ELSEVIER

Speech Communication 41 (2003) 71–80

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

# The end of categorical perception as we know it

Bert Schouten [a,*], Ellen Gerrits [a], Arjan van Hessen [b]

[a] *UiL-OTS, Utrecht University, Trans 10, 3512 JK Utrecht, The Netherlands*

[b] *University of Twente, Room 2039, P.O. Box 217, 7500 AE Enschede, The Netherlands*

**Abstract**

Comparing phoneme classification and discrimination (or "categorical perception") of a stimulus continuum has for a long time been regarded as a useful method for investigating the storage and retrieval of phoneme categories in long-term memory. The closeness of the relationship between the two tasks, i.e. the degree of categorical perception, depends on a number of factors, some of which are unknown or random. One very important factor, however, seems to be the degree of bias (in the signal-detection sense of the term) in the discrimination task. When the task is such (as it is in 2IFC, for example) that the listener has to rely heavily on an internal, subjective, criterion, discrimination can seem to be almost perfectly categorical, if the stimuli are natural enough. Presenting the same stimuli in a much less biasing task, however, leads to discrimination results that are completely unrelated to phoneme classification. Even the otherwise ubiquitous peak at the phoneme boundary has disappeared. The traditional categorical-perception experiment measures the bias inherent in the discrimination task; if we want to know how speech sounds are categorized, we will have to look elsewhere.



*Keywords:* Categorical perception

## 1. Introduction

There can be hardly any doubt that, as listeners, we spend a great deal of our waking time perceiving speech categories. The number of these categories (let us call them "phonemes") is very limited, but the acoustic variation within each category as it occurs in the speech signal is very large. Since this variation does not seem to hinder normal speech communication, we must have developed (or inherited) a mechanism that manages to extract the phonemes the speaker intended from the highly variable acoustic signal. This mechanism deals with the acoustic variation so efficiently and so quickly that we are usually not aware of it. We may notice differences in pitch, loudness, and the prosodic features based on them; we may even detect a speaker's mood or accent, but these things seem to reside on different levels: they do not affect the "segmental" verbal message in any major way.

It would seem appropriate to call this mechanism, which is capable of extracting 10–15 phonemes per second from the messy speech signal, "categorical perception of speech" (there are probably many such categorization mechanisms for various perceptual domains in addition to speech). However, since the mechanism itself is not directly accessible to research, the term "categorical perception" has come to stand instead for the

* Corresponding author. Tel.: +31-30-2536069; fax: +31-30-2536000.

*E-mail address:* bert.schouten@let.uu.nl (B. Schouten).

0167-6393/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
doi:10.1016/S0167-6393(02)00094-8

experimental method generally used to investigate it. This method consists of creating a stimulus continuum between two phonemes and asking subjects (1) to classify the resulting stimuli by giving them phoneme labels (classification) and (2) to discriminate between the same stimuli. The extent to which discrimination performance is predictable from classification, i.e. the strength of the relationship between the two tasks, is what is usually referred to as categorical perception. The aim of this paper is to find out whether the categorical-perception method has anything to do with the categorization mechanism it purports to investigate.

## 2. The experimental definition of categorical perception

First, however, we have to get a different hurdle out of the way: the fact that *almost any* relationship between classification and discrimination of a stimulus series has, at one time or another, been regarded as categorical perception. The term has been used loosely ever since Liberman et al. (1957) showed that discrimination of synthetic stop-consonant stimuli was influenced by the categories to which the same stimuli were assigned by the subject: if two stimuli were heard as belonging to different (phoneme) categories, they were relatively easy to discriminate (leading to a "peak" in the discrimination function), whereas if they were heard as belonging to the same category, they were more difficult to discriminate (a "trough" in the discrimination function). The results obtained by Liberman et al. (1957), however, were not in agreement with their own explicit definition of categorical perception, given in the same paper (and repeated by Studdert-Kennedy et al., 1970), which requires that *discrimination should be completely determined by categorization*: there were considerable differences between classification ("predicted" discrimination) and "obtained" discrimination. Note that this definition of categorical perception makes any mention of "phoneme boundaries" or "discrimination peaks" entirely superfluous: even without a phoneme boundary, stimuli can be perceived categorically, for example, if all stimuli are consistently assigned to a single category and discrimination is at chance levels.

