Perceptual dimension of openness in vowels

Hartmut Traunmüller

Institutionen för Lingvistik, Stockholms Universitet, S–106 91 Stockholm, Sweden

(Received 10 March 1980; accepted for publication 10 November 1980)

The role of intrinsic factors determining perceived degree of vowel openness was examined. In order to determine the role of F1 and F0, one-formant vowels, covering a wide range of fundamental and formant frequencies, were identified by 23 subjects who were native speakers of a Bavarian dialect in which five degrees of openness occur distinctively. The significance of intrinsic factors other than F1 and F0 was also studied using synthetic versions of natural vowels with F1 and/or F0 systematically displaced in frequency. It was found that, generally, the tonality distance between F1 and F0 is decisive for openness, while the higher formants contribute marginally. It was further found that the distance between widely spaced formants, as between F2 and F1 in front vowels, is not crucial for vowel identification. The results are evaluated in terms of a psychoacoustic model of identification by pattern matching. The model incorporates two basic assumptions. First, a certain pattern of excitation along the basilar membrane is recognized as a given feature regardless of position along the membrane. Second, there is an integration band with a width of 3 Bark effective in spectrum envelope recognition.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

Most of the distinctive features of speech sounds apparently constitute dimensions along which only a binary distinction is used. Along the dimension of "openness" in vowels, however, more than two distinctions occur in many languages. In the present investigation, we attempt to examine how this dimension is perceived. The investigation is limited to the role of intrinsic factors, i.e., to features that are acoustically present within the speech segment in question, as opposed to extrinsic contextual factors (cf. Ainsworth, 1975).

"Openness" is an articulatory term, referring to the degree of mouth opening or mandible depression. Its use here to describe a perceptual dimension is justified by a close correlation between that dimension and mandible depression (Kim and Fujisaki, 1974). This correlation is not compulsory, however, since subjects can produce, apparently without difficulty, a closed vowel with the mandible fixed in a position for a very open one (Lindblom and Sundberg, 1971; Lindblom et al., 1979). Instead of "openness" some phoneticians use the notion of vowel height, referring to the highest point of the tongue, but this feature, if taken literally, does not admit some important phonological generalizations, since the steps in height of that point are not equal between corresponding front and back vowels (Ladefoged, 1987).

Anyway, phonemes have to be perceived in order to be classified or distinguished, and if something is perceived as an /a/, it is an /a/, regardless of its production. Only the features perceptible to the listener can play a distinctive role in the process of speech communication; articulatory, or even proprioceptive features cannot. Distinctive features, then, should be described in perceptual terms. A description in acoustic terms is possible, but complicated by the transformations occurring in the organ of hearing.

An acoustic feature correlated with openness is the frequency of the first formant (F1). However, because of the overlapping of formant frequency data of different vowel phonemes produced by men, women, and children (Peterson and Barney, 1952), it is evident that perceived openness is not solely related to F1. It has been shown that the frequency of the fundamental (F0) also plays a certain role (Miller, 1953; Fujisaki and Kawashima, 1968; Ainsworth, 1975).

Potter and Steinberg (1950, p. 812) first formulated the hypothesis "that within limits, a certain spatial pattern of stimulation along the basilar membrane may be identified as a given sound regardless of position along the membrane." This is suggested by the displacement of the formant frequency pattern as observed in a vowel spoken by a man, a woman, and a child. This displacement is by and large uniform on a scale of auditory critical bands, i.e., a scale representing critical bands with unit width (1 Bark), which also may be considered as a tonality scale. Besides its proportionality with perceived tonality this scale is, in good approximation, proportional to distances along the basilar membrane (1 Bark = 1.5 mm = 150 haircells, according to Zwicker and Feldtkeller, 1967). There are, however, some deviations in vowels spoken by men, as compared to those spoken by women, which do not directly fit this hypothesis. The female–male differences in tonality are shown in Table I for a sample of vowels. The values in Table I are based on a summary of investigations in several languages (Fant, 1975). Similar differences were found in the various languages, but the spread was quite large. Therefore deviations less than 0.5 Bark in Table I should be disregarded. Values for F0 and F4 were not published in Fant (1975). The corresponding differences are roughly 1 Bark, equal to 100 Hz for F0 and at 18% of the frequency of F4.

