Available online at www.sciencedirect.com

SciVerse ScienceDirect

Speech Communication 55 (2013) 33–39

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Detecting categorical perception in continuous discrimination data

Paul Boersma [*], Kateřina Chládková

Amsterdam Center for Language and Communication, University of Amsterdam, The Netherlands

Available online 13 June 2012

Abstract

We present a method for assessing categorical perception from continuous discrimination data. Until recently, categorical perception of speech has exclusively been measured by discrimination and identification experiments with a small number of different stimuli, each of which is presented multiple times. Experiments by Rogers and Davis (2009), however, suggest that using non-repeating stimuli yields a more reliable measure of categorization. If this idea is applied to a single phonetic continuum, the continuum has to be densely sampled and the obtained discrimination data is nearly continuous. In the present study, we describe a maximum-likelihood method that is appropriate for analysing such continuous discrimination data.

Keywords: Categorical perception; Dense sampling; Discrimination; Maximum likelihood

1. Introduction

In speech perception research, categorical perception of vowels and consonants has been assessed by experiments that involve identification and discrimination tasks. In an identification task, sounds that belong to the same category receive the same label. Identification has mostly been tested by means of a multiple-forced choice experiment in which listeners label each stimulus as one of the phonemes of their native (or second) language.

In categorical perception, discrimination of sounds across a category boundary is easier than discrimination of sounds within a category (Liberman et al., 1957; Eimas, 1963). To test the discrimination of speech sounds, various laboratory tasks have been designed and utilized, which may slightly differ in the extent to which they exhibit categorical perception effects (Gerrits and Schouten, 2004); among these are the classical AX (“same”–“different”) task, in which listeners indicate whether the sounds of a pair are the same or different (e.g. Pisoni, 1973), the AXB task, in which listeners identify the second sound of a stimulus triplet either with the first sound or with the last sound (Liberman et al., 1957), or the 4IAX (four-interval “same”–“different”; or ABAA) task, in which listeners have to indicate whether the first or the second pair of a stimulus quadruplet contained a deviant sound (Pisoni, 1975).

Discrimination experiments reported in the vast majority of previous studies have used a relatively small number of stimuli that were repeated multiple times within a single experiment: in one of their experiments, Liberman et al. (1957) used 12 different stimulus pairs, and it is hard to find studies that employ even a slightly larger number. The use of such a small number of different stimuli comes with a problem. In order to have a sufficient amount of data to determine the existence of a discrimination peak, each stimulus pair has to be repeated multiple times (42 in Liberman et al.’s case), and this is problematic: Rogers and Davis (2009) showed that “stimulus repetition reduces discrimination of within-category differences, and enhances between-category discrimination” (p. 379). They compared the results of a discrimination task with eight different stimuli repeated 208 times each to the results of a discrimination task with 384 different stimuli that occurred eight times each, and found that having numerous repetitions of a small number of different stimuli introduces a bias towards perceiving the stimuli categorically. Although Rogers and Davis used many different phonetic continua in their

[*] Corresponding author. Address: University of Amsterdam, NL-1016CG Amsterdam, The Netherlands.
E-mail addresses: paul.boersma@uva.nl (P. Boersma), k.chladkova@uva.nl (K. Chládková).

0167-6393/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.specom.2012.05.003

"non-repeating" task (namely, 96), we like to apply their idea to the case of a single phonetic continuum, for reasons of ecological validity. That is, discrimination experiments should be designed with a large number of non-repeating stimuli even if the experiment is performed along a single continuum; the stimuli will then have to be densely sampled along that continuum, a situation that corresponds to how humans learn from phonetic continua in the real world. The explicit laboratory task for the listener (AX, AXB, 4IAX) can then stay the same as in earlier experiments.

A densely sampled (i.e. effectively "continuous") design poses a problem for the analysis of the data that earlier experiments did not face. In one of the experiments by Liberman et al. (1957), for instance, there were only 12 stimulus pairs along the phonetic continuum, so that each of those 12 points could be measured enough times (namely, 42) to produce a reliable measure of "percentage correct"; together these 12 percentages formed a response curve that could be visually inspected for whether it showed a discrimination peak. Statistical corroboration of the existence of a discrimination peak typically involved performing an analysis of variance on the heights of the curve near the boundary and away from the boundary (e.g. Best and Strange, 1992). These methods are not immediately available if the number of repetitions of each stimulus is low. In the discrimination experiment reported below, for instance, each stimulus pair was measured only twice (namely, once in each order of its members), so that any raw response curve would look quite noisy. In this paper we therefore introduce an analysis method that is appropriate for densely sampled discrimination data: we show how to represent the raw data as a continuous curve for visual inspection, and how to statistically establish the existence of a discrimination peak by a maximum likelihood method.

