Fully Optimized Discrimination of Physiological Responses to Auditory Stimuli

Authors: Stepan Y. Kruglikov [1,2], Sharmila Chari [1], Paul E. Rapp [3], Steven L. Weinstein[1,6], Barbara K. Given [1,4], Steven J. Schiff [1,5.7]

Affiliations: [1]Krasnow Institute for Advanced Study, [2]School of Computational Sciences, [4]Graduate School of Education, and [5] Department of Psychology and Program in Neuroscience, George Mason University, Fairfax, VA 22030, USA

[3]Department of Pharmacology and Physiology, Drexel University, Philadelphia, PA 19129, USA

[6]Department of Neurology, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010, USA

[7] Center for Neural Engineering, Departments of Neurosurgery, Engineering Science and Mechanics, and Physics, The Pennsylvania State University, University Park, PA 16802 USA

Key Words: Event Related Potential, Mismatch Negativity, Phoneme, Tone, Perception

Journal of Neural Engineering
In Press, 2008

Correspondence: Steven J. Schiff
212 Earth-Engineering Sciences Building
Pennsylvania State University
University Park, PA 16802
814-863-4210
815-550-2150(fax)
sschiff@psu.edu

ABSTRACT

The use of multivariate measurements to characterize brain activity (electrical, magnetic, optical) is widespread. The most common approaches to reduce the complexity of such observations include principal and independent component analysis (PCA and ICA), which are not well suited for discrimination tasks. We addressed two questions: first, how do the neurophysiological responses to elongated phonemes relate to tone and phoneme responses in normal children, and second, how discriminable are these responses. We employed fully optimized linear discrimination analysis to maximally separate the multi-electrode responses to tones and phonemes, and classified the response to elongated phonemes. We find that discrimination between tones and phonemes is dependent upon responses from associative regions of the brain apparently distinct from the primary sensory cortices typically emphasized by PCA or ICA, and that the neuronal correlates corresponding to elongated phonemes are highly variable in normal children (about half respond with neural correlates of tones, and half as phonemes). Our approach is made feasible by the increase in computational power of ordinary personal computers, and has significant advantages for a wide range of neuronal imaging modalities.

INTRODUCTION

The use of multivariate measurements to characterize brain activity (electrical, magnetic, optical) is now widespread. The most common approaches to reduce the complexity of such observations often include principal component analysis (PCA) (Pearson, 1901; Hotelling, 1931; Flury, 1997), and more recently, independent component analysis (ICA) (Bell A. J. and Sejnowski, 1995, Hyvarinen and Oja, 1997). These methods are designed for separating orthogonal (PCA) or statistically independent (ICA) components within multivariate data. Nevertheless, such approaches may not be well suited for certain classification tasks. Such failures might be expected to be found when the neural components generating the largest correlated signals are not, but smaller dependent combinations of signals are, the neuronal components critical for perceptual discrimination.

Linear discrimination analysis (LDA) was described by Fisher (1936) to separate species based on morphometric measurements. Fisher worked out the fundamental mathematics required to incorporate more than one variable (measurement) in each data sample (multivariate), and to admit more than 2 groups in the analysis. Once a data set whose group memberships are known have been used as a training set for LDA, one can examine unknown samples and perform classification. Classification literally involves determining in which group is a sample most likely to be a member of. In its more modern implementation (Schiff et al 2005), discrimination quality is determined by relying on normal distributional assumptions and covariance constraints which are rarely present in neurophysiological data. As computational power increases, it is now possible to fully optimize such discrimination analysis by, for instance, testing all combinations of electrodes to see which are best at discriminating responses. We will here apply such fully optimized analysis to an important problem in the cognitive physiology of dyslexia – the use of modified phonemes to retrain children with language impairments.

The neural circuitry that the brain employs to processes tones and phonemes is asymmetric. In normal right-handed individuals, the perception of language phonemes is associated with neural correlates in the left temporoparietal region, while pitch discrimination is biased to right frontotemporal structures (Kayser et al 1998). Phoneme discrimination selective to native language may form early during childhood (Kuhl, 2000), and by adulthood, native phoneme selectivity has asymmetrically left sided temporoparietal neural correlates (Naatanen et al 1997).

It has been proposed that the origin of dyslexia in certain children may be caused by deficits associated with phoneme discrimination. Tallal et al (1996) suggested that three to six percent of children are language learning impaired (LLI) due to phonological deficits in the rapid processing of speech. This deficit would lead to an inaccurate perception of fast components of speech, such as short or 'stop' consonants 'b' or 'd' in phonemes 'ba' versus 'da' in English (Tallal, 1980).

Mismatch negativity (MMN), first described in Naatanen et al (1978), is the difference between the late event related potential (ERP) amplitude detected with an occasionally presented stimulus (target, infrequent event) and the response amplitude to a commonly presented stimulus (non-target, frequent event). Kraus et al (1996) showed that the MMN associated with /da/-/ga/ auditory oddball experiments is decreased in language learning disabled children with normal intelligence.

It has been reported that children with LLI can show significant improvements in phoneme discrimination following training (Merzenich et al 1996, Tallal et al 1996) using software which elongates and emphasizes the fast transitions in speech (Fast ForWord (FF) Scientific Learning Corp, Berkeley). Despite the widespread use of such language training, we know little about the physiology of elongated phoneme perception. Although one might suppose that an individual with a deficit in processing fast components of speech might perceive elongated phonemes as 'phonemes', these computer synthesized sounds differ substantially from normal language components.

To learn more about the comparative physiology of tones, phonemes and elongated phonemes, we conducted measurements of mismatch negativity for these stimuli incorporating phonemes actually employed in FF training software (with permission of Scientific Learning Corp.). We specifically addressed two questions: first, how do the neurophysiological responses to elongated phonemes relate to tone and phoneme responses in normal children, and second, how discriminable are these responses. Our supposition was that the artificial nature of elongated phonemes might contribute to highly variable responses, with some subjects perceiving them as tones, and others as language phonemes. If this were so, then perhaps the difference in response would be related to the efficacy of training, and the use of a physiological screening test might be predictive of training effectiveness.

Previous attempts to contrast tone and phoneme responses have used techniques such as PCA to establish significant hemispheric asymmetries in ERP responses (Kayser et al 1998). We here employ LDA (Fisher, 1936) to maximally separate the multielectrode responses to tones and phonemes, and then use this discrimination to classify the response to elongated phonemes.

In a recent study, using a different algorithmic strategy, a sub-exhaustive search of sensor pairs was employed in order to discriminate magnetoencephalography correlations reflective of a variety of disease states (Georgopoulos et al 2007). To our knowledge, no one has employed fully optimized strategies to apply LDA to such problems. With exhaustive searching, we find that discrimination between tones and phonemes best relies upon responses from electrodes over associative regions of the brain apparently distinct from the primary sensory cortices typically selected by PCA or ICA, and that the neuronal correlates corresponding to elongated phonemes are highly variable in normal children (about half respond with neural correlates of tones, and half as phonemes). Our fully optimized approach is one made feasible in recent years by the increase in computational power of ordinary personal computers, and has significant applicability for a range of neuronal imaging modalities.