In a previous perceptual study, it was found that the tonality distance between the first formant and the fundamental, rather than the position of the first formant alone, is decisive for the identification of one-formant vowels, as long as this distance remains within the range found in natural vowels (Traunmüller, 1978).
TABLE I. Mean female–male tonality differences in vowels of several languages. Values in Bark calculated from Fant (1975). "*" means that the difference is estimated to be roughly 1.0 Bark.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>$F_0$</th>
<th>$F_1$</th>
<th>$F_2$</th>
<th>$F_3$</th>
<th>$F_4$</th>
<th>Number of languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>0.2</td>
<td>1.3</td>
<td>0.8</td>
<td>*</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$e, i$</td>
<td>0.4</td>
<td>1.4</td>
<td>1.1</td>
<td>*</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>$a, u$</td>
<td>0.8</td>
<td>0.7</td>
<td>0.9</td>
<td>*</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>$o$</td>
<td>0.5</td>
<td>0.3</td>
<td>1.1</td>
<td>*</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$u$</td>
<td>0.2</td>
<td>0.0</td>
<td>1.4</td>
<td>*</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>$y$</td>
<td>0.0</td>
<td>1.2</td>
<td>1.1</td>
<td>*</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.2</td>
<td>1.0</td>
<td>1.0</td>
<td>*</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

This conclusion was, however, based on only two listeners, while the results obtained from two other listeners were not sufficiently clear.

The aim of the first of the present experiments is to establish more reliably the decisive criteria for the identification of one-formant vowels. The experiment was planned to cover the whole range of variation in $F_0$ (50 to 700 Hz) and in $F_1$ (150 to 1480 Hz) that can be observed in natural speech.

It is an advantage to use listeners whose native dialect employs many distinctive degrees of openness without concomitant distinctions such as tense/lax or diphthongization. Such subjects are compelled by their language to make careful distinctions along the dimension of openness. Moreover, interference from secondary features will be minimal in their responses. In the eastern Central-Bavarian dialect, used by the subjects in this experiment, as spoken in lower Austria (Weigl, 1924/25), and by the older generation in Vienna (Kranzmayer, 1953; Koekkoek, 1955), there appear to be five distinctive degrees of openness.

Table II shows the vowels occurring in stressed positions in the dialect of the subjects. All the vowels occur in the sequences 'V:C and 'VC: with durations as described by Bannert (1976) for western Central-Bavarian. There is no perceptible difference in timbre associated with a difference in duration, although formant frequencies are slightly different (cf. Bannert, 1976). In syllables that cannot be stressed, there are a few lax (centralized) vowels and syllabic nasals and laterals. Table III shows the feature analysis of the non-nasal monophthongs in terms of which the experimental results will be presented. Identification responses of vowels sharing the same feature will be collapsed.

Experiments 2, 3, and 4 are designed to find out whether the conclusions drawn from the identifications of one-formant vowels can be generalized to the perception of natural vowels. These experiments are intended to determine the possible role of the second and higher formants, for the perception of openness. To this end, $F_1$ and $F_0$ were systematically displaced in frequency in synthetic versions of some natural vowels in three ways: First, $F_1$ and $F_0$ simultaneously, keeping their tonality distance constant, second, $F_1$ only, and third, $F_0$ only. All other characteristics were kept at their original positions in frequency. The particular natural vowels were chosen so that some information about the perceptual distinction between front rounded and unrounded vowels could also be obtained.

I. METHOD

A. Subjects

12 male and 12 female secondary school pupils from Amstetten, Lower Austria, served as subjects. They were between 16 and 18 years of age and had grown up in the environment of the dialect spoken in the western region of Lower Austria. The subjects were divided into four groups of six participants each. All 24 subjects participated in each of the four experiments. The results obtained from one of them were excluded from evaluation in all experiments because of several response omissions.