2. The experiment

We will illustrate our analysis method with example data obtained from real listeners. In this section we therefore report on a small perception experiment that addressed discrimination within the continuum between [i] and [ɛ]. In Section 3 we try to infer categorical perception along this continuum on the sole basis of the discrimination data obtained in this experiment. We did not elicit identification data, because the experiment was a part of a larger experiment that included continua on which the participants' language had no categories. That larger experiment, which has a research question on feature generalization, will be reported elsewhere; the subject of the present paper is only the establishment of the appropriate analysis method.

2.1. Stimuli

The stimuli were isolated steady vowels that differed only along an F1 continuum. They were synthesized with the Klatt synthesizer (Klatt and Klatt, 1990), as built into the program Praat (Boersma and Weenink, 1992–2011), and modeled after a female voice. The vowels all had the same F2 value, namely 2700 Hz, and F3 through F10 were fixed at 3300, 3850, 4950, 5950, 6950, 7950, 8950, and 9950 Hz, respectively. The vowels had a rising-falling pitch contour: F0 was 220 Hz at the start of the vowel, went up linearly to 270 Hz at one third of the duration, and from there fell down linearly to 170 Hz at the end. The voicing amplitude was maximal at the start of the vowel and fell linearly by 13.5 percent towards the end. Along the F1 continuum, which ranged from 6.93 erb (280 Hz) to 12.86 erb (725 Hz), we created 130 stimulus pairs that were equally spaced along the continuum; this resulted in 260 different vowels (as summarized in Fig. 1).

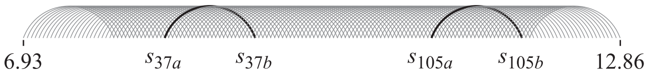

Fig. 1. The 130 stimulus pairs on an F1 continuum in erb. Each pair consists of two points along the horizontal axis, connected here by an arc. The distance between the members of a pair is constant, i.e. $s_{37b} - s_{37a} = s_{105b} - s_{105a} = 0.9$ erb.

The F1 distance between the two vowels *within* a stimulus pair was 0.9 erb, so that the 130 low members of the pairs ranged from 6.93 to 11.96 erb, and the high members from 7.83 to 12.86 erb. The number of 0.9 erb was chosen because in a pilot experiment a difference of 0.9 erb was just small enough to make the number of spontaneous "same" judgments comparable to the number of spontaneous "different" judgments. The number of 0.9 erb is also comparable to the just-noticeable difference for formants as measured by Mermelstein (1978).[1]

The F1 distance *between* two neighboring stimulus pairs was more than a factor of 20 smaller than the distance *within* a pair, namely (11.96 – 6.93)/(130 – 1) = approximately 0.039 erb. By thus oversampling the difference limen, we render the stimulus set effectively continuous.

As Fig. 1 illustrates, both the within-pair F1 distance and the between-pairs F1 distance were kept the same for all the 130 stimulus pairs along the continuum.

2.2. Procedure

Vowel discrimination was tested by means of a traditional AX task. The inter-stimulus interval (i.e. the time interval between the two members of a pair) was 500 ms, and the trial-initial silence (i.e. the time interval between the participant's mouse click and the first member of the next pair) was 600 ms. Each of the 130 stimulus pairs occurred twice, that is, in one trial the pair member with the lower F1 was played first, while in the other trial with the same pair the member with the higher F1 was played first; this was to factor out any stimulus-order effects that

[1] Kewley-Port (1995) found much smaller difference limens for formants, namely around 0.2 erb for /ɪ/. This is still much greater than the step between our pairs.

had been reported in previous vowel discrimination experiments (Polka and Bohn, 2003). The complete set of 260 pairs of stimuli was presented in random order.

