euronoise

Signal densities and criterion variance in speech and nonspeech perception

Luis E. Lopez-Bascuas

Universidad Complutense Madrid, Facultad Psicologia, Campus Somosaguas, 28223 Madrid, Spain
lelopezb@psi.ucm.es

The actual shape of signal densities has become an important issue when studying speech perception within the framework of Signal Detection Theory (SDT). Different authors have found different results as for such density shapes in the cases of speech and non-speech signals. While some report Gaussian densities only for non-speech, others find the same Gaussian shape for both, speech and non-speech acoustic signals. Some of the claims concerning non-gaussian signal densities have been made on the basis of finding "aberrant" Receiver Operating Characteristic (ROC) curves. In this paper we try to show that "aberrant" ROCs are not sufficient evidence for postulating non-gaussian signal densities; rather, unequal criterion variances can also be the cause of deviant ROCs. Our findings seem to indicate that, for non-speech (noise+square wave continuum), unequal criterion variances underlie the "aberrant" ROCs, but for speech (ba-pa continuum), non-gaussian signal densities must also contribute to this effect.

1 Introduction

Analyzing the shape of signal densities evoked by auditory stimulus is interesting both methodologically and theoretically. On the methodological side, it must be remembered that a common practice in speech perception research is to study how subjects identify and discriminate different speech sounds. Early research had emphasized the use of proportions as the independent variable of the experiments, either the percentage of assigning a specific label in identification or the proportion of correct responses in discrimination. The use of proportion correct has threshold implications and therefore the assumption that using this measure is not theory bound is not tenable [1], [2]. Being this true, there seems to be no reason to avoid the use of more powerful psychophysical models as those derived from Signal Detection Theory (SDT). As a matter of fact, different researchers have followed this track [2], [3], [4], [5]. However, all the aforementioned studies, framed within SDT, have assumed signal distributions of Gaussian shape and equal variance, along with zero criterion variance, without careful checks of these assumptions. Thus, from this methodological perspective, the correctness of their conclusions is, at least, partially bound to the correctness of such assumptions.

Theoretically, the shape of signal densities has been used in the debate concerning the possible existence of processing routines especially designed to handle speech signals [6], [7], [8]. For instance, López-Bascuas found that an improved signal detection model allowing for estimates of unequal criterion variances could fit a noise-buzz continuum but not a speech /ba/-/pa/ continuum. Further analysis indicated that the failure of the model with the speech sounds could be attributed to the non-Gaussian shape of the speech densities [7]. The implication is that a fundamental difference between the processing of speech and non-speech signals might be reflected in the shape of the underlying sensory distributions.

Schouten and van Hessen tested explicitly the gaussian assumption for a /pak/-/tak/-/kak/ speech continuum and for a non-speech continuum varying in intensity [9]. They measured response distributions by means of a non-numerical magnitude estimation procedure and concluded that all underlying densities for each continuum were Gaussian. However, Pastore and Macmillan reanalyzed Schouten and van Hessen´s data and their analyses point to a different conclusion [10]. In particular, Pastore and Macmillan constructed Receiver Operating Characteristics curves (ROCs), an analysis that assumes an ordinal relationship between the rating responses and the decision axes. According to the standard (Gaussian-equal variance) SDT model the theoretical ROC must run from point (0,0) to point (1,1), keep above the major diagonal, be concave downward and it must also be symmetric about the minor diagonal. More simply stated: Z-transformed ROCs must be linear [11]. The empirical ROC curves for non-speech seemed to accommodate well to the theoretical predictions and, therefore, the results seemed to warrant a Gaussian shape for the underlying distributions. However, ROCs for speech turned out to be "aberrant". The empirical ROCs were well described as two intersecting linear segments (compatible with threshold models) and thus, the underlying densities seemed not to be Gaussian.

A question that arises after Pastore and Macmillan analysis is whether "aberrant" ROCs are sufficient evidence for underlying non-Gaussian densities in rating scale experiments. The point is that SDT typically assumes that criteria have zero variances. Therefore, aberrant ROCs could emerge not only due to non-Gaussian density shapes but also due to nonzero criterion variances. In this work we try to demonstrate that non-linear z-transformed ROCs are compatible with underlying Gaussian densities under the assumption of unequal criterion variances.

We tested this hypothesis for the /ba/-/pa/ speech contunuum and the noise-buzz continuum employed in [7]. We constructed ROC curves (double-probability plots) for each pair of adjacent stimuli for the two different acoustic continua. We expected the speech ROCs to be nonlinear since their underlying densities are non-Gaussian [7]. Non-speech continua can also yield non-linear z-transformed ROCs [12]. Thus, if our z-tranformed ROCs for the noise-buzz turn out to be nonlinear, then we must conclude that nonlinear z-transformed ROCs are not a sufficient condition for postulating non-Gaussian densities, as long as we have demonstrated that a Gaussian model (with unequal criterion variances) can fit the data of this non-speech continuum. We will then try to show that the non-linear shape of the Z-transformed ROCs for non-speech has to do with criterion densities having unequal variances. To this end, we will try to fit a restricted Thurstonian SDT-like model (i.e., a model assuming equal criterion variances) to the non-speech data.

