. **Transitive**: Let $[y_m]_\alpha$ be a member of the criterion $c_k \in C$ and let $w \in W$ be the importance given to the criterion according to the process P, being $w_\rho > w_\lambda$, denoted as per Eqs. 23,21,22.

$$\forall y_j \in [y_j]_\alpha, \exists c_i \in C : w_\lambda \mu_{c_i}([y_j]_\alpha) > 0 \tag{21}$$
$$\forall y_l \in [y_l]_\alpha, \forall c_i \in C : \mu_{c_i}([y_l]_\alpha) = 0 \tag{22}$$
$$\forall y_m \in [y_m]_\alpha, \exists c_k \in C : w_\rho \mu_{c_k}([y_m]_\alpha) > 0 \tag{23}$$

If $[y_m]_\alpha \succ [y_j]_\alpha$ and $[y_j]_\alpha \succ [y_l]_\alpha$ hold, then also $[y_m]_\alpha \succ [y_l]_\alpha$.

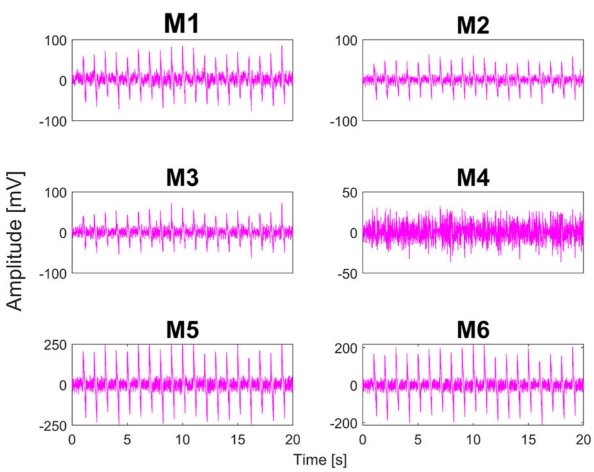
Complying with these properties implies that the relation is a *partial order*.



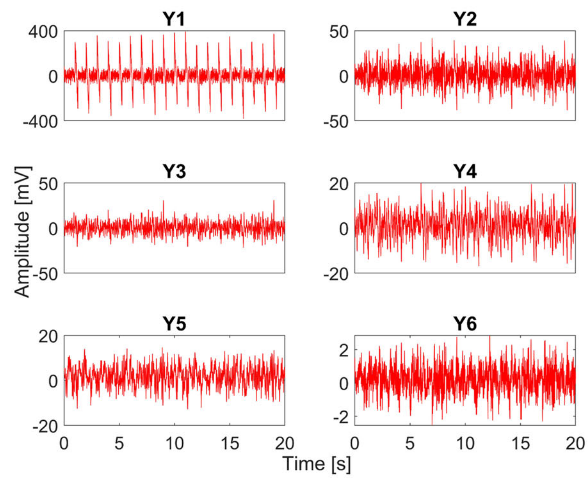
(a) Original signals X simulating the sources of interest.

Mixture matrix						
	y_1	y_2	y_3	y_4	y_5	y_6
c_1	1	0	0.8	0	0.1	0.3
c_2	0.2	1	-0.6	0	0	0.15
c_3	-0.1	0.7	1	0.2	0	0
c_4	0	0.6	0	1	-0.2	0.4
c_5	0	0.6	0	0.5	1	0
c_6	-0.5	0	0.1	0	0.75	1

(b) Example of a mixture matrix A used to generate the mixed signals.



(c) The mixed signals resulting from the mixing AX.



(d) Approximation to the components found by ICA. These form our synthetic set Y.

Fig. 2 Steps taken in the generation of synthetic data

5 Experiments and results

5.1 Synthetic data

We generated synthetic data emulating EEG signals. The raw synthetic signals in the dataset correspond to the expected sources of an EEG measurement (Fig. 2a). The frequency bands were formed by adding sinusoidal signals in the frequency ranges of interest, whereas the blink eyed and heart beat time courses were generated based on literature models (Chambayil et al. 2010; McSharry et al. 2003). A detailed description of the process for generating these synthetic signals has been given in Burgos-Madrigal (2018). These sources were later mixed using a mixture matrix (Fig. 2c) resulting in the mixed signals (Fig. 2c). The mixture matrix is varied across the experiment replications by adding a random matrix with a slack value of 0.63 to the base mixing matrix. Finally, to simulate synthetic components Y, the mixed signals are separated by the infomax ICA algorithm of Bell and Sejnowski (1995) already implemented in MATLAB® (R2017A, MathWorks, USA) as `runica`.

5.2 Experimental data

Datasets from two previously published neuroscientific experiments were considered. The first experiment is a brain computer interface exercise guided by *imagined speech* Torres-García et al. (2016). The second experiment studies brain response to an *attention* task in human–computer interaction research Soto et al. (2014). Details about these datasets

Algorithm 1 Ordering with WFMCA. In blue, changes to the original FCMA algorithm.

Inputs: Components $Y = \{y_1, ..., y_j\}$, Criteria $C = \{c_1, ...c_i\}$, Memberships functions (μ_{c_i}), Weighting strategy: W_p
for all $c_i \in C$ **do**
  for all $y_j \in Y$ **do**
    $\tilde{\mu}(i, j) \leftarrow \mu_{c_i}(y_j)$ {Evaluate memberships. Membership functions proposed: Knowledge based $\mu^K_{c_i}(y_j)$, Prototype based $\mu^P_{c_i}(y_j)$ and Distribution based $\mu^D_{c_i}(y_j)$. Other membership functions can also be used.}
  end for
  $\tilde{\mu}_{c_i} \leftarrow equalizeMembershipValues(\tilde{\mu}_{c_i})$
end for
$\tilde{\mu}_{c_i,\perp} \leftarrow lift(\tilde{\mu}_{c_i}, \perp)$
$\bar{\mu}_{\perp} \leftarrow W_p(C; \tilde{\mu}_{\perp})$ {Alternative weighting strategies suggested here}
for all $\bar{y}_j \in Y$ **do**
  for all $\bar{y}_l \in Y$ **do**
    $SH(j, l) \leftarrow subsethood(\bar{y}_j, \bar{y}_l; \tilde{\mu}_{\perp})$ {The subsethood matrix is not symmetric}
  end for
end for
$SHT \leftarrow transitiveClosure(SH)$
$\alpha \leftarrow estimateOptimalAlpha(SHT)$ {By Otsu method}
$SHT_\alpha \leftarrow defuzzification(\alpha, SHT)$
$EquivalenceClasses \leftarrow r_P(Y; SHT_\alpha)$
$[\bar{y}]_\alpha \leftarrow extractEquivalenceClasses(SHT_\alpha, EquivalenceClasses)$
$SHTE_\alpha \leftarrow calculateFuzzyZetaMatrix([\bar{y}]_\alpha)$
$J_r(Y) \leftarrow G_{DAG,r_P} < Y, SHTE_\alpha >$

can be found in the original publications but we describe them here briefly.

BCI imagined speech dataset (Torres-García et al. 2016). EEG signals belonging to 21 subjects were recorded during an imagined speech task in Spanish. Recordings were conformed of 33 repetitions of each word: 'derecha', 'izquierda', 'arriba', 'abajo' and 'seleccionar'. EEG signals were acquired at 128 [Hz] using a typical block design at 14 channels from 10–20 standard locations. The signals coming from those channels were filtered using a band-pass FIR filter at 4-25 [Hz], and the block data windows were adjusted to 256 samples from stimulus onsets. ICA was applied, allowing the components to be ordered. Noise-related components were discarded using the Hurst method (Torres-García et al. 2016), in essence an ordering operation, before signal reconstruction. During the experiment, related to classification, features were obtained from the denoised reconstructed signals using discrete wavelet transform (DWT) to train and assess the classifiers Random Forest (RF), Bagging and Boosting. Results were evaluated with common measures classification performance measures: accuracy, precision, specificity and sensitivity.

Attention dataset (Soto et al. 2014). The dataset corresponds to Chilean children from fifth to seventh grade while solving math problems using an intelligent tutoring system. EEG signals were obtained using a Biosemi ActiveTwo system with 32 scalp electrodes in a 10–20 extended setup at 2048 [Hz]. In contrast to the imagined speech dataset, these EEG recordings are not block designed but rather they correspond to continuous monitoring during unstructured stimulation. During the execution of the task, subjects got distracted occasionally. Two coders blind to each other manually labeled the moments when subjects were paying attention and when they were not. The number of instances of attention is considerably greater than the instances of distraction. Hence, to compensate for class unbalance during the classification exercise below, the instances of distraction considered included cases with 1 or more seconds of duration, while the attention task was considered from 2 seconds. In the original work, a time-frequency-topography (TFT) plot suggests that this was the analysis carried out, and classification of attention/distraction episodes was not attempted. Here, we proceed with ICA.

5.3 Statistical analysis

In all cases, statistical hypothesis testing was carried out in R (i386 3.3.2, R Foundation for Statistical Computing, Vienna, Austria). Significance level was set at $\alpha = 5\%$. Further details are given below for each experiment.

