GRADIENT AND CATEGORICAL PITCH DIMENSIONS IN DUTCH: DIAGNOSTIC TEST

Bert Remijsen and Vincent J. van Heuven
Phonetics Laboratory, Leiden University and
Holland Institute of Generative Linguistics, The Netherlands

ABSTRACT

It is not clear, a priori, that the categorical perception (CP) paradigm is sufficiently powerful to uncover CP in intonational contrasts even in cases where these exist beyond doubt. Until recently CP has only been applied to segmental contrasts. The basic applicability of CP to (in)ton(ation)al contrasts, therefore, remains to be shown. We applied the CP-method to the contrast between the high (H%) and low (L%) boundary tone in Dutch, an unchallenged categorical-linguistic distinction. Our results clearly show CP for L% vs. H%; for individual subjects the location of the category boundary correlates with a peak in discriminatory precision. We conclude that tonal dimensions can be perceived categorically, and that the CP-technique is sensitive enough to determine categoriality in tonal dimensions.

1. INTRODUCTION

1.1. Categorical nature of intonational contrasts

By intonation or speech melody we mean the pattern of rises and falls in the course of the pitch of spoken sentences. Melodic patterns in speech vary systematically across languages, and even within languages across dialects. The crosslinguistic differences can be parametrized and described in much the same way as has been done for the segmentals in language: a set of distinctive features defines an inventory of abstract units, which can be organised in higher-order units subject to wellformedness constraints. Moreover, intonational contrasts are used to perform grammatical functions that can also be expressed by lexico-syntactic means, such as turning statements into questions or into commands, and putting constituents in focus. For these reasons it has become widely accepted that intonation is part of the linguistic system [1: 8]. Yet, there have always been adherents of the view that speech melody should be considered as something outside the realm of linguistics proper, i.e., that intonation is a paralinguistic phenomenon at best, to be treated on a par with the expression of attitudes or emotions. Typically, the communication of emotions (e.g., anger, fear, joy, surprise) or of attitudes (e.g., sarcasm) is non-categorical: the speaker shows himself more or less angry, fearful, or sarcastic in a continuously gradient fashion.

A relatively recent insight is that a division should be made in melodic phenomena occurring in speech between linguistic versus paralinguistic contrasts. Obviously, only the former but not the latter type of phenomena should be described by the grammar and explained by linguistic theory. This, however, begs the question how the difference can be made between linguistic and paralinguistic phenomena within the realm of speech melody. A methodology to decide on the linguistic status a a melodic phenomenon, i.e., to decide whether the phenomenon should or should not be treated in the grammar, has only recently begun to develop. This situation challenges us to try to develop a suitable heuristic for establishing intonational categories.

Consequently, the objective of the present study is to gain insight into the the psycholinguistic nature of intonational distinctions and to apply this knowledge to the development of formal procedures to establish the linguistic nature of hypothesized categories in intonational systems. We will report an experiment that involves the use of the categorical perception paradigm to a contrast of boundary tones in Dutch. Assuming that these boundary tones constitute a undisputed binary categorical contrast in the language, we can determine whether the categorical perception paradigm is sensitive at all to serve as a diagnostic tool in settling the linguistic status of a melodic parameter.

1.2. Categorical perception of bondary tones

For segmentals, the categorical perception (CP) paradigm [2] can be used to falsify hypothesized phonemic distinctions. This paradigm involves two perceptual tasks: first, listeners classify utterances as exponents of one the two hypothesized categories. After that the listeners are subjected to a discrimination task: they are asked to discriminate stimuli that are either the same or slightly different. Language users are more sensitive to acoustic differences between categories than within categories. Perception is deemed to be categorical when the boundary between the two hypothesized categories in the classification responses coincides with a peak in discriminatory precision.

Ladd and Morton [3] were the first to apply this method to intonation; they investigated a contrast between two pitch-accents in English, viz. the regular high pitch-accent and the emphatic high, which involves a larger excursion size. This distinction can be analysed as a matter of paralinguistic variation within the single phonological category of accent [4]. Alternatively, the emphatic high is a linguistic category distinct from the normal high. Their results did indeed reveal S-shaped curves; however, the subsequent discrimination experiments failed to show the required peak in discriminatory precision at the category boundary.