Liberman et al. (1957) did not claim that their results met their own definition of categorical perception. Still, since that time, anything resembling a discrimination peak anywhere near a phoneme boundary has generally been enough to claim categorical perception of the stimuli concerned, although the results were rarely, if ever, closer to the definition than the Liberman et al. results. Despite an auspicious beginning with a clear experimental definition, therefore, categorical perception has in practice remained an ill-defined or even undefined concept, which could be used to underpin a variety of sometimes mutually exclusive claims, for example for or against the motor theory (Lane, 1965; Studdert-Kennedy et al., 1970) and for or against the more general idea of a special speech mode of perception (Schouten, 1980). Explicit criteria were rarely used, although there have been exceptions. For example, Cutting (1982) implemented the definition of categorical perception by using two statistical criteria: a significant correlation between "predicted" and "obtained" discrimination, and no significant difference between the means of the two functions. Between them, these two criteria require that the two functions approximate each other: they follow parallel trajectories and they are very close together.

This state of affairs has made it easy for sceptics, such as Massaro (1987), to dismiss the concept entirely. Stimuli from different categories will always be more discriminable than stimuli from the same category, but there is no need to give a special name and a special status to this phenomenon, unless it can be shown that the connection between categorization and discrimination is unusually strong for a particular type of stimulus. If the connection was perhaps a little stronger for speech stimuli than for other stimuli, this was sometimes attributed to the poor quality of the synthetic speech stimuli—if quality is low, stimuli may be hard to discriminate, so subjects may have to fall back on category information, thus producing an artificial effect of categorical discrimination.

We would like to return to a strict enforcement of the original experimental definition of categorical perception as proposed by Liberman et al. (1957): perception is fully categorical only if there is no significant difference between phoneme categorization and discrimination. If this criterion is not met, categorical perception is incomplete.

A way of quantifying degree of categorical perception proposed by Van Hessen and Schouten (1999) is the "categorical-perception index", calculated in one of the two following ways, depending on whether correct-response percentages or $d'$-measures are used:

$$\mathrm{CP} = \frac{r}{1 + 2[p(\mathrm{obt}) - p(\mathrm{pred})]} \times 100, \text{ or:}$$

$$\mathrm{CP} = \frac{r}{1 + 0.2[d'(\mathrm{class}) - d'(\mathrm{discr})]} \times 100,$$

in which CP is the degree of categorical perception, ranging from 0 to 100, $r$ is the coefficient of correlation between the classification function and a discrimination function, and the denominator contains a term determined by the averaged differences between the data points of the classification and discrimination functions, multiplied by a constant chosen in such a way that the full range can be used. These equations express degree of categorical perception as a function of the resemblance (numerator) and proximity (denominator) of the two functions.

In this section we have been talking about categorical perception as an experimental phenomenon, as the relationship between performance on two different experimental tasks. We should now remind ourselves of the main aim of this paper. We start from the assumption that categorization takes place continually during normal speech communication; the question is whether experiments involving classification and discrimination of a stimulus continuum tell us anything about the mechanism that makes this possible. If it can be shown that discrimination of a particular speech sound continuum is consistently determined to a significant extent by classification of the same stimuli, the answer will be yes. However, if discrimination performance is determined largely by experimental factors that have nothing to do with phoneme classification, the answer will have to be no.

## 3. Stimulus naturalness

For the first time in the history of research into the relationship between classification and discrimination, Schouten and Van Hessen (1992) obtained nearly perfect categorical perception of a 15-stimulus continuum spanning three stop-consonant categories /p/ (stimulus 1), /t/ (stimulus 8), and /k/ (stimulus 15). The results are shown in Fig. 1, where the abscissa represents stimulus pairs (e.g. stimulus pair 2 represents the comparison of stimuli 1 and 3). The four functions in the figure stand for four different tasks: classification (or "predicted discrimination"; this task is here referred to as "identification") and two discrimination tasks: AX (two intervals, same/different) and 2IFC (two intervals, order judgment). AX was presented only in a "fixed" context, i.e in blocks corresponding to one stimulus pair at a time, whereas 2IFC occurred both in a fixed and in a roving context (random presentation of stimulus pairs). All discrimination tasks involved correct-response feedback after each trial; the interval between stimuli in a trial was 300 ms.

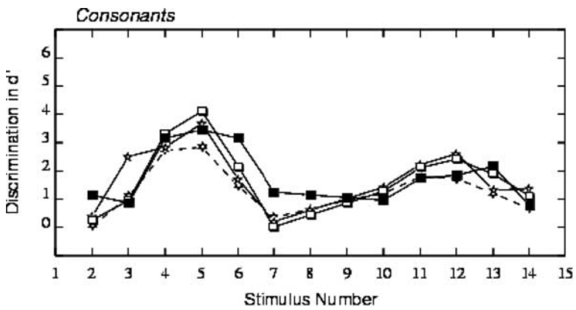


Fig. 1. Stop-consonant results from Schouten and Van Hessen (1992). Phoneme classification: star-of-David, fixed AX: (■), fixed 2IFC: (✡), roving 2IFC: (□). Classification distance expressed in $d'$ was calculated from the proportions of the *same* response (e.g. "/t/") given to two *different* stimuli (e.g. stimuli 1 and 2): one of these proportions was defined as "hits", the other as "false alarms". Stimulus 1 is a /p/, stimulus 8 is a /t/, and stimulus 15 is a /k/. The other stimuli were made by spectral interpolation between these three 'prototypes'.