5.4 Assessment membership level functions

Convergent (high when it should be high) and divergent (low when it should be low) validity of membership functions was examined. Synthetic functions with known high and low loads of the construct were generated for each criteria as described in Sect. 5.1. Eighty replications were executed using a variation from the mixture matrix shown (Fig. 2b). Qualitative comparison between ideal sources and exemplary synthetic functions (e.g., components) in the frequency domain is shown (Fig. 3) for one replication. During the experiment, matching between ideal sources and synthetic

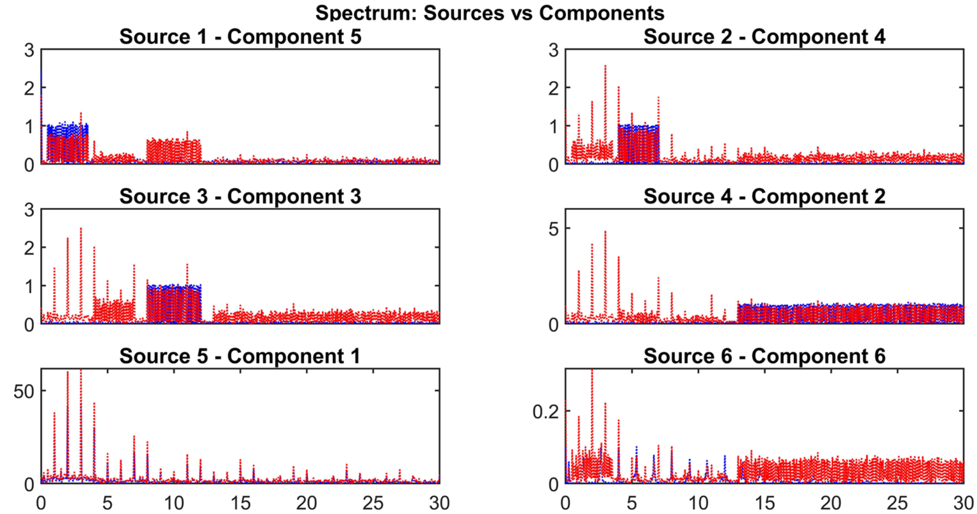


Fig. 3 Visual comparison between synthetic sources and functions (i.e., ICA-retrieved components). The *Sources* $S = \{s_1, ..., s_6\}$ are marked in blue, and the components $Y = \{y_1, ..., y_6\}$ are in red. Some functions y_j exhibit a high contribution from a particular source. In a real scenario, the functions or components are scrambled and optimal match is established using the RMSE across permutations

components was resolved by permutations observing which has the less root mean square error (RMSE) to the known mixing matrix construct. That is, the first signal source was compared to the 6 components, then the second signal source was compared to the 5 remaining components, and so on, until all the components are aligned. Figure 4a shows the construct after the 80 replications, while Fig. 4b illustrates the membership levels after the alignment done by permutations. Finally, the equalized membership levels are observed (Fig. 4c).

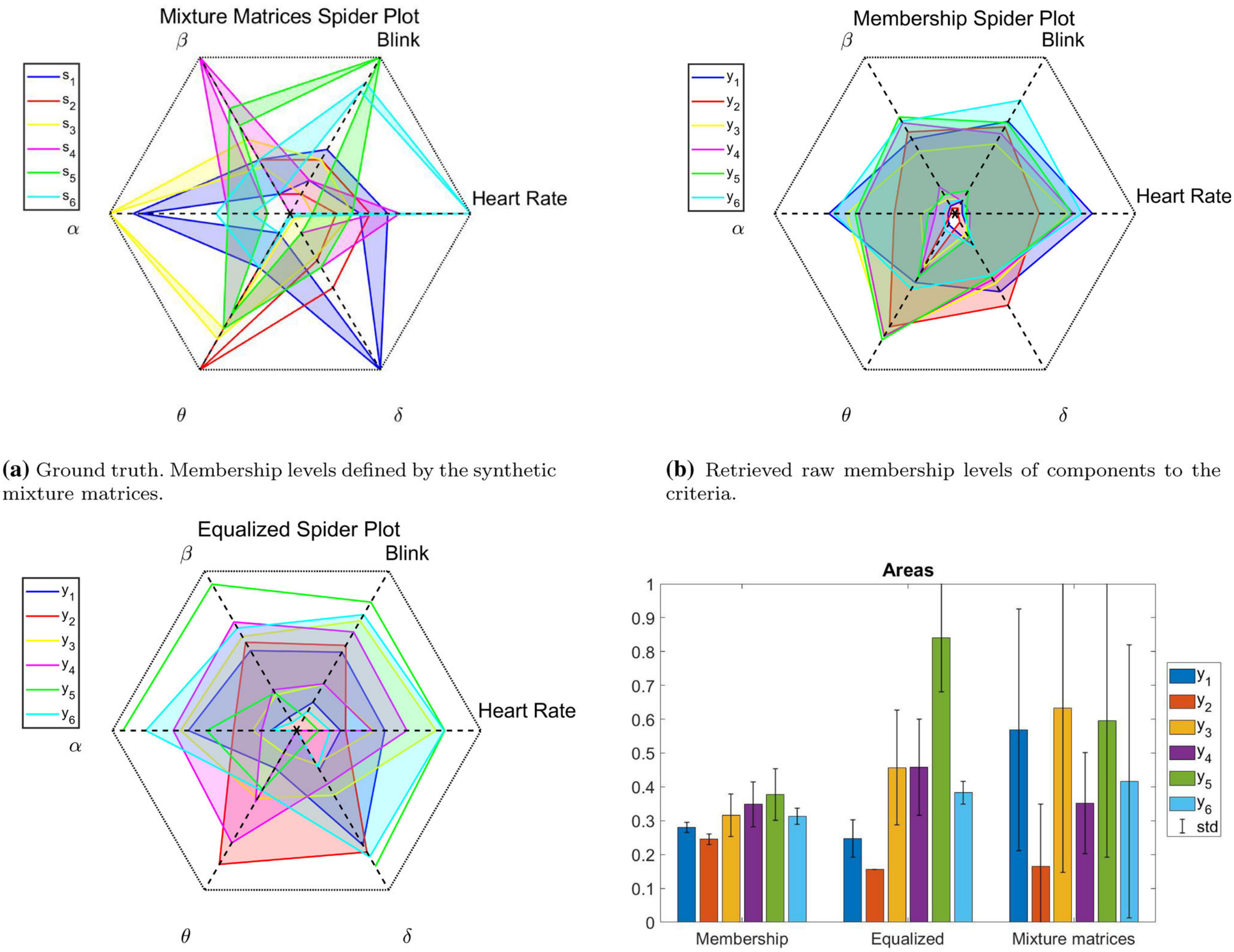
Although the components depend on how the ICA method separates the sources, Fig. 4 indicates that when the mixture matrix marked a source to be highly present in the components, this will also happen in the separation. Also, it is shown that the equalization permits to accentuate the differences between the signals in each criteria. It can be appreciated that the expected high or low membership is actually found, e.g., y_6 is expected to be lower to the δ band, and the components comply with this. To evaluate the recovered de-mixture of components with the ground truth, memberships values were plotted in spider plots (Fig. 4). The areas enclosed using the membership levels in the spider plots by the components were compared with the area enclosed by the known ground truth (Mann–Whitney U: W(6)=29, p-value=0.09307) and with the equalized values (Mann–Whitney U: W(6)=25, p-value=0.3095). Although the different components exhibited memberships according to expectations, the method of separation of the components does not permit controlling the quantity of sources truly present in each component. The total area covered by the signals already aligned shows correspondence with the mixture matrices (Fig. 4d). The standard deviation during the mixture matrices is caused by the slack value of 0.63 added to the base mixing matrix. This is diminished after the alignment during the membership values but with the equalization the deviation increases.

5.5 Tolerance to noise

Membership of the functions to the fuzzy criteria are the guiding force behind the ordering. Their values have an impact on the relation of order and thus departure from theoretical values due to noise have the potential to alter the ordering $J_r(Y)$. We tested the tolerance of our membership functions and the ordering algorithm to noise in the functions.

A synthetic signal that is originally a full member of each fuzzy set was systematically contaminated adding 5% (noise level 0.05 per signal unit) of white noise in increasing steps up to 50% (0.5 per signal unit). Twenty replications per noise level group (220 in total) were considered for each membership function. Following the procedure in Sect. 5.1, synthetic signals were created and then ordered (considering only as a criteria the source altered with white noise). The order procedure is executed adding noise to the source in every execution, and then, we observed when did the output (the resulting order) changed.

Figure 5 summarizes the effect of noise over the membership values estimated by the membership functions. It further indicates the noise threshold at which the retrieved ordering $J_r(Y)$ is altered empirically establishing an upper limit to noise tolerance of the relation of order due to noise in the different membership functions. The $\mu_{c_i}^D$ shows less stability, whereas the $\mu_{c_i}^P$ is the one less affected by the noise, decreasing slowly the membership found and changing the order until 30% of added noise. These results are affected by how the sources and prototypes are created from the same function. We have not explored other types of noise, or mixed effects when noise affects more than one membership functions at once, nor experimented beyond the criteria of interest for the domain at hand. Further experimentation may be needed for other fuzzy sets.



(a) Ground truth. Membership levels defined by the synthetic mixture matrices.

(b) Retrieved raw membership levels of components to the criteria.

(c) Equalized membership levels of components to the criteria.

