Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Detectability of changes of level and spectral slope in vowels

Carlson, R. and Granström, B.

journal: STL-QPSR
volume: 17
number: 2-3
year: 1976
pages: 001-004

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

I. SPEECH PERCEPTION

A. DETECTABILITY OF CHANGES OF LEVEL AND SPECTRAL SLOPE IN VOWELS

R. Carlson and B. Granström

Abstract

It has been argued that spectral slope could be used to signal vocal effort and to emphasize words and syllables. Even if the data on these subjects are not clear, there is still an important question: What are the limits in the perception of spectral slope?

Thus, we have tried by means of listening tests to estimate intensity and spectral slope detectability in different vowels. The result shows that the perception of spectral level and slope is highly vowel dependent.

Introduction

Changes in articulation and phonation will have a variety of effects on the acoustic output of the vocal tract. The most noted effects are changes in the frequency domain described for example as movements of formants. However, a number of especially phonatory changes results primarily in differences in spectral level. The linguistic factors that could be mediated by such changes include stress, emphasis, and possibly, position within a phrase. Increase in voice effort could result in change in the level of a few harmonics or in a decrease in the slope of the entire spectrum according to a study of Sundberg (1970) including both sung and spoken vowels. We are interested in the perceptual effects of such changes. Data have been gathered on the detectability of changes in the SPL of vowel sounds by Flanagan (1955). In that study a difference limen (DL) of about 1.5 dB in a neutral vowel was found. The DL of individual harmonics in vowel sounds has been estimated in a study by Kakusho et al (1971), who found DL's down to less than 2 dB. However, the DL's were noticeably greater for the [u] vowel.

Subjective loudness of spoken vowels has been estimated in a number of studies (e.g. Mendel et al (1969); Brandt et al (1969)). One result of of these studies is that subjective loudness increases more rapidly as a function of SPL when the intensity is increased by means of increased vocal effort than if only the overall level is increased. This might indicate a greater sensitivity to speaker controlled intensity changes.

In this study we have tried to estimate intensity- and spectral slope-detectability in different vowels.

Method

To allow for maximum flexibility in experimentation we produced our vowel stimuli by adding sinusoids from simulated frequency generators. This meant that all stimuli were produced at a constant pitch (of 100 Hz). Four vowels were used [i], [a], [u], and the "neutral" vowel [N], each specified by its four lowest formants, see Table I-A-I. In Fig. I-A-1 spectra for these reference sounds can be seen. The calculated amplitude for each partial is indicated.

Vowel	F_1	F_2	F_3	F_4
i	255	2065	2960	3400
a	580	940	2480	3290
u	310	730	2280	3300
N	500	1500	2500	3500

TABLE I-A-I.

For each vowel three sets of stimuli were produced in the following ways:

A. A decrease in overall amplitude by 1, 2, 3, 4, 5, and 6 dB (as a "bridge" to earlier data).

B. An increase in spectral slope by . 5, 1, 2, 3, 4, and 5 dB/oct starting at 200 Hz.

C. The same as "B" but with the level adjusted to the same SPL as the original vowels. This is approximately the same as making the first formants equal in amplitude.

These stimuli were arranged in ABX triads, each containing only one vowel type where either A or B was an unchanged reference vowel. Each stimulus had a natural on- and off-set and was 250 msec long. The interval between the stimuli was 950 msec and the pause between triads 3 sec.

A discrimination test was given in three test sessions, where each session contained different vowel stimuli but with only one kind of manipulation (A, B, or C). Three groups of five listeners were given the sessions in rotated order. The test was presented from tape through earphones (HD 414) and responses recorded manually.