As described above, the two members of a stimulus pair were never identical, and in fact the auditory distance between the two members of a pair was the same for every trial. Despite the fact that the two sounds were always different, we asked the listeners to indicate whether the sounds were different or the same; as noted in Section 2.1, roughly a 50 percent "same" judgment was expected.

In line with the definition of categorical speech perception, our listeners (whose language has at least two segmental phonemes along the presented vowel continuum) were expected to perceive stimulus pairs in some regions of the F1 continuum as different (i.e., stimuli across a category boundary) and stimulus pairs in other regions of the F1 continuum as identical (i.e., stimuli that lie within one category). We can find (a gradient form of) categorical perception if our listeners have more "different" responses for stimulus pairs in some regions along the vowel continuum than for stimulus pairs in other regions. The location of the category boundary will lie between the sounds that elicit the largest number of "different" responses.

2.3. Participants

The subjects in the whole experiment (which will be reported elsewhere) were a large group of young Czech monolinguals. In the present paper, which only addresses the analysis method, we discuss only three of these participants; we choose these three people because they seem to reflect the three most common strategies found in the larger group.

3. Analyzing a listener

The analysis of the data of a single listener runs as follows. The listener is confronted with N (here: 130) different stimulus pairs. The nth stimulus pair ($n = 1 \ldots N$) is repeated K_n (here: always 2) times. Of these K_n replications, the listener judges a pair as "same" s_n times, and as "different" d_n times, with $s_n + d_n = K_n$.

Fig. 2 shows the raw data of three listeners. For every stimulus pair, the possible number of "different" responses was 0, 1, or 2, and the figure shows that the listeners indeed used all three possibilities. Since the visualization of the raw data by poles is not very informative with respect to where the discrimination peaks lie, Fig. 2 also shows smoothed versions of the data, obtained by convolving the raw data with a unit-area Gaussian (Babaud et al., 1986) with a standard deviation of 10 steps (i.e. 0.39 erb) along the continuum; an edge correction is obtained by dividing the resulting curve by the convolution of that same Gaussian with data consisting of all ones (analogously to the window correction for autocorrelation in Boersma, 1993). The first purpose of having these smoothed curves is to help us in visually inspecting the data: they suggest, for instance, that participant 1 has a constant probability of judging "different", that participant 2 has a single discrimination peak around stimulus pair 49, and that participant 3 could have discrimination peaks around stimulus pairs 53 and 113. Whether these visual suggestions are correct, e.g. whether the small right-hand bump of listener 2 is indeed irrelevant and the taller right-hand bump of listener 3 is not caused by random variation, remains to be seen. The following three subsections therefore submit these data to several maximum-likelihood analyses, each of which