(d) Mean of total area covered by the components retrieved.

Fig. 4 The components membership level to every fuzzy subset confirming the convergent and divergent capacity of the membership functions. The colored area shows the standard deviation

5.6 Effect of different weighting strategies

Next, the effect of different weighting strategies was evaluated. A target order r'_P based on the original sources of the process was generated as a comparing reference. The order is generated by calculating the similarity (cross-correlation) of a set of given components to a set of known sources considering the importance of the sources by a vector of weights (unknown to the latter ordering methods). The different weighting strategies were compared to the target order using the Jaccard index and the element-wise XOR. The Jaccard or Tanimoto index (Dehmer and Varmuza 2015) compares sets using the ratio of the intersection over the union. The XOR operation is a basic binary that returns true when input operands are different, and false otherwise. To compare matrices, it can be applied element wise.

Trends in membership values after weighting are presented (Fig. 6) showing relation between the not expert-guided (the Unweighted and the Contrast intensification) and the expert-guided (Additive model and the Linguistic hedges). Note how expert-guided weightings are similar among them.

Ordering differences are quantified (Fig. 7). The better matching was obtained for the Additive model (likely consequence of how the target order was generated), followed by the Linguistic hedges which we showed (Fig. 6). Differences were deemed significant for the Jaccard index (ANOVA: $F(2, N{=}80) = 5.86, p{=}3.029e^{-8}$) and the element-wise XOR (ANOVA: $F(2, N{=}80) = 6.48, p{=}1.997e^{-10}$) using the adjacency matrices associated with these order graphs. Pairwise Tukey post hoc analysis suggests that the significant differences are between the Additive model vs the Contrast

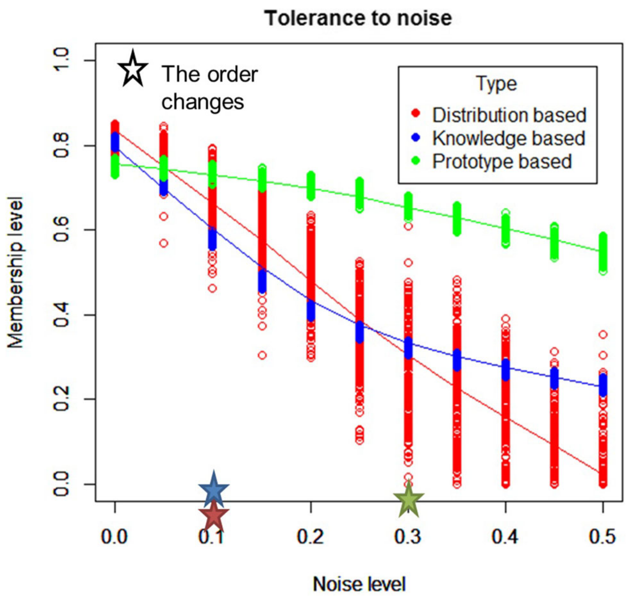


Fig. 5 Decay of membership values with increasing noise. The star symbol indicates when the original ordering is first altered

intensifier and the Additive model vs the Unweighted strategies, while results for the Additive model are not significantly different from those of the Linguistic hedges. The fuzzy modifiers (Contrast intensifier and Linguistic hedges) are more stable in the presence of noise with smaller inter-quantile ranges. Taking into account both aspects, Linguistic hedges appear to be the better choice.

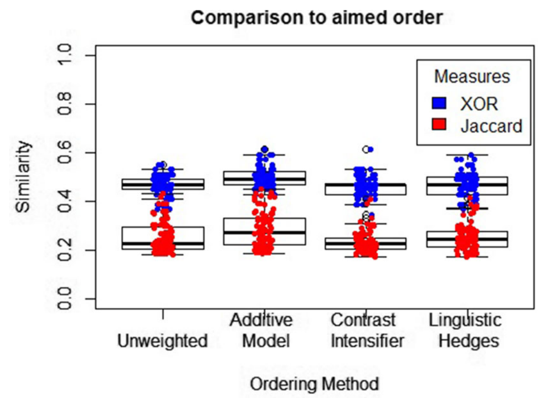


Fig. 7 The best ordering (higher median) was obtained for the Additive model (Van de Walle et al. 1995) but this is not significantly different from the Linguistic hedges. Fuzzy modifiers exhibit lower dispersion suggesting higher stability under noise

5.7 Performance in a EEG denoising application

The relation of order has been evaluated on a real application in electroencephalography. Datasets are described in Sect. 5.2. Evaluation of the performance of the new relation of order in experimental datasets was established according to two criteria; first, using a wrapped evaluation in terms of the usefulness for a subsequent classification exercise which many readers may be interested in, and second and more inherently importantly for the relation of order itself, in terms of the interpretability/expressivity of the Hasse diagram.

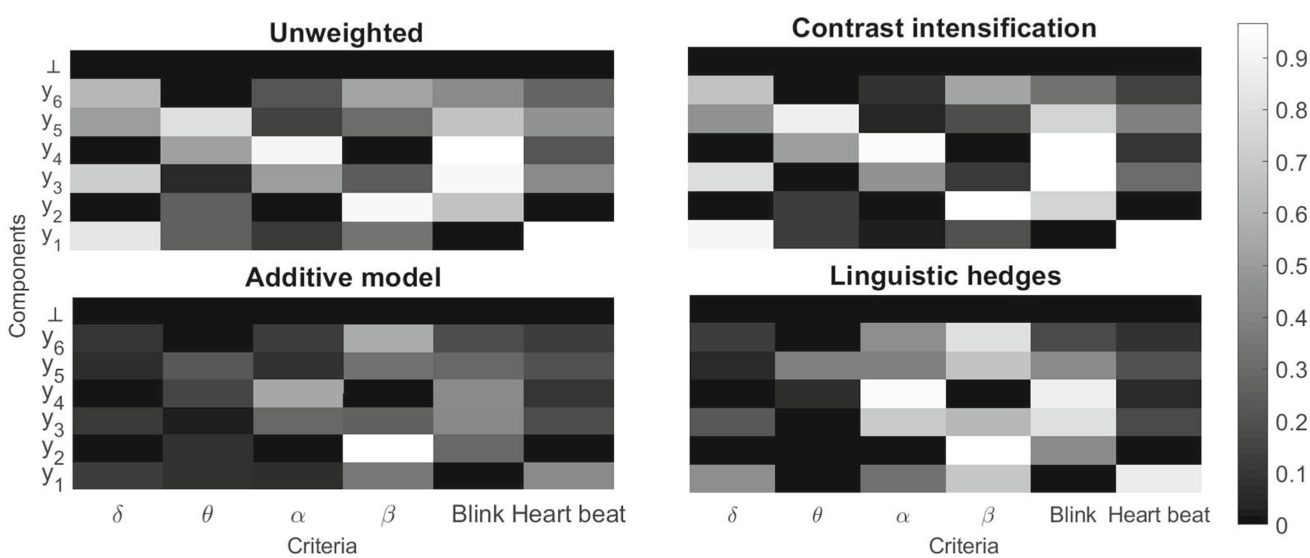
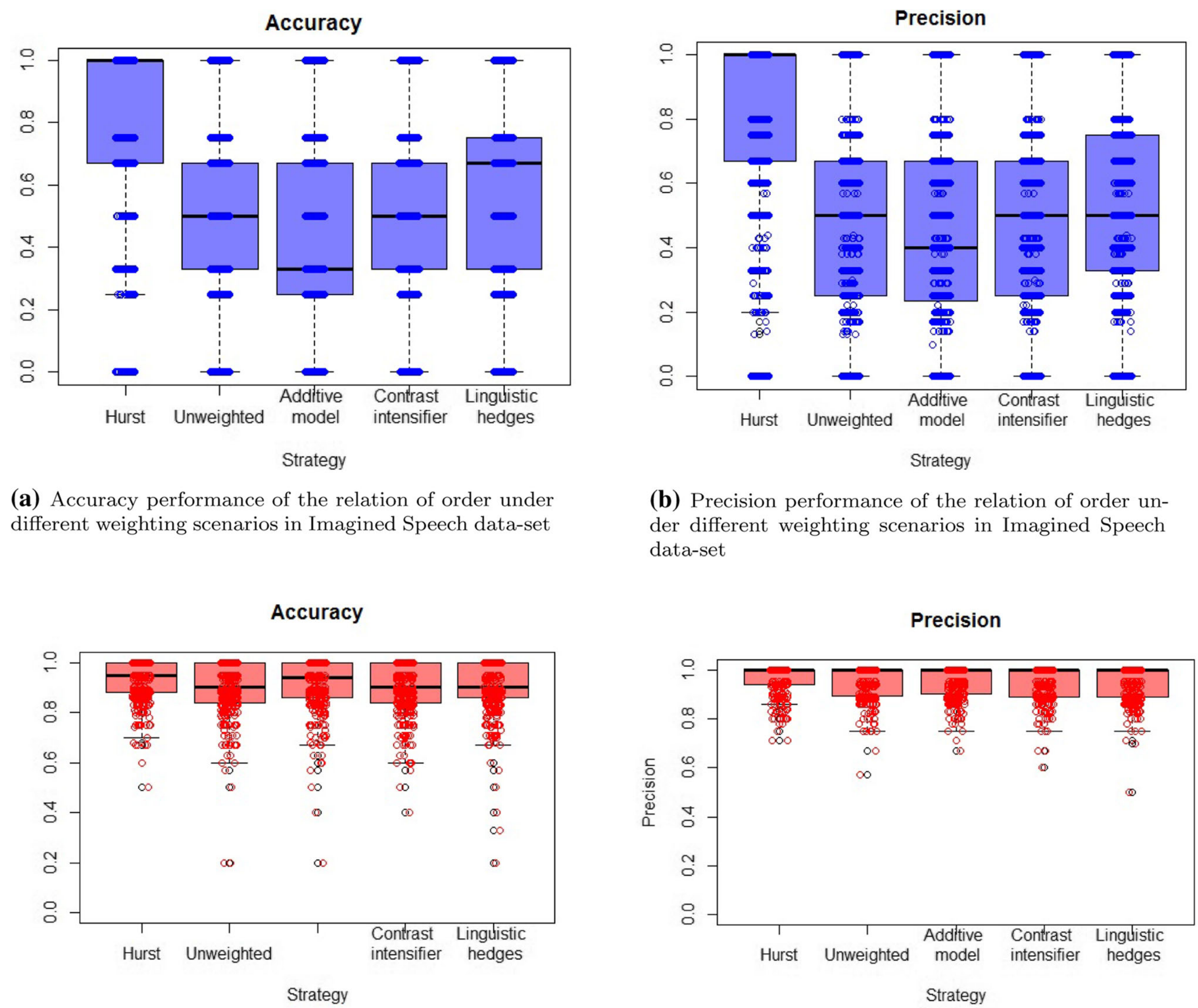


