Symposium on
AUDITORY ANALYSIS AND PERCEPTION OF SPEECH
August 21-24, 1973 . Leningrad, USSR

AUDITORY ANALYSIS OF VOWEL-LIKE SOUNDS

Tammo Houtgast

Institute for Perception TNO, Soesterberg, the Netherlands

SUMMARY

The results of traditional masking experiments do not fit in with the effect of "two-tone inhibition" revealed by single-nerve-fibre studies. This is considered a serious objection against the interpretation of a masking pattern as reflecting the auditory spectrum of a masker. The results of a different masking paradigm, based on the pulsation threshold, do reveal an effect similar to "two-tone inhibition". Therefore, such pulsation-threshold patterns migh reflect more accurately the auditory spectrum of the masker. Pulsation-threshold patterns of three vowel-like sounds were determined. These were compared with the spectral patterns of the "vowels" after analysis with a 1/3 octave-band filter with running centre frequency. It appeared that the pulsation-threshold patterns are sharper than the 1/3 octave-band spectra and, in some cases, even more pronounced than the original vowel spectra. This has interesting implications with regard to vowel-perception theories, such as theories on "formant extraction".

1. Introduction

The auditory processing of sound spectra plays an important role in speech perception. There are two main sources of information concerning these processes: (1) psychophysical experiments and (2) electrophysiological single-unit studies. The best documented result of single-nerve-fibre studies is the high degree of frequency selectivity in the auditory processing of a simple (pure) tone, as disclosed by tuning curves. There appears to exist a similarity between these tuning curves and psychophysical masking curves. This has led to the well-accepted idea that the masking pattern of a stimulus (or, derived from that, its loudness pattern) gives an adequate description of the effect of (peripheral) auditory processing (Chistovich, 1971). Since in many masking experiments a "critical bandwidth" of about 1/3 octave is found, it is often felt that 1/3 octave-band analysis of speech is a good approximation of the speech processing in the first stages of the auditory channel (Pols, 1973).

However, there are some arguments against the assumption that a masking pattern does reflect correctly the auditory spectrum of a masker after (peripheral) auditory processing. The activity of primary auditory units in response to *two* simultaneous tones clearly reveals an effect which has become known as two-tone inhibition (Sachs and Kiang, 1969; Arthur *et al.*, 1971). This effect might be considered as an example of the more general phenomenon of lateral inhibition or lateral suppression. The important point is that this well-documented effect is *not* revealed by the masking pattern of a stimulus. In this respect there appears to exist a discrepancy between the traditional psychophysical masking pattern and the auditory spectrum as expected on the basis of single-nerve-fibre studies.

Recently, a number of masking paradigms were applied to a two-tone masker (Houtgast, 1973). Indeed, the results of the traditional masking paradigm, with the test tone superimposed on the masker, did not fit in with the "two-tone inhibition" effect. However, it appeared that other paradigms gave results which did agree well with the "two-tone inhibition" effect found in primary unit's responses. A masking pattern of a stimulus obtained with such a paradigm, rather than the traditional masking paradigm, might be considered a more appropriate candidate to disclose the auditory spectrum of the stimulus. Such a paradigm will be applied to some vowel-like sounds with typical formant-like spectral structures. The main question is: can the relation between such masking patterns and the original sound spectra be understood on the basi of a traditional filtering process (with a bandwidth of about 1/3 octave), or does the relation suggest that other processes, like lateral suppression, play a role as well?

2. A new masking paradigm: pulsation threshold

The basic condition for the interpretation of a masking pattern as reflecting the auditory spectrum of the masker is the agreement with well-documentec features of the neural response to one-and two-tone stimuli. Recently, a type of experiment was introduced which appears to fulfill this condition: a methoc based on the pulsation threshold . I will give a short description of that method and an illustration of the agreement between results obtained with that method and the two-tone inhibition effect. (For a full report, see Houtgast 1973.)

