Show me what you listen to! Auditory classification images can reveal the processing of fine acoustic cues during speech categorization.

Léo Varnet[1,4], Kenneth Knoblauch[2,4], Fanny Meunier[3,4], Michel Hoen[1,4]

[1]Lyon Neuroscience Research Center, CNRS UMR 5292,
Auditory Language Processing (ALP) research group, Lyon, France.
[2]Stem Cell and Brain Research Institute, INSERM U 846,
Integrative Neuroscience Department, Bron, France.
[3]Laboratoire sur le Langage le Cerveau et la Cognition, CNRS UMR 5304,
Auditory Language Processing (ALP) research group, Lyon, France.
[4]Université de Lyon, Lyon 1, Lyon, France

leo.varnet@isc.cnrs.fr, ken.knoblauch@inserm.fr, fanny.meunier@isc.cnrs.fr, michel.hoen@inserm.fr

Abstract

An essential step in understanding the processes underlying the general mechanism of perceptual categorization is to identify which portions of a physical stimulation modulate the responses of our perceptual system. More specifically, in the context of speech comprehension, it is still unclear what information is used to categorize a speech stimulus as one phoneme or another. Here we propose to adapt a Generalized Linear Model with smooth priors, already used in the visual domain for estimation of so-called *classification images*, to auditory experiments. We show how this promising approach can be applied to the identification of fine functional acoustic cues in speech perception.

Index Terms: phoneme categorization, psychophysics, acoustic phonetics, classification images, speech, speech-in-noise, Gaussian noise, fine acoustic cues, GLM.

1. Introduction

A general objective in perceptual science is to establish what exact parts of a complex physical stimulation modulate the percept it induces in an observer and constrain its behavior towards that stimulus. This issue has been a long-standing challenge in the particular context of speech comprehension where determining which among the auditory primitives that are coded at the neural acoustic/phonetic interface are actually used to recognize and categorize phonemes still constitutes an important open debate [1]. One of the current potential limits in this field is the lack of a method allowing direct estimation of which parts of the speech signal are used by listeners to understand speech in natural settings.

One way often used to identify relevant acoustic cues in speech is to proceed by progressive signal reductions, i.e., eliminating certain cues in order to demonstrate which ones are required. In the 50's, phoneme recognition was extensively studied by Liberman and colleagues for example, using the systematic variation of features in the time-frequency domain, usually along a continuum of synthetic speech [2], [3] and [4]. More recent work has involved artificially degraded-speech such as noise-vocoded [5], sine-wave [6], or band-pass filtered speech [7]. These approaches however, can only offer a very limited account of the problem as it is known that the speech comprehension system shows very fast and efficient plasticity, rapidly modifying relevant cue extraction in the face of drastic signal reductions or even stronger manipulations [8].

An alternative way to proceed would be to develop a method allowing experimenters to directly "see" where humans listen inside natural speech signals, without having to modify them. In the present paper, we show how we think a methodological solution to this issue can be provided by new developments in the domain of so-called *classification images* (CIm), see [9] for a historical-perspective review. In 1971, A.J. Ahumada Jr. first developed a correlational technique to estimate the frequency weighting-function of observers detecting a 500-Hz tone-in-noise [10]. The basic idea lying behind the CIm approach is that if one can determine how the presence of background-noise at each point of a stimulus interferes with the decision of the observer, one can derive a map of the perceptual cues relevant to achieve a specific categorization task. CIm experiments often use a two-alternative forced choice (2AFC) paradigm, involving for example a pair of stimuli to categorize ($\underline{t}_0$ and $\underline{t}_1$). Stimuli are systematically masked by additive noise, randomly varying at each trial. The best known and most intuitive method for then calculating a CIm is to average all noise fields eliciting response $\underline{t}_0$ and subtract the average of the noise fields eliciting response $\underline{t}_1$; this method was termed reversed-correlation [11].