Fig. 6 Membership levels by the weighting strategies Unweighted (Bruggemann et al. 2011), Additive model (Van de Walle et al. 1995), Linguistic hedges



(a) Accuracy performance of the relation of order under different weighting scenarios in Imagined Speech data-set

(b) Precision performance of the relation of order under different weighting scenarios in Imagined Speech data-set

(c) Accuracy performance of the relation of order under different weighting scenarios in Attention data-set

(d) Precision performance of the relation of order under different weighting scenarios in Attention data-set

Fig. 8 Performance of the relation of order under different weightings (including the Hurst method used in (Torres-García et al. 2016), the Unweighted strategy (Bruggemann et al. 2011) and the Additive model taken from (Van de Walle et al. 1995)) as support for a denoising preprocessing step for subsequent classification for Imagined speech (in blue) and Attention task (in red) dataset

5.7.1 Performance within a classification problem

The performance of the relation of order as a guidance for a de-noising preprocessing step for subsequent classification was established using a wrapping strategy. Performance metrics (accuracy, precision, sensitivity and specificity) correspond to the classification exercise—not the relation of order or the output of the ordering itself.

Before classifying, the components retrieved are ordered by modifying the mixing matrix W^{-1} by multiplying the matrix with the importance I_i determined for the i component, depending on the order retrieved. Then, the signals are reconstructed using the resulting order. To do so, Eq. 24 measures the importance based on the degree of centrality C^D and depth d of the elements:

$$I_i = \frac{\frac{\sum_{j \in G} a_{ij}}{N-2} + d}{2} \tag{24}$$

where G corresponds to the graph, a_{ij} the adjacency matrix and $N-2$ considers the i node and the $\perp$ element.

Figure 8 summarizes the classification accuracy and precision results using Random Forest with the configuration given for default (bag size percent of 100, batch size of 100, 1 execution slot and 100 iterations and 5 seeds) in Weka (3.8.1, The University of Waikato, Hamilton, New Zealand) following $k = 10$ cross-folded replications per subject. In the imagined speech scenario, the Hurst method (*ad-hoc* for this dataset) obtained the best classification rates. Among the more generic sorting of the components, the Linguistic hedges exhibit the higher classification rates than the other weighting strategies.

Figures 9 , 10 summarize the ROC analysis using different classifiers for both datasets: Random Forest, Bagging (based on REPtree classifier with 10 iterations) and Boosting (based on Decision Stump with 10 iterations) for both replication schemes; cross-folding and LOOCV. The Hurst method is included for reference, and its natural advantage can be appreciated (remember that it was specifically picked for classification, not for ordering). Since discriminative power may be guided by information unrelated to the process, higher classification rates do not equate to a better description of the phenomenon (as shown in Sect. 5.7.2). Among the ordering-specific methods, the Linguistic hedges proved to be the one that better selected information of interest that actually helped also to discriminate.

Changes in accuracy were significant for both experimental datasets, the Imagined speech (ANOVA: $F(5, 1350) = 240.9$, $p = 2e^{-16}$) and the Attention scenario (ANOVA: $F(5, 400) = 5.128$, $p = 4.13e^{-4}$). Analogously, differences in precision were found to be significant for either scenario: Imagined speech (ANOVA: $F(5, 1350) = 301.9$, $p = 2e^{-16}$) and Attention (ANOVA: $F(5, 400) = 4.753$, $p = 8.11e^{-4}$). The Tukey post hoc comparison indicated that the groups exhibiting significant differences were Contrast modifier vs Unweighted strategy for the Imagined speech and the Contrast modifier vs Additive model as well as Linguistic hedges vs Unweighted, and Additive model vs Contrast modifier in the Attention dataset.

In summary, in datasets similar to the Imagined speech which are well controlled with balanced classes and similar tasks (all of them are imagined speech), the difference among the weighting strategies is low because they depart from (reasonably) homogeneous patterns. For less controlled recordings, such as the Attention dataset, differences become less pronounced.

5.7.2 Interpretability of the order

Orders obtained during the analysis of the experimental datasets are exemplified (Fig. 11). The resulting order is discussed qualitatively and quantified in terms of its richness R proposed as an indicator of spread of a tree affected by the number of leaves l, number of nodes N, the out-degree D considered as the maximum number of children of any given node and the height h of the tree (Thareja 2011). It is operationalized using Eq. 25 where $T = 4$ is the total number of metrics that were consider to conform R, and c is the number of components needed to normalize the results of each metric:

$$R = \frac{\frac{l}{(c+1)c} + \frac{N}{c+1} + \frac{D}{c} + \frac{h}{c+1}}{T} \tag{25}$$

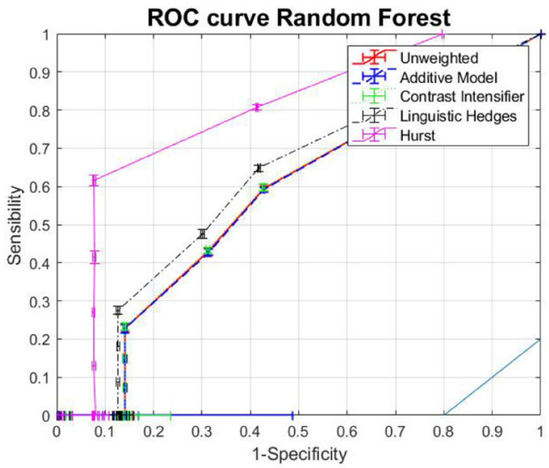
R is illustrated and quantified as an example (Fig. 11). The criteria that influence a branch is specified. Observe that when the criteria have higher priority, it generates more branches than the ones with less priority. Later, (Fig. 12) we show the statistics for all the orders generated for each dataset. We included the orderings retrieved with the Hurst approach for reference purposes. The higher expressiveness (proxy of interpretability) of the orderings retrieved with WFMCA over the Hurst-based ordering is self-evident.

In the Attention dataset (ANOVA: $F(4, 2114) = 1063$, $p = 24e^{-16}$), pairwise Tukey post hoc analysis suggests there are significant differences between all the methods over $1e^{-4}$. The equivalence classes had more difficulties to identify differences among the components. This is close related to the high number of components retrieved for this dataset, having less differences among them. The highest R was achieved by the Linguistic hedges which are weighted by the expert guided prioritization of the criteria but not limited in an arbitrary weighting vector. Nevertheless, it is also the strategy with highest dispersion which can be a consequence of finding more problems; errors of measures, distractions of the subjects or involuntary movements.

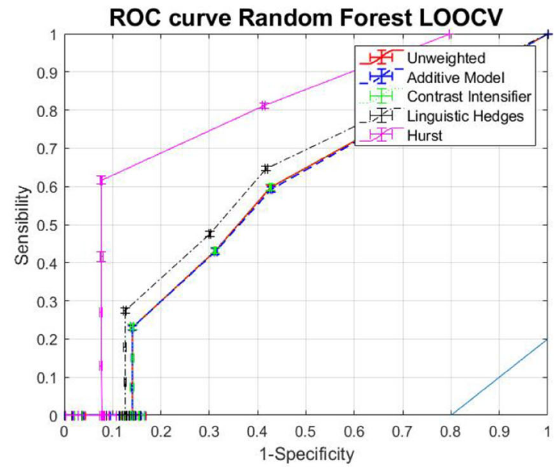
In the Imagined speech dataset (ANOVA: $F(4, 1350) = 18.22$, $p = 2.74e^{-11}$), pairwise Tukey post hoc analysis suggests that the significant differences are Additive Model$<$ (Contrast Modifiers, Unweighted) $<$ Linguistic Hedges. The dispersion incremented due to a number of factors including inherently low signal-to-noise ratio (SNR) and the presence of artifacts which can dominate and obscure the actual cortical signals (Brigham and Kumar 2010).