METHODS

Subjects

Thirty-four child volunteers 11 to 13 years old were recruited for the study from public schools located in a suburban area of Washington, D.C. All children were psychometrically tested for IQ, phonemic awareness, receptive language, and basic reading skills. Eighteen right-handed (Edinburgh inventory) children (10 females), having all standardized psychometric clustered or total scores above 84 were selected for further investigation. Subjects were paid 20 $US per visit. This protocol was approved by the Institutional Review Board of George Mason University.

Experimental Procedures

Following parental informed consent and assent, subjects had a 32-channel EEG cap applied (Neuroscan Inc.), including two bipolar (vertical and horizontal) extraocular electrodes as well as linked ear lobe references. Conductive gel (Quick-Gel, Neuroscan Inc.) was inserted beneath each electrode, and gentle abrasion used to obtain impedances below 10 kOhm. EEG signals were amplified 1000 times (Neuroscan amplifiers), bandpass filtered between 0.1 to 200 Hz, and digitized at 1000 Hz.

Matched earphones, calibrated for loudness, were inserted into the auditory canals. A screening test was performed for normal auditory acuity defined as pure-tone threshold <25dB sound pressure level for frequencies of 500 and 1000 Hz (Bellis et al, 2000). Possible ocular and muscular artifacts were demonstrated to the subject by showing the effects on real time EEG. EEG caps were disinfected with 2.5% glutaraldehyde solution between uses.

Protocol

An active attention (motor response required) tonal oddball experiment employing 440 Hz and 880 Hz tones was presented using 300 ms duration stimuli. Hanning windowing with 5 ms rise/fall was applied to the stimuli. A total of 200 presentations, with 20% randomly interspersed deviant target stimuli, were presented followed by another block of 200 presentations where the targets and non-target frequencies were reversed. In each block, the subject was told to press a button when hearing a deviant stimulus with the right hand index finger for the first 100 presentations, and after a short break, respond to the target with the left hand index finger. Interstimulus interval was randomly jittered between 1.8 and 2.2 sec. The duration and the average loudness root mean square (RMS) of the tones, phonemes and elongated phonemes were set to be equal (Matlab™ was used for signal processing). This was achieved by calculating RMS of all stimuli and then multiplying the waveforms of each stimulus by the ratio of RMS of one of the stimuli (we chose elongated ABA) to the RMS of the stimulus being adjusted. We show in Figure 1 examples of the tones, normal phonemic, and elongated phonemic stimuli, as well as difference plots to illustrate the differences between the normal and elongated phonemes. The difference plots show the differing lengths of the initial vowel in normal and elongated phonemic stimuli, which were used to set the varying intervals to start the mismatch negativity averaging below – it is only the differing consonants which create the 'oddball' perception we seek.

Normal phoneme oddball presentation of /aba/ and /ada/ was performed in a similar manner to the tonal oddballs. Phonemic stimuli were matched for duration, and again, an active balanced design with two blocks of 200 presentations (100 for each hand), using 20% targets with the targets counterbalanced between the two blocks, was employed as above.

Elongated phoneme oddball presentation of /aba/ and /ada/ was performed using actual elongated /aba/ and /ada/ from FF software, adjusted for RMS, with the identical protocol procedures used for normal phoneme oddball task.

<u>Data reduction</u>

Epochs were defined as segments of EEG 200 ms before and 823 ms after the stimulus onset, so the total epoch length was 1024 ms. Epochs were subject to artifact rejection (if any electrode exceeded ±75μV), band-pass filtered from 1 Hz (-24 dB/oct) to 30 Hz (-24 dB/oct), and mean baseline adjustment performed to zero the prestimulus mean.

After the above-mentioned procedures, we calculated average waveforms separately for the targets and the non-targets. We then subtracted non-target averages from target averages, creating difference waves for each electrode and each subject. A plot of the grand average difference waves for all subjects is shown in Figure 2.

MMN magnitude for each subject was determined as average voltage difference in the window from 180 to 280 ms after stimulus onset for tones. Since normal phonemic stimuli start with the vowel ‘a’, whose duration is about 100 msec, the neural discrimination processes cannot begin before 100 msec after the stimulus onset. Elongated phonemes start with a short, approximately 50 ms, ‘a’ at the beginning of the stimulus. Therefore the MMN window for normal and elongated phonemes were shifted by 100 ms and 50 ms respectively, yielding windows for MMN between 280 to 380 for normal and 230 to 330 for elongated phonemes (Figure 1).

Analysis of variance was performed in order to identify laterality of responses for tones, phonemes and elongated phonemes. PCA was implemented using a singular value decomposition procedure (Golub, 1996; Flurry, 1997). For ICA we employed FastICA (Hyvärinen 2001).

In classification analysis, one can ask whether an unknown sample (elongated phonemes) is closer in multivariate characteristics to well defined groups (tones or phonemes). In this study we use linear discrimination analysis.

Discrimination methods

Linear Discrimination Analysis (LDA) creates a linear combination of variables (electrode data) that maximally separate the standard distance between two sets of measurements (see Appendix). A complete discussion of the theory of LDA can be found in Flury (1997), and a set of efficient matrix algorithms for the LDA implementation we employed can be found in the code archived in the supplementary material from Schiff et al. (2005).

LDA does not provide, in general, a means of determining which variables are more important for the classification. Although it is possible to calculate the significance of each variable using normal theory, this calculation requires data assumptions such as normality and equality of the covariance matrices, which may not apply to our data (Flurry, 1997). Because these assumptions may not accurately estimate the significance of variables in our data set, another approach was pursued.

We chose to calculate all possible combinations of variables and ascertain which variables (electrodes) are optimal for the classification. This procedure may also be viewed a search for a projection, where the observations of the two groups are maximally separated in terms of standard distance as defined in (Flury, 1997).

The optimal classifier (projection) was found by optimizing two types of error rates: plug-in and leave-one-out. Plug-in error rate is defined as a fraction of misclassified observations from the training set based on the classification constructed from all the samples in the training set. In other words, the same observations are used to create and test the classifier. Leave-one-out error rate is calculated by leaving out one observation, constructing the classifier and then testing the classification of the omitted observation. The procedure is repeated for all observations in the data set. Thus, leave-one-out error rate is defined as the fraction of observations that would be misclassified omitting each observation in turn.

RESULTS

Electroencephalographical MMN data were collected during an oddball experiment with tones, phonemes and elongated phonemes as described in Methods.

Before calculating the discriminant functions it is useful to confirm that there is a statistically significant difference between the two groups. We found that the significance of MANOVA analysis using 36 observations (averaged MMN amplitude in a window) from 18 subjects with stimulus as a factor (2 levels: tones, phonemes) varied as a function of the number of electrodes used. With all 25 electrodes as dependent variables, the difference was marginal ($F=2.29$; $df=25, 10$; $p=0.085$), whereas with a more optimal number of electrodes (13 electrodes), the differences are highly significant ($F=3.97$; $df=13, 22$; $p=0.002$).

We further investigated the physiological asymmetry in the responses to tones, phonemes, and elongated phonemes. All 18 subjects completed the initial tones and phonemes protocol, and 8 of 18 completed the elongated phoneme protocol. We used repeated measures ANOVA with two factors: side and electrode. We indexed each electrode on one side with the corresponding electrode on the other side. Midline electrodes were excluded from this asymmetry analysis similarly to Kayser et al (1998). Our results were analogous to the results reported earlier by Kayser et al (1998). The responses were asymmetric, with responses to tones more negative in electrodes located on the right, and with responses to phonemes more negative in electrodes located on the left. Interestingly, there was no significant asymmetry in responses to elongated phonemes.