2 Method

2.1 Participants

Two Spanish monolingual subjects (Castilian dialect) with no reported hearing defects participated in these experiments.

2.2 Stimuli

A /ba/-/pa/ continuum was used for speech. We created it by varying the voice-onset-time (VOT) in 10 ms steps. The continuum ranged from –35 ms to +55 ms of VOT.

The noise-buzz continuum was modeled after Miller et al. stimuli [13]. Each stimulus consisted of a 100 Hz square wave (a buzz) and a wide-band noise. The level of the noise was 15 dB below that of the buzz. For one end point the buzz led the noise by 35 ms. To make the continuum, leading portions of the buzz were removed in 10 ms steps until the noise led by 55 ms.

2.3 Procedure

Each subject was tested in six individual sessions, three for the /ba/-/pa/ continuum and three for the noise-buzz continuum. The initial session for each was solely for training. Within this session the sequence was as follows:

1. Presentation of endpoints ("ba" or "pa"; "buzz alone" or "noise-buzz")

2. Identification test on the endpoints with feedback (only for nonspeech). This part ended after the subject had correctly identified 90% of the stimuli.

3. Familiarization with the rating procedure. The subject was told that different sounds would be presented during the experiment requiring two different judgments for each one. The stimuli presented in the first part of training were to serve as prototypes. The subject first had to give each sound one of the two labels used previously and then had to rate how well the sound fitted the chosen category. A four-point rating scale was provided, where *1* meant a 'poor exemplar', *4* a 'good exemplar', and *2* and *3* denoted intermediate values.

The second and third sessions were devoted to data collection, with a common pattern for all three continua. The subject identified and rated each stimulus, following the procedure learned during training. At the beginning of a session, the endpoint stimuli were presented as a reminder. Across the two data collection sessions, each subject had 1500 experimental trials per continuum. Therefore, for a given continuum, each subject gave 150 categorization and rating judgments per stimulus for data analysis.

3 Results

We constructed z-transformed ROC curves (double-probability plots) for each pair of adjacent stimuli. To do this, we first calculated de cumulative probabilities for the different criteria involved in the experiment and obtained the corresponding z-transforms for a given pair of stimuli. Since all our continua contained ten stimuli, we generated nine ROCs for each subject in each condition. Thus, a total of 36 (9x2x2) curves were computed.

In our experiments subjects had eight possible response categories and, therefore, seven points were available to construct each curve. As an example we present in table 1 the calculations for the (-5, 5) /ba/-/pa/ pair corresponding to our first subject.

	CUMULATIVE FREQUENCIES							
	BA4	**BA3**	**BA2**	**BA1**	**PA1**	**PA2**	**PA3**	**PA4**
-5	54	125	138	139	140	145	147	150
5	26	63	98	115	139	144	145	150
	CUMULATIVE PROBABILITIES							
	BA4	**BA3**	**BA2**	**BA1**	**PA1**	**PA2**	**PA3**	**PA4**
-5	0.36	0.833	0.92	0.927	0.933	0.987	0.98	1
5	0.173	0.42	0.653	0.767	0.927	0.96	0.967	1
	Z-TRANSFORMED SCORES							
	BA4	**BA3**	**BA2**	**BA1**	**PA1**	**PA2**	**PA3**	**PA4**
-5	0.358	-0.967	-1.405	-1.451	-1.501	-1.834	-2.054	-3
5	0.541	0.202	-0.394	-0.728	-1.451	-1.751	-1.834	-3

Table 1

For illustration purposes we plot here several ROCs (see Figure 1). The first two correspond to speech data (one per subject) while the other two correspond to the noise-buzz condition**.**

To test for nonlinearity we proceeded as follows:

1. We did not use values were the $p(R_j/S_i)$ was zero or unity.

2. For a given set of data we used ROC curves for adjacent stimuli where at least four paired z-values were available. Curves with fewer points were ignored.

3. We fitted a linear and a quadratic function to each usable ROC and obtained the residuals for each fit.

4. We combined the linear residuals across all curves in the set, combined the quadratic residuals across all curves in the set, and ran an F-test on the sums of the squared residuals.

5. If most or all ROC curves in the set were linear, F should not be significant.

The results are summarized in table 2.