It is unclear, however, what we are to expect when applying CP to intonation: categorical perception in its strongest form is characteristic for stop consonants only; there the peak in discriminatory precision can be predicted from the classification results [5]. When CP is applied to vowel contrasts, the improvement of discriminatory precision at cross-over is less clear, e.g. taking the shape of a distributed plateau rather than that of a local peak. Acoustically, intonational units are similar to vowel phonemes in that they are encoded over relatively long time intervals; as such they may be equally perceived in continuous rather than categorical manner. Consequently, when CP is applied to an intonational contrast, various correspondence patterns between classification and discrimination results may emerge: a peak or a plateau at the crossover signals categorical perception. In the case of [3], however, neither pattern was encountered, which leaves two possibilities: (i) the normal vs. emphatic distinction could be considered as paralinguistic, or (ii) it may be that the CP paradigm is inadequate for distinguishing intonational contrasts. Option (ii) is the position taken by [3] themselves, who maintain that the distinction should be considered as linguistic. In order to decide between the alternative interpretations, we should first find out whether the CP paradigm is applicable to intonation. We will do this by investigating whether the criterion for categorical perception – crossover point corresponding with a peak in discriminatory precision – is satisfied when the CP paradigm is applied to an intonational contrast that is linguistic beyond suspicion. If such a contrast is perceived categorically, then the normal vs. emphatic distinction should be interpreted as paralinguistic. If not, Ladd and Morton's [3] interpretation of this contrast may be relevant to intonational contrasts in general. In our experiment we will study the contrast between a low (L%) vs. a high (H%) boundary tone in Dutch. H% is the prototypical question marker [6,7]; L% marks the end of a statement.

2. METHOD

2.1 Stimuli

The utterance selected *De Dennenlaan* /də ˈdɛnəlaːn/ 'Pine Lane' is segmentally ambiguous between a statement and a question, the contrast being encoded by F_0. In both sentences, there is a high pitch-accent on /dɛn/, the lexically stressed first syllable of the noun. While in the statement F_0 remains low (L-L%), the question realisation features a late rise in the final syllable (L-H%):

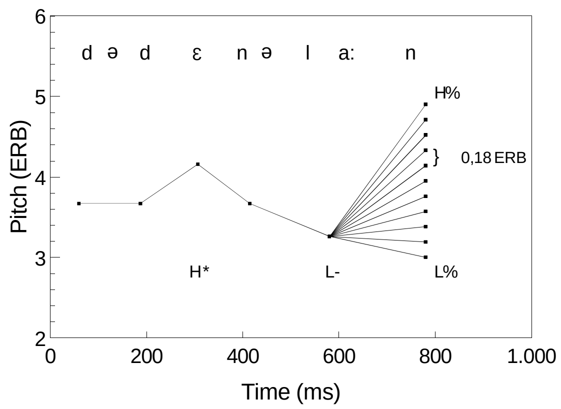

Figure 2. F_0 continuum (11 steps) between L-L% and L-H%.

The source sentence was realised by a state-of-the-art Dutch diphone text-to-speech system [8] with a male voice, and subsequent manipulation of F_0 in order to vary the pitch contour from a question realisation to a statement (figure 1). Though in natural speech intensity may co-vary with F_0, we kept intensity constant in order to isolate the perceptual effect of F_0, which we consider the primary correlate of intonation. The F_0 contours were generated using straight-line interpolation in the PSOLA resynthesis option provided by the Praat speech analysis and resynthesis software [9]. An 11-step continuum was created, at equidistant points along the ERB scale [10]. The stimuli were numbered 1 (clear statement) to 11 (clear question).

3.2 Tasks and experimental procedures

The tasks and the experimental procedures follow [3]: the classification task is a forced choice setup in which the listeners respond to a stimulus from the continuum by classifying it either as a question or a statement. There were ten repetitions of each of the 11 points from the continuum. These 110 stimuli were presented in random order.