B. Stimuli

The stimuli were generated by digital simulation of a terminal-analog speech synthesizer. Four repetitions of each stimulus were recorded on tape with a high-frequency limit of 8 kHz and with due precautions for high-quality reproduction. Stimulus durations and intervals between repetitions and between different stimulus quadruplets were as shown in Fig. 1. The fundamental frequency as a function of time was as shown in Fig. 2 for all stimuli. This contour and the durations were based on isolated vowels spoken by the author, who was familiar with an eastern Central-Bavarian dialect.

FIG. 1. Intervals between four repetitions of each stimulus and between different stimulus quadruplets in experiments 1 through 4.
The formant bandwidths $B(F)$, defined at $3\, \text{dB}$, obeyed the function $B(F) = 0.05 f(F) + 50\, \text{Hz}$, where $f(F)$ is the center frequency of the formant. The voice source spectrum had a 100 Hz, 6 dB/oct low-pass contour, obtained by low-pass filtering a train of pulses with a duration of 50 μs each.

C. Procedure

The stimuli were presented binaurally via headphones, type Sennheiser HD 414, from a calibrated tape recorder, type Uher 4000. Loudness was set at a comfortable listening level. The subjects were instructed to answer by encircling the symbol of the perceived vowel on answer sheets. Each of the 13 vowels was given as a permitted response to each stimulus in all experiments. The subjects were allowed to encircle several vowel symbols in cases of doubt. Processing the results, polysymbolic responses were then fractionally attributed to the different symbols. Before presenting the stimuli, roughly 10 min were used to instruct the subjects about the phonetic value of the symbols appearing on the answer sheets by telling them and questioning them about words containing the corresponding vowels. Care was taken to check that the subjects could use correctly the symbols for $\text{e}/[\text{e}]$ and for $\text{s}/[\text{e}]$. These distinctions lack consistent correspondences in standard High German and are mostly neglected in dialect writing. The symbols, their arrangement, and corresponding phonetic values were as shown in Table IV. The symbol "a" was taken from dialect writing, the other symbols reflect standard High German correspondences and letter names. To each stimulus quadruplet was assigned one frame on the answer sheets containing all the different vowel symbols. The subjects were told not to have any expectations concerning the frequency of appearance of different sounds and that the same vowel might be heard many times in sequence.

The stimuli were 13 series of one-formant vowels. Within each series, the tonality of the formant, here denoted $F'$, was increased in steps of 0.5 Bark, beginning with a position 0.5 Bark above the nominal tonality of $F_0$, and ending at 7.0 Bark above $F_0$. The tonality scale was taken from Zwicker and Feldtkeller (1964, p. 74). The range of $F'$ was restricted to 150 Hz (1.5 Bark) minimum and 1480 Hz (11.0 Bark) maximum. The fundamental frequency was subsequently increased by 0.5 Bark for each series, beginning with 50 Hz (0.5 Bark), and ending at 700 Hz (6.5 Bark). The number of different stimuli thus obtained was 166, located in 13 series with at most 14 different stimulus quadruplets. The sixth series, with $F_0 = 300\, \text{Hz}$, was presented twice, the first time being in the beginning of the experiment, before the series with $F_0 = 50\, \text{Hz}$. This first presentation served to allow the subjects to get accustomed to their task.

B. Results

The results, pooled over the 23 subjects and in classes of the vowels with the same degree of openness, are displayed in Fig. 3(a) for each series of stimuli with the same $F_0$. The initial presentation of the series with $F_0 = 300\, \text{Hz}$ is not shown. Figure 3(b) shows the corresponding results concerning perceived roundedness and "place of articulation" (front/back).

Figure 4 shows the dominantly perceived degrees of openness and regions of at least 50\% agreement among subjects for the whole range of $F_0$ and $F'$ used in this experiment, together with the tonality positions of the first four partials.

The results were reasonably uniform across subjects with respect to openness, and the pooling therefore seems justified. There were, however, systematic differences between subjects with respect to place of articulation. One subject heard no front vowels at all, and two other subjects heard no front vowels at $F_0 = 630$ and 700 Hz, and at 150 Hz, respectively. All other subjects heard some front vowels in each $F_0$-series. It can be seen in Fig. 3(b) that the agreement among subjects on perceived place of articulation of rounded vowels was quite low.