Condition	F	p
Speech (1)	48.18	2.2929e-13
Speech (2)	92.44	5.3595e-12
Noise-buzz (1)	20.58	1.1460e-4
Noise-buzz (2)	30.02	1.0217e-5

Table 2

As can be seen we found significant differences for both subjects on the two conditions (speech and non-speech). Deviations from linearity were clearer for the speech (ba-pa) continuum than for the noise-buzz (broad-band noise + square wave) continuum. Nevertheless, on the basis of this analysis, it seems that no data set has mainly or entirely linear ROC curves.

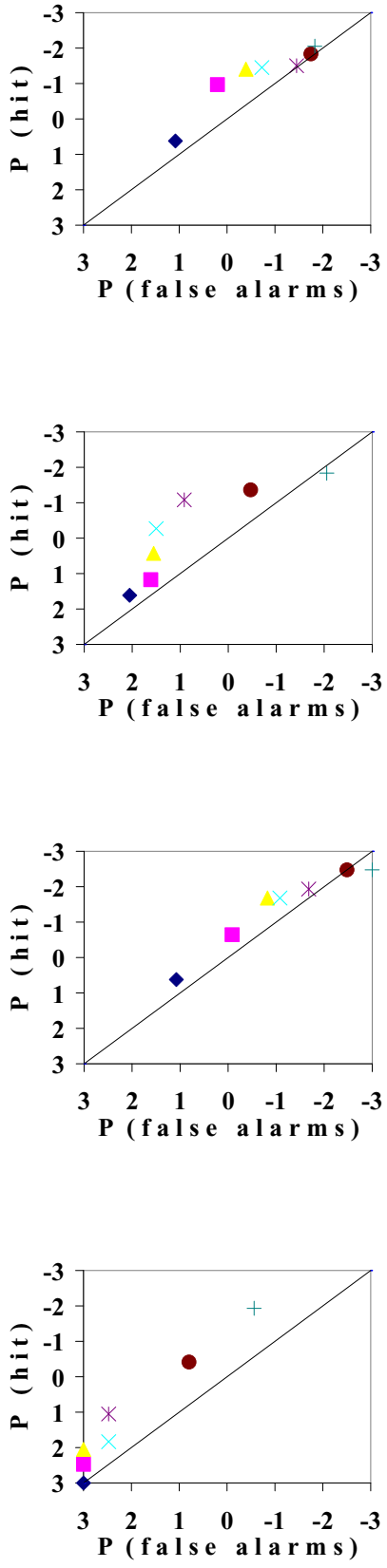

Fig. 1 Some examples of Z-transformed ROCs. The first two correspond to speech stimuli and the last two correspond to non-speech stimuli. Different points on each curve correspond to different criteria.

We then undertook the task of fitting the non-speech data to the constrained Thurstonian SDT-like model. Thurstone´s Law of Categorical Judgment is basically a continuous psychophysical model that allows estimates of signal and criterion parameters for rating scale experiments. Torgerson designated the general (unconstrained) equation Condition A (this condition assumes unequal signal and criterion variances). We are interested in fitting our rating data to what Torgerson called Condition B: under this condition only signal variances are unequal)[14]. In particular, maximum likelihood estimates were obtained using numerical optimization routines (quasi-Newtonian algorithm). Obtained parameters were then used to predict a theoretical matrix of frequencies for the rating experiment. Chi-square tests were used to assess the goodness of fit of our predictions.

For subject one χ^2=17.629 (5 d.f.), and for subject two χ^2=54.817 (4 d.f.). In both cases, the statistical tests show that our estimated parameters cannot fit the experimental data (α=0.01). Therefore, it is concluded that criterion variances need to be unequal in order to predict our rating data.

4 General discussion and conclusions

Z-transformed ROC curves from both the speech ba-pa continuum and the non-speech (noise-buzz) continuum seem to depart significantly from linearity. Under the standard (Gaussian-equal variance) SDT model this would imply that signal densities are not Gaussian. However, in [7] it was shown that only the speech signals did not accommodate to the Gaussian assumption. The non-speech data provided acceptable fits under Condition A of the Law of Categorical Judgment. This means that the noise-buzz generates non-linear z-transformed ROCs and yet, it evokes Gaussian densities on the sensory-decision axis. Therefore, it seems that non-linear z-transformed ROCs are not sufficient evidence for inferring non-gaussian signal densities.

In a second step, we tried to figure out the possible causes of the non-linear plots obtained for non-speech given that this departure from linearity is not connected to the shape of the signal densities. Treisman and Williams had already proposed that subjects try to optimize their performance in any psychophysical task by adjusting the location of criteria from trial to trial [15]. Thus, one possibility is that, contrary to the standard assumption, criterion variances might be non-zero and unequal.