Nevertheless, it reaches highest R than the Attention dataset which can be a consequence of the activity being studied. In the imagined speech, the components analyzed are all the same type, while in the Attention dataset the components from the "No" class are too different between them and with fewer samples than the components from the "yes" class.

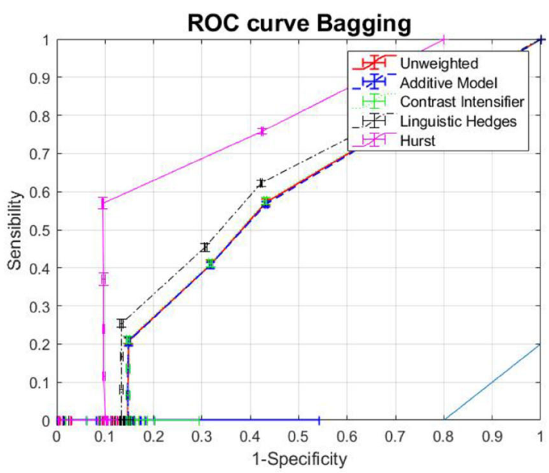
The highest R was achieved by the Additive model being one of the expert guided strategies closely followed by a non-expert-guided strategy and Linguistic hedges at the end despite been strategies that give the same importance to the criteria. That is, the R retrieved resulted high but the importance of each criteria actually has no influence in the decision. This gives indications that the criteria to be considered should be extended and studied in more detail. In fact, some studies



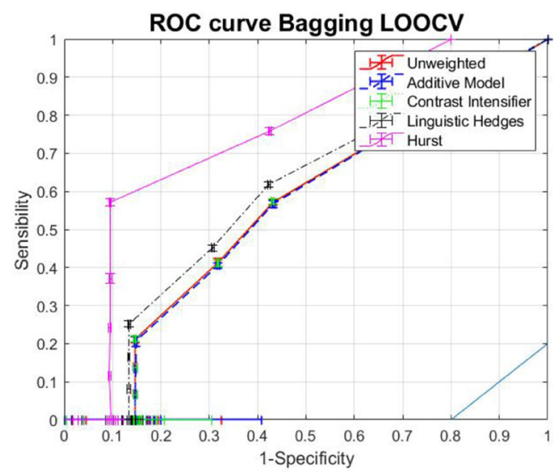
(a) ROC curve using as classifier Random Forest with the replication scheme Cross-folding (k=10)



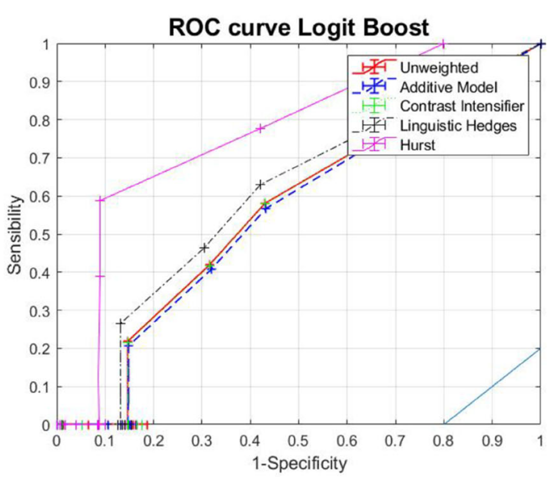
(b) ROC curve using as classifier Random Forest with the replication scheme LOOCV



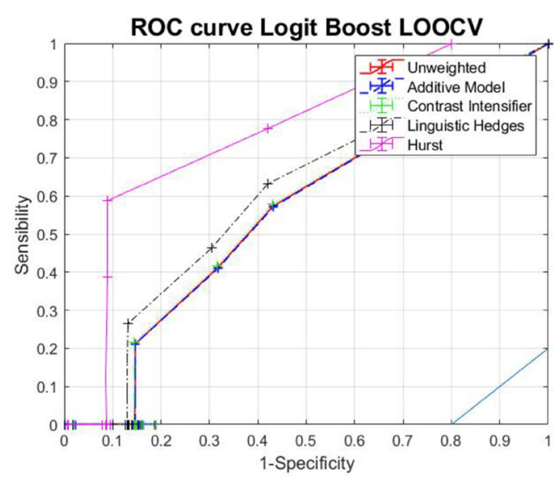
(c) ROC curve using as classifier Bagging with the replication scheme Cross-folding (k=10)



(d) ROC curve using as classifier Bagging with the replication scheme LOOCV

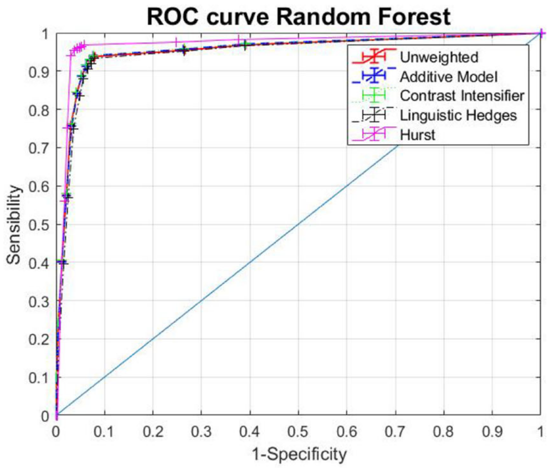


(e) ROC curve using as classifier Logit Boost with the replication scheme Cross-folding (k=10)

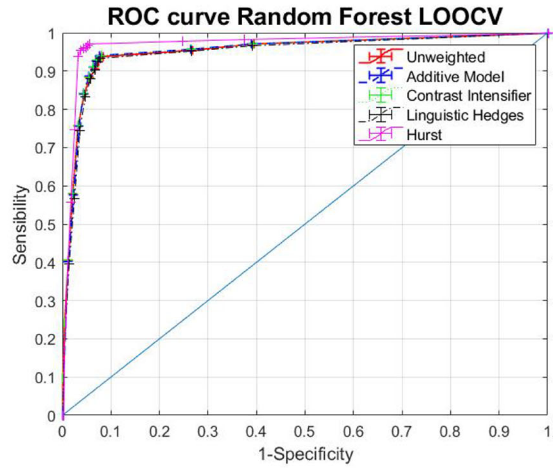


(f) ROC curve using as classifier Random Forest with the replication scheme LOOCV

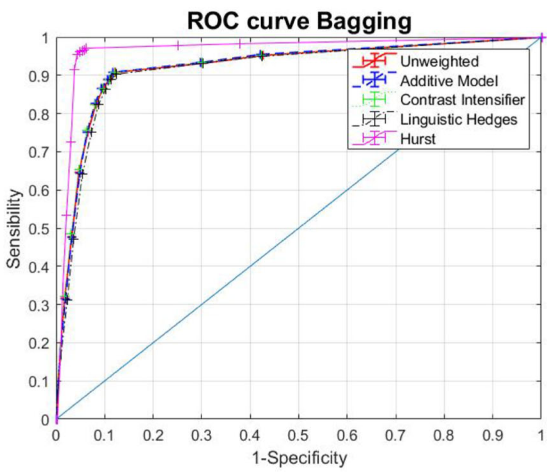
Fig. 9 Sensitivity and specificity in Imagined speech using Random Forest, Bagging and Boosting as wrapping classifiers for two replication schemes. Left column: cross-folding (k=10). Right column: LOOCV



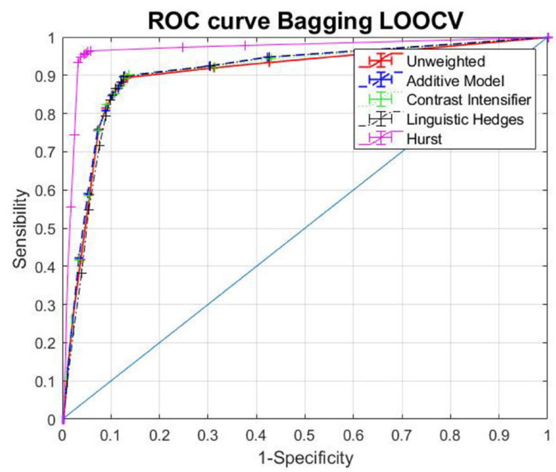
(a) ROC curve using as classifier Random Forest with the replication scheme Cross-folding (k=10)



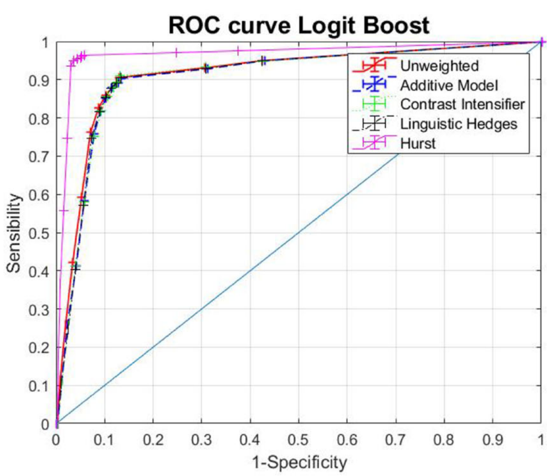
(b) ROC curve using as classifier Random Forest with the replication scheme LOOCV