We then investigated the optimal number of electrodes for classifications. For *n* electrodes, there are $\sum_{k=1}^{n} \frac{n!}{k!(n-k)!} = 2^n - 1$ unique combinations. Therefore we studied 2^{25}-1=33,554,431 combinations of 25 electrodes. For each combination we calculated linear discriminator functions, and computed plug-in error and leave-one-out error rates as described in Methods. In order to compute leave-one-out error for a particular combination it was necessary to calculate 36 discriminators (tones and phonemes for 18 subjects), for a total of (36+1)* 33,554,431= 1,241,513,947 discriminators. Using 2 laboratory grade personal computers, these calculations were completed in about 3 weeks (about 650 hours of processor time).

Only a few combinations of electrodes yielded the best classification. An example of optimization for 9 out of 25 electrodes is shown in Figure 3. In this Figure the vertical axis represents leave-one-out error and the horizontal axis plug-in error. Each point corresponds to the number of errors from different combination of electrodes. Note that since some combinations produced exactly the same errors, some points represent multiple combinations. Only one combination out of 2,042,975 gives the optimal classification with 2 plug-in errors and 3 leave-one-out errors. The results for all combinations of electrodes are presented in Table 1. The optimal leave-one-out errors are all higher than the plug-in errors as would be expected. The utility of employing both leave-one-out and plug-in errors for optimization is illustrated in the results of Figure 3 – there were 4 combinations with 3 leave-one-out errors, of which 3 had 3 plug-in errors, and one, the optimum combination, had 2 plug-in errors.

Discrimination between tone and phoneme observation is not feasible using only one electrode. Using any one electrode, the LDA yielded the leave-one-out error rate of

12 observations (33%). The minimum leave-one-out error rate using two electrodes dropped to 8 observations (22%).

In Table 1 we also summarize the classification of elongated phonemes. The average number of elongated phoneme observations classified as tones is 3 versus 5 classified as phonemes. Normal subjects have no clear propensity to process elongated phonemes as either tones or phonemes. An example of creating a classifier from an optimal combination of 13 electrodes between tones and phonemes is shown in Figure 4. In Figure 4A, selected electrodes are shown as solid squares and unselected electrodes are shown as hatched squares. The magnitude of the coefficients from the linear combination for each selected electrode is coded in gray scale. We used this classifier to assign group membership (tones, phonemes) to the data obtained in response to elongated phonemes. The results indicate that four out of eight elongated phoneme results were classified as tones (Figure 4B). Note that the optimal electrodes are scattered spatially over the scalp. The electrodes that are important for classification are not constrained to lie directly above the auditory cortices, but include strong contributions from frontal and parietal association areas.

In order to further assess the statistical validity of the linear discrimination we randomized the assignment of tone versus phoneme for each observation consisting of electrodes and repeated the search for the optimal combination of 13 out 25 electrodes. The randomization procedure was repeated 100 times (Figure 5). Most randomizations of the data set resulted in optimal combinations of electrodes with minimum leave-one-error around 15, whereas optimal leave-one-out error for the original dataset was 4. This bootstrap result confirms that the original LDA optimization was not a result of multicomparison overfitting.

Multivariate methods such as PCA, ICA and LDA provide a way of combining the data from multiple electrodes into a linear combination, and it is instructive to compare these combinations. Only five PCA components of a total of 25 are required to account for 95% of the variance in our data, whereas 20 out of 25 ICA components are necessary to account for the same amount of variance. This result was expected, since PCA is designed specifically for dimensionality reduction, whereas ICA is designed for unmixing linearly mixed signals.

To compare PCA, ICA and LDA, we first rewrite the data models for these methods so that the data is factorized into a product of 2 matrices, in the form

$$\mathbf{X}_{m,n}=\mathbf{U}_{m,n}\,\mathbf{B}^{-1}_{n,n}$$

or

$$\mathbf{U}_{m,n}=\mathbf{X}_{m,n}\,\mathbf{B}_{n,n}$$

where **X** is an *m*-by-*n* data matrix with *n* electrodes and *m* observations, **U** is an *m*-by-*n* matrix of linear combinations of electrode values, and **B** is an n-by-n set of linear combination coefficients. The data mean across time was removed prior to calculating the components. We then calculated the linear combinations (columns of the matrix **B**, only one column for LDA) and plotted the coefficients (weights) of linear combinations in Figure 6 in the spatial arrangement of the corresponding electrodes (and in corresponding linear bar charts).

The first component of PCA reflects the spatial mean of the voltages from the scalp electrodes. Other PCA components separate different spatial areas, such as front and back (2nd component), left and right (3rd component) and center versus sides and edges (4th component). ICA components and LDA coefficients do not have such well-defined spatial patterns.

We compared the differences in the separation of the linear combinations for PCA, ICA, and LDA, using data from tones and phonemes using a t-test. Only the LDA combination and the 3rd PCA component separated the two means, but LDA was far more effective (3rd PCA component: $p=0.0007$; LDA: $p<0.0001$; and the standard distances separating the transformed values are much greater for LDA as seen in the projected means in Figure 6). The third PCA component reflected right from left sided asymmetry, as has been noted previously (Kayser et al, 1998). Since the third PCA component separated left and right electrodes, the ability of this PCA component to separate tones and phonemes is another demonstration of the laterality of the responses to tones and phonemes.

DISCUSSION

Although LDA is in common use in analyzing multivariate data (Schiff et al 2005), we are not aware of full optimization, using all combinations of electrodes, for neuronal multielectrode data as we have described. We have demonstrated that fully optimized LDA can be highly effective in separating physiological responses to tones versus phonemes in individual subjects. Interestingly, the response to elongated phonemes is mixed among a population of normal children.

LDA shows a very different spatial pattern than PCA or ICA. In discriminating phonemes from tones, the associational cortices heavily contribute to the neural correlates of discrimination. PCA, highlighting the large asymmetric discrepancies between receptive language sites in posterior temporal left versus right brain, is not an effective discrimination tool for individual subjects. Perhaps it is the disentangling of linear combinations (blind source separation) that ICA optimizes is the reason for its poor performance in this setting – it is in the optimal linear combinations of electrode data for this task, and presumably for the relevant underlying neuronal generators, that the best discrimination is found.

We have shown that the response to elongated phonemes optimally segregates into responses that are divided between tone and phonemic responses in the subjects. Such widely varying responses to these stimuli may help explain variability in the response to elongated training paradigms for children. Our findings lead us to the hypothesis that examining a population of LLI dyslexic children for physiological responses to elongated phonemes might help identify which subjects will benefit from such training.

The use of optimal electrodes improves the classification quality, lowering the dimensionality of the data set. The electrodes that were optimal were not located directly over auditory processing areas, where the sensory evoked responses are largest bilaterally. Our findings suggest that the neural correlates from other areas are critical for the classification; indeed, responses from frontal, parietal, and central midline areas were necessary to perform this classification well. Perhaps the associational regions of cortex, where the actual perceptions are being processed, are critical signals for us to quantify discrimination from in such perceptual tasks. This said, we also recognize that our anatomical associations of brain structures without source localization limits our interpretation. Studying such subjects with high density EEG or MEG, along with using anatomically valid head models to estimate current sources, would be quite valuable.

