



HAL
open science

What d' scores can hide and what decision space representations should reveal: an account on asymmetries in vowel perception

Charalampos Karypidis

► **To cite this version:**

Charalampos Karypidis. What d' scores can hide and what decision space representations should reveal: an account on asymmetries in vowel perception. Fechner Day 2009, Oct 2009, Galway, Ireland. pp.347-352. hal-00502791

HAL Id: hal-00502791

<https://hal.science/hal-00502791>

Submitted on 15 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WHAT d' SCORES CAN HIDE AND WHAT DECISION SPACE REPRESENTATIONS SHOULD REVEAL: AN ACCOUNT ON ASYMMETRIES IN VOWEL PERCEPTION

Charalampos Karypidis

LPP - UMR 7018, Université de la Sorbonne Nouvelle - Paris 3 / C.N.R.S.

ch.karypidis@gmail.com

Abstract

The introduction of the signal detection theory has been a significant step in psychophysics, and d' score estimations have helped us understand cognitive processing to a significant degree. However, researchers in the field of speech perception have been using this sensitivity measure while assuming that the decision spaces are symmetrical, neglecting thus the internal structure of categories. In a two-interval 'same-different' paradigm, for example, the areas covered by the S_1S_1 , S_1S_2 , S_2S_1 and S_2S_2 distributions are not necessarily equivalent. In this paper, we will contest the assumption of symmetry by revisiting some papers which provide evidence in favour of asymmetrical vowel perception. Perceptual distance is dependent on the salience of the acoustic properties of the stimuli, with salient stimuli either decaying slower or interfering more with the processing of adjacent stimuli. We will then attempt to provide a hypothetical decision space for a 'same-different' task implementing two vowels, <i> and <e> and will discuss the usefulness of decision spaces as tools for understanding the perceptual robustness of stimuli.

The importance of the signal detection theory in psychophysics has been undeniable. Its application, however, to the field of speech perception seems to neglect a portion of our knowledge of speech acoustics.

Vowels, upon which this article focuses, are highly complex auditory objects and defining the dimensions along which listeners make judgements is tenuous, if not impossible. Reducing this complexity to less than a handful of acoustic or perceptual dimensions (see Klein, Plomp, & Pols, 1970; Nearey, 1989 for earlier data) has proven to be illusive, which renders the estimation of decision boundaries problematic. Whereas boundaries for two-dimensional stimuli are often represented by quasi-straight lines on the Euclidean plane, it would need an enormous amount of data covering the totality of the representational plane for real-speech segments for us to understand the inner structure and equilibrium of categories. Unaware of these difficult-to-pinpoint discriminant parameters and wanting to control as many unknown variables as possible, many researchers have been using synthetic stimuli. Our view is that a study using speech-like stimuli is more relevant to psychoacoustics and mnemonic capacities than to the way listeners actually decode speech.

The issue of perceptual complexity will be discussed in a future paper. In the following section, we will discuss another a priori assumption often present in most studies: the symmetry of the decision space.

Symmetries in decision spaces and asymmetries in vowel perception

Let us begin with a simple example of a 'same-different' experiment using one-dimensional stimuli differing in intensity. Figure 1 depicts the decision space for stimuli S_1 and S_2 . The four areas covered by S_1S_1 , S_2S_2 and S_1S_2 , S_2S_1 respectively, appear to be symmetrical and

tantamount. When the perceptual difference between two auditory traces falls within the grey areas, the subject is expected to give a ‘different’ answer.