For the discrimination task, we again followed [3] in using the AX discrimination paradigm. Stimuli were presented in pairs that are either the same or one step apart on the continuum. In the latter case, the second can be higher or lower than the first (hereafter resp. AB and BA). The ten AB stimulus types ran from pair {1,2} to {10,11}; the ten correponding BA types from {2,1} to {11,10}. There were five tokens of each pair. This yielded 200 discrimination stimuli (50 AB, 50 BA, 100 AA).

Seventeen native Dutch listeners, 10 males and 7 females, took part in the experiment on a voluntary basis. Participants were university students and staff from various departments at Leiden University. None of them reported any perceptual deficiencies. They were paid for their services.

The experiments were run individually or in small groups of subjects, who listened to the stimuli over headphones, while seated in a quiet lecture room. Subjects marked their responses on printed answer sheets provided to them, always taking the identification task first and the discrimination task later.

3. RESULTS

3.1. Classification task

The responses from three listeners were excluded as the perceptual threshold proved too high for them: they obtained less than 10 percent hits in the discrimination task. One more subject was excluded because he spontaneously reported more and other pragmatic meanings than 'statement' versus 'question'.

Figure 2 illustrates the perceptual cross-over between the two categories. The psychometric function is unmistakably S-shaped, going from less than 20 to more than 80 % 'statement' responses within three steps of the 11-point continuum. A probit analysis [11] suggests that exact cross-over point between 'question' and 'statement' lies at an (interpolated) terminal F_0-value of 122 Hz.

In general, the identification curves of individual subjects are very steep, crossing over in just one or two steps (figure 3). This precision is obscured in figure 2, where the data are averaged over subjects: clearly, there is substantial between-subject variability in the location of the cross-over; a stimulus of which all tokens are marked as questions by one subject may be consistently associated with declarative meaning by another subject. This phenomenon has considerable influence on the data analysis, since it predicts by necessity a high-discrimination plateau rather than a local discrimination peak.

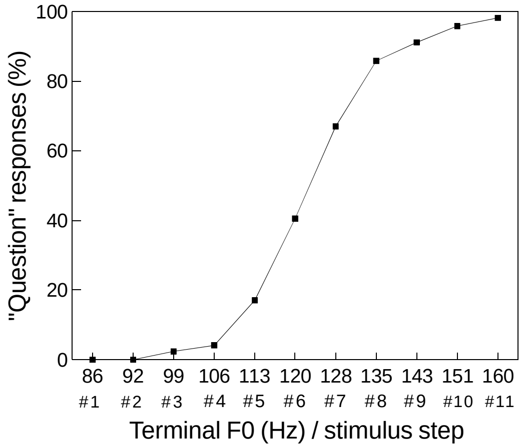

Figure 2. Percent ‘question’ responses as a function of stimulus step number. Data have been accumulated across all listeners.

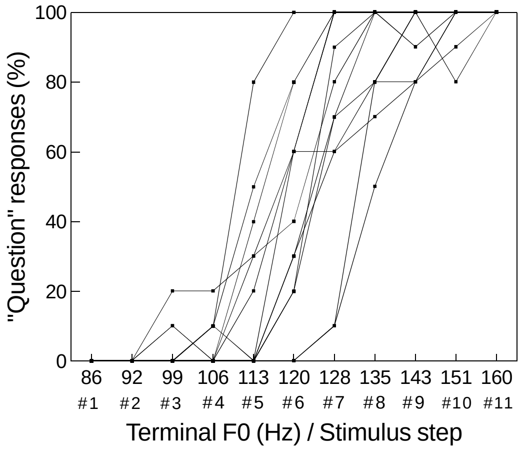

Figure 3. Percent ‘question’ responses as a function of stimulus step number. Data are presented for individual listeners.

3.2 Discrimination task

Figure 4 presents the mean percentage of succesfully discriminated different stimuli (hereafter ‘hits’), broken down by the two different orders (AB versus BA) as well as averaged over orders, and the the percentage of false alarms (‘different’ responses to – identical – AA stimuli). Various aspects of these results are discussed in the following paragraphs.