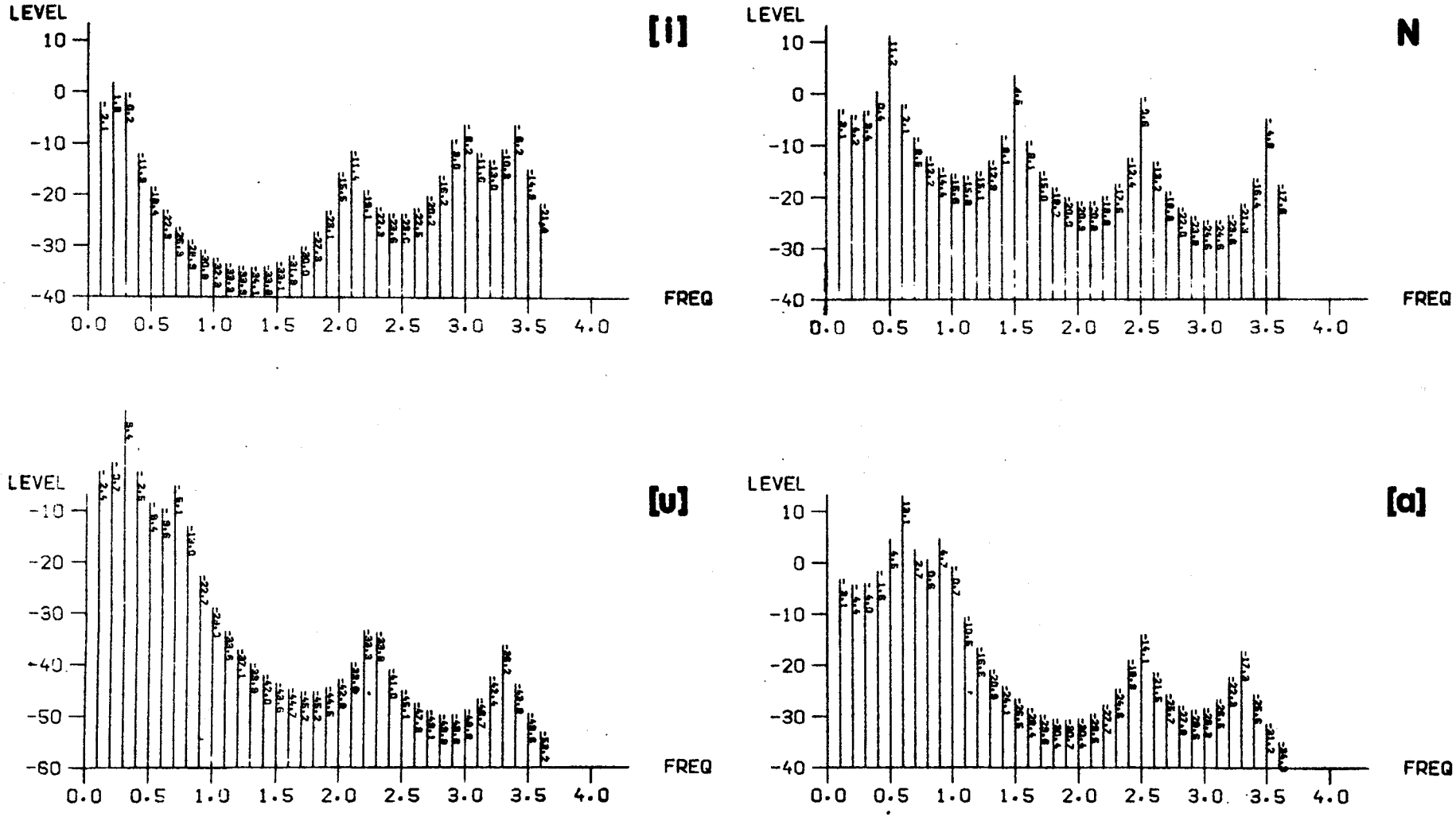

Fig. I-A-1. Spectra for the reference vowels used in the experiment. Relative level is noted on each harmonic.