Although it has been primarily conceived as an answer to a question that arose in the auditory modality [10], [11], this powerful tool has surprisingly been mostly exploited up to now in studies on visual psychophysics. Despite the strong potential of applying this method in the auditory domain, attempts currently remain rare and have up to now had limited impact (see however [12]) even though very similar correlational procedures have been used to determine spectral weighting functions of speech stimuli [13], [14], [15]. This situation can be partially explained by two severe limitations of the methods usually employed to derive CIm. Firstly, the several thousands of trials (up to 11400 [16]) typically needed to compute CIm accurately make it less adaptable to the auditory- than to the visual-modality. Methods minimizing this number do exist but they impose restrictions or simplifications on the noise used that are well adapted to visual stimuli, but not to experiments involving such complex signals as speech, varying in time and frequency. Secondly, the strong assumptions about the statistical distribution of the noise imposed by statistical theories limit their application to the study of auditory processes. Since its theoretical background has mostly been developed assuming additive Gaussian-noise, methods such as reverse-correlation are not the most suitable statistical framework to deal with non-Gaussian noise-fields.

In the present approach we propose, firstly: that the calculation of an auditory CIm should not be based on the amplitude of the noise samples, but rather on the power of the time-frequency bins of the power-spectrum, which have non-Gaussian distributions and secondly: to apply recent advances in the comprehension and computation of CIm introduced in recent works [17] and extended in [9], [18] and [19]. These studies proposed to fit the data using Generalized Linear Models (GLMs) with or without penalty term. Here we propose to implement and test this new method on the categorization of two VCV logatomes: ABA (/aba/) and ADA (/ada/).

2. Materials and Methods

In the following we use underlined symbols to indicate vectors, double underlined symbols to indicate matrices, and non-underlined symbols to indicate scalars.

2.1 Experimental Procedure

Two French native-speakers with normal hearing took part into this exploratory study: LV and MH. Target-sounds ($\underline{t}_0$ and $\underline{t}_1$), were two natural-speech samples (/aba/ and /ada/), obtained by cross-splicing the same utterance of /a/ with an utterance of /ba/ or /da/ (Figure 1). Original sounds were recorded in a soundproof chamber by the same female speaker and digitized at a sample rate of 44.1 kHz. The sound samples were 680 ms long, and their average power was normalized. Each stimulus $\underline{s}_i$ consisted of one target-sound ($\underline{t}_0$ or $\underline{t}_1$), embedded in Gaussian additive-noise using equation (1).

$$(1)\ \underline{s}_i = \alpha_i \cdot \underline{t}_{k_i} + \underline{n}_i$$

In (1), i is the trial number, k_i the target number (0 or 1) associated with this trial, $\underline{n}_i$ the noise field drawn from a normal distribution, and α_i a factor allowing the adjustment of signal-to-noise ratio (SNR) as a function of the participant's behavior, see 2.2 below. The experiment consists in listening to a set of 10.000 noisy stimuli (5000 for each target) in random order. Participants were instructed to listen carefully to the stimulus and indicate by a button press whether the masked signal was $\underline{t}_0$ or $\underline{t}_1$, response denoted by r_i (0 or 1). Listeners could complete the task over a period of 1 week, during office hours, at their leisure.

2.2 Adaptive stimulus-delivery procedure

During the course of the experiment, the noise level was continuously readjusted to ensure a correct response rate around 75%, as in most CIm experiments [20]. The SNR was varied from trial to trial on the basis of a local rate of correct responses calculated on a 10-trials window, with an adaptation of 0.2, 0.4, 0.6 or 0.8 dB for differences of 5, 10, 15, or 20% between intended and actual scores. SNR variations of the SNR were limited to the range -20 dB to 0 dB.

2.3 Deriving auditory classification images

Each stimulus noise $\underline{n}_i$ is characterized by its power spectrogram, whose components are entered as predictor variables in the model. Power spectrograms were calculated using a Short-Time Fourier Transform with a moving 512-pts Hamming window with no overlap, resulting in an 86.13 Hz frequency resolution and an 11.6 ms time-resolution. We limited our analysis to a time-range of 0 to 340 ms and a frequency range of 0 to 4048 Hz, ensuring that the size of the data-set would not exceed computational limits. The resulting 46-by-30 matrix is reshaped into a 884-by-1 vector of time-frequency bins, labeled $\underline{X}_i$. A similar treatment is applied to both targets, resulting in vectorized power spectrograms $\underline{T}_0$ and $\underline{T}_1$ (Fig. 1).