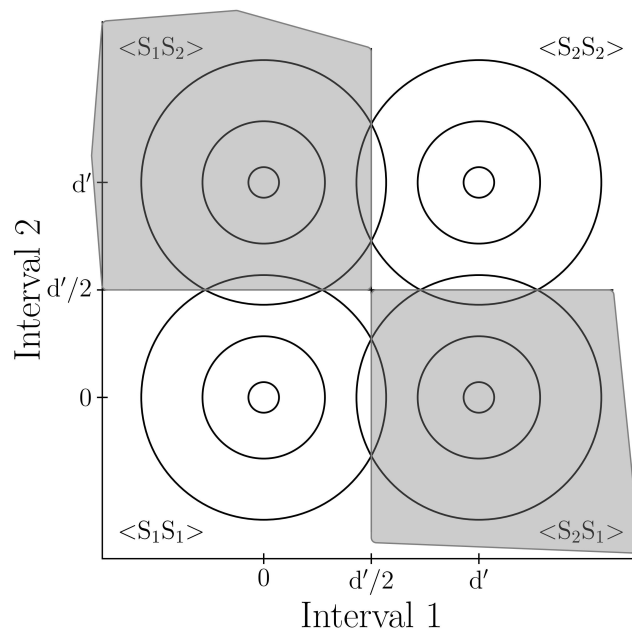


Figure 1. Decision space for a ‘same-different’ task (from Macmillan & Creelman, 2005, p.215). The listener must respond ‘different’ in the shaded area.

This representation has been deemed to be valid for continua of virtually any nature and, as far as this paper is concerned, also speech categories. This parallelism between semantic categories stored in long-term memory and *ad hoc* categories (Barsalou, 1983) created by the subject during an experimental task is equivalent to neglecting a primary characteristic of semantic categories: internal structure.

Within most theoretical frameworks dealing with perception, categories are considered to be internally structured. Rosch’s take on Gestalt theory leads her to the concepts of ‘cognitive reference points’ and ‘natural prototypes’ (Rosch & Mervis, 1975; Rosch, 1975, 1973); Kuhl (1991) speaks of long-term memory prototypes whereas Massaro’s (1989) fuzzy-logical model assumes the use of prototypes created on-the-spot for the needs of a certain task and to which incoming stimuli are compared. A significant amount of evidence appears to corroborate the idea that this structure has an impact on the perceived distance between stimuli.

In a two-interval paradigm, we can postulate either that the processing of the two stimuli occurs in a discrete/independent way – no interaction – or that the two auditory traces interact. The first hypothesis would more or less correspond to the categorical perception model (Lieberman, Harris, Hoffman, & Griffith, 1957), where each stimulus is processed independently and the result of cognitive processing is not an auditory trace but a label, that is, the category to which the stimulus belongs to. This does not appear to be the case (Schouten, Gerrits, & Hessen, 2003) even when the same paradigm is used for identification and discrimination (Repp, Healy, & Crowder, 1979). We will posit the independence of stimuli in the following paragraph.

Repp, Healy, & Crowder (1990) use an *ad hoc* stimulus set to examine the neutralization hypothesis formulated by Cowan & Morse (1986) and according to which the auditory trace of a stimulus ‘drifts’ towards the centre of the vowel space while stored in memory. The more a trace is retained in short-term memory, the greater its ‘displacement’ towards this central region corresponding to the vowel [ə], found in the weak form of “the”. The prediction is that in the case of stimuli 1 and 2 (Figure 2), the perceived distance between these vowels will be

greater when they are presented in the order 2-1 than in the order 1-2. Repp, Healy, & Crowder conclude that stimuli do not decay towards $[\text{ə}]$ but rather towards the centre of the area covered by a stimulus set. Given the model illustrated in Figure 2, the decision space will be asymmetrical, with the area of S_1S_2 being more restricted and possibly closer to S_1S_1 and S_2S_2 than the area of S_2S_1 . As we can see, the neutralization hypothesis does not necessarily assume any interaction between the two stimuli but focuses on the acoustic nature of the stimuli and on short-term-memory issues.

The nature of the stimulus is the core of another hypothesis, that of peripherality. Polka & Bohn (2003), summarizing previous literature on infant perception (Best & Faber, 2000; Polka & Bohn, 1996; Polka & Werker, 1994, among others) support the idea that listeners are induced to give more 'different' answers when the vowel change in a pair is directed towards a more peripheral region of the vowel space (see Figure 3). Karypidis's (to appear) reanalysis of Repp & Crowder's data and his adaptation of their original protocol for French vowels provide further evidence in favour of the peripherality hypothesis.

The origin of this asymmetry has yet to be elucidated. We believe two approaches can be adopted for this issue to be dealt with. We will take the pair /i/-/e/ as an example. The first point of view consists in considering that *more 'different' answers are given in the presentation order <ei>*. Peripheral vowels are perceptually more salient and thus more probable to serve as perceptual anchors (reference points). Rosch (1975) reports that subjects find the phrase '52 is greater than 50' more natural than the phrase '38 is greater than 32', 50 playing the role of a reference point because it is a multiple of 10 and thus 'salient in the decimal system'. In order for the perceptual distance between the two stimuli to be substantial, the reference point has to be placed second in a pair, the salience of the anchor becoming accentuated by means of recency effects. If the anchor is placed first, the decay of its properties caused by the volatility of the short-term memory will not allow for a contrast effect to occur. Something similar may come about with virtually any type of stimuli where reference points may be found or constructed.