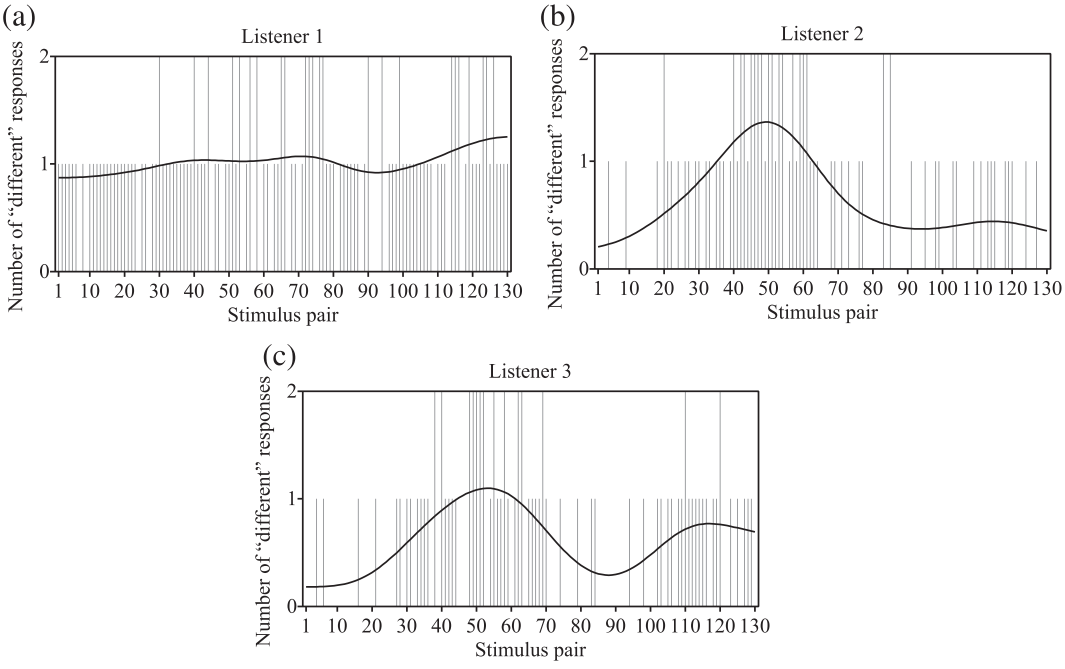

Fig. 2. Raw data (grey poles) and smoothed data (solid curves) of three participants with apparently zero, one, and two discrimination peaks, respectively.

corresponds to a different model of what the listener is doing. The models are compared in Section 3.4.

3.1. First model: no discrimination peaks

Our first, simplest, model assumes that the listener has no categorical perception along the continuum but instead only has an acoustic discrimination strategy. Since we used constant distances along the auditorily uniform erb scale, an ideal acoustic listener has a constant probability p_{const} of judging any stimulus pair as "different". In other words, the probability p_n that the nth stimulus pair is judged as "different" is simply

$$p_n = p_{const} \tag{1}$$

Although an estimate of the parameter p_{const} could simply be computed by dividing the total number of "different" judgments by the total number of trials (260), we here provide a more general method of estimation, which can also be used for more complicated formulas for p_n, as we do in Sections 3.2 and 3.3.

The general maximum-likelihood method (Fisher, 1922) for finding the best underlying model p_n (for any formula for p_n, not just the model in (1)) runs as follows. The probability that both the first and the second presentation of the first pair are judged as "different" is p_1^2, and the probability that they are both judged as "same" is $(1 - p_1)^2$; the probability that the first is judged as "different" and the second as "same" is $p_1(1 - p_1)$, and the probability that the first is judged as "same" and the second as different is $(1 - p_1)p_1$. In general, the probability of a certain observed sequence of d_1 "different" judgments and s_1 "same" judgments is $p_1^{d_1}(1 - p_1)^{s_1}$. Given the values of all d_n and s_n of the listener, the probability of the total observed data sequence of the listener is

$$L = \prod_{n=1}^{N} p_n^{d_n}\ (1 - p_n)^{s_n} \tag{2}$$

This probability is denoted as L, because it is the *likelihood* associated with the parameter(s) of p_n. The logarithm of this is the "log-likelihood"

$$\begin{aligned} LL &= \ln \prod_{n=1}^{N} p_n^{d_n}\ (1 - p_n)^{s_n} \\ &= \sum_{n=1}^{N} (d_n \ln p_n + s_n \ln(1 - p_n)) \end{aligned} \tag{3}$$

The best underlying model is now the set of parameters for p_n that maximize LL. This is true for any formula for p_n; in the case of the parametrized model in (1), the only parameter is p_{const}, so we have to find the value of p_{const} that maximizes LL.

The maximization of LL for the model in (1) runs as follows. We initially assign to the parameter p_{const} a random value between 0 and 0.5 and subsequently add small positive or negative values to it (starting with a uniformly distributed random number between −0.1 and +0.1, under the constraint that p_{const} stays between 0 and 1), always checking whether LL improves (becomes less negative) according to formulas (1) and (3). Whenever LL improves, we keep the changed p_{const} as our new best value of p_{const}, and we subsequently start again from this new value. After 1000 iterations, in which the maximum change gets exponentially smaller (after 1000 iterations it has decreased by a factor of 100, i.e. to a uniformly distributed random number between −0.001 and +0.001), we arrive near *the* best value of p_{const}. We then home in on the best value in 10,000 more steps, in which the changes gradually decrease by another factor of 100), and thus arrive at the best value for p_{const}. For listener 1 it is 0.508, for listener 2 it is 0.319, and for listener 3 it is 0.304. The top row of Fig. 3 shows these values, together with the best LL values obtained. As expected, the optimized p_{const} values are indeed identical to the overall fraction of "different" responses. The Figure suggests that the constant model of equation (1) fits the data well for listener 1 but not for listeners 2 and 3 (this is corroborated in the model comparison of Section 3.4).