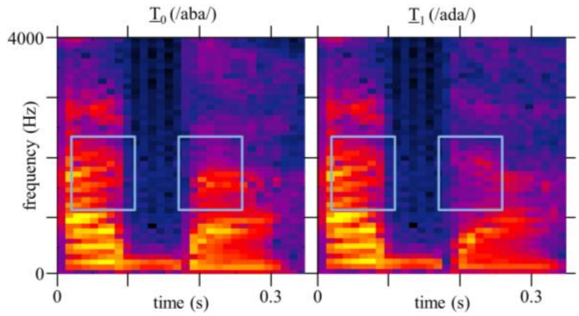

Figure 1: *Spectrograms of target-signals t_0 (/aba/) and t_1 (/ada/). Blue rectangles show the F2 transitions.*

In general agreement with the literature on CIm [21], we assume the observer to perform the detection of acoustic cues linearly by template matching, a longstanding model for decision-making. First, an internal decision variable d_i is computed by convolving the input with a weighting function referred to as the observer's template, and adding a random variable ε_i representing the internal noise of the system (accounting for the fact that the observer does not necessarily give the same response when presented with the same stimulus twice). In (2), the errors ε_i are assumed to have a zero mean symmetric distribution and to be independent from trial-to-trial.

$$(2)\ d_i = \left(\underline{X}_i + \underline{T}_{k_i}\right)^T * \underline{w} + \varepsilon_i$$

Then the response variable is given by (3):

$$(3)\ r_i = \begin{cases} 1 & if\ d_i > c \\ 0 & otherwise \end{cases}$$

c is a fixed criterion that determines the bias of the observer toward one alternative.

Knoblauch and Maloney [17] reformulated this very simple model in terms of a Generalized Linear Model, by expressing the probability that the observer gave the response $r_i = 1$, given the data $\underline{X}_i$, when the target number k_i was present as in (4) and (5):

$$(4)\ P(r_i = 1 | k_i = 0) = \Phi(\underline{X}_i{}^T * \underline{\beta}_0 + \beta_{0,0})$$

$$(5)\ P(r_i = 1 | k_i = 1) = \Phi(\underline{X}_i{}^T * \underline{\beta}_1 + \beta_{1,0})$$

with Φ the cumulative distribution function associated with ε, $\underline{\beta}_0$ and $\underline{\beta}_1$ the weighting functions corresponding to the presentation of target number 0 and 1 respectively, and $\beta_{0,0}$ and $\beta_{1,0}$ the constant terms reflecting the bias (see [9], for details). In line with the psychophysics literature, we could assume that ε is taken from a logistic distribution (a common choice for modeling binomial data). The associated psychometric function Φ will then be the inverse of the logit function. The structure of equations (4) and (5), with a linear combination of parameters linked to the dependent variable via a psychometric function, is the typical form of a Generalized Linear Model ([22], [23]). At this stage we could thus determine the values of the model parameters $\underline{\theta} = \left\{\underline{\beta}_0, \underline{\beta}_1, \beta_{0,0}, \beta_{1,0}\right\}$ that best fit the empirical data, by simply maximizing the log-likelihood by a standard maximization algorithm as in (6):

$$(6)\ \underline{\hat{\theta}}_{ML} = \underset{\underline{\theta}}{\operatorname{argmax}}\ L(\underline{\theta}) = \underset{\underline{\theta}}{\operatorname{argmax}}\ \log\left(\prod_i P\left(r_i \middle| \underline{\theta}, k_i, \underline{\underline{X_i}}\right)\right)$$

This would provide us maximum likelihood estimates of the two classification images, $\underline{\hat{\beta}_0}$ and $\underline{\hat{\beta}_1}$, and of the two intercept terms $\hat{\beta}_{0,0}$ and $\hat{\beta}_{1,0}$.

These estimates would presumably be too noisy to be decipherable. Indeed, GLMs, as reverse-correlations, when comprising a large number of predictors are prone to overfitting, which means that the model will describe the trial-dependent noise as well as the underlying classification mechanism. Estimates of the observer's template by GLM can therefore be quite noisy, and the model will not be able to generalize to novel data. One solution has been developed in the Generalized Linear Model framework under the name "Penalized Likelihood", which has been widely used for estimating the receptive fields of single neurons [24], [25], [26] and adapted to Classification Images by Knoblauch and Maloney [17] see also [18] and [27]. The aim of this method is to incorporate prior knowledge about the smoothness of the intended CIm. To do so, we associate with each value of the model parameters $\underline{\theta}$ a probability (7):

$$(7)\ P\left(\underline{\theta}\middle|\lambda_1, \lambda_2\right) = \lambda_1 \underline{\theta}^T \underline{\underline{L_1}}\ \underline{\theta} + \lambda_2 \underline{\theta}^T \underline{\underline{L_2}}\ \underline{\theta}$$