Results and discussion

The pooled results with the stimuli sets A, B, and C can be seen in Fig. I-A-2a, I-A-2b, and I-A-2c, respectively. Each data point depends on 60 judgments. The DL for intensity decrease (Fig. I-A-2a) crudely estimated as the intersections with the 25 % error line ranges from less than two to just more than three dB. This is about twice the figure given by Flanagan. This is not surprising considering the relatively long inter-stimulus interval and the fact that different vowels were mixed in the test. There is a tendency to [i] and "N" to have smaller DL's than [a] and particularly [u]. One may speculate that this depends on the number of spectrally well separated regions that are audible enough to be used for level comparison.

In Fig. I-A-2b the results from spectral slope changes can be seen. The slope change is confined to frequencies above 200 Hz and could be thought of as a change in the slope of the glottal source. The discriminability is highly vowel dependent. This could at least in part be explained by differences in the first formant amplitude, since a given spectral slope results in different first formant amplitudes depending on the F1 frequency. Fig. I-A-2c shows the results from the stimulus set where these differences are compensated for. Again a strong vowel dependence is apparent, the estimated DL's for spectral slope changes ranging from about .75 dB/oct for [i] to nearly 3 dB/oct for [u].

It is clear that spectral slope per se cannot explain our data. To explain our results we have plotted the data in Fig. I-A-2c in a somewhat different form, Fig. I-A-3. The abscissa is now level difference in F3. The rational behind this choice of parameter is that for all vowels the level difference is great in this region and the audibility is comparatively high. As can be seen the DL expressed in level difference at F3 is about 4 dB for all vowels except [u]. The result with [u] could be understood considering the low audibility of F3 as seen e. g. in an analysis based on Zwicker's loudness model (Zwicker and Feldtkeller, 1967) or by noting the great DL for attenuation of partials in the F3 region of [u] reported by Kakusho et al (1971).

Final remarks

In our study we have found a considerable vowel dependence for discrimination of vowel level and spectral slope. The perceptual processes