An alternative to the theory of contrast effect is that of forward masking. This time, the idea is that *less 'different' answers are given in the order <ie>*. This simply consists in a primacy effect: the more salient the first stimulus is, the more stable its auditory trace will be and the more it will interfere with the second stimulus's processing.

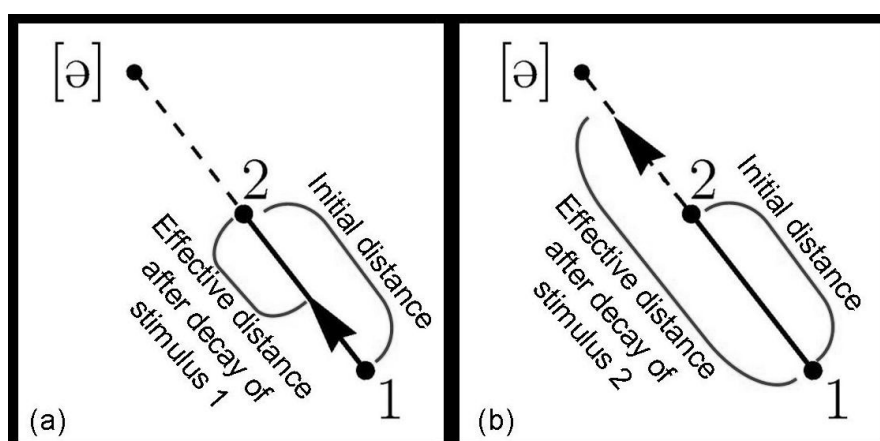


Figure 2. Direction of decay predicted by the neutralization hypothesis for stimuli 1 and 2 presented in the order 1-2 (a) and 2-1 (b). In (a), 1 moves towards $[\text{ə}]$ and, at the same time, towards stimulus 2; in (b), stimulus 2 moves towards $[\text{ə}]$ but this time, further away from 1.

The actual mental representations of stimuli 1 and 2 are located at the endpoint of each arrow. After Karypidis (2007).

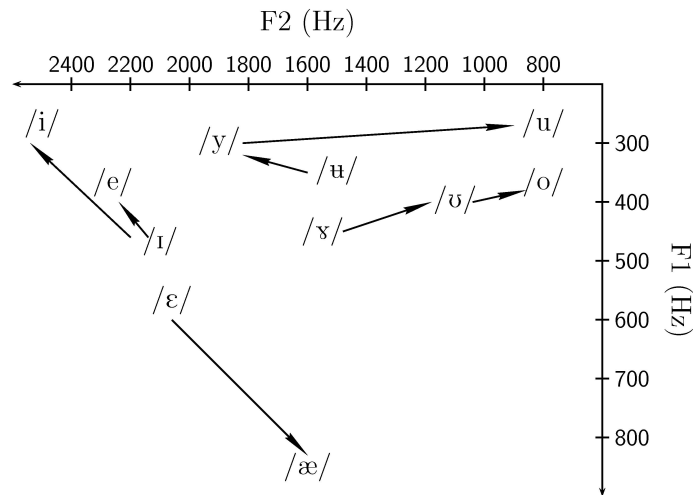


Figure 3. F1/F2 representation of the asymmetries found in infant data (after Polka & Bohn, 2003). Arrows point towards the direction in which more ‘different’ answers were given.

The two approaches, focusing on different processes, provide different estimations of decision spaces. Let /i/ and /e/ correspond to S_1 and S_2 of Figure 1 respectively. The contrast effect will therefore result in a low-variance distribution for S_2S_1 with more hits whilst the second calls for a higher-variance S_1S_2 distribution with more misses.

Apart from the grey ‘areas’ in Figure 1, the white – ‘same’ – areas are also of great interest. If we consider S_1 as more salient than S_2 and given the postulate that salient stimuli are more robust in short-term memory, the distributions for the two stimuli will be unequal, with $\sigma_{11} < \sigma_{22}$. This means that fewer false-alarms will be found for <ii> than for <ee>. An illustration of the above is presented in Figure 4. It should be noted that the non-rectangular shape of the area between the distribution means does not allude to a perceptual integration.