3.2. Second model: one discrimination peak

Our second model assumes that the listener mixes an acoustic discrimination strategy with a categorical perception strategy based on the existence of two categories along the continuum. We assume, therefore, that the probability of a "different" judgment shows one peak somewhere along the continuum. If we assume that the peak has a Gaussian shape, is centered at μ, has a height of p_+ and a width of σ, and that the height of the tails of the peak far away from μ is p_-, the formula for the probability of a "different" judgment for the nth pair is

$$p_n = p_- + \ (p_+ - p_-) \exp\left(-\frac{(n - \mu)^2}{2\sigma^2}\right) \tag{4}$$

In terms of an underlying categorical perception model, p_- can be regarded as the probability of judging two 0.9-erb-distant stimuli *within* a category as "different" (this is 0 for perfect categorical perception), and p_+ can be regarded as the probability of judging two 0.9-erb-distant stimuli *across* a category boundary at their midpoint as "different" (this is 1 for perfect categorical perception across an infinitely crisp boundary).

The optimization procedure again starts with random values of the four parameters p_- (between and 0 and 0.5), p_+ (between p_- and 1), μ (between 1 and 130) and σ (between 0 and 100), and randomly changes these parameters 1000 times (initially by at most ±0.1 for p_- and p_+, and ±10 for μ and σ) so as to increase the value of LL according to (4) and (3), under some constraints (p_- and p_+ have to stay between 0 and 1, p_+ can never become less than p_-, μ has to stay between 1 and 130, and σ has to stay positive). Since this procedure can arrive in a non-global local optimum, it is repeated 100 times from different random starting conditions, yielding a set of 100 best LL values. The