In 90.3\% of all cases, subjects responded by encircling the symbol of a single phoneme, 9.5\% were two-
symbol, and 0.1% were three-symbol responses. There were 0.2% response omissions.

C. Discussion

In Fig. 4, boundaries exclusively determined by the tonality distance between $F'$ and $F_0$ would be parallel to the $F_0$ axis, while boundaries determined by $F'$ alone would be parallel to the lines restricting the range of $F'$ in the upper right and the lower left corner of that figure. Figure 4, then, shows quite clearly that only the distinction between openness 4 and 5 at fundamental frequencies above 350 Hz is determined by the frequency of the formant alone. This boundary between [a] and less open vowels apparently cannot be pushed further up than roughly 1.2 kHz. The distinction between openness 1 and 2 is determined by the distance between the formant and the fundamental. The boundary remains at roughly 1.2 Bark for the whole range of $F_0$. The boundaries between higher degrees of openness roughly follow the trend observed for the boundary between openness 1 and 2 up to a $F_0$ of 350 Hz. Between $F_0 = 350$ Hz and $F_0 = 400$ Hz, an abrupt change in response behavior is observed. At $F_0$'s above 400 Hz, the distance between $F'$ and $F_0$ is again decisive, but vowels with intermediate openness (degree 3 or 2) are not often heard. The change in response behavior at $F_0$ between 350 and 400 Hz can be explained in the light of a psychoacoustic model, incorporating the concept of a "second integration band" which will be described in Sec. IV.

On the basis of the hypothesis that the stimuli were
identified by matching their total pattern of excitation along the basilar membrane with stored reference patterns for each vowel, it might be expected that one-formant vowels would be identified as back vowels, since they are dissimilar to front vowels, in that they do not have any prominent energy at high frequencies. One of the subjects, in fact, never perceived any front vowels. All the others heard, besides back vowels, also front vowels, particularly front rounded ones. In the previous investigation (Traunmüller, 1976), the subjects also heard front vowels. It was then suspected that this might be due to a higher "formant" caused by distortion. To test this possibility, the stimuli were also heard under low-pass filtering; nevertheless the front vowel identifications remained. Thus we have no reason to doubt the reliability of this result. This is further supported by the results of Agefors and Gräslund (1979) who found that when low-pass filtered at 470 Hz, 36 dB/oct, the vowels [i, y, u, e, ə, o] were all identified predominantly as front vowels by Swedish listeners.

In an investigation by Ainsworth and Millar (1972), testing the effect of relative amplitude in two-formant vowels, it was found that most but not all of their English subjects gave predominantly back vowel responses when the amplitude of the second formant was reduced to zero. It appears, then, that there are two possible perceptual strategies for the distinction of front versus back rounded vowels, namely to rely on a second formant close to the third (present in front, absent in back vowels), or to rely on prominent components at higher tonalities, naturally achieved by a second formant close to the third (present in front, absent in back vowels). One-formant vowels, then, contain no direct cue to place of articulation since there are no prominent components above F' at all, and they contain conflicting indirect cues, the absence of a close second formant indicating frontness, while the absence of higher formants indicates backness. The lack of agreement in perceived place of articulation can thus be understood. The subjects, however, appear not to have been aware of missing information or conflicting cues, otherwise, the number of multiple-phoneme responses should have been larger.

III. EXPERIMENTS 2 TO 4: VARIATION OF F0 AND F1 FREQUENCY IN NATURAL VOWELS

In these experiments, the stimuli were synthetic versions of natural vowels in which F1 and/or F0 were displaced. The purpose was mainly to test the perceptual saliency of the F1-F0 cue to openness against imaginable conflicting cues based on the higher formants. The higher formants F2, F3, and F4, as well as the nondisplaced F1 or F0 were fixed at the values obtained spectrographically from two careful isolated pronunciations of the vowels by the author. The formant positions and F0 values were as shown in Table V. There was further a fifth formant, always fixed at 4.5 kHz. Front vowels were chosen in order to see whether the tonality distance between F2 and F1 has something to do with perceived roundness.