The peaks in the four functions in Fig. 1 are generally assumed to involve stimulus pairs spanning a category boundary. Although an analysis of variance revealed that the factor tasks had a small effect, there were no significant differences between any two tasks. A very similar experiment involving a vowel continuum from /ɪ/ (stim. 1) to /ɛ/, (stim. 8), and /ɑ/ (stim. 15) revealed a different pattern (Fig. 2): although the shapes of the four functions were very similar, the functions themselves lie far enough apart for every task to differ significantly from every other task. According to our definition of categorical perception, the consonant continuum is perceived much more categorically (almost perfectly so) than the vowel continuum.

Schouten and Van Hessen (1992) attributed this unprecedentedly high degree of categorical perception to the nature of their stimuli, which had not been made in the usual manner by varying one or two crucial parameters in equal steps between the endpoints, but by varying the full spectrum in equal steps over a number of time windows. This made it difficult, if not downright impossible, for subjects to listen analytically to the varying parameter; moreover, since the full spectrum was preserved, the stimuli sounded quite human.

It would seem, then, that stimulus naturalness could be a factor determining degree of categorical perception. This was at least partly confirmed in an experiment carried out by Van Hessen and Schouten (1999), in which four different synthesizers were used to make four different vowel and consonant continua of strongly varying quality. Figs. 3 and 4 show the results for the two most

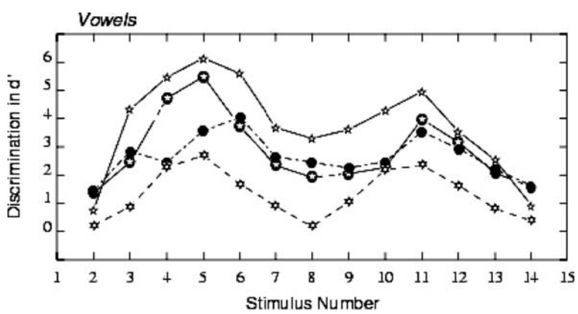


Fig. 2. Vowel results from Schouten and Van Hessen (1992). Phoneme classification: star-of-David, fixed AX: (●), fixed 2IFC: (✡), roving 2IFC: (○). Classification distance expressed in $d'$ was calculated from the proportions of the *same* response (e.g. "/t/") given to two *different* stimuli (e.g. stimuli 1 and 2): one of these proportions was defined as "hits", the other as "false alarms". Stimulus 1 is an /ɪ/, stimulus 8 is a /ɛ/, and stimulus 15 is a /ɑ/. The other stimuli were made by spectral interpolation between these three 'prototypes'.

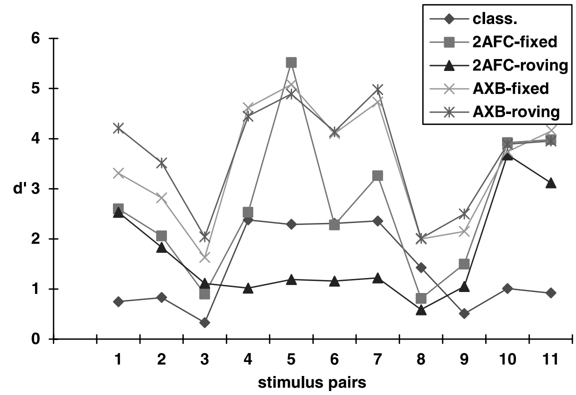


Fig. 3. Perceptual distance ($d'$) between the consonant stimuli obtained by means of linear predictive coding with 18 parameters (LPC18) of the stop consonants /p/, /t/, and /k/. There were 13 stimuli; the numbers $n$ along the abscissa indicate comparisons between stimuli $n$ and $n+2$. The tasks used for these stimuli are: classification (◆), 2AFC fixed context (■), 2AFC roving context (▲), AXB fixed discrimination (×), and AXB roving discrimination (✳).