3.2.1. Unexpected peak in discriminatory precision. Both AB and BA hit curves reach a maximum at the low end of the stimulus range (figure 4). Yet, this phenomenon should not be interpreted as an artifact of ERB scaling. Had this been the case, the increase would have been gradual rather than sudden. Therefore, the explanation is to be sought in the domain of linguistics rather than that of psychophysics. Stimuli that are consistently marked as statements are discriminated relatively better. Possibly, the continuum we presented actually reflects three linguistic categories instead of two. The evidence is inconclusive as the continuum ends within the maximum. Moreover, we would need other category labels in the classification task in order to test this distinction by means of the CP paradigm. The hypothetical categorical distinction may crucially depend on the amount of final lowering.

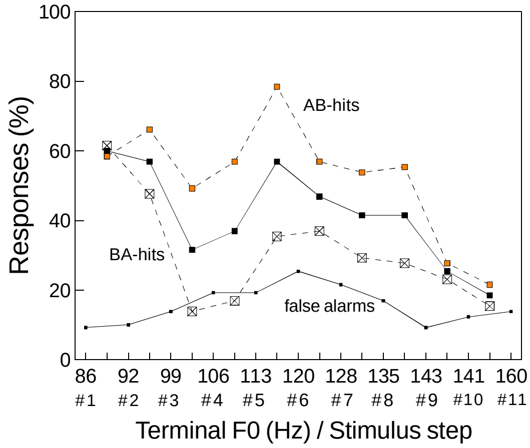

Figure 4. Percent hits (AB and BA-orders presented separately as well as averaged) and false alarms as a function of stimulus step number. (terminal F_0 of boundary tone).

3.2.2. Order of presentation effect. There is an order of presentation effect in the discrimination data (figure 5): two differrent contours are more succesfully discriminated when the second one has the higher terminal pitch (AB sequence). The same effect has been observed for pitch-accents by [3]. They found that the size of the effect depends on step size: when the difference between A and B is minimal, the hit rate for the BA stimuli was equal to the false alarm rate (i.e., chance performance). As step size increased, more BA stimuli could be distinguished, but never as many AB stimuli. In our data, the BA hit curve lies between the curve for false alarms and the one for AB hits; the phenomenon is probably related to declination [3: 331].

3.3.3. The category cross-over and the discrimination peak. Ignoring the peak at the low end, both the AB and BA curves show a second maximum in the middle of the continuum, between 113 and 128 Hz (figure 4). The BA curve reaches a plateau between 116 and 124 Hz, with the value at the 124 Hz data point slightly higher. The maximum of the AB is located at 116 Hz.

There are various ways to check whether these peaks correspond with the crossover between the categories in the classification task. One is to apply the so-called Haskins formula [12],

$P(c)=0.5 \times [1 + (p_1 - p_2)^2]$, resulting in discrimination curve as predicted by the classification results. This curve can then be compared with the actual discrimination responses, which comparison forms the basis for deciding whether the contrast can be interpreted as categorically perceived. The predicted discrimination curve features a plateau-like maximum, with a peak at 124 Hz. This corresponds to the BA maximum; also, the plateau extends one step to the left, where the peak of the AB curve is located. An alternative procedure is used by [3]: they compute the point by means of a probit analysis. In our case, this results in a value of 122 Hz, which is in between the AB and BA maxima.

When discussing the classification results, we found a large amount of between-subject variability. From this fact we may derive yet another heuristic for categorical perception of the boundary tone contrast under investigation; if the crossover varies between subjects, then the CP theory predicts the location of individual maxima in discriminatory precision to vary accordingly. Therefore, we performed a regression analysis with the individual cross-over points as predicted by probit analysis as the predictor variable and the individual peak in the discrimination results (AB and BA hits collapsed), i.e., the point on the continuum for which most hits were marked, ignoring the responses below 106 Hz, as the criterion. Figure 5 shows the correlation in a scatterplot. (For two subjects there was no clear maximum, as two non-adjacent points have the highest value; they were excluded. This reduced the dataset to 11 cases.) The correlation is significant ($r=0.68$, $p=0.02$). Therefore, while subjects vary in perceiving the location of the hypothesized category boundary (figure 2), the responses of the individual subjects feature a peak in discriminatory precision that significantly correlates with the location where they perceive the category boundary.