In this study, we sought an optimal combination of electrodes to use for discrimination with an entire population of subjects as a potential screening tool. Nevertheless, our strategy is applicable to individual subjects, for instance, in brain machine interface paradigms, where one seeks the optimal electrode combination to use for cognitive discrimination for an individual subject. We will show the utility of such individual subject optimization in future reports.

Our use of full optimization has become feasible because the power to perform this computation is now practical using routine laboratory grade personal computers. Although many theoretical methods exist to test for discrimination quality (Flury 1997), data from neuronal experiments will always deviate from the constraints on distribution and covariation required for normal theory assumptions. In contrast, our use of empirical error rates and brute force full optimization yields electrode combinations from among millions of possibilities, which are most suitable for a given situation. Our results find that such discriminators are powerful enough to work on individual subjects, with very low error rates. Such an approach is of general applicability to a wide range of multivariate neuronal data such as multielectrode electrical, magnetic encephalography, optical measurements, or functional MRI.

ACKONOWLEDGMENTS

This research was supported by NIH (S.K. F31DC05651), (S.J.S. K02MH01493), and the U.S. Department of Education National Institute on Disability and Rehabilitation Research (S.K., S.C., P.R., and B.G. H133G000142).

APPENDIX. Methods of linear discrimination

The univariate standard distance Δ_Y between two numbers y_1 and y_2 with respect to random variable Y is defined as (see Flury 1997)

$$\begin{aligned} &\Delta_Y(y_1, y_2) = \frac{|y_1 - y_2|}{\mathrm{var}^{1/2}(\mathbf{Y})} \\ &\mathrm{var}(\mathbf{Y}) = E\left[(\mathbf{Y} - \bar{y})^2\right]. \\ &\bar{y} = E\left[\mathbf{Y}\right] \end{aligned} \tag{A1}$$

where E represents expectation. Standard distance may be interpreted as a one-dimensional Euclidian distance in units of standard deviation of random variable **Y**. Generalization to the p-variate case is accomplished by considering all linear combinations of p variables, i.e. reducing the problem to the univariate case. The p-variate standard distance is then defined as maximum of all univariate standard distances of linear combinations of p variables, or

$$\begin{aligned} &\Delta_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2) = \max_{\substack{\mathbf{a} \in \Re^p \\ \mathbf{a} \neq 0}} \frac{|\mathbf{a}'(\mathbf{y}_1 - \mathbf{y}_2)|}{(\mathbf{a}'\psi\mathbf{a})^{1/2}} \\ &\psi = E\left[(\mathbf{Y} - \mu)(\mathbf{Y} - \mu)'\right] \qquad , \\ &\mu = E\left[\mathbf{Y}\right] \end{aligned} \tag{A2}$$

where **a**, $\mathbf{y}_1$, $\mathbf{y}_2$ are p-variate vectors and ψ is the covariance matrix of p-variate random variable **Y**. The maximum of $\Delta_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2)$ is found by using the extended Cauchy-Schwartz inequality

$$(\mathbf{u}'\mathbf{v})^2 \leq (\mathbf{u}'\mathbf{M}\mathbf{u})(\mathbf{v}'\mathbf{M}^{-1}\mathbf{v}) \tag{A3}$$

where **u**, **v** are p-variate nonzero vectors and **M** is a positive definite symmetric matrix. The equality holds when $\mathbf{u}=c\mathbf{M}^{-1}\mathbf{v}$ or $\mathbf{v}=c\mathbf{M}\mathbf{u}$ with c, a real number. Maximizing $\Delta_{\mathbf{Y}}(\mathbf{y}_1,$

$\mathbf{y}_2$) is equivalent to maximizing $\Delta^2_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2)$ Therefore by setting $\mathbf{u}$=$\mathbf{a}$, $\mathbf{v}$=$\mathbf{y}_1$-$\mathbf{y}_2$ and $\mathbf{M}$=ψ, the maximum is found at

$$\mathbf{a} = c\,\psi^{-1}(\mathbf{y}_1 - \mathbf{y}_2) \tag{A4}$$

Conventionally constant c is set to unity, thus the p-variate standard distance is equal to

$$\Delta_{\mathbf{Y}}(\mathbf{y}_1, \mathbf{y}_2) = [(\mathbf{y}_1 - \mathbf{y}_2)'\psi^{-1}(\mathbf{y}_1 - \mathbf{y}_2)]^{1/2}\,. \tag{A5}$$

If the means of random variable $\mathbf{Y}$ are $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ in populations 1 and 2 respectively and the covariance is equal in two populations, the linear combination $\mathbf{a}$ that maximizes the standard distance between the two means is defined as the linear discriminant function (LDF) and is equal to

$$\boldsymbol{\beta} = \psi^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)\,. \tag{A6}$$

LDF may be viewed as a projection from a p-variate space to a one-dimensional real line ($\Re^p \to \Re$) that maximizes the separation of the two populations on the real line.

In practice covariance matrix ψ is estimated as a weighted sum of the covariance matrices in two populations or pooled covariance S, i.e.

$$\begin{aligned} \bar{y}_j &= \frac{1}{N_j}\sum_{i=1}^{N_j} y_{ij}, \\ S_j &= \frac{1}{N_j - 1}\sum_{i=1}^{N_j}(y_{ij} - \bar{y}_j)(y_{ij} - \bar{y}_j)',\ j = 1,\ 2 \\ S &= \frac{1}{N_1 + N_2 - 2}[(N_1 - 1)S_1 + (N_2 - 1)S_2] \end{aligned} \tag{A7}$$

where $\bar{y}_j$ and S_j are the estimates of the mean and covariance matrix in the jth population.

After the calculation of the LDF, we may choose the midpoint between the means of the projections of the data from the two populations as a cutoff point. This point corresponds to a straight line (in the case of bivariate data) or a hyper-plane (for multivariate data set) in the original variable space.

An alternative approach requires that the data from the populations follows a normal distribution, that is $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_1, \psi)$ in population 1 and $\mathbf{Y} \sim N_p(\boldsymbol{\mu}_2, \psi)$ in 2.

The probability distribution function (PDF) of the multivariate normal distribution is equal to

$$N_p(\mu, \psi) = (2\pi)^{-p/2}(\det \psi)^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mu)'\psi^{-1}(\mathbf{y}-\mu)\right], \tag{A8}$$

where $\det(\psi)$ is the determinant of ψ.