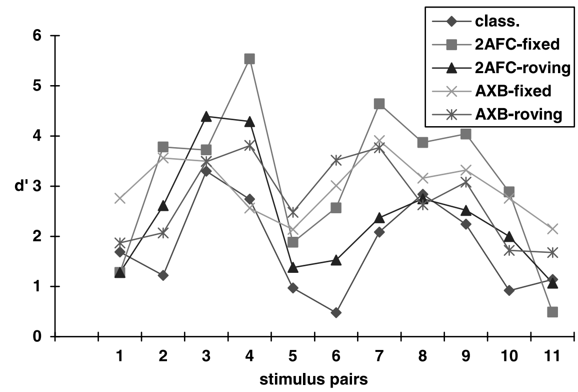


Fig. 4. Perceptual distance ($d'$) between the consonant stimuli obtained by means of SWG of the stop consonants /p/, /t/, and /k/. There were 13 stimuli; the numbers $n$ along the abscissa indicate comparisons between stimuli $n$ and $n+2$. The tasks used for these stimuli are: classification (◆), 2AFC fixed context (■), 2AFC roving context (▲), AXB fixed discrimination (×), and AXB roving discrimination (✳).

natural consonant continua, that obtained by means of LPC (linear predictive coding) using 18 parameters (Fig. 3) and that made in the same way as the continua of Figs. 1 and 2, i.e. by means of sinewave generation (SWG). There were 13 stimuli, ranging from /p/ (stim. 1) via /t/ (stim. 7) to /k/ (stim. 13). In Figs. 3 and 4, stimulus pair 1 along the abscissa means a comparison of stimuli 1 and 3. The tasks were classification, 2IFC (here referred to as 2AFC) and AXB (three intervals, which of A and B is identical to X?). All inter-stimulus intervals were again 300 ms.

Figs. 3 and 4 show fairly clearly that the more natural form of stimulus generation (SWG in Fig. 4) leads to a greater degree of categorical perception, although the effect is not as convincing as it was in Fig. 1. The classification function in Fig. 3 (diamonds) indicates that the average listener divided this relatively poor stimulus series into two categories, with a very gradual transition (spanning stimuli 4–9) between them; most of the discrimination functions follow this division. Fig. 4 indicates a division of the SWG series into three categories, but although the stimuli underlying Figs. 1 and 4 were highly similar, and the results in the two figures should therefore have been nearly identical, the degree of categorical perception in Fig. 4 is much lower. Obviously, stimulus quality is not the only determinant of categorical perception—the difference between Figs. 1 and 4 is probably due to the very different groups of subjects: 14 completely naive ones in Figs. 1 and 2, and eight very experienced students of phonetics in Figs. 3 and 4.

The full results from Van Hessen and Schouten (1999) are summarized in Table 1, which shows two additional types of synthesis: SbR (synthesis by rule) and LPC with six parameters. Table 1 shows the percentages of variance in the data explained by the two main factors: tasks and stimuli. Task variance should be inversely related to degree of categorical perception, indicated by means of the CP index in the table (see the introduction above). Stimulus variance may, as a consequence, be expected to increase with the CP index, provided it is the only other systematic source of variance.

Table 1
Percentages of variance explained by the task and the stimulus factor, plus categorical-perception index

| | Stops | | | Vowels | | |
|---|---|---|---|---|---|---|
| | Tasks | Stimuli | *CP-index* | Tasks | Stimuli | *CP-index* |
| *SbR* | 6.9 | 6.9 | 32 | 63.8 | 4.7 | 25 |
| *LPC*-6 | 27.9 | 18.1 | 46 | 29.3 | 16.0 | 35 |
| *LPC*-18 | 25.5 | 20.2 | 34 | 24.2 | 19.9 | 41 |
| SWG | 11.8 | 27.0 | 48 | 19.5 | 23.1 | 50 |

It is clear that stimulus quality is related to degree of categorical perception, but the relationship as not as strong as Fig. 1 from 1992 would seem to suggest. The subjects in Table 1 seem to perceive the vowels even slightly more categorically than the stop consonants.

If the relationship between classification and discrimination had turned out to be strongly dependent on stimulus quality, this would have provided support for the idea that categorization in everyday speech communication can be investigated by comparing classification and discrimination of stimulus continua: such a result would have meant that only natural, very speech-like stimuli are perceived categorically. However, the fact that this does not equally apply to every group of listeners undermines this conclusion. In addition, it should be borne in mind that the results and the conclusions to be drawn from them would have been rather different if we had opted for a different set of decision models. The $d'$ scores are not straight z(H)−z(FA) subtractions, but have been modified to take account of decision models for each of the different tasks. These models may contain false assumptions about the behaviour of some or all listeners.