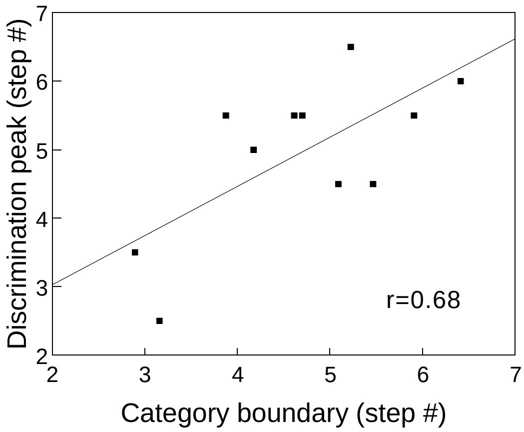

Figure 5. Location of peak in the individual discrimination function (averaged over AB and BA orders) against location of individual cross-over in identification task

4. CONCLUSION

We have accumulated evidence for the phonological status of the contrast between boundary tones in Dutch. In accordance with CP theory [2], discriminatory precision is better at the category boundary. The large amount of between-subject variation has lead us to diagnose this correspondence in an alternative way, namely by means of a regression analysis. If the large between-subject variation observed is a general characteristic for intonational categorical distinctions, this heuristic may prove fruitful for future research. For example, it may be that Ladd and Morton's [3] results can be interpreted in a different manner by taking into account between-subject variability. If not, we would conclude that their emphatic versus normal contrast is non-linguistic. Also, if there appears to be large variability in the location of the category boundary, this constitutes psycholinguistic support for the claim that intonational categories are to be represented in terms of continuous parameters.

Ladd and Morton [3] concluded that the normal versus emphatic contrast may be categorically interpreted while still being perceived continuously (§ 1.2). Our data are in disagreement with this hypothesis: our results provide a clear instance of categorical perception of an intonational contrast. Our data, therefore, show that the classical combination of identification and discrimination provides a feasible heuristic after all to decide whether a melodic contrast is linguistic (i.e., categorical, quantal) or paralinguistic (i.e., continuous, scalar).

ACKNOWLEDGEMENTS

We thank the following people, whose comments and advice have been very helpful: Johanneke Caspers, Judith Haan, Bob Ladd and Bert Schouten. This research was funded by the Royal Netherlands Academy of Arts and Sciences (KNAW) under project # 95-CS-05 (principal investigators W.A.L. Stokhof and V.J. van Heuven).

REFERENCES

[1] Ladd, D.R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

[2] Liberman, A. M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psych.*, 61, 379-388.

[3] Ladd, D.R. and Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *J. Phon.* 25, 313-342.

[4] Ladd, D.R. and Terken, J. (1995). Modelling intra- and inter-speaker pitch range variation. *Proc. 13th ICPhSc*, Stockholm, 386-389.

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

[6] Hart, J. 't, Collier, R. and Cohen, A. (1990) *A perceptual study of intonation. An experimental-phonetic approach to speech perception*. Cambridge University Press, Cambridge.

[7] Gooskens, Ch. and Heuven, V.J. van (1995). Declination in Dutch and Danish, *Proc. 13th ICPhSc*, Stockholm, 374-377.

[8] Fluency Speech Technology (1997). Fluent Dutch Text-To-Speech. Professional version 1.0. Utrecht.

[9] Boersma, P. and Weenink, D. (1996). PRAAT: A system for doing phonetics by computer. Report of the Institute of Phonetic Sciences of the University of Amsterdam 132. [http:// fonsg3.let.uva.nl/praat].

[10] Hermes, D.J. and Gestel, J.C. van (1991). The frequency scale of speech intonation. *J. Acoust. Soc. Am.*, 90, 97-102.

[11] Aldrich, J.H. and Nelson, F.D (1984). *Linear probablility, logit, and probit models*. Sage Publications: Quantitative Applications in the Social Sciences Vol. 45.

[12] Cutting, J.E. (1982). Plucks and bows are categorically perceived, sometimes, *Perc. Psychoph.*, 31, 462-476.