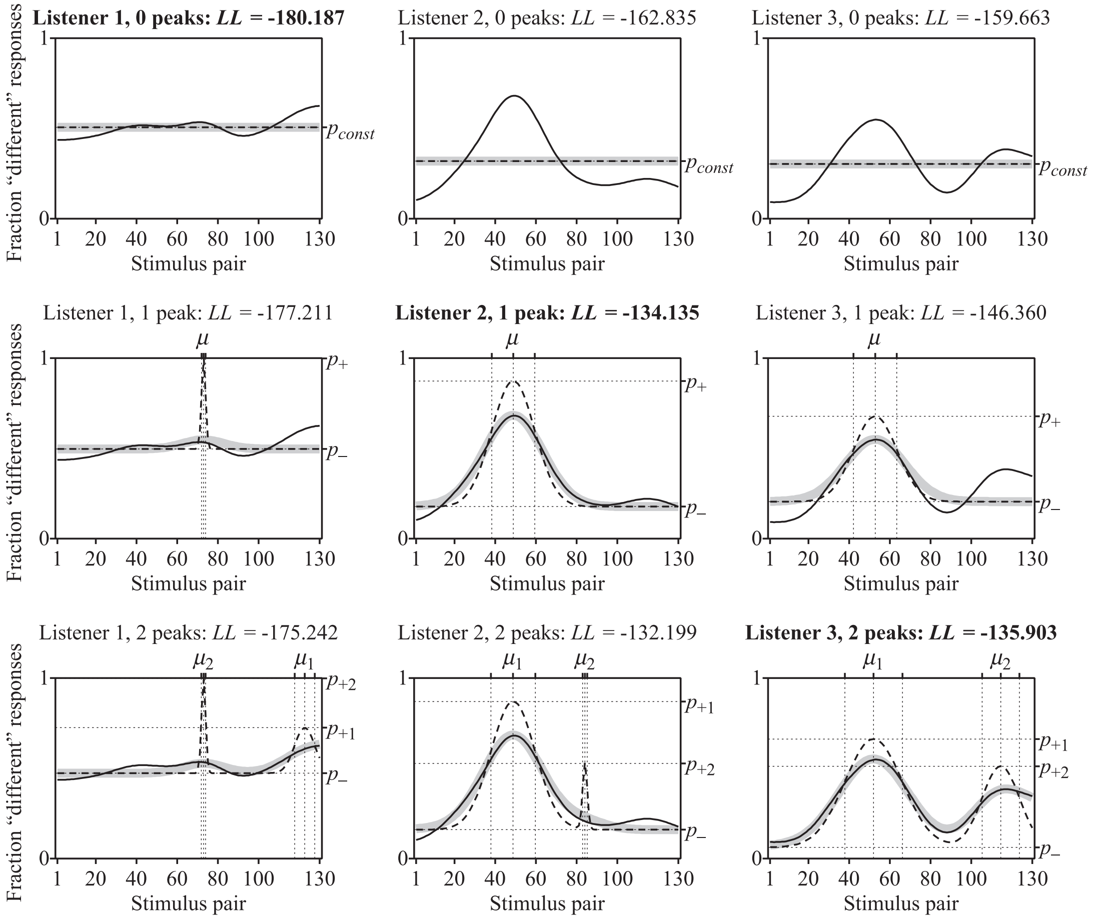

Fig. 3. Maximum-likelihood fitting of three listeners, each with zero, one, and two peaks. Solid curves: smoothed data (copied from Fig. 2). Dashed curve: fit (unlabelled vertical dotted lines: $\mu \pm \sigma$). Thick grey curve: smoothed fit.

highest of these 100 values is chosen as our first approximation of the "best best *LL*"; starting from the parameter values associated with this approximation, we make the approximation more precise with a single set of 10,000 decreasing changes, as before.[2] The resulting best best *LL* values are shown in the middle row of Fig. 3. The figure shows both the fitted p_n itself and its smoothed version, which ought to be close to the smoothed data because it is obtained by convolving the fitted p_n with a unit-area Gaussian with a standard deviation of 10 (and edge correction), just as was applied to the raw data above. We see that visually, the smoothed fit for listener 2 is indeed very close to her smoothed data (this comparison is the second purpose of showing the smoothed raw data).

3.3. Third model: two discrimination peaks

Our third model assumes that the listener has three categories along the continuum, and therefore two discrimination peaks:

$$p_n = p_- + (p_{+1} - p_-)\exp\left(-\frac{(n-\mu_1)^2}{2\sigma_1^2}\right) + (p_{+2} - p_-)\exp\left(-\frac{(n-\mu_2)^2}{2\sigma_2^2}\right) \qquad (5)$$

When we then optimize the seven parameters p_-, p_{+1}, p_{+2}, μ_1, μ_2, σ_1, and σ_2 for the maximum *LL*, in an iterative process analogous to Section 3.2, we obtain the bottom row of Fig. 3. For each listener, the smoothed fit is now close to her smoothed data.

3.4. Comparison of the three models

Instead of judging visually how an increase in the number of model parameters improves the fit or not, we should ask the question: does the likelihood rise significantly with each addition of parameters? The table below summarizes the values of *LL* for the three listeners, together with ΔLL, which is the increase in *LL* from the next simpler model. Following a known property of maximum likelihood estimation (Wilks, 1938), the *p* values in the table are derived from performing a χ^2 test on $-2\Delta LL$ (with 3

[2] There are faster methods for finding local optima (Press et al., 1992, ch. 10.6), and they can be used as well. Repeating the search multiple times to find the *global* optimum, however, cannot be prevented.

degrees of freedom, which is the number of parameters added to the model with each peak).

We see that the data of listener 1 show no evidence for any discrimination peak, i.e. that they are consistent with the idea that he listens acoustically (with a probability p_{const} of hearing the difference) or that he has only one category (with a bias p_{const} toward responding "different"); a mix of these two strategies is also possible. The data of listener 2 indicate that she has at least one discrimination peak, which suggests that she has at least two categories (again, p_- reflects the success of acoustic listening and/or a bias toward responding "different"); there is no evidence for more categories than two. The data of listener 3 indicate that she has at least two discrimination peaks, suggesting that she has at least three categories, without evidence for more. Summing up, the *minimal* model for listener 1 has zero peaks, the minimal model for listener 2 has one peak, and the minimal model for listener 3 has two peaks. Deciding what is the *best* model for each speaker is more speculative: Akaike's Information Criterion (Akaike, 1974) favors a model over another if the improvement in LL is greater than the increase in the number of parameters (i.e. 3), and would therefore indeed select zero peaks for listener 1 (although having one peak is only 2.3% less likely), one peak for listener 2, and two peaks for listener 3; several Bayesian criteria (for an overview, see Pitt et al. (2002)) could also be used.