## 4. Inter-stimulus interval

If discrimination of a speech continuum is determined by classification, the interval between the stimuli (ISI) to be discriminated should play no more than a minor role. This hypothesis was tested by Van Hessen and Schouten (1992), who used three different intervals: 100, 300, and 2000 ms.

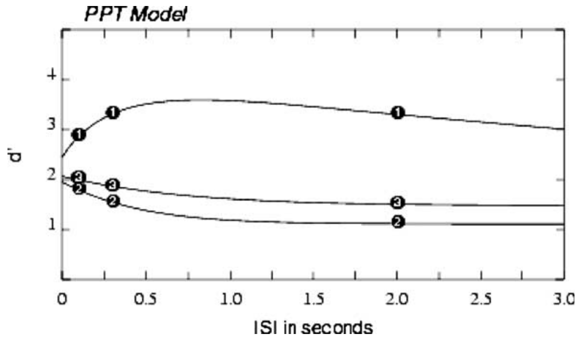


Fig. 5. Effect of inter-stimulus interval on discrimination. Lines fitted by the PPT model through three data points from the AX fixed discrimination task. 1 = between-category discrimination, 2 = within-category discrimination, 3 = overall mean.

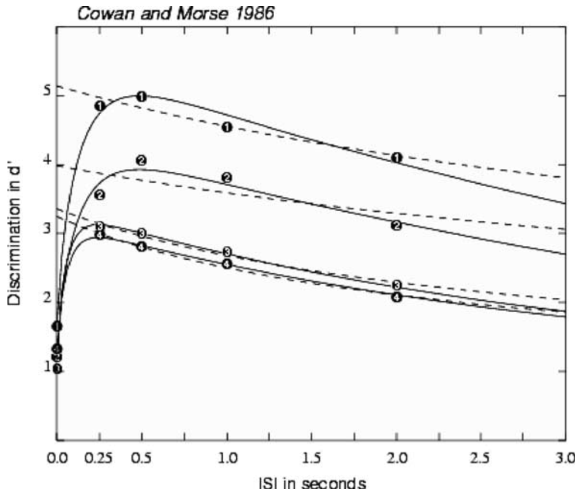


Fig. 6. Effect of inter-stimulus interval on discrimination. The AX vowel data from Cowan and Morse (1986), fitted by PPT (—) and by TCT (- - -). The data points marked "1" represent fixed between-category discrimination, those marked "2" represent roving between-category discrimination, "3" stands for fixed within-category discrimination, and "4" for roving within-category discrimination.

Their results for the fixed AX task involving stop consonants are shown in Fig. 5, where the abscissa represents ISI (only three values), and the ordinate average $d'$ over all stimulus pairs (data points labelled as 3), over the between-category pairs (1), or over the within-category pairs (2). Criterion for membership of the set of between-category pairs was a classification $d'$ of 2 or more. The lines fitted through the data points are from a model (PPT, or phoneme perception theory) that assumes time dependencies for the fading of auditory traces (within-category discrimination), and for the synthesis and subsequent fading of category labels (between-category discrimination).

Fig. 5 shows that time does not have a great effect on overall discrimination. When phoneme labels are relatively useless for discrimination (i.e. when both stimuli in a pair receive the same phoneme label), time has the expected negative effect; when phoneme labels are useful, however, it takes time to analyze the signal sufficiently for label determination, but this additional time does lead to much better discrimination around the phoneme boundaries.

The subjects in Fig. 5 were, on the whole, the same naive listeners used for Figs. 1 and 2, who showed almost perfectly categorical perception. It is therefore of great interest to examine other data, like those reported by Cowan and Morse (1986) and shown in Fig. 6. The pattern is not identical, but it is very similar. Cowan and Morse had used AX on vowel stimuli, in both fixed and roving contexts. The continuous lines in Fig. 6 are again the estimates of our PPT model; the dashed lines represent a much simpler model, which incorporates trace fading as its only time dependency.

These ISI data do not provide negative evidence for categorical perception defined as the relationship between classification and discrimination, but it should not be forgotten that the relevant data have only been gathered from a group of subjects who tended to use only phoneme labels during the discrimination tasks.

## 5. Response bias

All discrimination tasks discussed so far are subject to bias effects. Subjects may, for whatever reason, prefer one response type to another, or they may refer the stimuli to an internal criterion of their own (internal to a subject, but external to the stimuli) before making a decision about what response to give. In both cases, decisions about the stimuli are reached on a predisposition on the part of the subject: on a bias.