The method is as follows. The masker is gated periodically, 125 msec on, 125 msec off, etc. The test tone is presented during the silent intervals. Thus, the test tone is gated in antiphase with the masker, 125 msec off, 125 msec on, étc. (all on- and offsets are smoothed with a 15-msec time constant). The level of the test tone is under the subject's controll. When the test-tone level is raised, starting at a low level, the sensation caused by the series of test-tone bursts typically passes three stages: (1) for low levels, the test-tone bursts are not perceived, (2) for a range of intermediate levels, the test-tone bursts are perceived as a *continuous* tone and (3) for high levels, the test-tone bursts are perceived as a pulsating tone. Subjects are instructed to adjust the test-tone level to the borderline between region (2) and (3): the highest level for which the series of test-tone bursts is just not perceived as a pulsating tone. This is called the *pulsation threshold level* of the test-tone. It can easily be demonstrated that the effect of continuity, and thus pulsation threshold, depends on the presence of the masker bursts. When the masker is removed, the effect of continuity no longer occurs; then, according to the definition, the pulsation threshold coincides with the detection threshold of the test tone. It has been argued that, in its most simple form, pulsation threshold might be considered as an indication of the apparent level of the maskers auditory spectrum in the frequency region of the test tone.

As an illustration I present some results obtained earlier with the pulsation-threshold method applied to a two-tone masker. The masker consisted of a weak tone (fixed at 1000 Hz, 40 dB) and a strong tone of variable frequency F_2 and level L_2. The test tone was centred at the weak tone, 1000 Hz; its level was subject-controlled. It is obvious that, when the masker does *not* contain the strong tone, pulsation threshold is essentially equal to 40 dB. The contours in Fig. 1 indicated the numerical deviation of the pulsation threshold from 40 dB caused by the addition of the strong tone to the weak tone. Interpreting pulsation threshold as the apparent 1000-Hz level in the auditory representation of the two-tone masker, the contours labeled -5 dB and -15 dB indicate the effect of suppression of the fixed weak tone by a strong tone F_2, L_2.

An example of two-tone suppression in primary auditory neurons is given in Fig. 2 (taken from Arthur *et al.*, 1971). The similarity between the two graphs greatly inspires the application of the pulsation-threshold method as a use-

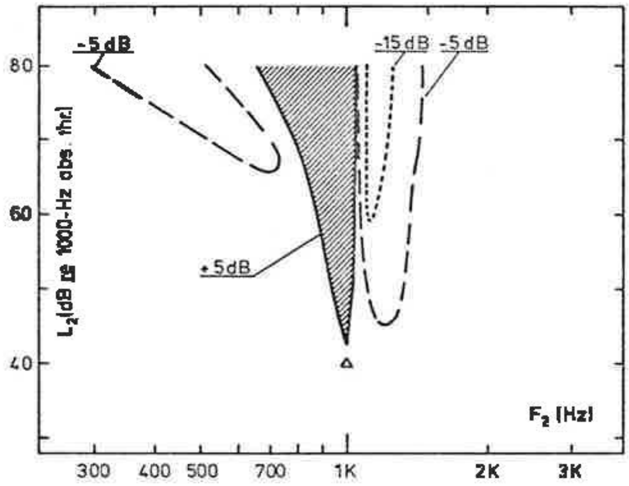

Fig. 1. Pulsation-threshold data with a two-tone masker: a fixed weak tone (triangle) and a variable strong tone F_2 L_2. The test tone is centred at the weak tone. For example, pulsation threshold caused by the weak tone *plus* a strong tone on the -5 dB contour is 5 dB *below* that caused by the weak tone alone (Houtgast, 1973).

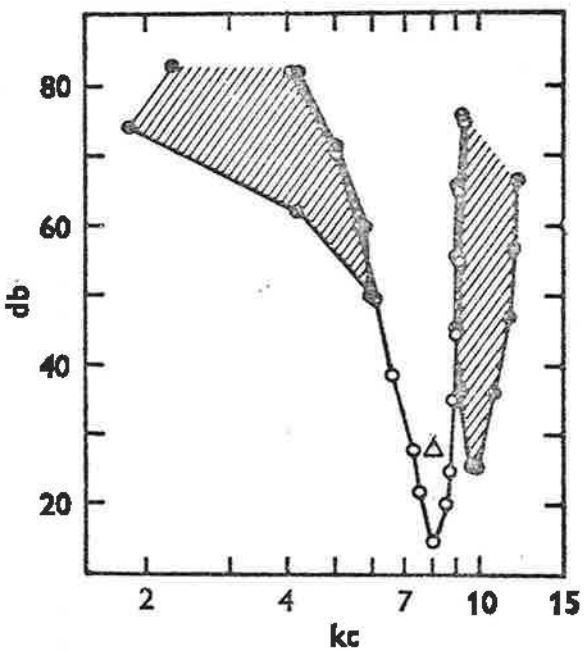

Fig. 2. An example of excitatory and inhibitory areas (hatched) of a primary auditory neurone. The response to a first tone at the best frequency (triangle) *plus* a second tone within the hatched area is more than 20% *below* the response to the first one alone (Arthur *et al.*, 1971).

full tool to disclose the effects of (peripheral) auditory processing.