Introducing the notion of prior probabilities π_1 and π_2, and the conditional PDF of $\mathbf{Y}$, $f_j(\mathbf{y})$, given population membership j=1, 2, the posterior probability of being in the population 1 or 2 given data $\mathbf{y}$ by Bayes theorem is equal to

$$\pi_{j\mathbf{y}} = \frac{\pi_j f_j(\mathbf{y})}{\pi_1 f_1(\mathbf{y}) + \pi_2 f_2(\mathbf{y})}. \tag{A9}$$

We then shall classify an observation $\mathbf{y}$ to population 1, if $\pi_{1y} > \pi_{2y}$. This condition is equivalent to $\pi_1 f_1(\mathbf{y}) > \pi_2 f_2(\mathbf{y})$, or $\ln \frac{\pi_1 f_1(\mathbf{y})}{\pi_2 f_2(\mathbf{y})} > 0$. Since

$$\ln \frac{\pi_1 f_1(\mathbf{y})}{\pi_2 f_2(\mathbf{y})} = \ln \frac{\pi_1}{\pi_2} + \boldsymbol{\beta}'\left[y - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)\right] = \alpha + \boldsymbol{\beta}'\mathbf{y} \tag{A10}$$

it follows that the posterior probability follows logistic function

$$\pi_{1\mathbf{y}} = \frac{\exp(\alpha + \boldsymbol{\beta}'\mathbf{y})}{1 + \exp(\alpha + \boldsymbol{\beta}'\mathbf{y})}, \text{ or}$$
$$\ln \frac{\pi_{1\mathbf{y}}}{1 - \pi_{1\mathbf{y}}} = \alpha + \boldsymbol{\beta}'\mathbf{y} \tag{A11}$$

We introduce a dummy variable

$$z = \ln \frac{\pi_1}{\pi_2} + \boldsymbol{\beta}'\left[y - \frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \right] \tag{A12}$$

and calculate the posterior probability $\pi_{1\mathbf{y}} = \frac{e^z}{1 + e^z}$. Posterior probability for the population 2 is found by subtracting the result from unity.

Efficient matrix codes written in Matlab (The Mathworks, Natick, MA) for performing the above LDA can be found in the archived supplementary material from Schiff et al (2005).

REFERENCES

Bell A J and Sejnowski T J 1995 An information-maximization approach to blind separation and blind deconvolution *Neural Computations* **7** 1129-1159

Belli T J, Nicol T and Kraus N 2000 Aging affects hemispheric asymmetry in the neural representation of speech sounds *Journal of Neuroscience* **20** 791-797

Fisher R A 1936 The use of multiple measurements in taxonomic problems *Annals of Eugenics* **7** 179-188

Flury B 1997 *A first course in multivariate statistics* New York: Springer

Georgopoulos A P, Karageorgiou E, Leuthold A C, Lewis S M, Lynch J K, Alonso A A, Aslam Z, Carpenter A F, Georgopoulos A, Hemmy L S, Koutlas I G, Langheim F J P, McCarten J R, McPherson S E, Pardo J V, Pardo P J, Parry G J, Rottunda S J, Segal B M, Sponheim S R, Stanwyck J J, Stephane M, and Westermeyer J J 2007 Synchronous neural interactions assessed by magnetoencephalography: a functional biomarker for brain disorders *Journal of Neural Engineering* **4** 349–355

Glantz S S and Glantz S A 2001 *Primer of Biostatistics* (p 376) New York: McGraw Hill

Golub G H and Loan Van C H 1996 *Matrix Computations* Baltimore: The John Hopkins University Press

Hotelling H 1931 Analysis of a complex of statistical variables into principle components *Journal of Educational Psychology* **24** 417-441

Hyvarinen A and Oja E 1997 A fast fixed-point algorithm for independent component analysis *Neural Computation* **9** 1483-1492

Hyvarinen A and Oja E 2000 Independent component analysis: algorithms and applications *Neural Networks* **13** 411-430

Kayser J, Tenke C E and Bruder G E 1998 Dissociation of brain ERP topographies for tonal and phonetic oddball tasks *Psychophysiology* **35** 576-590

Kraus N, McGee T, Carrell T and Sharma A 1995 Neurophysiological basis of speech discrimination *Ear and Hearing* **28** 240-244

Kraus N McGee T J, Carrell T D, Zecher S G, Nicol T G and Koch D B 1996 Auditory neurophysiologic responses and discrimination deficits in children with learning problems *Science* **273** 971-973

Kuhl P K 2000 Language, mind and brain: experience alters perception In M S Gazzaniga (Ed), *The New Cognitive Neurosciences* (2nd ed) Cambridge: MIT Press

Makeig S, Jung T P, Bell A, Ghahremani D and Sejnowski T J 1997 Blind separation of auditory event-related brain responses into independent components *Proceedings of the National Academy of Sciences* **94** 10979-10984

Merzenich M M, Jenkins W M, Johnston P, Schreiner C, Miller S L and Tallal P 1996 Temporal processing deficits of language-learning impaired children ameliorated by training *Science* **271** 77-81

Naatanen R, Gaillard A W and Mantysalo S 1978 Early selective-attention effect on evoked potential reinterpreted *Acta psychologica* **42** 313-329

Naatanen R, Schoger E, Karakas S, Tervaniemi M and Paavilainen P 1989 Development of a memory trace for complex sounds in the human brain *NeuroReport* **4** 503-506

Naatanen T, Lehtokoski A, Lennes M, Cheoru M, Houtilainen M, Livonen A, Vainio M, Alku P, Llminiemi R H, Luuk A, Allik J, Sinkkonen J and Alho K 1997 Language-specific phoneme representations revealed by electric and magnetic brain responses *Nature* **385** 432-434

Nagarajan S, Mahncke H, Salx T, Tallal P, Roberts T and Merzenich M M 1999 Cortical auditory signal processing in poor readers *Proceedings of the National Academy of Sciences* **96** 6483-6488

Pearson K 1901 On lines and planes of closest fit to systems of points in space *Philosophical Magazine Series* **6** 559-572

Saber G A F 1984 *Multivariate observations* New York: Wiley

Schiff S J, Sauer T, Kumar R, Weinstein S L 2005 Neuronal spatiotemporal pattern discrimination: the dynamical evolution of seizures *Neuroimage* **28** 1043-55

Tallal P 1980 Auditory temporal perception, phonics and reading disabilities in children *Brain and Language* **9** 182-198

Tallal P, Miller S L, Bedi G, Byma G, Wang X, Nagarajan S S, Schreiner C, Jenkins W M and Merzenich M M 1996 Language comprehension in language-learning impaired children improved with acoustically modified speech *Science* **271** 81-84

TABLES

Number of Electrodes	Number of Combinations	Optimal Plug-in Error	Optimal Leave-One-Out Error	Number of Optimal Combi-nations	Average Elongated Phonemes Classified as Tones	Average Elongated Phonemes Classified as Phonemes
1	25	12	12	1	3	5
2	300	6	8	1	2	6
3	2300	5	6	1	2	6
4	12650	4	4	1	1	7
5	53130	2	4	1	1	7
6	177100	2	4	2	2	6
7	480700	2	3	5	1.8	6.2
8	1081575	2	3	4	2	6
9	2042975	2	3	1	2	6
10	3268760	3	3	2	1	7
11	4457400	3	3	1	1	7
12	5200300	4	4	1	2	6
13	5200300	2	4	1	4	4
14	4457400	2	4	1	4	4
15	3268760	2	5	2	4	4
16	2042975	0	6	6	4	4
17	1081575	0	6	3	4	4
18	480700	0	7	3	3.67	4.33
19	177100	0	8	3	4.67	3.33
20	53130	1	8	1	4	4
21	12650	1	9	1	4	4
22	2300	0	10	3	4.67	3.33
23	300	1	10	1	4	4
24	25	1	11	1	4	4
25	1	0	13	1	5	3
Average					**3.0**	**5.0**

Table 1. Results of search for optimal electrodes.

For a given number of electrodes sampled (column 1), we give the total number of all unique combinations of electrodes (column 2), the resulting optimal plug-in and leave-one-out errors (columns 3 and 4), the number of optimal combinations (column 5) the average number of elongated phonemes misclassified as tones (column 6), and the average number of phonemes misclassified as phonemes (column 7). On the bottom row, we give the average number of misclassifications for all electrode combinations.