Figure 5 shows the kind of stimuli and their order of presentation in the three experiments. In each experiment the displaced characteristic was shifted successively in steps of 0.5 units on the modified tonality scale described below and shown in Fig. 6.

Several careful identifications, made by the author, of one-formant vowels of essentially the same kind as those used in experiment 1 led to the suggestion that the dependence of perceived openness on the distance between F1 and F0 could be described uniformly if the Bark scale of tonality was modified at its low-frequency end. This modification is shown in Fig. 6. On the basis of this modified scale, the comparatively small female-male differences in F1 of the same vowels can also be understood in terms of the hypothesis of movable spatial patterns on the basilar membrane as mentioned in the Introduction. At frequencies below 200 Hz, the dependence of this modified tonality measure on frequency is weaker than at higher frequencies. Therefore a formant at, e.g., 350 Hz in a vowel with F0 at 200 Hz has to be lowered by only 50 Hz if F0 is lowered by 100 Hz and the tonality distance is to remain the same. Within the frame of the model to be described later in Sec. IV, the difference between the two scales is explained as an "end of scale effect."

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Characteristic frequencies</th>
<th>Characteristic tonalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel</td>
<td>F0</td>
<td>F1</td>
</tr>
<tr>
<td>i</td>
<td>200</td>
<td>250</td>
</tr>
<tr>
<td>e</td>
<td>150</td>
<td>350</td>
</tr>
<tr>
<td>a</td>
<td>140</td>
<td>455</td>
</tr>
<tr>
<td>o</td>
<td>130</td>
<td>700</td>
</tr>
<tr>
<td>y</td>
<td>190</td>
<td>275</td>
</tr>
<tr>
<td>ə</td>
<td>140</td>
<td>400</td>
</tr>
<tr>
<td>a</td>
<td>130</td>
<td>470</td>
</tr>
<tr>
<td>ə</td>
<td>130</td>
<td>600</td>
</tr>
</tbody>
</table>

TABLE V. F0 and formant positions in the original natural vowels and their synthetic versions used in experiments 2 through 4. Values in Hz (left) and in Bark (right).
A. Experiment 2

In this experiment F0 and F1 were shifted in the vowels [i, e, ë] from their original positions (Table V) in such a way that the original distance between F1 and F0 remained unchanged on the modified tonality scale (Fig. 6).

The identification results are displayed in Fig. 7. Responses were grouped in classes of the vowels with the same degree of openness (left column) and in classes with the same roundedness (right column). The original vowel [ë] was apparently a slightly too open variant. Several subjects perceived it as an [æ].

In all four vowel series the majority of judgments on openness remained the same at all positions of the F0 and F1 configuration. This fact implies that either the subjects relied on the distance between F1 and F0, or they relied on information contained in the higher formants, or both. Experiments 3 and 4 were designed to select among these possibilities. The results mitigate against the hypothesis that the distance between F2 (or any higher formant) and F1 (or F0) has anything to do with perceived openness in these vowels.

Although several subjects at some point jumped from perceived openness 2 to openness 4 in the e and in the ë series at increasing F0, the sudden change in response behavior observed at increasing F0 in experiment 1 did not appear in the e series in this experiment, where it could be expected in analogy with the results of the one-formant vowel identifications. This is likely due to in-
formation conveyed by the higher formants.

With respect to the hypothesis of vowel recognition by pattern matching, at large shifts, the overall patterns evoked by the vowels in this experiment would not match internalized templates derived from natural vowels, no matter how movable these templates were. Such an attempt would inevitably result in a striking mismatch in the pattern of either F0 and F1, or of the higher formants, if not of both. The observed front vowel responses to one-formant stimuli and to low-pass filtered natural vowels also appear incompatible with such an overall match. These results are, however, compatible with the hypothesis that relevant parts of the evoked pattern are matched with stored reference patterns for the F1–F0 cue alone. The results thus suggest a feature-detection process. We should, then, reformulate the hypothesis of Potter and Steinberg (1950): Sounds are not identified, but certain features of sounds are recognized on the basis of spatial patterns of excitation along the basilar membrane regardless of position along the membrane.