3. Application to vowel-like sounds

One period of each of three vowel-like sounds was stored digitally by 167 samples of 12 bits. Each signal consisted of the first 32 harmonics of the fundamental frequency 125 Hz, all added in sine-phase relation. The line spectra are indicated in Fig. 3. The common feature of the three spectra is a slope of -10 dB/1000 Hz. With respect to this line, the level of the harmonics in some frequency regions was raised slightly, thus simulating a formant-like structure. According to the positions of these "formants", the three "vowels" are labelled "a", "ε" and "i". These signals were subjected to 1/3 octave-band analysis and to experiments with the pulsation-threshold method.

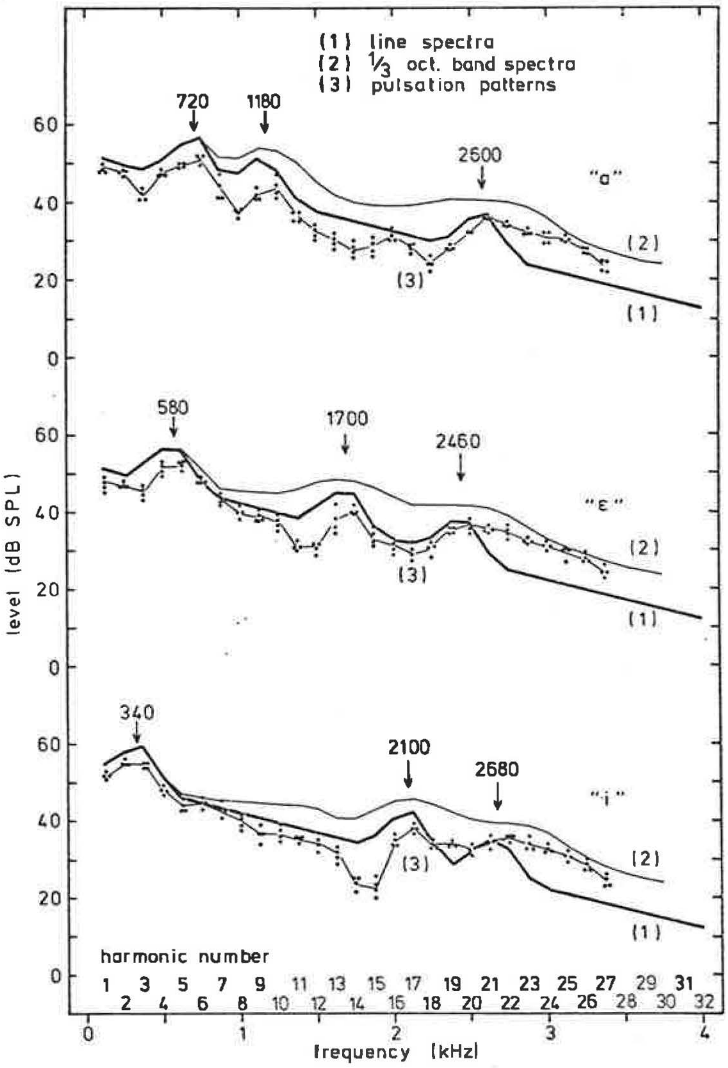

Fig. 3. For each of the three vowel-like sounds three curves are given: (1) the level of each of the 32 harmonics (note the formant-like structure), (2) the output level of a 1/3 octave-band filter centred at the first 30 harmonics and (3) the pulsation-threshold levels for a test tone centred at the first 27 harmonics (one subject).