The most common form of response bias is a preference for one of the possible responses over the other one(s). This was investigated by Thomassen (1993), who took stop-consonant stimuli 1–8 from Schouten and Van Hessen (1992), i.e. half the range shown in Fig. 1: /p/ to /t/. He performed a classification and four discrimination experiments: fixed and roving ABX and 2IFC. The $d'$ results were quite similar to those in Fig. 1—perception was quite categorical. Here we are mainly interested in response bias, however: does it affect the two discrimination tasks differently?

The answer is given in Fig. 7. Since for all stimulus pairs all conditions occurred equally often (i.e. in ABX, X was as often identical to A as it was to B, and in 2IFC both stimulus orderings occurred equally often), an unbiased subject would have given both responses ("first" or "second" in both experiments) equally often. In Fig. 7 such unbiased behaviour would have resulted in a bias of 0%. Unbiased responding was only found among stimulus pairs straddling the phoneme boundary: pairs 3–5 and 4–6. Away from the boundary, however, where stimuli within a pair tend to get the same label, some bias was present. In ABX, this took the form of a majority responses of the type "second", i.e. "X = B". In 2IFC, there was a preference for "first" responses ("order /p/-/t/") at the /p/-end of the continuum, and a preference for "second" responses ("order /t/-/p/") at the /t/-end of the continuum.

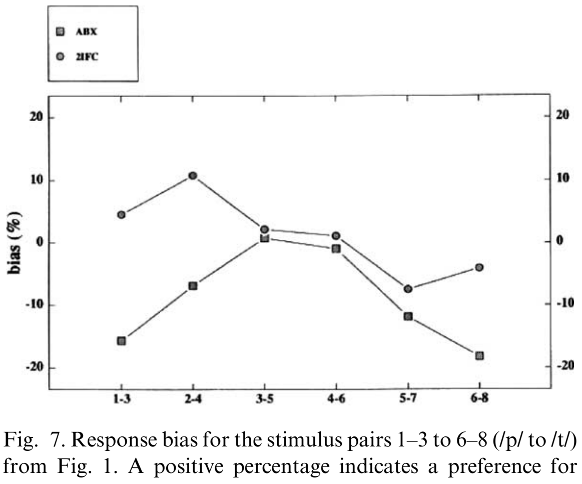


Fig. 7. Response bias for the stimulus pairs 1–3 to 6–8 (/p/ to /t/) from Fig. 1. A positive percentage indicates a preference for responses of the type "first". In ABX such a positive percentage would indicate a preference for "X = A" over "X = B"; in 2IFC it would mean a preference for "ordering /p/-/t/" over "ordering /t/-/p/".

There is a fairly strong task-induced response bias, then, away from the phoneme boundary. However, what is potentially the strongest response bias may be expected for stimuli on either side of the boundary, which is nothing but a personal criterion for response assignment. This bias is not explicit in Fig. 7, but it is very likely to be present; the question is how much of the response is determined by it. If it dominated a subject's response, stimuli would only be discriminated if they were separated by this subject's boundary between two phoneme categories. The danger of this happening is particularly great in 2IFC, in which subjects are usually explicitly instructed in terms of phoneme labels—and even when they are not, they quickly develop a strategy based on phoneme labels.

In the other task investigated by Thomassen (1993), ABX, the same risk is present, although it is probably smaller. The interval between A and B was again 300 ms, but the one between B and X was 500 ms. A and X were therefore separated by 800 ms plus the duration of B—too much for a direct auditory comparison, even without the intervention of B itself. It would not be surprising, therefore, to find subjects using a labelling strategy to discriminate stimuli in ABX, too.

This brings us to the heart of this paper: if the nature of the task compels subjects to use a labelling strategy, categorical perception will be pretty much a foregone conclusion. In order to make sure about this, we decided to carry out a series of discrimination experiments with a relatively "unbiased" task, i.e. a task that is not dominated by a subjective internal criterion. Such a task is 4IAX, where each trial consists of three identical stimuli and one different one; the subject's task is to indicate whether the different one occurred in the first or in the second pair (which are usually separated by a longer ISI). This task involves the comparison of two differences and deciding which is the greater one; subjective criteria are unlikely to play a dominant part. If a 4IAX experiment yields results that are predicted by classification of the same stimuli, we can be pretty certain of having encountered a genuine case of categorical perception: stimuli falling within the same category are auditorily indistinguishable, but stimuli falling into two different phoneme categories can be discriminated.