FIGURE CAPTIONS

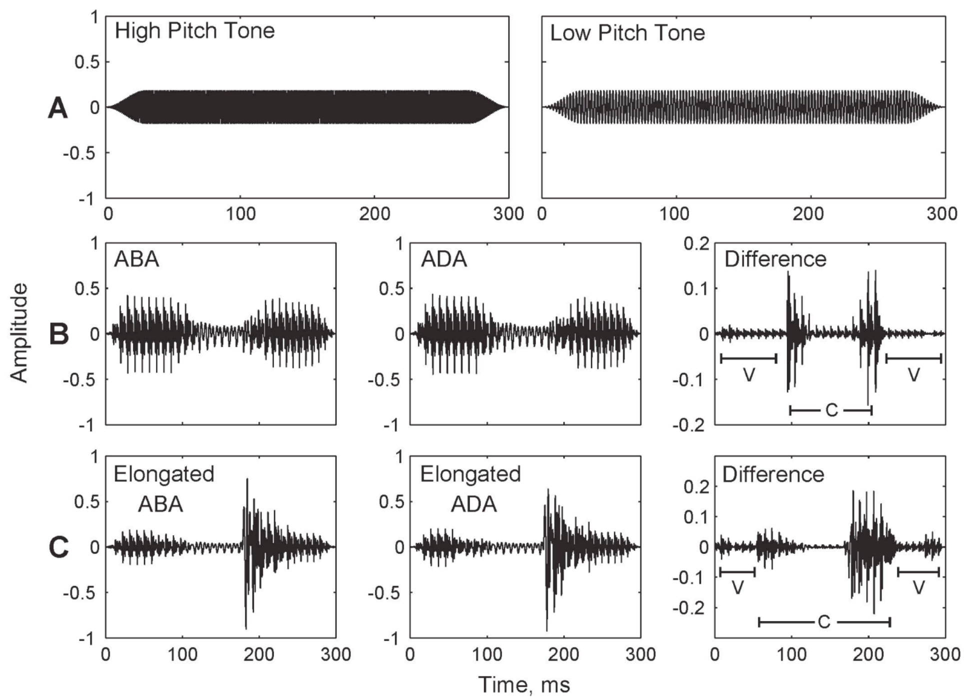
1. Stimuli used in oddball tasks. A. Tones. B. Phonemes. C. Elongated phonemes. The RMS power was adjusted so that it was equal for all stimuli (see Methods). Differences are ADA-ABA, where v and c indicate vowel and consonant respectively.
2. Difference wave grand averages for tones, phonemes and elongated phonemes from all subjects. We aligned all recordings to start at 0 ms which was the onset time of each 300 ms stimuli. Each plot corresponds to the geometric alignment of the 10-20 nomenclature for the 30 EEG electrodes (F, FT, T, TP, P, FC, C, CP, O, FZ, FCZ, CZ, CPZ, PZ, OZ), along with horizontal extraoculogram electrodes (HEOG) and vertical extraoculogram electrodes (VEOG). Grand averages from all subjects for all similar trials given for tones (dark lines), phonemes (dotted lines), and elongated phonemes (gray lines).
3. Example of optimization of plug-in and leave-one-out errors for 9 out of 25 electrodes. We plot all plug-in and leave-one-out error rates as open circles. There are many overlapping combinations of error rates. But there is only one combination, indicated by a star, which gives the minimum combined error rates of 3 leave-one-out errors and 2 plug-in errors.
4. Example of optimal classification. A. Thirteen optimal electrodes selected from 25 electrodes for linear discrimination (unselected electrodes are shown as hatched squares). Gray scale indicates the value of the weights of the linear coefficients corresponding to each electrode. B. One-dimensional plot of transformed data using optimal classifier. Vertical dashed line indicates

classification boundary. Two observations from combined phonemes and tones set are misclassified. Four out of eight observations of elongated phonemes are classified as tones (observations are spread vertically for clarity).

5. Results of the dataset bootstrap randomization for 13 electrodes to further assess the validity of the LDA.
6. Comparison of PCA, ICA and LDA.

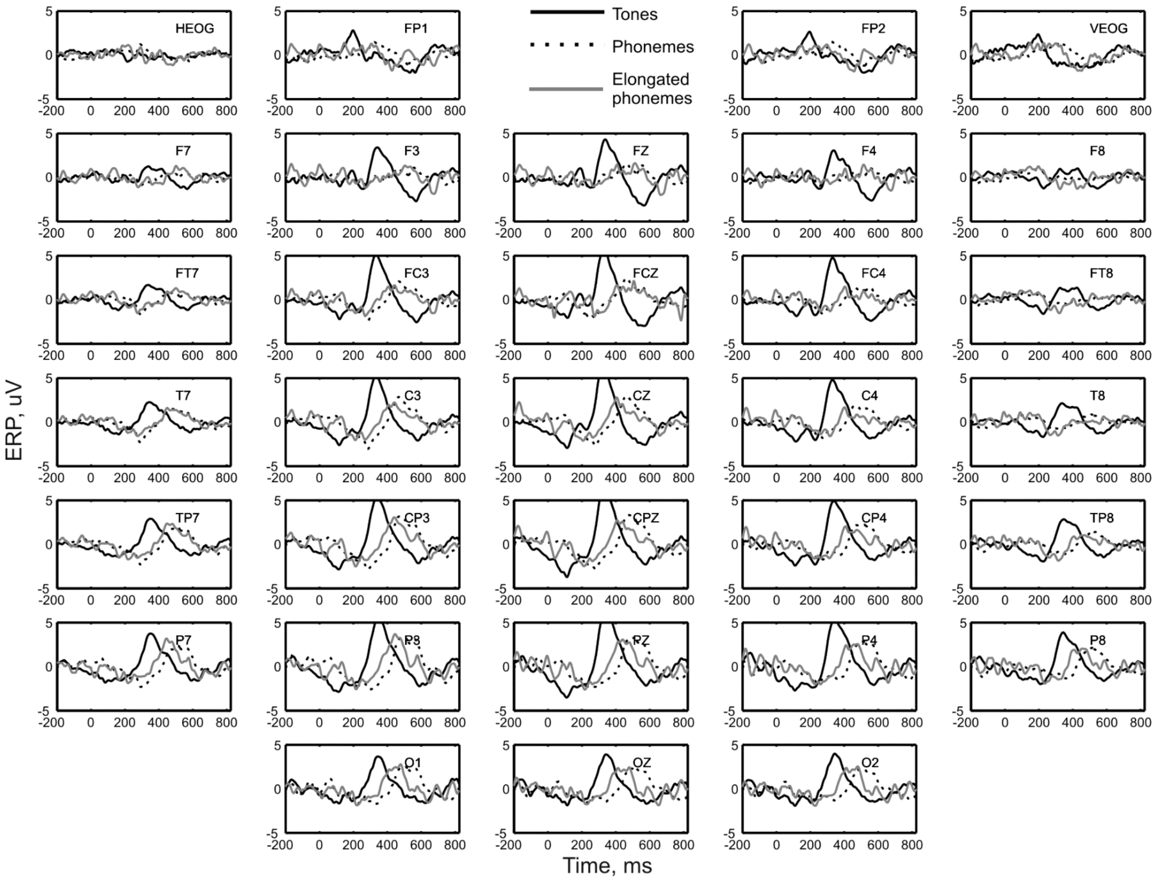
First 5 components (linear combinations) from PCA and ICA for combined data matrix of tone and phoneme measurements plotted as magnitude of component coefficients in spatial arrangement of electrodes (color scale) and linear bar charts. Electrode arrangement for array is shown at right. At the top right is shown a plot for the LDA coefficients. To the left of each array plot is shown the projection of mean MMN values for tones (18, blue circles) and phonemes (18, red asterisks). Red and blue bars represent the means of these projections. The PCA first component reflects overall spatial mean, second component contrasting front to back, third right to left. No such clear spatial patterns are evident in ICA components or LDA coefficients. Only two transformations significantly separated the means significantly, the third PCA component and LDA (3rd PCA component: $p=0.0007$; LDA: $p<0.0001$). The third PCA component reflected right from left sided asymmetry, as has been noted previously (Kayser et al, 1998). Nevertheless, the discrimination power of LDA was substantially more powerful than the other transformations.