Delving into this possibility, a numerical optimization procedure was used in order to try to fit the non-speech data to the Condition B of the Law of Categorical Judgment. This condition assumes equal criterion variances and, thus, if acceptable fits are found, the idea that non-linear z-transformed ROCs are a consequence of unequal criterion variance can be ruled out. However, no such statistically acceptable fits were found for the noise-buzz data. Therefore, we find, at least, partial support to the claim that, for these stimuli, unequal criterion variances underlie the non-linear shape of the z-transformed ROC curve.

So, as we have shown, the same non-linear plot might arise from two very different situations. On the one hand, it might come from signal densities not being Gaussian. On the other hand, it might come from criterion variances being unequal. Interestingly, our data suggest that speech and non-speech fall at different places, which provides some evidence for modular theories of speech perception

[16], [17]. The non-linear ROCs for non-speech might be explained in terms of criterion variability; however, those for speech seem to be directly connected to the very shape of the signal densities they evoke. This seems to lend some support to the claim that speech perception requires some kind of domain specific processing routines in order to cope with the inherent problems associated to this task.

Acknowledgments

Burton S. Rosner and Jose E. García-Albea provided critical insights for all this work. Part of this work was presented at the 147th meeting of the Acoustical Society of America [18]. This research was partly supported, by grant SEJ2006/11955 from the Spanish Ministry of Education and Science.

References

[1] J.A. Swets, "Indices of discrimination or diagnostic accuracy: their ROCs and implied models", *Psychological Bulletin* 99, 100-117 (1986)

[2] N.A. Macmillan, R.F. Goldberg, L.D. Braida, "Resolution for speech sounds: basic sensitivity and context memory on vowel and consonant continua", *J. Acoust. Soc. Am.* 84, 1262-1280 (1988)

[3] D.B. Pisoni, "Auditory and phonemic codes in discrimination of consonants and vowels", *Perception and Psychophysics* 13, 253-260 (1973)

[4] B. S. Rosner, "Perception of voice-onset-time continua: a signal detection analysis", *J. Acoust. Soc. Am.* 75, 1231-1242 (1984)

[5] M.E.H. Schouten, A.J. van Hessen, "Modeling phoneme perception I: categorical perception", *J. Acoust. Soc. Am.* 92, 1841-1855 (1992)

[6] L.E. López-Bascuas, "Speech and nonspeech signal densities for the perception of temporal order", *Proc. Eurospeech´95,* 2281-2283 (1995)

[7] L.E. López-Bascuas, "Speech signals might ignore auditory processors", In W. Ainsworth, S Greenberg (Eds.) *Auditory basis of speech perception,* 158-161 (1996)

[8] L.E. López-Bascuas, B.S. Rosner, J.E. García-Albea, "Voicing and temporal order perception by Spanish speakers", *Proc. 15th International Congress of phonetic sciences,* 1, 403-405 (2003)

[9] M.E.H. Schouten, A.J. van Hessen, "Response distributions in intensity resolution and speech discrimination", *J. Acoust. Soc. Am.* 104, 2980-2990 (1998)

[10] R.E. Pastore, N.A. Macmillan, "Signal detection analysis of response distributions for intensity and speech judgments", *J. Acoust. Soc. Am.* 111, 2432 (2002)

[11] N.A. Macmillan, C.D. Creelman, "*Detection theory: a user´s guide*", Cambridge: Cambridge University Press (1991)

[12] M. Treisman, A. Faulkner, P.L.N. Naish, B.S. Rosner, "Voice-onset time and tone-onset time: the role of criterion-setting mechanisms in categorical perception", *Quarterly Journal of Experimental Psychology,* 48A, 334-366 (1995)

[13] J.D. Miller, C.C. Wier, R.E. Pastore, W.J. Kelly, R.J. Dooling, "Discrimination and labeling of noise-buzz sequences with varying noise-lead times: an example of categorical perception", *J. Acoust. Soc. Am.* 60, 410-417 (1976)

[14] W. S. Torgerson, "*Theories and methods of scaling"*, New York: Wiley (1958)

[15] M. Treisman, T.C. Williams, "A theory of criterion setting with an application to sequential dependencies", *Psychological Review,* 91, 68-111 (1984)

[16] A.M. Liberman, I.G. Mattingly, "The motor theory of speech perception revised", *Cognition,* 21, 1-36 (1985)

[17] I.G. Mattingly, A.M. Liberman, "Speech and other auditory modules", In G.M. Edelman, W.E. Gall, W.M. Cowan (Eds.), *Signal and Sense: global and local order in perceptual maps,* New York: Wiley (1990)

[18] L.E. López-Bascuas, B.S. Rosner, J.E. García-Albea, "Voice-onset time and buzz-onset time identification: a ROC analysis, *J. Acoust. Soc. Am.* 115, 2465 (2004)