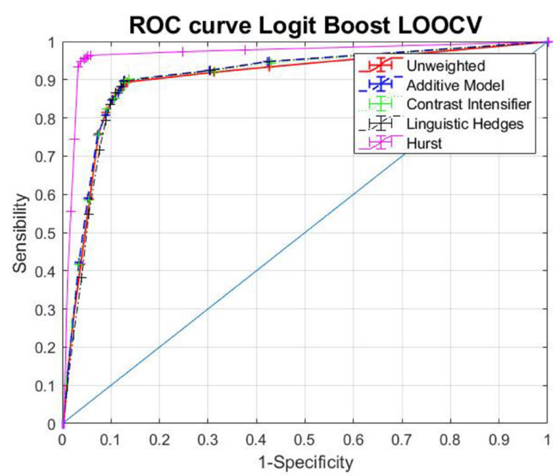
(c) ROC curve using as classifier Bagging with the replication scheme Cross-folding (k=10)



(d) ROC curve using as classifier Bagging with the replication scheme LOOCV



(e) ROC curve using as classifier Logit Boost with the replication scheme Cross-folding (k=10)



(f) ROC curve using as classifier Logit Boost with the replication scheme LOOCV

Fig. 10 Sensitivity and specificity in Attention using Random Forest, Bagging and Boosting as wrapping classifiers for two replication schemes. Left column: cross-folding (k=10). Right column: LOOCV

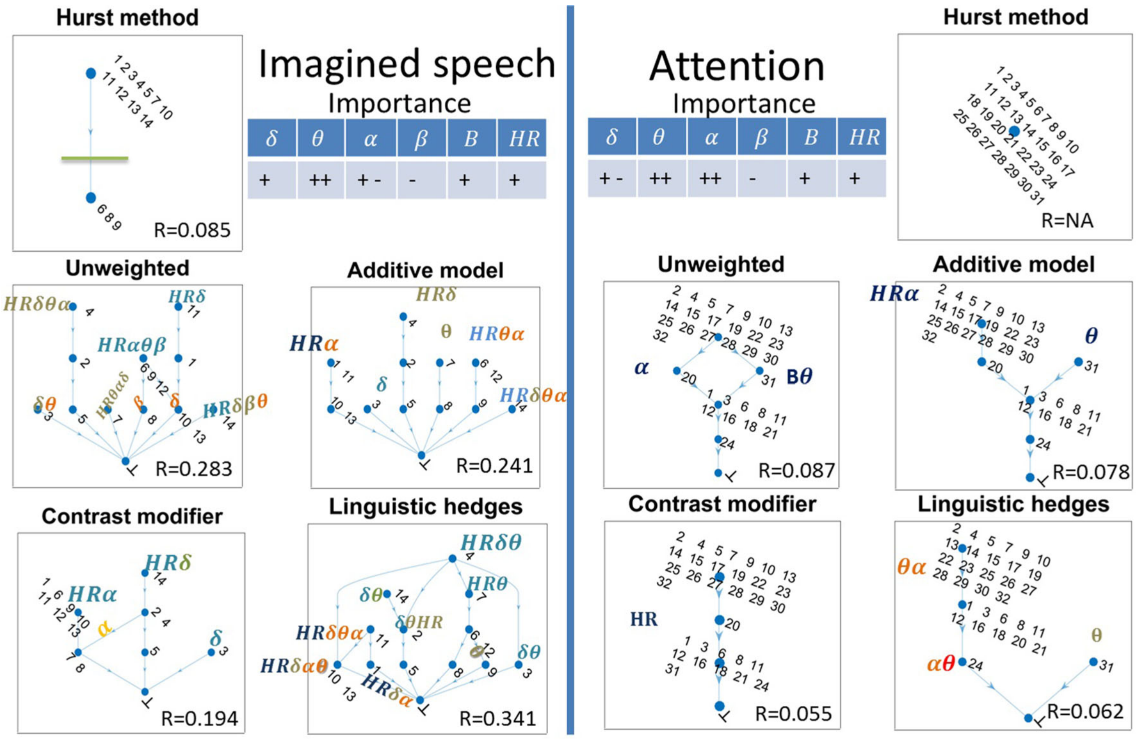


Fig. 11 Hasse diagrams for different scenarios and strategies. The criteria dominating the subtree is highlighted. Left: data from the Imagined speech dataset (Subject 10; class 'right'; instance 33). Right: data from the Attention dataset (Subject 4; class; 'attention present'; instance 12)

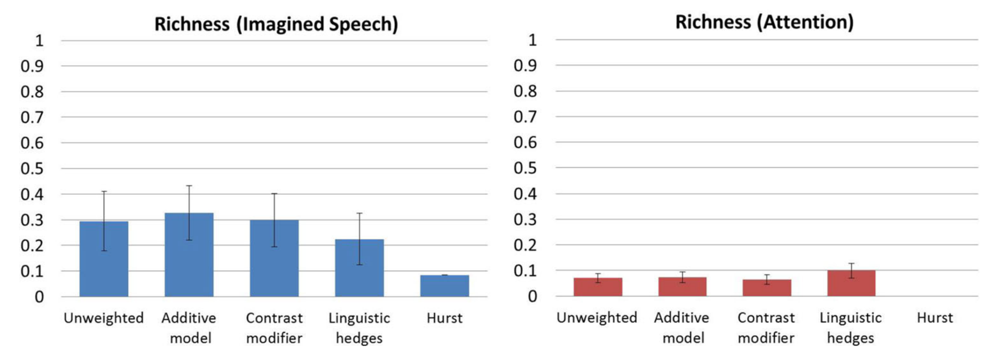


Fig. 12 Richness of resulting orders across datasets by the different methods

suggested that the frequency band activated during Imagined speech is from 50 to 150 [Hz] (ranges not considered during this study) (Perrone-Bertolotti et al. 2014) or for perception of speech (Lachaux et al. 2007) which can be involved with inner speech or imaginary speech.

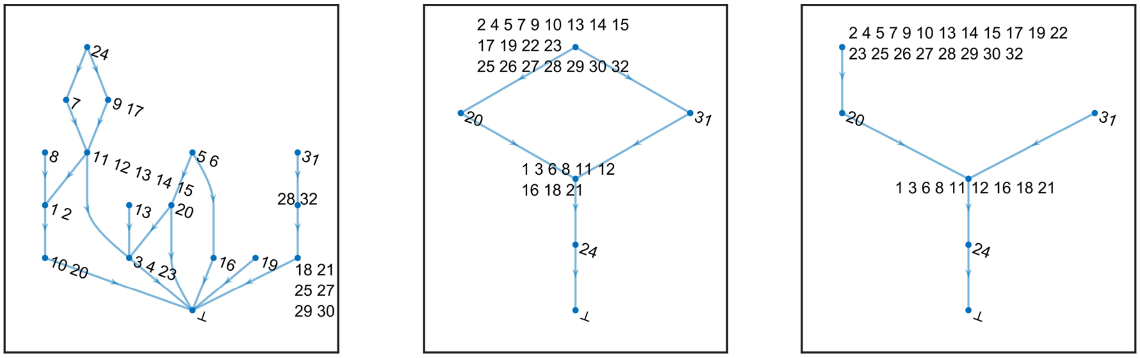
When the Hurst method succeeds in discriminating components, it produces a straightforward order with two possible equivalence classes: the ones to be discarded and the ones to be kept. This is regardless of their semantics in the domain at hand, for instance the physiological meaning in our case. This may be convenient for classification purposes, but may be inadequate in many other applications.

The prioritization of criteria helps to create groups of components during the ordering that appear to be meaningful for the domain.

5.7.3 Comparison against ordering from a human expert

Ordering components is a tangled task even for experts. When an expert attempts to order a set of instances, he defines some rules, criteria and ranges for himself and tries to apply these as objectively as possible. We further carried out a single test where a human expert in neuroimaging was asked to manually rank ICA components obtained from one of the EEG recordings.