In (7) $\underline{\underline{L_1}}$ is the Laplacian matrix along dimension 1 (time), $\underline{\underline{L_2}}$ the Laplacian matrix along dimension 2 (frequency) [24]. The quadratic form $\underline{\theta}^T \underline{\underline{L_i}}\ \underline{\theta}$ thus provides a measure of the smoothness of $\underline{\theta}$ over dimension i, and this equation represents our a priori beliefs about the true underlying template (a smoother CIm will be more expected, and therefore have a higher prior probability). This prior is defined by a distribution and a set of hyperparameters $\underline{\lambda} = \{\lambda_1, \lambda_2\}$. Then, instead of maximizing the log-likelihood as before, we maximize the log of the posterior $P\left(\underline{\theta}\middle|\underline{r}, \underline{k}, \underline{\underline{X}}, \underline{\lambda}\right)$ that takes into account the likelihood and prior information. The Maximum a posteriori estimate of the model parameters is then given by (8):

$$(8)\ \underline{\hat{\theta}}_{MAP} = \underset{\underline{\theta}}{\operatorname{argmax}}\left[\log\left(P\left(\underline{\theta}\middle|\underline{r}, \underline{k}, \underline{\underline{X}}, \underline{\lambda}\right)\right)\right]$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}}\left[\log\left(P\left(\underline{r}\middle|\underline{\theta}, \underline{k}, \underline{\underline{X}}\right)\right) + \log\left(P\left(\underline{\theta}\middle|\underline{\lambda}\right)\right)\right]$$

$$= \underset{\underline{\theta}}{\operatorname{argmax}}\left[L(\underline{\theta}) + R(\underline{\theta})\right]$$

The last equation can be seen as the same maximization of the log-likelihood as before, with an additional penalty term, $R(\underline{\theta})$, that biases our estimate towards model parameters with higher a-priori probability. The optimal estimate is a tradeoff between fitting the data well and satisfying the constraints of the penalty term. Therefore a prior on smoothness will favor CIm with slow variations in time and frequency. Actually, for large (> 1) values of λ_1 and λ_2 we put a strong disadvantage on sharp CIm, and for $\lambda_1 = \lambda_2 = 0$ we recover the initial maximum likelihood solution.

The standard method for setting the value of the hyperparameters is cross-validation. This approach involves a partition of the data between a "training" and a "test" set. For each given couple of hyperparameters, we can estimate the model parameters on the "training" set by maximum a posteriori, as explained previously. It becomes thus possible to assess how the model parameters would generalize to an independent dataset by comparing the predicted responses on the test-set to the actual responses. When the model predicts most accurately unseen data, the strength of priors is considered as accurate. Therefore, the optimal hyperparameters are found by choosing the models that yield a maximum in cross-validation rate. In more simple terms, this technique yields to a form of Automatic Smoothness Determination [28].

3. Results

The SNR was manipulated across trials to maintain the percentage of correct answers roughly equal to 75% during the course of the entire experiment. Nevertheless, variations of SNR provide an overview of observers' performances in the phoneme categorization task. Figure 2A plots the evolution of SNR during the experiment and the mean SNR for each participant. The psychometric functions are then estimated on all available data for each listener (figure 2B). The linear relationship signal contrast and detectability index in the 0 to 15% range is in agreement with our assumption of a (at least locally) linear model for the observers [29].

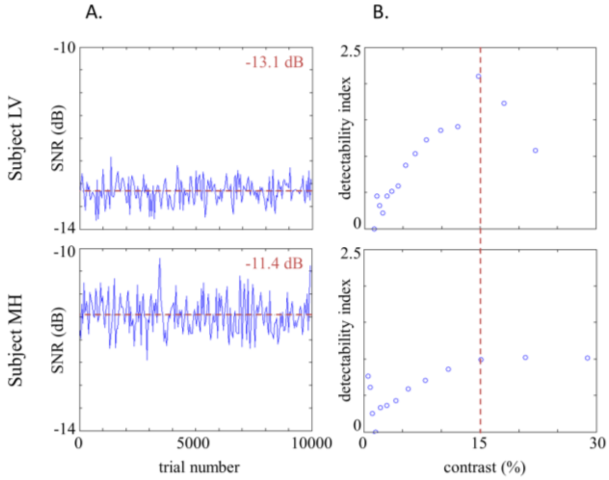

Figure 2: *A. Evolution of SNR across trials (mean SNR by blocks of 50 trials), and mean SNR for the whole experiment. B. Detectability index as a function of signal contrast.*