3.5. Interpretation of models

As has been illustrated in Section 3.4, the number of peaks in the last significantly-improving model can be interpreted as the number of reliably detected category boundaries that the listener has along the continuum. In the absence of response bias and decision noise, the value of p_- can be interpreted as the probability that the listener is listening acoustically within a category, and the value of $p_+ - p_-$ can be interpreted loosely as the degree of categorical perception. The value of μ expresses the location of the category boundary (here, a real number between 1 and 130), and σ expresses the crispness of the boundary.

Several other models are possible. If your idea of the underlying mechanism of the listener's responses is such that p_{+1} has to be equal to p_{+2}, you can eliminate one parameter in (5) by forcing p_{+1} and p_{+2} to be identical, leading to a model with only 6 parameters:

$$p_n = p_- + (p_+ - p_-) \times \left(\exp\left(-\frac{(n-\mu_1)^2}{2\sigma_1^2}\right) + \exp\left(-\frac{(n-\mu_2)^2}{2\sigma_2^2}\right)\right) \qquad (6)$$

When we apply this model to the data of listener 3, the result is very similar to the third row of the third column of Fig. 3, but with a value for p_+ intermediate between p_{+1} and p_{+2} in that figure. The resulting log-likelihood is -136.532, which is significantly better than the value for one peak ($\chi^2(df = 2) = 2(146.360 - 136.532)$, $p = 0.00054$) but not significantly worse than the value for two peaks with different heights ($\chi^2(df = 1) = 2(136.532 - 135.903)$, $p = 0.26$), which means that the model in (6) has reliably detected the second peak and that allowing p_+ to be different for the two peaks, as in (5), has not reliably been shown to improve the fit.

The models in (4)–(6) could be criticized on the basis that they are just "fitting" models, i.e., they produce a fit with Gaussian response curves without supplying an underlying mechanism for the shape of these curves. An alternative model could be based on production: it may be argued that the value of F1 produced for the categories A and B in the listener's prior learning environment were distributed as the Gaussians $P_A(x) = \text{norm}(\mu_A, \sigma_A)$ and $P_B(x) = \text{norm}(\mu_B, \sigma_B)$, where x is the value of F1 in erb. A probability-matching strategy for the listener (as is predicted by some perception learning algorithms, e.g. Boersma, 1997) would lead her to identify the auditory value x as the category A with a probability of $I_A(x) = P_A(x)/(P_A(x) + P_B(x))$. In the perfect categorical perception case (Liberman et al., 1957), the probability of a "different" judgment would then be $D_{AB}(x) = I_A(x - d/2)I_B(x + d/2) + I_A(x + d/2)I_B(x - d/2)$, where d is the distance between the members of a pair (0.9 erb). Mixing this with the simplest model of acoustic versus categorical listening strategies (for two categories) would yield

$$p_n = p_{acoustic} + p_{categorical} D_{AB}(x_n) \qquad (7)$$

where $p_{acoustic}$ is the probability of listening purely acoustically times the probability of detecting a 0.9 erb acoustic difference, $p_{categorical}$ is the probability of listening purely categorically, and x_n is the erb value in the center of the nth stimulus pair. The six parameters of the model in Eq. (7), i.e., $p_{acoustic}$, $p_{categorical}$, μ_A, σ_A, μ_B, and σ_B, can be computed just as easily as the four parameters in Eq. (4).

4. Discussion

Previous research has argued that discrimination tasks with a large number of different non-repeating stimuli form a more "naturalistic" environment for measuring categorical perception than tasks with a small number of repeating stimuli (Rogers and Davis, 2009). When this finding is applied to a single phonetic continuum, the need arises for a method of analysis suitable for continuous discrimination data; devising such a method was our aim in this study. We introduced a maximum-likelihood method that is appropriate for "continuous" (i.e. densely sampled) discrimination data, analogously to the way in which another maximum-likelihood method, namely the usual method of finding the optimal values of the parameters in a logistic regression, is appropriate for continuous *identification* data (Nearey, 1990). Our method thereby contributes to the validity of any claims about categorical perception made on the basis of continuous data. We illustrated how the method works on continuous discrimination data of three real listeners.

Table 1
Development of log-likelihood as a function of the number of modeled distribution peaks. Bold = statistically significant improvement.