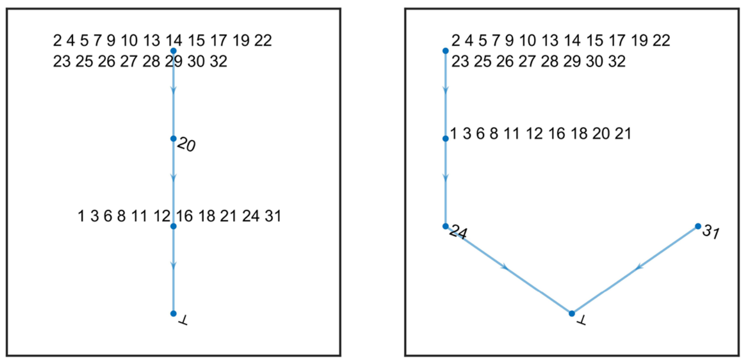
Figure 13 illustrates the Hasse diagram created manually by an expert in the Attention dataset and compares it to the orders retrieved by the automatic strategies. The process done by the expert took a couple of hours while he evaluated the components observing the signal in the frequency/time domain (plots provided), generated equivalence classes and



(a) Expert ordering

(b) Unweighted ordering (Bruggemann et al., 2011) achieved with XOR:0.4086 and with Jaccard: 0.2388

(c) Additive model ordering achiever with XOR: 0.4224 and with Jaccard: 0.2416 (Van de Walle et al., 1995)



(d) Contrast modifier ordering achieved with XOR: 0.3930 and with Jaccard: 0.2356

(e) Linguistic hedges ordering achiever with XOR: 0.4215 and with Jaccard: 0.2408

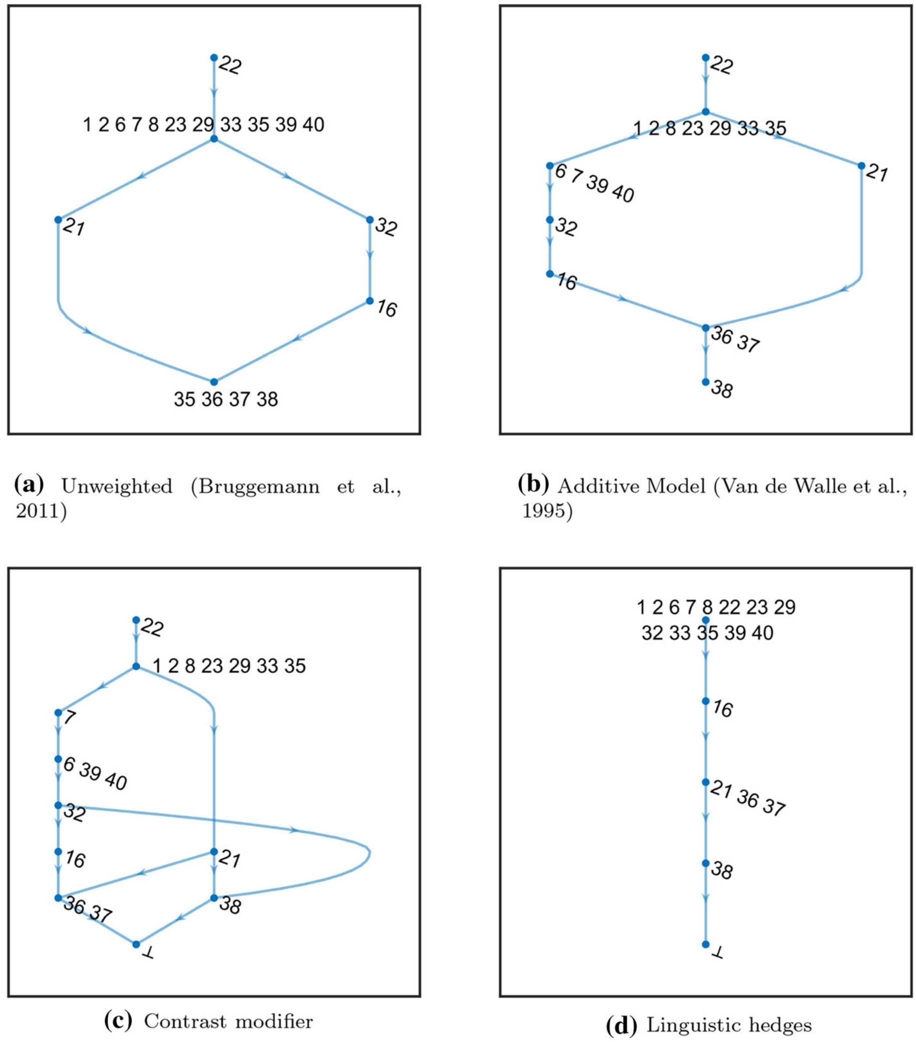
Fig. 13 Comparison between the Hasse diagram generated automatically by the proposed methods and the one generated manually by an expert

generated the corresponding Hasse diagram that represents the decided order. This contrasts with the orders retrieved in minutes (processor: Intel (R) Core (TM) i5-3337U CPU 1.8 [GHz] Memory 4096 [MB] RAM) by the strategies explained in this work. Further, in personal communications the expert declared difficulties taking hard decisions whilst carrying out the ranking ultimately introducing some bias due to subjectiveness on applying his personal decision rules. The higher values for the XOR and Jaccard metrics were achieved by the weightings proposed based on the expert-guided prioritization of the criteria showing that adding knowledge about the process helps to improve the automatic orders generation. Although the automatic strategies retrieved orders with less branches and less differences detected between the components, the difficulties may be caused because of the criteria considered.

5.8 Generalization capabilities

Despite the original motivation arising from electroencephalographic analysis, the development of the theoretical solution was independent of such application. The formal abstraction and the mathematical proof of the properties of ordering made no assumption about the domain. Hence, the proposed method should be in principle useful in other problems of the same nature. This section hints the capability of generalization by testing the approach in a domain different from its motivating scenario.

In Bruggemann et al. (2011) research, the application is to evaluate refrigerants according to its ozone depletion potential (ODP), its global warming potential (GWP) and its atmospheric lifetime (ALT)—sorting criteria. We ordered the data from Bruggemann et al. (2011) applying α-cut at 0.3. Bruggemann et al. (2011) departed from normalized values, rather than memberships. In order to mimic such condition, here the membership values are not calculated, and instead are calculated by the Otsu method (Otsu 1979). To mitigate potential degeneracies from membership values being too close, these were spread out by equalizing the memberships acquired in each criteria. Further, lifting was applied. In Bruggemann et al. (2011), all the criteria are considered to have the same importance. On the other hand, Van de Walle



(a) Unweighted (Bruggemann et al., 2011)

(b) Additive Model (Van de Walle et al., 1995)

(c) Contrast modifier

(d) Linguistic hedges

Fig. 14 Ordering on a refrigerants domain application (Bruggemann et al. 2011) (α-cut= 0.3). Observe that the bottom element that we proposed to add is only integrated in (c) and (d)

et al. (1995) shows an Additive model to prioritize the criteria.

Figure 14 shows the application of our approach to a domain of refrigerants. The Additive model and the Linguistic hedges are giving preference to the criteria as follows: ODP(+),GWP(+/-) and ALT(-), while the Contrast modifier estimates the best threshold to prioritize the higher values. The element 22 is very high in the criteria of ODP and, as a result, in all the methods is on the top. Something similar happens with the elements 1, 2, 8, 33 and 35. Other elements such as 16, 21, 32 changed their relation depending on which criteria prioritization takes action.

The Additive model does not have many differences with the Unweighted ordering (observe that the weighting strategy changes the importance to elements such as 6, 7, 39 and 40). The Linguistic hedges are finding relation between all the elements generating a total order. Finally, the highest richness in this scenario was achieved by the Contrast modifier because of the closeness between too many elements. The best method ultimately depends on the application.

6 Discussion

The relation of order proposed is based on a model of the process being studied. Applicability is exemplified on a specific domain, namely electroencephalography. But because formalization of the problem is generic, applicability to other domains is also possible. We have hinted this above. Domain knowledge about the process produces the specific sorting criteria, and in turn determines which associated membership functions should be used depending on the information available. Prioritizing the criteria depending on an expert knowledge has been shown here to help in the distinction of functions and their ordering. The Linguistic hedges weighting has shown better performance overall, both under synthetic testing and later with experimental data in the EEG domain exhibiting higher R, but this apparent dominance cannot be considered universal as suggested by results in a different domain.

We studied how does the related works of Unweighted (Bruggemann et al. 2011) and the Additive model (Van de

Walle et al. 1995) perform compared to the proposed fuzzy weighting strategies. Linguistic hedges and the Additive model are two strategies that discriminate information by understanding the importance of each criteria and it resulted not being significantly different between them, but the fuzzy weighting strategies are more stable to noise and more flexible. Nevertheless, the application also contributes to decide which technique should be better. Imagined Speech is a dataset where the differences between classes are very dim and, not surprisingly, it looks to perform better the Linguistic hedges but, in the Attention dataset where it is easy to identify when the subject loses attention, it seems to be better the Contrast modifiers due to a better separation with ICA. Also, we observed in the ROC curves that having the best classification performance it does not always reflects describing better the problem. The Hurst threshold only separates the data in two, and it is the one with best performance. However, Linguistic hedges outperforms the other weighting strategies in the Imagined Speech scenario highlighting the potential of understanding the problem and the relevance of the signals. This results can be improved by integrating more criteria to the process description.

Classification is an interesting application for ordering procedures (but by no means the only one!). Classification itself can be boosted with ad hoc efforts, but these may lack explicative power or lack clear interpretation as shown here. The relation of order introduced here yields an output which encodes process-related information. For other applications, ordering does not discard any information (there is no loss), but instead it only modulates importance depending on its relevance to the process.

The proposed method has three main advantages. First, membership functions are defined based on the information available. Other membership functions can be developed, but the method will not be affected in essence. Second, the fuzzy weighting proposed here is more flexible than simply applying a scalar as done in other works, in other words, the Additive model. Note that a scalar is just a specific case of a function. Last, the procedure of ordering permits to discern and prioritize between more than one options in multicriteria conditions which is also a difficulty for a human expert.