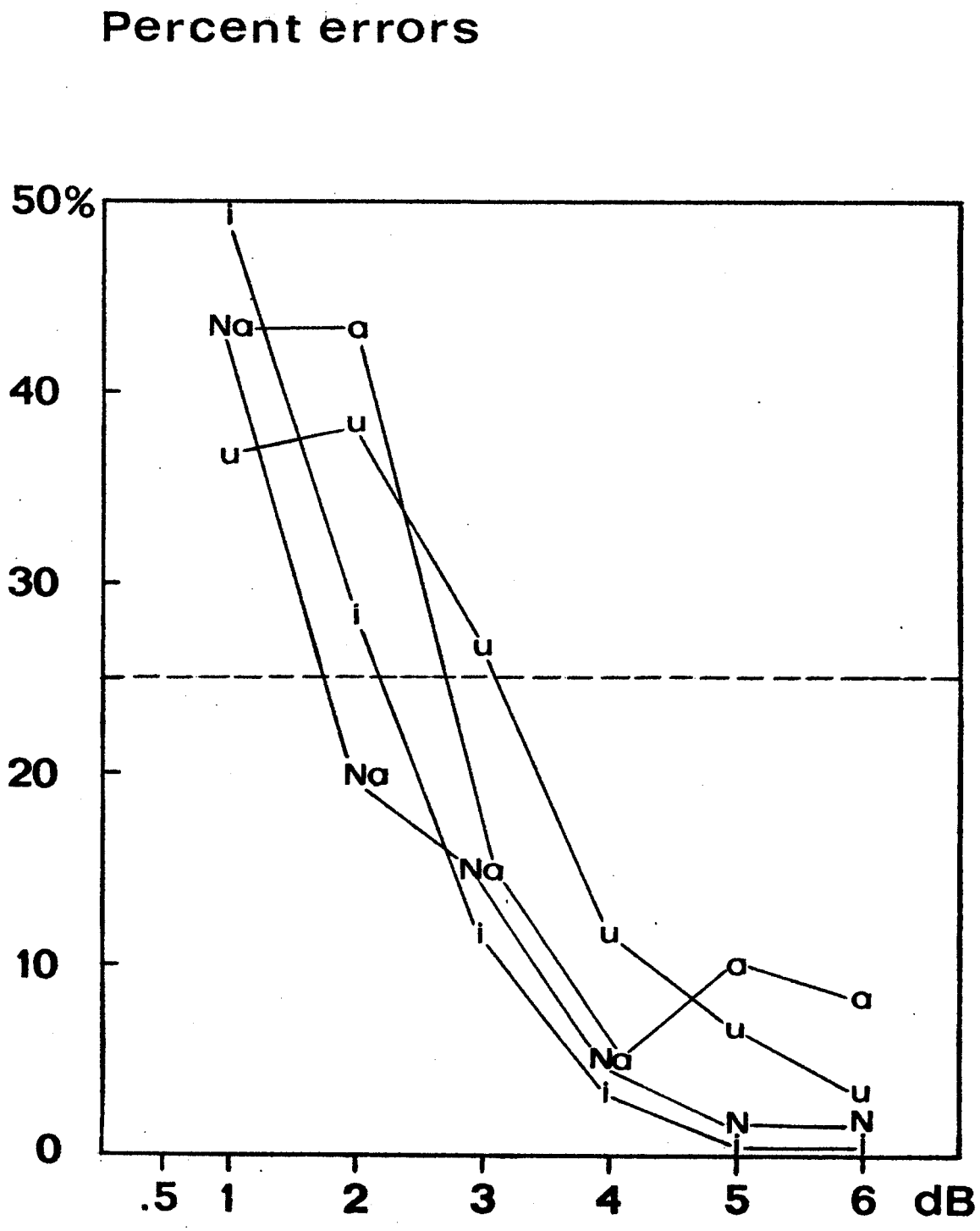

Fig. I-A-2a. Result of ABX test using overall amplitude as parameter.