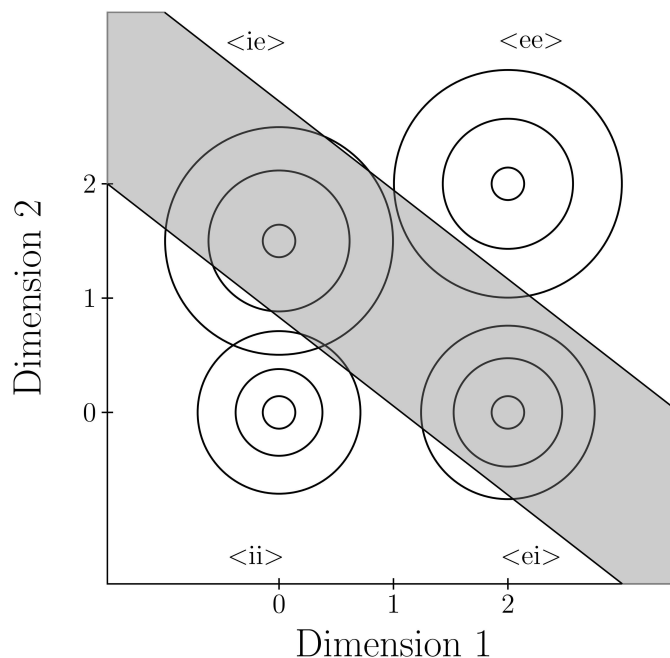


Figure 4. The hypothetical decision space for a ‘same-different’ task exploring <i> and <e>. Variance for <ii> is relatively low, triggering fewer false-alarms than <ee> does. Variance inequality is also found for <ei> and <ie>, with the former yielding more hits.

Conclusion

Asymmetries in perception are a part of cognitive processing and deserve a more extensive study, given that they can possibly reveal the degree to which a stimulus serves as a reference point. We can achieve this by investigating the effect that a stimulus has on adjacent stimuli. If, on the other hand, asymmetries are triggered by memory decay, we might be able to come up with a model predicting the evolution of auditory traces during their decay. Which of their properties 'disappear' first? Is it the redundant ones? If so, a hierarchisation of category attributes may be tangible. And what if stimuli decay in a chaotic way? Does that mean that a stimulus is an indivisible unit that gives a unique sensation and is not the sum of the sensation of each of its components, as Professor James has eloquently stated more than a hundred years ago?

Although d' scores may give us an insight into the sensitivity of a listener to a specific change in one or more dimensions, we believe that the careful study of the former – hits, misses, etc. – and of their distribution can provide us with rich information on a stimulus's sensation.

References

- Barsalou, L. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211-227.
- Best, C. T., & Faber, A. (2000). Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel. In *International Conference on Infant Studies*. Brighton, UK.
- Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, *79*, 500–507.
- Karypidis, C. (2007). Order effects and vowel decay in short-term memory: the neutralization hypothesis. In *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 657-660). Saarbrücken, Germany.
- Karypidis, C. (to appear). Les effets d'ordre de présentation et l'hypothèse de neutralisation. La voyelle neutre et les points de référence. In *Cognito - Les Cahiers Romains des Sciences Cognitives*.
- Klein, W., Plomp, R., & Pols, L. C. W. (1970). Vowel Spectra, Vowel Spaces, and Vowel Identification. *The Journal of the Acoustical Society of America*, *48*(4B), 999-1009.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*, 93–107.
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, H. S. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358–368.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Lawrence Erlbaum Associates.
- Massaro, D. W. (1989). Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology*, *21*(3), 398-421.
- Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, *85*(5), 2088-2113.
- Polka, L., & Bohn, O. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221-231.
- Polka, L., & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, *100*, 577-592.

- Polka, L., & Werker, J. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 421-435.
- Repp, B., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 129-145.
- Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *Journal of the Acoustical Society of America*, 88(5), 2080–2090.
- Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, 7(4), 532–547.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4(3), 328–350.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573-605.
- Schouten, B., Gerrits, E., & Hesses, A. V. (2003). The end of categorical perception as we know it. *Speech Communication*, 41(1), 71-80.