7 Conclusions

This research has presented a new relation of order for ordering functions under multiple fuzzy criteria and extended an algorithm to generate the orderings. We have operationalized all the necessary elements including new membership functions and modeled an assumed generative model. We have further given evidence of compliance with a partial relation of order properties, and studied some of its properties (tolerance to noise, richness, etc.) under several scenarios. The new membership functions were validated. The proposed ordering algorithm permits ranking a set of functions capitalizing on information known about the process studied. Applicability is limited by the need of knowledge about the process, but the given formalization is domain agnostic. Although we have exemplified the use of the new relation of order in electroencephalography, it is potentially useful and applicable in other domains. Future work shall consider moving towards total orders, outputting fuzzy orderings, and supporting ordering of fuzzy functions. All of these have particular relevance for electrophysiology.

Funding This study was funded by CONACyT (426901) and also been partly funded by CONACYT under grant CB-2014-01-237251.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval It does not apply directly to our study but we thank the authors of the original experimental datasets, Torres-García et al. (2016) where informed consent was obtained from all individual participants included in the study and Soto et al. (2014) where all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee, for allowing us to use them here. For replicability, codes are available at: https://github.com/burgosmad/FuzzyOrder

References

Abramsky S, Jung A (1994) Domain theory. In: Handbook of logic in computer science, Oxford University Press

Annoni P, Fattore M, Bruggemann R (2008) Analyzing the structure of poverty by fuzzy partial order. Partial orders, ambiguities and applied issues, Multicriteria ordering and ranking, pp 107–124

Bell AJ, Sejnowski TJ (1995) An information-maximization approach to blind separation and blind deconvolution. Neural Comput 7(6):1129–1159

Bellman RE, Zadeh LA (1970) Decision-making in a fuzzy environment. Manag Sci 17(4)B–141

Bortolan G, Degani R (1985) A review of some methods for ranking fuzzy subsets. Fuzzy Sets Syst 15(1):1–19

Box GE, Cox DR (1964) An analysis of transformations. J Royal Stat Soc Series B (Methodological) pp 211–252

Brigham K, Kumar BV (2010) Imagined speech classification with eeg signals for silent communication: a preliminary investigation into synthetic telepathy. In: Bioinformatics and biomedical engineering (iCBBE), 2010 4th International Conference on, IEEE, pp 1–4

Bruggemann R, Kerber A, Restrepo G (2011) Ranking objects using fuzzy orders, with an application to refrigerants. MATCH Commun Math Comput Chem 66:581–603

Burgos-Madrigal A (2018) Ordering of functions according to multiple fuzzy criteria: application to electroencephalography. Instituto Nacional de Astrofísica Óptica y Electrónica

Chambayil B, Singla R, Jha R (2010) Virtual keyboard BCI using Eye blinks in EEG. In: Wireless and mobile computing, networking and communications (WiMob), 2010 IEEE 6th International conference on, IEEE, pp 466–470

Cheng CH, Yang KL, Hwang CL (1999) Evaluating attack helicopters by ahp based on linguistic variable weight. Eur J Op Res 116(2):423–435

Chuang KS, Tzeng HL, Chen S, Wu J, Chen TJ (2006) Fuzzy c-means clustering with spatial information for image segmentation. Comput Med imag Gr 30(1):9–15

Dalal MK, Zaveri MA (2014) Opinion mining from online user reviews using fuzzy linguistic hedges. Appl Comput Intell Soft Comput 2014:2

Dehmer M, Varmuza K (2015) A comparative analysis of the tanimoto index and graph edit distance for measuring the topological similarity of trees. Appl Math Comput 259:242–250

Figueroa-García JC, Chalco-Cano Y, Román-Flores H (2018) Yager index and ranking for interval type-2 fuzzy numbers. IEEE Trans Fuzzy Syst 26(5):2709–2718

Früh L, Kampen H, Kerkow A, Schaub GA, Walther D, Wieland R (2018) Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations. Ecol Model 388:136–144

Garg H, Kumar K (2019) Improved possibility degree method for ranking intuitionistic fuzzy numbers and their application in multiattribute decision-making. Granul Comput 4(2):237–247

Gunter CA (1985) A universal domain technique for profinite posets. In: International colloquium on automata, languages, and programming, Springer, pp 232–243

Hameed IA (2017) Enhanced fuzzy system for student's academic evaluation using linguistic hedges. In: Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on, IEEE, pp 1–6

Klir G, Yuan B (1995) Fuzzy sets and fuzzy logic, vol 4. Prentice hall New Jersey

Klir GJ, Yuan B (1996) Fuzzy sets and fuzzy logic: theory and applications. Possibility Theory versus Probab Theory 32(2)

Knyazev GG (2013) EEG correlates of self-referential processing. Front Hum Neurosci 7:264

Kosko B (1986) Fuzzy entropy and conditioning. Inform Sci 40(2):165–174

Kumar K, Garg H (2017) Connection number of set pair analysis based topsis method on intuitionistic fuzzy sets and their application to decision making. Applied Intelligence pp 1–8

Kundu S (2000) A representation theorem for min-transitive fuzzy relations. Fuzzy Sets Syst 109(3):453–457

Lachaux JP, Jerbi K, Bertrand O, Minotti L, Hoffmann D, Schoendorff B, Kahane P (2007) A blueprint for real-time functional mapping via human intracranial recordings. PLoS One 2(10):e1094

Leal W, Restrepo G (2019) Formal structure of periodic system of elements. Proceedings of the royal society A 475(2224):20180,581

Martinez L, Liu J, Yang JB (2006) A fuzzy model for design evaluation based on multiple criteria analysis in engineering systems. Int J Uncertain Fuzziness Knowl Based Syst 14(03):317–336

McSharry PE, Clifford GD, Tarassenko L, Smith LA (2003) A dynamical model for generating synthetic electrocardiogram signals. IEEE Trans Biomed Eng 50(3):289–294

Moctezuma LA, Torres-García AA, Villaseñor-Pineda L, Carrillo M (2019) Subjects identification using eeg-recorded imagined speech. Exp Syst Appl 118:201–208

Niazy RK, Beckmann CF, Iannetti GD, Brady JM, Smith SM (2005) Removal of fmri environment artifacts from eeg data using optimal basis sets. Neuroimage 28(3):720–737

Otsu N (1979) A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern 9(1):62–66

Perrone-Bertolotti M, Rapin L, Lachaux JP, Baciu M, Loevenbruck H (2014) What is that little voice inside my head? inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behav Brain Res 261:220–239

Santana F, Santiago R (2013) Interval metrics, topology and continuous functions. Comput Appl Math 32(3):459–470

Smith SW, et al. (1997) Digital signal processors. In: The scientist and engineer's guide to digital signal processing, California Technical Pub. San Diego

Soto A, Orihuela-Espina F, Cosmelli D, Alcholado C, Heyer P, Sucar LE (2014) Intelligent tutors exploiting novel sensing modalities for decoding students' attention. In: Workshop on Utilizing EEG Input in Intelligent Tutoring Systems (ITS2014 WSEEG), p 35

Sucar LE (2015) Probabilistic graphical models: principles and applications. Springer, New York

Thareja R (2011) Data Structures Using C. Oxford University Press, Inc

Torres-García AA, Reyes-García CA, Villaseñor-Pineda L, García-Aguilar G (2016) Implementing a fuzzy inference system in a multi-objective eeg channel selection model for imagined speech classification. Exp Syst Appl 59:1–12

Vigário RN (1997) Extraction of ocular artefacts from eeg using independent component analysis. Electroencephalogr Clin Neurophysiol 103(3):395–404

Van de Walle B, De Baets B, Kerre EE (1995) Fuzzy multi-criteria analysis of cutting techniques in a nuclear reactor dismantling project. Fuzzy Sets Syst 74(1):115–126

Wang H, Xu Z, Zeng XJ (2018a) Hesitant fuzzy linguistic term sets for linguistic decision making: current developments, issues and challenges. Inform Fus 43:1–12

Wang H, Xu Z, Zeng XJ (2018b) Linguistic terms with weakened hedges: a model for qualitative decision making under uncertainty. Inform Sci 433:37–54

Wang X, Kerre EE (2001) Reasonable properties for the ordering of fuzzy quantities (ii). Fuzzy Sets Syst 118(3):387–405

Wibowo S, Grandhi S (2018) Fuzzy multicriteria analysis for performance evaluation of internet-of-things-based supply chains. Symmetry 10(11):603

Wu D, Mendel JM (2009) A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. Inform Sci 179(8):1169–1192

Xing Z, Xiong W, Liu H (2018) A euclidean approach for ranking intuitionistic fuzzy values. IEEE Trans Fuzzy Syst 26(1):353–365

Xiu Z, Wang W, Wang Y (2012) An approach to rank targets using fuzzy partial orderings. In: System science and engineering (ICSSE), 2012 International conference on, IEEE, pp 470–475

Young VR (1996) Fuzzy subsethood. Fuzzy Sets Syst 77(3):371–384

Zadeh LA (1965) Fuzzy sets. Inform Control 8(3):338–353

Zadeh LA (1971) Similarity relations and fuzzy orderings. Inform Sci 3(2):177–200

Zadeh LA (1972) A fuzzy-set-theoretic interpretation of linguistic hedges. J Cybernet 2(2):4–34

Zeng C, Yang L, Zhu A et al (2017) Construction of membership functions for soil mapping using the partial dependence of soil on environmental covariates calculated by random forest. Soil Sci Soc Am J 81(2):341–353

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.