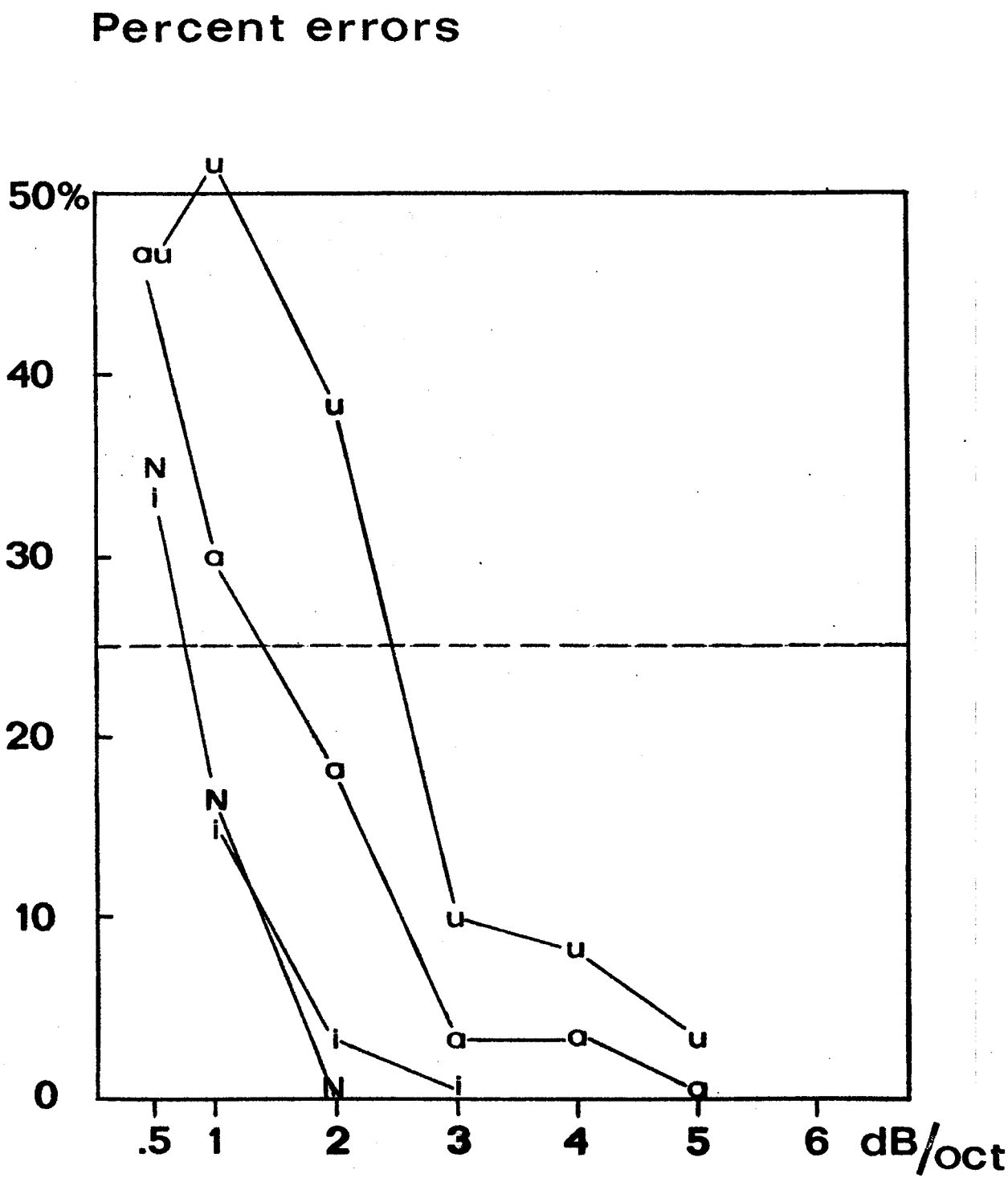

Fig. I-A-2b. Result of ABX test using spectral slope as parameter.