Figure 3 shows the CIm obtained by the GLM method with smooth priors, as well as the values of λ_1 and λ_2. For each participants, two CIm $\underline{\hat{\beta}_0}$ and $\underline{\hat{\beta}_1}$ are computed, based on the trials where the target signals $\underline{t_0}$ or $\underline{t_1}$ were presented, providing a measure of the strength of the relation between the noise at different time-frequency locations and the speech identification scores in the case where the masked signal was $\underline{t_0}$ or $\underline{t_1}$ respectively. In that sense, the templates may be regarded as a measure of the contribution of each time-frequency bins to categorization, with high absolute values for locations at which the power of the noise influences the decision of the observer. As can be seen from Figure 3, CIm often exhibit both positive (yellow) and negative (black) weights corresponding to areas where the presence of noise respectively increases or decreases the probability of the stimulus to be identified as signal $\underline{t_1}$. The differences between the two estimates are generally interpreted as evidence of the nonlinearities of the auditory system (the template used for detection depending on the input signal).

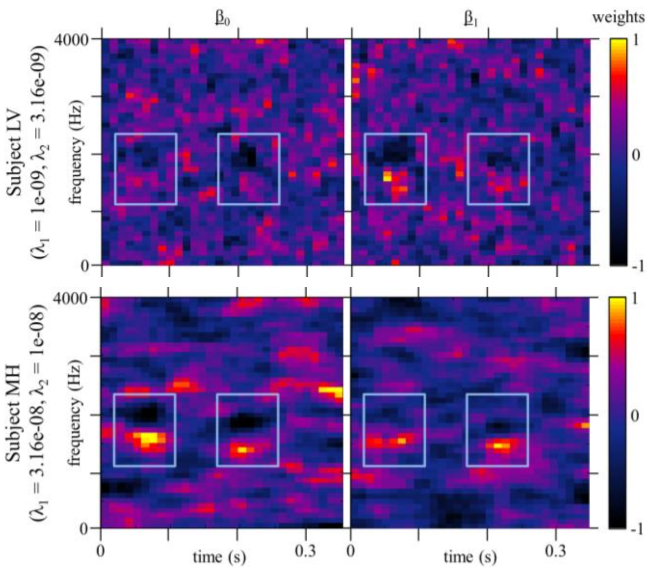

Figure 3: *Classification Images $\underline{\beta}_0$ and $\underline{\beta}_1$ for both participants. Blue rectangles show the regions surrounding the F2 transitions in the original signals. The color scale represents the weights divided by their maximum absolute value.*

For a better understanding of these CIm, we ran a similar test performed by an ideal template matcher. This classifier simulates the optimal observer with the linear model presented in equation (2) and (3) by taking $\underline{w} = \underline{T}_1 - \underline{T}_0$. Therefore the template matcher observer bases its classification strategy on the time-frequency bins where the spectrograms of the two signals differ most. As the performances of the algorithm do not vary over time, the SNR for stimulus presentation was set to -25 dB, for a resulting percentage of correct answers of 68%. The obtained CIm are plotted in Figure 4. We can notice that, unlike in Figure 3, $\underline{\hat{\beta}}_0$ and $\underline{\hat{\beta}}_1$ are very close because the linear observer algorithm involves a single template.

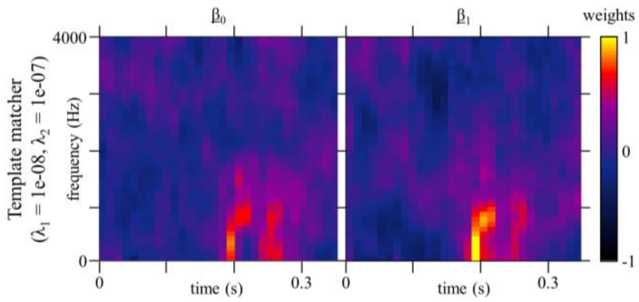

Figure 4: *Classification Images $\underline{\beta}_0$ and $\underline{\beta}_1$ for the template-matcher.*