As for the perception of roundedness, we can conclude from the results in Fig. 7 that the distance between F2 and F1 (or between F2 and F0) may play a certain role, but it is not usually decisive, otherwise, we should expect only /yɪkə/ responses when F1 (F0) is increased by 3 Bark or more, and F2–F1 (F2–F0) consequently decreased by the same amount, in the vowels based on /iɛz/. There may be some differences in the degree to which different subjects relied on the F2–F1 (F2–F0) cue in roundedness recognition, but observe that in the i series no rounded vowel was heard at all. Thus the distance between F2 and F1 is not the major cue for the distinction of front rounded versus front unrounded vowels.

B. Experiment 3

In this experiment F1 was shifted in the vowels [e, i, a, ð, œ] from Table V. The identification results are displayed in Fig. 8. No systematic differences in response behavior appeared across subjects in this experiment.

The boundaries between degrees of openness appear to be located approximately at the same positions of F1 in all four series of stimuli. This conformity is particularly pronounced within the pairs of series based either on rounded or on unrounded vowels only.

The region of F1 positions for which /e/ responses (openness 3) were given in the series based on unrounded vowels is larger than in the series based on rounded vowels and in experiment 1. Since the difference is accounted for by the higher formants, it may be concluded that these within narrow limits do have an influence on perceived openness if there is a difference in roundedness. If there is no difference in roundedness, however, the influence of the higher formants appears to be negligible. If there were such an influence, it would be expected that the regions of F1 positions in which the original vowels were heard would be enlarged in each series.

FIG. 8. Identifications of vowels in which only F1 had been shifted. Each frame represents results for a series of stimuli with F0 and higher formant frequencies as in the indicated original vowels. Left: Perceived openness, right: Perceived roundedness.

Since the possibility of F1 to determine openness without reference to F0 has been ruled out by the results of experiments 2 and 1, these results strongly suggest that the tonality distance between the first formant and the fundamental is indeed the most important cue in perception of openness.

About the perception of roundedness, not much can be said. Anyway, the results obtained with the stimuli based on [e] appear to indicate that F1 plays some kind of role.

C. Experiment 4

In this experiment F0 was shifted upwards in the vowels [e, ð, œ] from Table V, up to the limit where the distance between F1 and F0 was diminished to less than 0.7 units on the modified tonality scale.

The identification results are displayed in Fig. 9. Most subjects perceived the original vowel "œ" as an [œ]. Apparently the positions of F3 and/or F4 in Table V are not representative for a typical /œ/. For this reason, the results allow conclusions about the F0 dependence of perceived openness only.

Vowels with formant positions of the original vowels [e] and [œ] were perceived as having a lower degree of openness at higher F0. This was also true for 10 subjects when listening to vowels with the formant positions of the original vowel "œ." They typically heard the series with increasing F0 as /œ–œ–œ–œ/. For another 5 of the subjects there was no substantial F0 dependence in perceived openness. They perceived predominantly /œ/ and a few vowels with openness 3 or 5 at all positions of F0. Another 5 of them followed the
IV. EXPLANATORY MODEL

A. General description

Some of the major findings in the present experiments can be explained via a general model of auditory pattern matching and identification, shown in Fig. 10 (cf. Traunmüller, 1979). Some other findings can be incorporated into such a model.

The acoustic stimulus reaching the ear, at point (1) in the model, can be completely described by \( p(t) \) until it reaches the tympanic membrane, and no further dimension is added until it reaches the oval window. Within the cochlea, however, a frequency-to-place transformation is performed. The resulting place dimension is largely maintained along the neural pathway up to the auditory cortex (e.g., at the collicular level according to Rose et al., 1983). Things are more complicated concerning magnitude. There is no close correspondence between the magnitude of neural activity in the 8th nerve and the psychoacoustical magnitude (loudness) of a stimulus or parts of it. It has been claimed (Traunmüller, 1979), that magnitude for the purpose of spectral matching and phonetic identification should be scaled on a linear scale of just noticeable differences. The first step on such a scale of prominence is the threshold of masking. The informative parts of a spectrum are those above that threshold. Accordingly, at point (2) in the model we have a prominence-density versus tonality representation of the stimulus, incorporating the effects of frequency resolution into critical bands.