The thin lines in Fig. 3 represent the 1/3 octave-band spectra for a filter with running centre frequency, when the filter output level is taken each time the centre frequency is a multiple of 125 Hz (thus, centred at one of the harmonics). Actually, these data were obtained by changing the sample-generation time of the signals such that the frequency of each harmonic could be matched to a centre frequency available in a bank of 1/3 octave-band filters.

The results of the pulsation-threshold measurements are presented in Fig. 3 as data points. A "vowel" was alternated with a test tone, as described before. The test tone was digitally-generated as well, with frequency and phase equal to one of the harmonics of the "vowel". The subject was instructed to adjust the test-tone level to the highest value at which the series of test tone bursts was just not perceived as a pulsating tone: pulsation-threshold level. For one subject, four such adjustments at the first 27 harmonics of the three "vowels" were obtained, in a number of properly randomized sessions and conditions. The curves, connecting the average pulsation thresholds at each of the harmoncis are referred to shortly as *pulsation patterns*.

Fig. 4 was constructed from the data in Fig. 3 in order to illustrate to what extent 1/3 octave-band spectra and pulsation patterns reflect the original spectral differences between two "vowels".

Finally, part of the measurements was repeated with three other subjects. The data of the four subjects are presented together in Fig. 5. This figure is analogous to the upper part of Fig. 4.

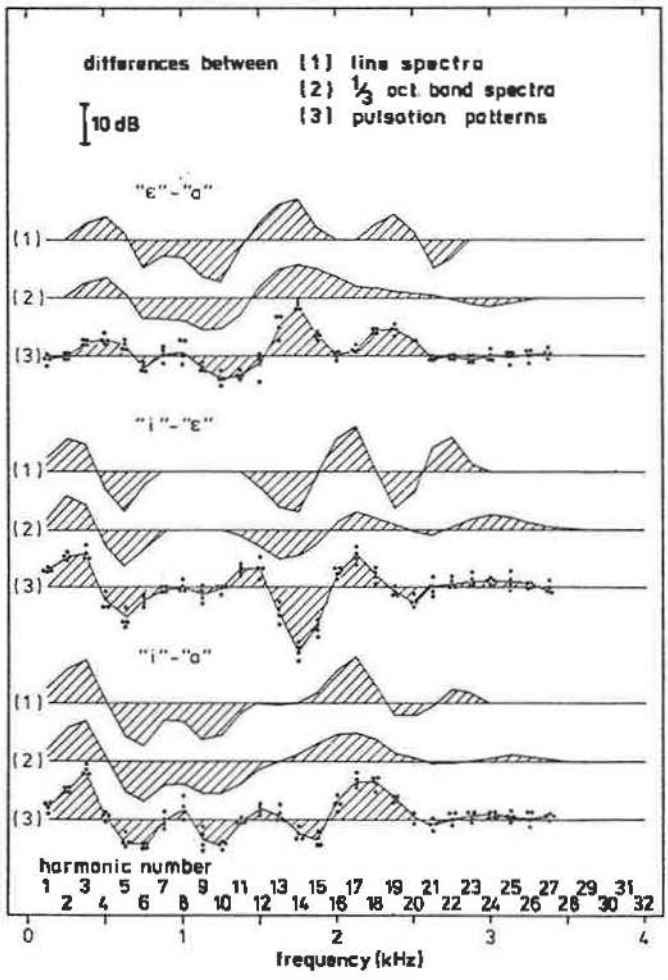

Fig. 4. This graph is constructed from Fig. 3. For each pair of "vowels", the graph gives the *differences* between the two line spectra, the two 1/3 octave-band spectra and the two pulsation patterns.

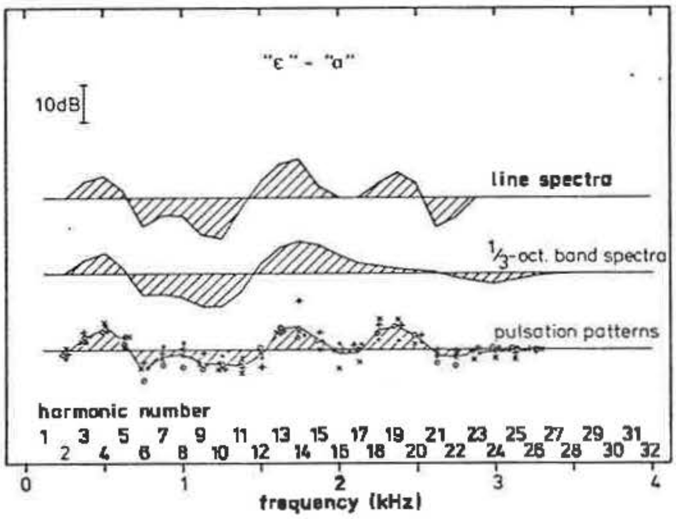

Fig. 5. As Fig. 4 upper part. This graph presents the data of four subjects.