4. Discussion

We have shown that the use of a GLM with a smoothness prior as a statistical method for the estimation of CIm in the auditory modality is a reasonable way of overcoming traditional limitations of this approach to auditory studies. First, this method allows the addition of prior knowledge on the smoothness of the expected CIm. By exploiting the dependencies between adjacent noise values, one can significantly reduce the number of trials required to obtain a reliable CIm. Secondly, unlike the reverse-correlation method, the GLM does not require the stimulus or the noise to be normally distributed. Accordingly, it can measure CIm using noise-fields with non-Gaussian distributions, such as the power spectrum of an acoustic-noise, in a similar way to the calculations of second-order CIm using GLM in [17, 24]. Therefore, Generalized Linear Models with priors provide a suitable and powerful framework to investigate the way in which the human system achieves fast and efficient categorization of phonemes in noise and to estimate how human observers differ from ideal template matchers.

We could further demonstrate that this method would be suitable for studying the use of fine-acoustic cues used during speech categorization. If we map the CIm obtained from both of our human listeners onto the original stimuli spectrograms, we can observe two main foci of high- and low-values weights, located in the time-frequency domain exactly over the second formant F2 (blue squares in figure 3). More precisely, our preliminary observations suggest that, unlike the template-matcher, the functional cue used for categorizing /aba/ and /ada/ stimuli is composed of the end of the second formant at the end of the vowel and of the onset of F2, on the consonant, just following the occlusion. This is in agreement with the strong hypothesis, formulated in [3], that the second formant transition would be a key for classifying phonemes into [ABA] or [ADA]. The pattern consistently observed at each time-frequency location of a second formantic transition, composed of a cluster of positive weight below a cluster of negative weights, supports the assumption that frequency information is coded in terms of relative difference across channels [6]. A similar pattern has been observed for a Vernier acuity task in the visual domain [11], highlighting the fact that our phoneme categorization task can be seen as the detection of the alignment of formants. In addition, the obtained CIm evidence the fact that the estimation of the second formant by the auditory system is a relative measurement, since the presence of noise masking the position of F2 in the preceding vowel influences the decision of the observer (even though this region contains no useful information for performing the task as the first syllable was obtained by cross-splicing the same utterance of /a/). This could be in line with theories conceptualizing phonemic perception as the interpretation of phonetic movements and trajectories, and indicates that the categorization mechanism involved in our experiment is not task-specific. Further work will be dedicated to studying in detail the relationship between CIm and phonetic discriminations.

5. Conclusions

In this study we have demonstrated the feasibility of studying the use of fine-acoustic cues relevant for phoneme categorization using auditory classification images (ACI). We have shown that applying new statistical developments, including GLM models with priors, to the methodology of classification images dramatically extends the possibility of using CIm in the auditory modality by increasing the power of the method and avoiding traditional limitations such as the high number of needed trials or assumptions on the Gaussian distribution of the noise used to mask target-stimuli.

6. Acknowledgements

The first author (L.V.) is funded by a PhD grant from the Ecole Doctorale Neurosciences et Cognition (EDNSCo), Lyon-1 University, France. This research was supported by a European Research Council grant to the SpiN project (no. 209234) attributed to F.M.

7. References

[1] Cutler, A. “Native listening: Language experience and the recognition of spoken words”, in Cambridge, MA: MIT Press, 2012.

[2] Liberman, A.M., and Delattre, P. C., “The role of selected stimulus-variables in the perception of the unvoiced stop consonants”, The American Journal of Psychology, 65(4):497-516, 1952.

[3] Liberman, A. M., Delattre, P. C., Cooper, F. S., and Gerstman, L. J., “The role of consonant-vowel transitions in the perception of the stop and nasal consonants”, Psychological Monographs: General and Applied, 68(8), 1-13, 1954.

[4] Liberman, A. M., Safford Harris, K., Hoffman, H. S., and Griffith, B. C., “The discrimination of speech sounds within and across phoneme boundaries “, Journal of experimental Psychology, 54, 5, 1957.

[5] Xu, L., Thompson, C. S., and Pfingst, B. E., “Relative contributions of spectral and temporal cues for phoneme recognition”, Journal of Acoustic Society of America, 117(5), 3255–3267, 2005.

[6] Loizou, P., Dorman, M., and Tu, Z., “On the number of channels needed to understand speech”, J. Acoust. Soc. Am., 106, 2097-2103, 1999.