After point (2) in the model, the signal is distributed to several channels for the extraction of specific qualities. Beside the channel for spectral matching, there is a similar channel for pitch extraction. In the box labeled "specific quality extraction" certain transformations are performed which result in a representation (3) that is largely freed from variation due to characteristics which are irrelevant to the particular specific quality, e.g., intensity in pitch extraction and partial resolution in spectral matching.

The representation at point (3), then, is entered into short term memory and may be compared with a representation previously entered into this memory and/or with permanently stored, i.e., learned, templates representing phonetically significant patterns. The stimulus will then be categorized with the template(s) most similar to the stimulus representation at point (3).

FIG. 10. Psychoacoustical model of spectral matching and identification described in text.
The results from the present experiments, indicating that *parts of vowel stimuli distant in tonality fuse into distinct perceptual gestalten*, demonstrate that these phonetically significant patterns do not correspond to whole phonemes, but possibly to distinctive features. The templates can be thought of as rigid but within certain limits movable in tonality.

The box labeled "specific quality extraction" represents the location of the spectrum envelope reconstruction. If we accept the view that the difference between spectral representations of matched stimuli in some way determines their perceived phonetic similarity, it is necessary to explain how high-pitched vowels are matched to low-pitched ones. This leads to a seemingly inevitable gross mismatch without spectrum envelope reconstruction or an equivalent process of spectral integration, since the lowest partials in high-pitched vowels ($F_0 > 150$ Hz) are clearly resolved at a peripheral stage.

In matching experiments where one-formant stimuli were matched to two-formant stimuli it has been shown (Chistovich et al., 1979) that the single formant will be located in the middle between both formants in the two-formant stimulus as long as the distance between the formants in the two-formant stimulus is less than the critical distance of 3.0 to 3.5 Bark, provided that their levels are approximately the same. If the formant separation is increased above the critical distance, the location of the matched single formant will suddenly approach one of the two formants. These results suggest the reality of spectral integration over a range of roughly 3 Bark.

The results published so far do not allow a determination of the shape of the integration function. Tentatively, a Gaussian shape can be assumed. This appears preferable to a rectangular integration window. The stimulus representation at point (3), then, is determined by a convolution of the prominence representation at point (2) with a Gaussian shaped spreading function with $\sigma = 1.5$ Bark.

If there is a second integration band of the described kind, with a width of about 3 Bark, we can expect some anomalies at both ends of the tonality scale, since the convolution cannot reasonably be extended further than the ends of that scale.

We have, then, to reckon with at least two integrative processes of different range: First, peripheral frequency resolution, the familiar critical band—width 1 Bark; second, the range of spectral integration in envelope reconstruction, i.e., a second integration band—width 3 Bark; and further, with a limited range of gestalt fusion.

**B. Evaluation of predictions**

Let us consider how the $F_1-F_0$ cue, which was found to be decisive for the perception of openness, is reflected in a blurred prominence-density representation as described above.

If the distance between the two partials closest to the formant is less than 3 Bark, there will be a peak in the representation close to the tonality position that corresponds to the resonance frequency of the formant. As can be seen in Fig. 4, at $F_0 > 370$ Hz, the distance between the first two partials is larger than 3.0 Bark. Then, there will be two peaks, one at the fundamental, and one at the second partial or above. The third partial will contribute to this peak, since the distance between sequential higher partials always is less than 3 Bark. Then, if the prominence of the fundamental is not as high as it is when the formant is very close to it, a very open vowel will be heard for the following reason: In a natural [æ] or an [e] with $F_0 < 370$ Hz there will be a peak at the fundamental, which is quite prominent in those vowels, and another peak close to the first formant, located at a distance larger than 3.0 Bark above the fundamental. If $F_0$ is larger than 370 Hz, this kind of pattern will result even if $F_1-F_0$ is less than 3 Bark, since the position of the second peak will then be determined by the second and third partials rather than by the theoretical $F_1$. The model predicts the sudden change in response behavior observed between $F_0 = 350$ and 400 Hz.