Gerrits and Schouten (2002) shied away from such a rigorously 'bias-free' task. They reasoned that subjects should at least have the option of using an internal phoneme-boundary criterion, especially if, as the authors expected, many subjects might not be able to discriminate stimuli from a continuum at all, without the benefit of phoneme labels. They therefore chose a different four-interval task: "2IFC with flanking stimuli", for which they used the slightly misleading name "four-interval oddity". In this task, three of the four intervals contain the same stimulus, but there is a different stimulus in either the second or the third interval. The subject has to indicate which of these two intervals contains the "oddball". This task can be performed in at least two ways: a subject could treat it as a 4IAX task, compare the differences within the two pairs, and respond "second" if the difference in the first pair exceeds that in the second pair, or "third" if the reverse is true. This would lead to a relatively bias-free response. Alternatively, an "ideal observer" would ignore the first and fourth intervals, the flankers, which are redundant, and treat the task simply as a case of 2IFC, with its inevitable bias. As it turned out, all 19 subjects used by Gerrits and Schouten underwent this task as a 4IAX-like task. This led to the results depicted in Fig. 8: phoneme classification has no relationship at all with discrimination. Without the benefit of phoneme labelling, these stimuli are very difficult to discriminate, but when subjects have to (or are allowed to) use their special categorization mechanism, as they are in the classification task, stimuli straddling the phoneme boundary are assigned to different categories fairly consistently. This applies to all subjects, the poor discriminators, whose scores are shown in Fig. 9, as well as the good discriminators, whose scores are shown in Fig. 10. Everyone is capable of using phoneme categorization to distinguish among stimuli along this continuum, but discrimination ranges from very poor to fairly good, without having any relation at all with phoneme classification.

The continuum in Figs. 8–10 was a vowel one, going from /u/ to /i/; inter-stimulus interval was 200 ms throughout, and the 19 naive subjects had to discriminate adjacent stimuli from the continuum. Subsequent tests with ISI's of 500 ms, greater physical differences between vowel stimuli to be discriminated, and with consonants rather than vowels, did not materially change this picture.

One change did make a difference, however: a change of task, from 2IFC *with* flankers to 2IFC *without* flankers, i.e. from a four-interval, poten-

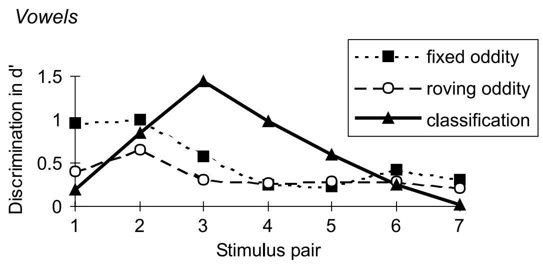


Fig. 8. Discrimination and classification from Gerrits and Schouten (2002). Stimulus 1 is the vowel /u/, stimulus the vowel /i/.

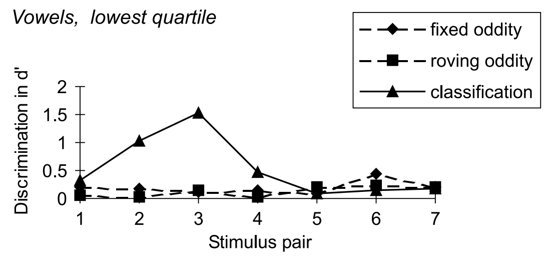


Fig. 9. Discrimination and classification results: lowest quartile. Stimulus 1 is the vowel /u/, stimulus the vowel /i/.

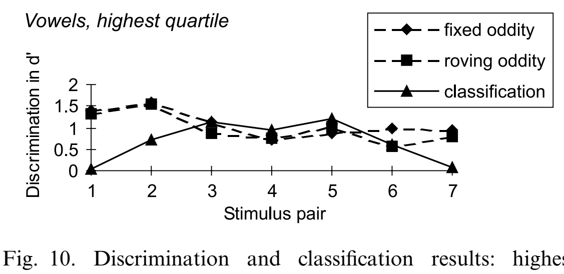


Fig. 10. Discrimination and classification results: highest quartile. Stimulus 1 is the vowel /u/, stimulus the vowel /i/.

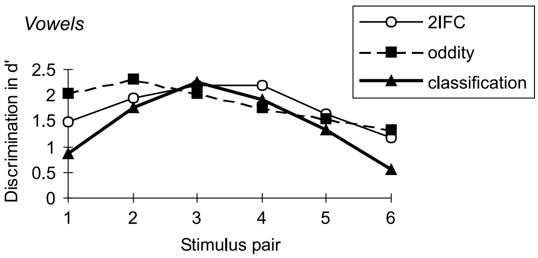


Fig. 11. "Oddity" and 2IFC discrimination results with a step size of 2. Stimulus 1 is the vowel /u/, stimulus the vowel /i/.

tially bias-free, task to a strongly biased one. The result of this change can be seen in Fig. 11, where the vowel continuum was presented to subjects in 2IFC with and without flankers, and where, moreover, step size between stimuli to be discriminated had been increased. The greater physical distance between the stimuli led to higher scores, but did not improve the relationship between classification and four-interval "oddity", as it is called here. However, 2IFC (without flanking stimuli) yielded a different picture: a clear positive relationship with classification.