[7] Apoux, F. & Healy, E.W., “On the number of auditory filter outputs needed to understand speech: Further evidence for auditory channel independence”, Hearing Research, 255, 99-108, 2009.

[8] Shannon, R.V., Zeng, F.-G., Wygonski, J., Kamath, V., and Ekelid, M., “Speech Recognition with Primarily Temporal Cues”, Science, 270, 303-304, 1995.

[9] Murray, R. F., “Classification images: A review”, Journal of Vision, 11(5):2, 1–2, 2011.

[10] Ahumada, A., and Lovell, J., “Stimulus Features in Signal Detection”, Journal of Acoustic Society of America, 49(6B), 1751-1756, 1971.

[11] Ahumada, A. J., “Perceptual classification images from vernier acuity masked by noise”, Perception, Vol. 25, ECVP'96 Abstracts, 1996.

[12] Ahumada, A., Marken, R., and Sandusky, A., “Time and frequency analyses of auditory signal detection”, Journal of Acoustic Society of America, 57(2), 385-390, 1975.

[13] Ardoint, M., Mamassian, P., & Lorenzi, C., “Internal representation of amplitude modulation revealed by reverse correlation”, in Abstract AR0 n 919. 30th ARO midwinter meeting, Feb 10-15, Denver, Colorado, USA, 2007.

[14] Doherty, K. A., and Turner, C. W., “Use of a correlational method to estimate a listener’s weighting function for speech”, Journal of the Acoustical Society of America, 100 (6), 3769-3773, 1996.

[15] Apoux, F., and Bacon, S. P., “Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise”, Journal of the Acoustical Society of America, 116(3), 1671-1680, 2004.

[16] Calandruccio, L., and Doherty, K. A., “Spectral weighting strategies for sentances measured by a correlational method”, Journal of the Acoustical Society of America, 123(4), 2367-2379, 2007.

[17] Barth, E., Beard, B. L., and Ahumada, A. J., “Nonlinear features in vernier acuity”, in B. E. Rogowitz and T. N. Pappas, [Ed] Human Vision and Electronic Imaging IV , SPIE Proceedings. 3644(8), 1999.

[18] Knoblauch, K., and Maloney, L. T., “Estimating classification images with generalized linear and additive models”, Journal of Vision, 8(16):10, 1–19, 2008.

[19] Mineault, P. J., Barthelmé, S., and Pack, C. C., “Improved classification images with sparse priors in a smooth basis”, Journal of Vision, 9(10):17, 1–24, 2009.

[20] Murray, R. F., “Classification images and bubbles images in the generalized linear model”, Journal of Vision, 12(7):2, 1–8, 2012.

[21] Gold J. M., Murray R. F., Bennett P. J., and Sekuler A. B., “Deriving behavioural receptive fields for visually completed contours.” Current Biology, 10, 663–666, 2000.

[22] Ahumada, A. J., “Classification image weights and internal noise level estimation”, Journal of Vision, 2(1), 121-31, 2002.

[23] Fox, J., “Generalized linear models”, in Applied Regression Analysis and Generalized Linear Models Second Edition, SAGE Publications, Chapter 15, 379-424, 2008.

[24] Knoblauch, K., and Maloney, L. T., “Modeling Psychophysical Data in R”, in Springer [Ed], chap 6, 173-202, 2012.

[25] Wu, M. C. K., David, S. V., and Gallant, J. L. “Complete Functional Characterization of Sensory Neurons by System Identification”, Annual Review of Neuroscience, 29, 477–505, 2006.

[26] Calabrese, A., Schumacher, J. W., Schneider, D. M., Paninski, L., Woolley, S. M. N., “A Generalized Linear Model for Estimating Spectrotemporal Receptive Fields from Responses to Natural Sounds”, PLoS ONE, 6(1), e16104, 2011.

[27] Schönfelder, V. H., and Wichmann, F. A., “Sparse regularized regression identifies behaviorally-relevant stimulus features from psychophysical data”, Journal of the Acoustical Society of America, 131(5), 3953-3969, 2012.

[28] Sahani, M., Linden, J. F., “Evidence Optimization Techniques for Estimating Stimulus-Response Functions” in Proceedings of the 2002 Conference (2002)

[29] Abbey, C. K., Eckstein, M. P., “Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer”, Journal of Vision, 6, 335-355, 2006.