Even if the distance between partials closest to a formant is less than the critical distance, due to the nonrectangular integration function we can expect a slight tendency for peaks to be attracted to partials. The effect of this would be that the boundaries between different degrees of openness would tend to follow lines in the middle between partials. Such a tendency can be traced in Fig. 4 for the boundary between openness 3 and 4 at $F_0 > 200$ Hz.

The question of perceptual interaction between single partials and $F_1$ had been investigated previously (Chistovich, 1971; Carlson et al., 1975; Florén, 1979), but the results were difficult to interpret and did not quite match the expectations. The concepts of movable spatial patterns, i.e., of movable templates for certain features, and of a second integration band appear to explain the consistent findings and leave some effects observed at small variations in $F_0$ to be interpreted as statistical or other noise. The concept of prominence-density might explain the discrepancy between the functions of equal loudness and equal phonetic significance of partials in the region of $F_1$ found by Mushnikov and Chistovich (1972). Within the frame of the present model, proportionality between these functions is not expected.

If $F_1-F_0$ is less than 3 Bark, the pattern will contain only one peak, and it can be described by its width or by the distance between the peak and the low-tonality flank of the representation. For $F_0 = 150$ Hz, this lower flank will be located at the end of the scale. If $F_0$ is decreased further than to 150 Hz in such a vowel with fixed $F_1$, the width of the representation will not increase any more. The $F_0$ dependence of openness will stop at a lower limit of 150 Hz. However, since the shape of the second integration band is not rectangular, there will be a transitional region of about 100 Hz. This explains most of the difference between the Bark scale and the modified scale shown in Fig. 6. To this extent,
then, the modification is not a real modification of the tonality scale. This end-of-scale effect also explains the comparatively small female–male differences in F1 of vowels with F1-F0<3 Bark (see Table I). The effect did not emerge very clearly in the results of experiment 1.

If the distance between two formants is less than 3 Bark, they will merge. In an [a] F1 and F2 are roughly equally prominent and just close enough to merge, but they are not in [ae] and Joel. The results of Chistovich et al. (1979) tell us that a single formant will be placed between F1 and F2 in an [a] in order to achieve the best match. Therefore the boundary between openness 4 and 5 in Fig. 4 does not coincide with the position of F1 in natural vowels. This boundary is shifted upwards by the influence of the natural F2 on the internalized phonetic template, and therefore the band of (F1-F0) values in which openness 4 was perceived is so much wider than those for openness 1, 2, and 3. The model assumes the task of identification to be equivalent to that of matching, in the sense that in both cases a stimulus is matched with a pattern in memory, although stored either permanently or just temporarily.

Although the present lack of a validated measure of prominence restricts a quantitative treatment, the model can be said to account satisfactorily for the experimental results. It remains, however, to be established whether any modifications are necessary in order to explain phenomena that might appear in the range above 1.5 kHz.

ACKNOWLEDGMENTS

I am grateful to civ. ing. Peter Branderud, who arranged the vowel generating program, and to fil.kand. Karin Holmgren for their assistance at the computer, and to Hofrat Dr. Leo Klingenberg and his staff at the Bundesrealgymnasium in Amstetten for making localities available and for recruiting subjects. I am further indebted to Professor Björn Lindblom and Dr. James Lubker for valuable comments on the manuscript.

1In a generative phonological description, the front rounded series is derived from underlying front unrounded vowels of the same degree of openness, followed by an /I/. This, together with the symmetric occurrence or non-occurrence of — diphthongs and of casualized phonemes among front and back vowels (Table II), appears already to provide enough support for the feature analysis given in Table III. If [a] is considered neither front nor back, as in Table III, it can be specified without use of a 5th degree of openness, but a description of its acoustic manifestation as the vowel with the highest F1 appears to call for a 5th degree.

2The "second integration band" should not be confused with the "second filter" involved in order to explain the high peripheral frequency selectivity of the ear (Zwicker, 1974; Dufhuis, 1974, 1977). The feature in common to these notions is that to both there is a corresponding, more peripheral, first thing, namely the 1 Bark critical band and the mechanical frequency selectivity of the basilar membrane, respectively.