## 6. Discussion

The question to be answered in this paper was how much the traditional relationship between phoneme classification and discrimination, which has, for a long time, been regarded as the hallmark of categorical perception, tells us about the process of phoneme categorization, which we have assumed to operate during everyday speech communication. The omens were unfavourable at the start of our investigations: despite the many claims to the contrary, a close relationship between classification and discrimination had seldom been found, so it seemed that discrimination did not have a lot to do with classification, although it used the outcome of the classification process to improve discrimination at the phoneme boundaries.

Things appeared to change when natural stimuli began to be used (Section 2). Suddenly, there was a very strong relationship between classification and discrimination, both in AX and in 2IFC, at least for stop consonants. For vowels, the relationship was also strong, but discrimination was rather better than had been predicted from classification. These results seemed to confirm an old claim: stops are perceived categorically, vowels are not.

It turned out to be very difficult, however, to replicate these results. An attempt to confirm the role of stimulus quality in the degree of categorical perception (Section 3) was successful in itself, but failed to reproduce the perfectly categorical perception that had been predicted for the most natural stimuli, those with the highest quality. A change from naive subjects to experienced students of phonetics seemed to be enough to bring this about. Inter-stimulus interval (Section 4) did not seem to affect the degree of categorical perception very much.

Response bias (Section 5) proved to be a very different story, however. Preference for one response type over another did not affect degree of categorical perception to any great extent, but the use of subjective criteria certainly did. A potentially bias-free discrimination task (2IFC with flankers) destroyed any relationship there might have been between phoneme classification and discrimination—a relationship that was quite strong in a 2IFC-experiment with the same stimuli and the same subjects.

Degree of categorical perception depends on many factors—more than we have investigated. The most crucial factor, however, is the discrimination task. A task that compels listeners to refer the stimuli to a subjective criterion is bound to produce results that indicate some degree of categorical perception; a task that is free of bias will produce no categorical perception at all. The conclusion must be that the relationship between classification and discrimination tells us very little about the process of phoneme classification as it occurs in everyday speech communication. Nevertheless, the experiments reviewed in this paper, especially Figs. 8–11, reveal one important aspect of this process: it enables listeners to distinguish fairly consistently between speech stimuli that they find it very hard, if not impossible, to discriminate. The phoneme categorization mechanism, when brought into play, enhances discrimination at the phoneme boundary. It is apparently not available in a relatively bias-free discrimination experiment.

## References

Cowan, N., Morse, P.A., 1986. The use of auditory and phonetic memory in vowel discrimination. J. Acoust. Soc. Amer. 79, 500–507.

Cutting, J.E., 1982. Plucks and bows are categorically perceived, sometimes. Percept. Psychophys. 31, 462–476.

Gerrits, E., Schouten, M.E.H., 2002. Categorical perception only occurs in the speech mode, submitted for publication.

Lane, H., 1965. The motor theory of speech perception, a critical review. Psychol. Rev. 72, 275–309.

Liberman, A.M., Harris, K., Hoffmann, H.S., Griffith, B., 1957. The discrimination of speech sounds within and across phoneme boundaries. J. Exp. Psychol. 54, 358–368.

Massaro, D.W., 1987. Categorical partition: a fuzzy-logical model of categorization behavior. In: Harnad, S. (Ed.), Categorical Perception. Harvard University Press, Cambridge, MA, pp. 254–283.

Schouten, M.E.H., 1980. The case against a speech mode of perception. Acta Psychol. 44, 71–98.

Schouten, M.E.H., Van Hessen, A.J., 1992. Modelling phoneme perception, I: Categorical perception. J. Acoust. Soc. Amer. 92, 1841–1855.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S., Cooper, F.S., 1970. The motor theory of speech perception: a reply to Lane's critical review. Psychol. Rev. 77, 234–249.

Thomassen, K., 1993. Categoriale perceptie bij 2IFC en ABX MA-thesis, Utrecht University.

Van Hessen, A.J., Schouten, M.E.H., 1992. Modelling phoneme perception. II: A model of stop consonant discrimination. J. Acoust. Soc. Amer. 92, 1856–1868.

Van Hessen, A.J., Schouten, M.E.H., 1999. Categorical perception as a function of stimulus quality. Phonetica 56, 56–72.