Model	Listener 1	Listener 2	Listener 3
No peaks	**−180.187**	**−162.835**	**−159.663**
One peak	−177.210	**−134.135**	**−146.360**
Improvement	+2.977	**+28.700**	**+13.303**
p	0.11	$\mathbf{2.1 \cdot 10^{-12}}$	$\mathbf{7.1 \cdot 10^{-6}}$
Two peaks	−175.242	−132.199	**−135.903**
improvement	+1.968	+1.936	**+10.457**
p	0.27	0.28	**0.00011**
Three peaks	−174.798	−131.987	−135.671
improvement	+0.444	+0.212	+0.232
p	0.83	0.94	0.93

The present method fits the obtained discrimination function with several models that assume different numbers of discrimination peaks. Given that a peak in the discrimination function corresponds to a category boundary (Liberman et al., 1957), this method determines a plausible (or at least minimum) number of categories along the stimulus continuum. The method also determines the locations and crispnesses of the boundaries. Of course one cannot divide the 62 listeners into three groups solely on the basis of the *p* values in Table 1 (one cannot prove that a listener does not have more peaks). Such a division may require adding latent variables to the model.

The method presented here is quite general. While we used it here for a case with dense sampling along the auditory continuum, it could have been used with any of the cases reported in the previous literature, which typically samples the continuum sparsely into 7 to 15 values. While we used the method here with even sampling, it applies equally well to other unskewed kinds of sampling such as random sampling. And while the method was applied here to an AX task, it can be applied with the same ease to an ABX task, a 4IAX task, or to any other task in which the participant has to choose from two response options, especially if an underlying decision mechanism, like the one presented here at the end of Section 3.5, can be formulated.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Babaud, J., Witkin, A.P., Baudin, M., Duda, R.O., 1986. Uniqueness of the Gaussian kernel for scale-space filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1), 26–33.

Best, C.T., Strange, W., 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. Haskins Laboratory Status Report on Speech Research SR-109/110, 89–108.

Boersma, P., 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In: Proc. Institute of Phonetic Sciences, University of Amsterdam, vol. 17, pp. 97–110.

Boersma, P., 1997. How we learn variation, optionality, and probability. In: Proc. Institute of Phonetic Sciences, University of Amsterdam, vol. 21, pp. 43–58.

Boersma, P., Weenink, D., 1992–2010. Praat: doing phonetics by computer [Computer program]. Version 5.1.30, retrieved 1 April 2010 from <http://www.praat.org/>.

Eimas, P.D., 1963. The relation between identification and discrimination along speech and non-speech continua. Language and Speech 6 (4), 206–217.

Fisher, R.A., 1922. On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. A 222, 309–368.

Gerrits, E., Schouten, M.E.H., 2004. Categorical perception depends on the discrimination task. Perception Psychophys. 66 (3), 363–376.

Kewley-Port, D., 1995. Thresholds for formant frequency discrimination of vowels in consonantal context. J. Acoust. Soc. Am. 97 (5), 3139–3146.

Klatt, D.H., Klatt, L.C., 1990. Analysis, synthesis and perception of voice quality variations among male and female talkers. J. Acoust. Soc. Am. 87 (2), 820–856.

Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C., 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. Human Perception Perform. 54 (3), 358–368.

Mermelstein, P., 1978. Difference limens for formant frequencies of steady-state and consonant-bound formants. J. Acoust. Soc. Am. 63 (2), 572–580.

Nearey, T.M., 1990. The segment as a unit of speech perception. J. Phon. 18, 347–373.

Pisoni, D.B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Percept. Psychophys. 13 (2), 253–260.

Pisoni, D.B., 1975. Auditory short-term memory and vowel perception. Memory Cognition 3 (1), 7–18.

Pitt, M.A., Myung, I.J., Zhang, S., 2002. Toward a method of selecting among computational models of cognition. Psychol. Rev. 109 (3), 472–491.

Polka, L., Bohn, O.-S., 2003. Asymmetries in vowel perception. Speech Commun. 41, 221–231.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1992. Numerical Recipes in C. Cambridge University Press.

Rogers, J.C., Davis, M.H., 2009. Categorical perception of speech without stimulus repetition. In: Proc. Interspeech, Brighton, pp. 376–379.

Wilks, S.S., 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. Ann. Math. Stat. 9 (1), 60–62.