Figure 1.



High Pitch Tone
Low Pitch Tone
A
B
C
Amplitude
ABA
ADA
Difference
V
C
Elongated ABA
Elongated ADA
Time, ms

Figure 2.



Tones
Phonemes
Elongated phonemes
HEOG
FP1
FP2
VEOG
F7
F3
FZ
F4
F8
FT7
FC3
FCZ
FC4
FT8
T7
C3
CZ
C4
T8
TP7
CP3
CPZ
CP4
TP8
P7
P3
PZ
P4
P8
O1
OZ
O2
ERP, uV
Time, ms

Figure 3

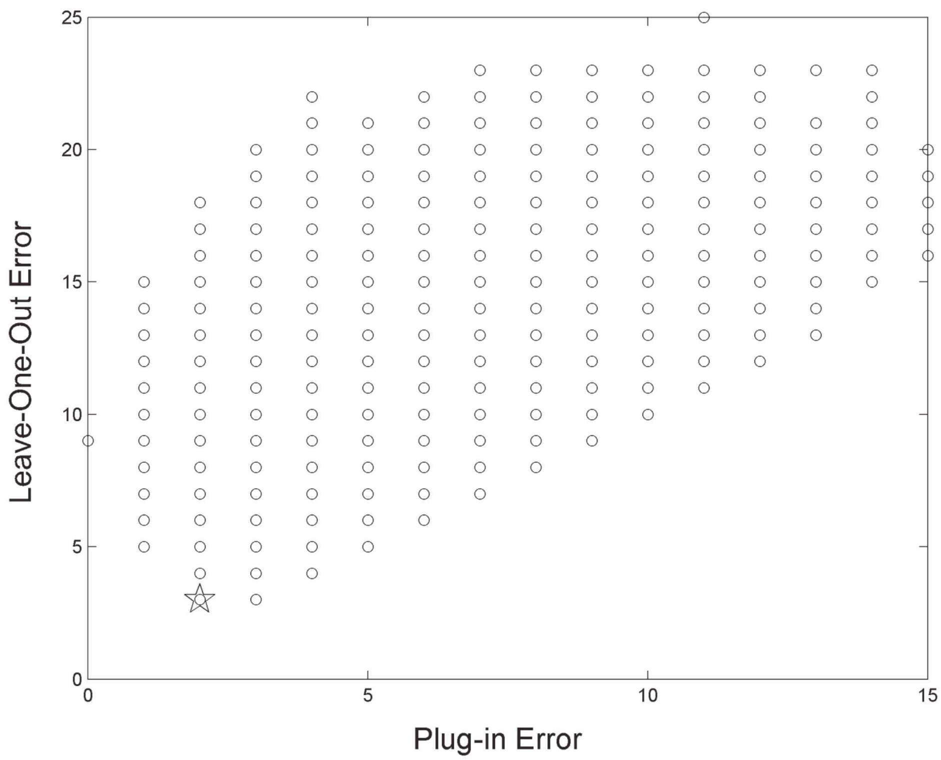


Figure 4

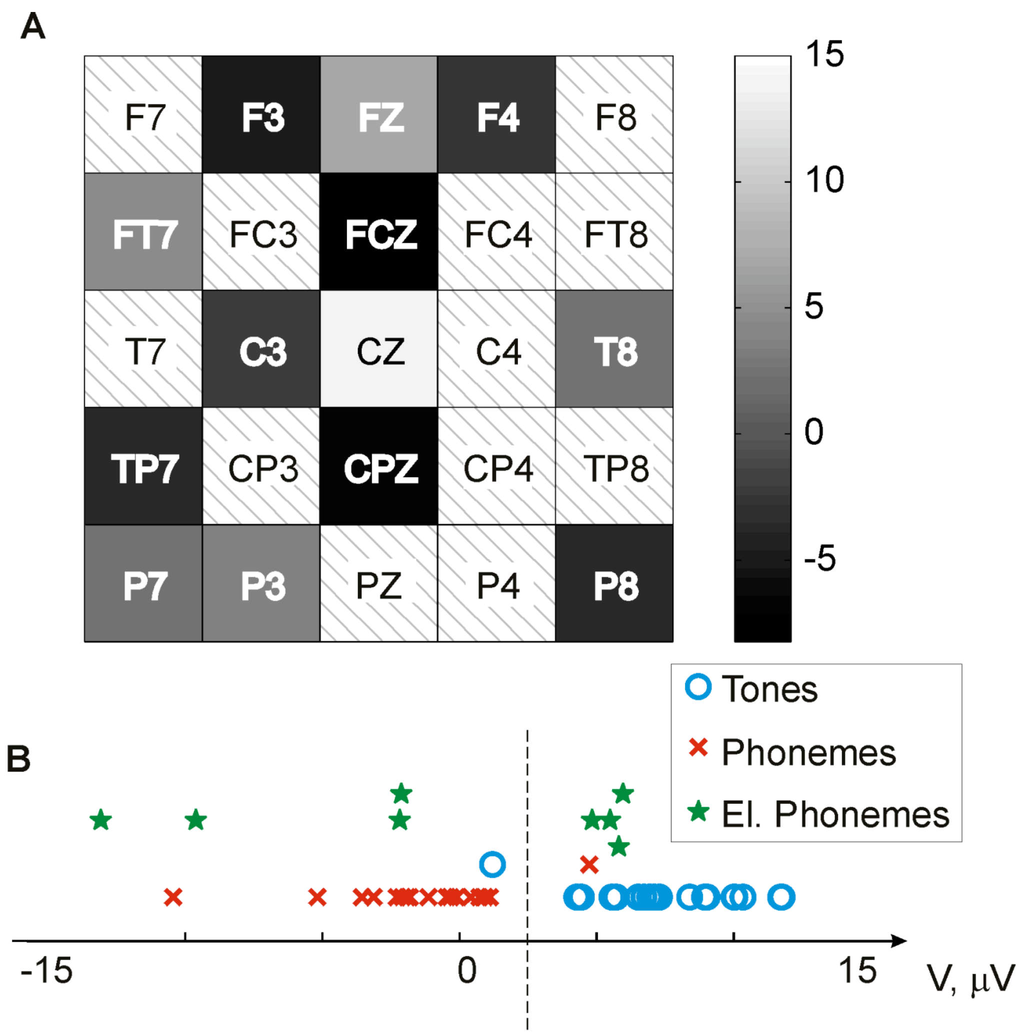

A
F7
F3
FZ
F4
F8
FT7
FC3
FCZ
FC4
FT8
T7
C3
CZ
C4
T8
TP7
CP3
CPZ
CP4
TP8
P7
P3
PZ
P4
P8
15
10
5
0
-5
B
Tones
Phonemes
El. Phonemes
-15
0
15
V, μV