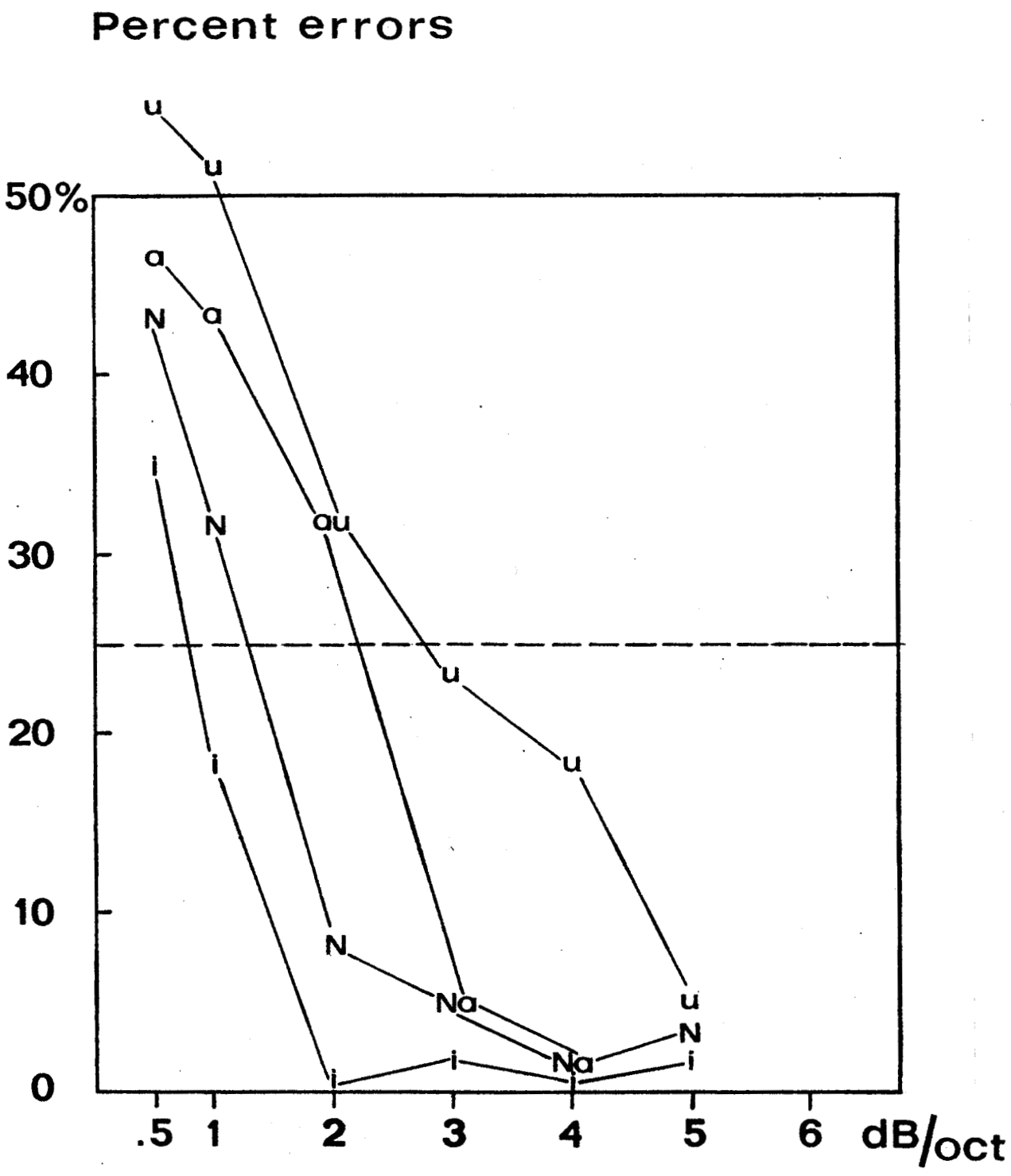

Fig. I-A-2c. Result of ABX test using spectral slope as parameter and all stimuli adjusted to the same SPL, i.e. approximately a normalization of the F1 level.