4. Discussion

The possible implications of the results reported here with respect to speech perception depend on accepting a direct relation between the pulsation-threshold pattern of a stimulus and its auditory spectrum . As argued before, the basic condition for accepting this relation is fulfilled. Since there appear to be no arguments against a generalization of the validity of the relation, let us, for a moment, identify a pulsation-threshold pattern as the auditory spectrum of the masker.

It is generally accepted that the auditory spectrum (reflecting the distribution pattern of neural activity in the ensemble of primary neurons) forms the basic information for vowel discrimination and recognition. However, there appears to be a major difference in opinion about the way in which these patterns are processed. Either the positions of the peaks in the auditory spectrum (formants) are considered important (*e.g.*, Chistovich and Mushnikov), or the shape of the whole pattern, without any specific weight ascribed to the peaks, is considered important (*e.g.*, Plomp, 1973).

It is often assumed that the relation between the shape of the actual sound spectrum and the shape of the auditory spectrum can be approximated basically by the effect of 1/3 octave-band filtering. The present data illustrate that the *peaks* in the auditory spectrum are more pronounced than after 1/3 octave-band filtering and also that the *differences* between the auditory spectra of two "vowels" are greater than those after 1/3 octave-band filtering. Thus, the neural projection of the actual sound spectrum appears to be sharper than is assumed traditionally. It should be noted that this does not favour one of the two conceptions exculsively. Both for a "formant extraction" type of process or for a more general type of pattern transformation, such a sharp and pronounced projection of the original spectrum seems to be a favourable starting position.

The data suggest that, when a physical analysis of a stimulus is performed specifically to simulate the effect of the auditory processing, a 1/3 octave-band analysis can serve only as a rough first approximation. By using narrower filters and some type of lateral interaction of filter outputs one might arrive at the sharp, and in some cases exaggerated, patterns disclosed by the pulsation-threshold method.

Conclusions

A model of 1/3 octave-band analysis is often considered to be a relevant approximation for relating the actual stimulus spectrum to the shape of the auditory spectrum (reflecting the distribution of neural activity in the ensemble of primary auditory fibres). Accepting the relevance of the pulsation-threshold method in disclosing the auditory spectrum of a stimulus it was found that:

* the peaks in the auditory spectrum of a "vowel" (formants) are more pronounced than the peaks in the 1/3 octave-band spectrum (with running centre frequency),
* the differences between the auditory spectra of two "vowels" are greater than the differences between the 1/3 octave-band spectra,
* simulating the (peripheral) auditory processing of stimuli by a process like 1/3 octave-band analysis can serve only as a rough first approximation, not accounting for the accurate neural projection of the original shape of the stimulus spectrum.

References

Arthur, R.H., Pfeifer, R.R. and Suga, N. (1971): Properties of "Two-Tone Inhibition" in Primary Auditory Neurons, J. Physiol. 212, 593.

Chistovich, L.A. (1971): Auditory Processing of Speech Stimuli-Evidence from Psychoacoustics and Neurophysiology, Proceeding of the 7e ICA, Budapest, 21 G 1.

Chistovich, L.A. and Mushnikov, V.N. (1971): Auditory Measurement of the First Formant, Proceedings of the 7e ICA, Budapest, 24 C 17.

Houtgast, T. (1973): Psychophysical Experiments on "Tuning Curves" and "Two-Tone Inhibition", Acustica 29, in press.

Plomp, R. (1973): Auditory Analysis and Timbre Perception, this symposium.

Pols, L.C.W. (1973): Analysis and Synthesis of Speech Using a Broad-Band Spectral Representation, this symposium.

Sachs, M.B. and Kiang, N.Y.S. (1969): Two-Tone Inhibition in Auditory-Nerve Fibres, J. Acoust. Soc. Amer. 43, 1120.