Figure 5

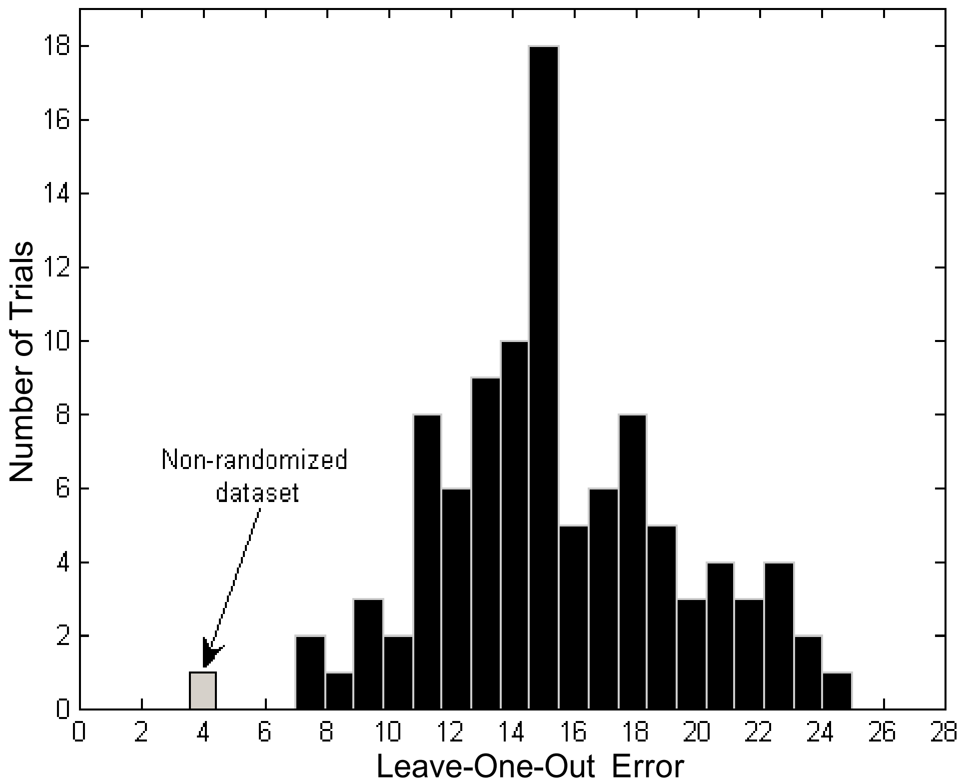
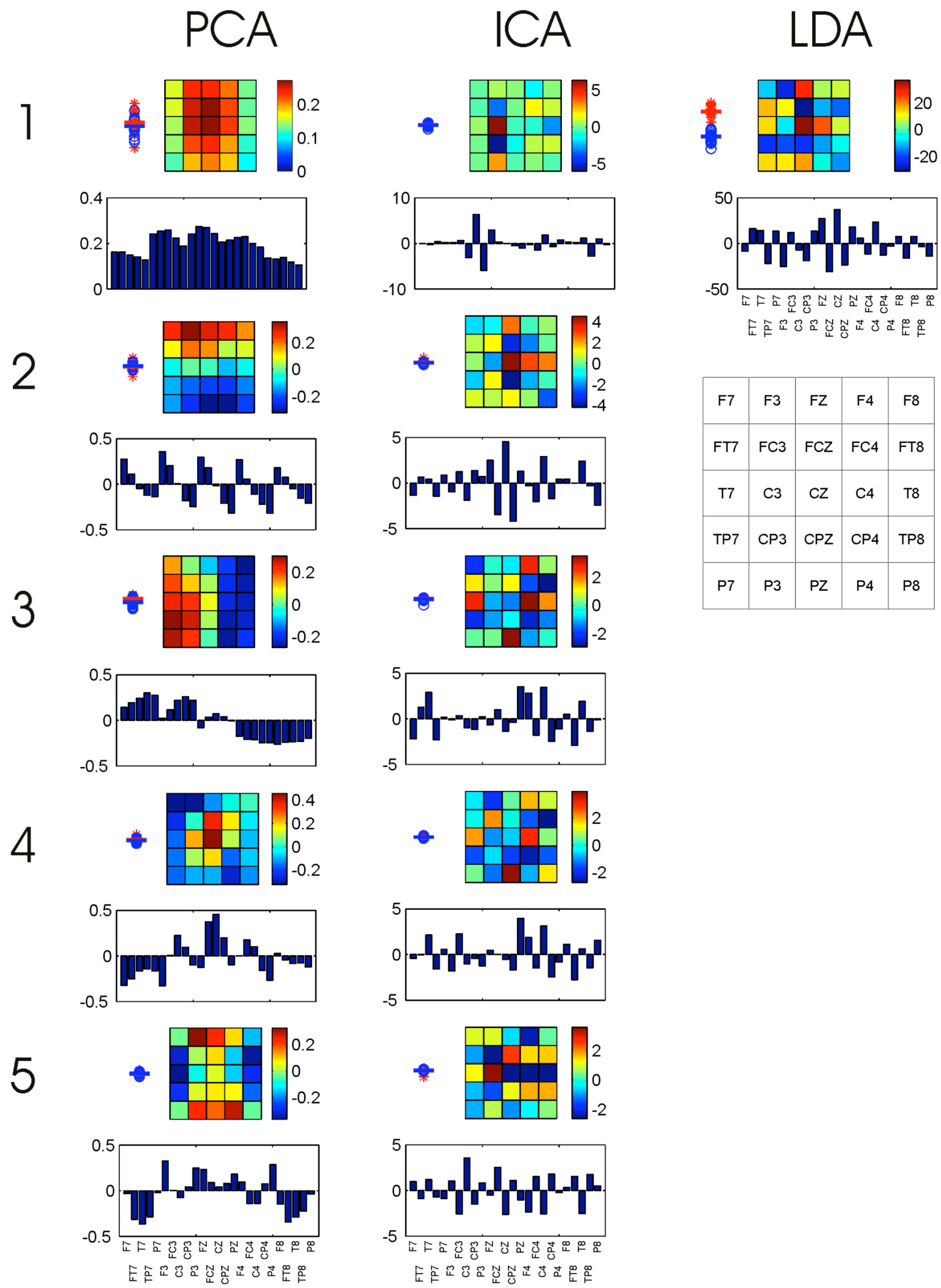

Number of Trials
Non-randomized dataset
Leave-One-Out Error
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28

Figure 6


PCA
ICA
LDA
1
2
3
4
5
F7 F3 FZ F4 F8
FT7 FC3 FCZ FC4 FT8
T7 C3 CZ C4 T8
TP7 CP3 CPZ CP4 TP8
P7 P3 PZ P4 P8
F7 FT7 T7 TP7 P7 F3 FC3 C3 CP3 P3 FZ FCZ CZ CPZ PZ F4 FC4 C4 CP4 P4 F8 FT8 T8 TP8 P8