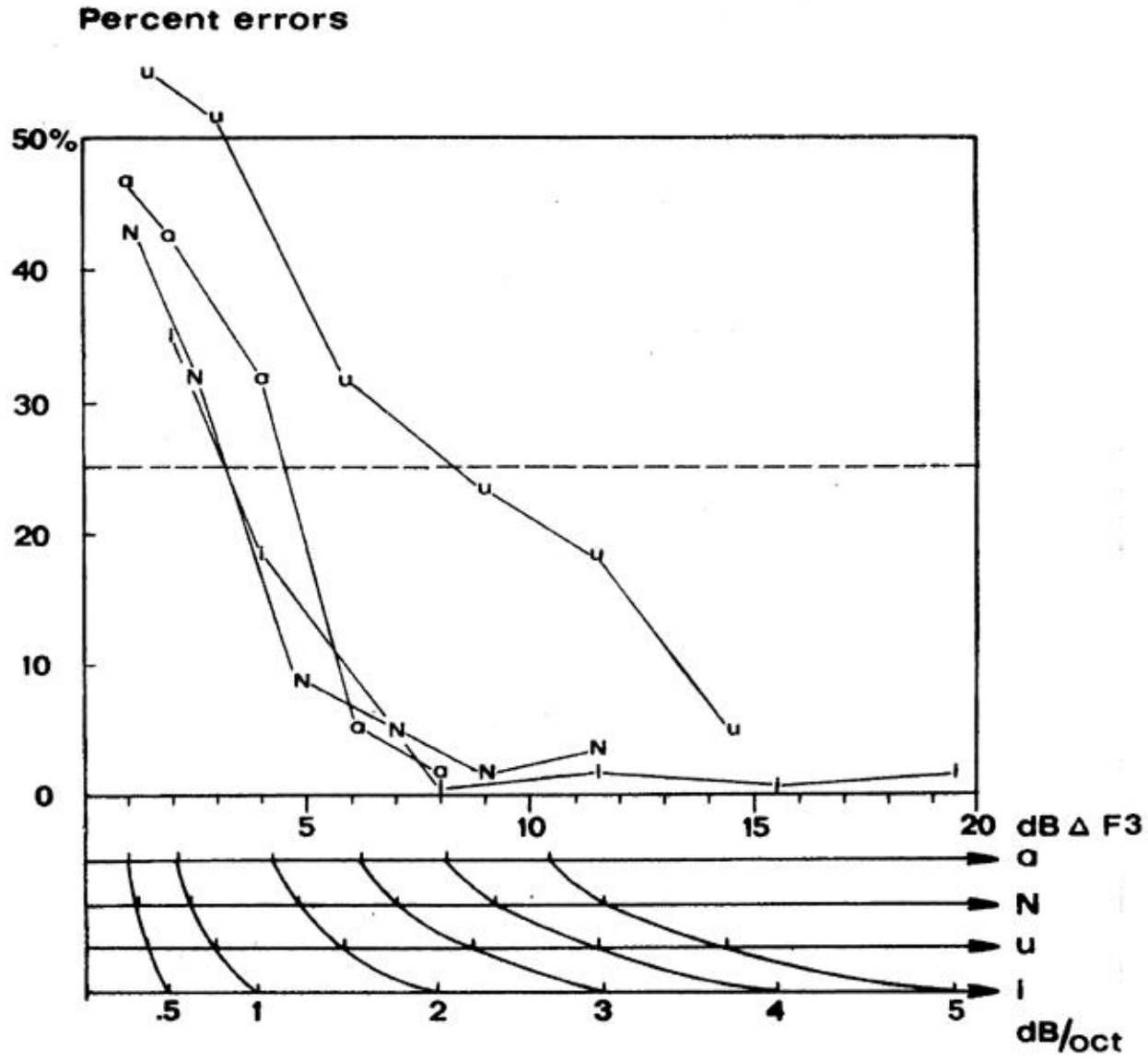

Fig. I-A-3. The same result as in Fig. I-A-2c plotted in another way using the level difference of F3 as variable on the x-axis. The axis below shows the slope variation according to each test vowel.

that account for this could only be speculated on and constitute a challenge for future research. It is, however, clear that the slope difference per se does not explain our data. The level difference in prominent spectral regions seems much more likely as a decision parameter. Of course it could be questioned to what extent our result could be used in estimating the importance of different cues in natural speech. Research on this path is however under way.

Acknowledgments

This work was supported by the Swedish Board for Technical Development under Contract STU 74-3547.

References:

BRANDT, J.F., RUDER, K.F., and SHIPP Jr., T. (1969): "Vocal Loudness and Effort in Continuous Speech", J.Acoust.Soc.Am. 46, pp. 1543-1548.

FLANAGAN, J.L. (1955): "Difference Limen for the Intensity of a Vowel Sound", J.Acoust.Soc.Am. 27, pp. 1223-1225.

KAKUSHO, O., HIRATO, H., KATO, K., and KOBAYASHI, T. (1971): "Some Experiments of Vowel Perception by Harmonic Synthesizer", Acustica 24, pp. 179-190.

MENDEL, M.I., SUSSMAN, H.M., MERSON, R.M., NAESER, M.A., and MINIFIE, F.D. (1969): "Loudness Judgments of Speech and Non-speech Stimuli", J.Acoust.Soc.Am. 46, pp. 1556-1561.

SUNDBERG, J. (1970): "The Level of the 'Singing Formant' and the Source Spectra of Professional Bass Singers", STL-QPSR 4/1970, pp. 21-39.

ZWICKER, E. and FELDTKELLER, R. (1967): Das Ohr als Nachrichtenempfänger, S. Hirzel Verlag, Stuttgart.