ARIPUC 16, 1982, p. 101-111

THE EFFECT OF HIGH AND LOW VOWELS ON THE FUNDAMENTAL FREQUENCY IN SINGING: SOME PRELIMINARY OBSERVATIONS

METTE GREIFFENBERG
AND
NIELS REINHOLT PETERSEN

Fundamental frequency was measured in high and low vowels in solo-singing. The effect of vowel height on Fo in singing was smaller than in speech, but it was shown to be highly significant. The results of a subsequent listening test indicated that listeners seemed to compensate perceptually for the inherent fundamental frequency differences when listening to solo-singing.

I. INTRODUCTION

There is evidence from a great number of languages that high vowels tend to have a higher Fo than do low vowels, everything else being equal (see e.g. Di Cristo and Chafkouloff 1977, Hirst, Di Cristo and Nishinuma 1979, Johansson 1976, Ladefoged 1964, Lehiste 1970, Lehiste and Peterson 1961, Neweklowsky 1975, Peterson and Barney 1952, Reinholt Petersen 1978).

Whether a similar effect is to be found in singing has - to our knowledge - not been subject to systematic research. One study touches upon this question, viz. Hagerman and Sundberg 1980, who investigated how accurately singers could treat fundamental frequency in barbershop singing, which is a type of male quartet singing, where vibrato is avoided and where great accuracy of intonation is required in order to avoid beats. The authors state (p. 42) that "*The accuracy with which the fundamental frequencies are chosen in barbershop singing is extremely high and does not seem to depend on the vowel to any great extent*". However, in the material the chords were made up by EITHER [mɑ] OR [mo] syllables, and the statement quoted refers to the finding that it made no difference whether a given interval was sung on [mɑ] or on [mo] syllables. It would have been interesting to see what had happened if different vowel qualities had been used in the same chord.

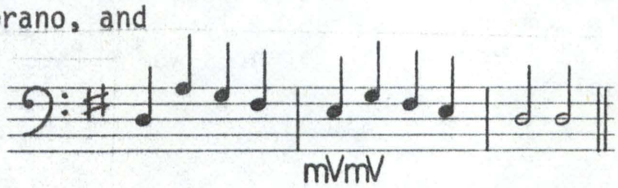
The purpose of the pilot experiment reported below was to investigate the effect of vowel height on the fundamental frequency in solo-singing and - for the sake of comparison - in spoken language. We chose solo singing as the object of study because we believed that this type of singing would provide the most favourable conditions for eliciting a possible effect of vowel height on Fo.

II. METHOD

The vowels *a* and *u* were inserted in the nonsense word *ˈmVmVmi* in all possible combinations (i.e. *u-u*, *u-a*, *a-u*, and *a-a*), and the words were embedded in the carrier phrase *stavelserne ..... forkortes* [ˈsɡ̊avəlsʌnə ····· fʌˈɡ̊ʰɒːd̥əs][1].

The carrier tune used for the singing of the material was

for the soprano, and



for the tenor, the musical interval between the vowels in the test sequences being the minor third between E_2 and G_2 and between E_3 and G_3 , respectively. The four test sentences were arranged in a list in twelve different randomizations. The first and last randomizations were used as dummies and not submitted to subsequent measurement.

Two subjects participated in the experiment, one tenor (BE) and one soprano (MG, co-author of the present paper). They are both trained, semi-professional singers and university students of musicology (MG also of phonetics).

In a preparatory recording session with the singer MG, a slight Fo downdrift of about one semitone was observed during recording. In order to counteract this a sinus tone of 440 Hz was played to the subjects as a reference before the singing of each sentence. The recordings took place in a sound treated room, and the subject was standing in front of the microphone at a distance of about 50 cm. The equipment used was a Sennheiser MD21 microphone and a semi-professional REVOX A77 tape recorder. The list was first sung, and after a short break spoken by the subjects, who were instructed in singing to use no vibrato and in reading to use a neutral declarative intonation.

After the recording session had been finished it was discovered that in speaking the list BE had made reading errors in the greater part of the *mamumi* words, mostly by substituting the word *mamamu*, which by the way does not occur in the list. These errors had passed unnoticed by the experimenter during recording. There had been no problems in singing, and it was, therefore, decided not to recall the subject for another recording session, and to make do with the remaining (correctly rendered) test words.

The following acoustic registrations were made of the material: duplex oscillogram, two intensity curves, and a fundamental frequency curve. Since the fundamental frequency was only in a few cases completely level all through the vowels but drifted slightly, Fo was measured at a point two thirds from vowel onset. The Fo at this point represents, according to Rossi 1971 and 1978, a good approximation to the perceived pitch of short Fo glides.

III. RESULTS AND DISCUSSION

All measurements were converted into semitones relative to 440 Hz, and means and standard deviations were computed. The data are tabulated in table I, and the means are displayed graphically in figure 1.

Table I

Mean fundamental frequencies (in semitones relative to 440 Hz) and standard deviations in the four test sequences in speech and singing. The numbers 1 and 2 above the columns refer to the first/stressed vowel and the second/unstressed vowel, respectively.

		MG		BE	
		1	2	1	2
u-u	speech	-14.71 0.323	-11.72 0.598	-17.39 0.375	-16.53 0.458
	singing	-4.96 0.170	-1.66 0.153	-16.78 0.259	-13.88 0.236
u-a	speech	-14.85 0.437	-12.25 0.430	-17.64 0.491	-17.48 0.301
	singing	-4.83 0.196	-2.08 0.290	-16.78 0.121	-14.22 0.130
a-a	speech	-17.92 0.420	-13.70 0.352	-21.57 0.419	-17.72 0.482
	singing	-5.30 0.200	-2.34 0.254	-17.16 0.177	-14.12 0.200
a-u	speech	-17.37 0.537	-12.96 0.385	- -	- -
	singing	-5.40 0.097	-1.93 0.267	-17.28 0.207	-13.91 0.240

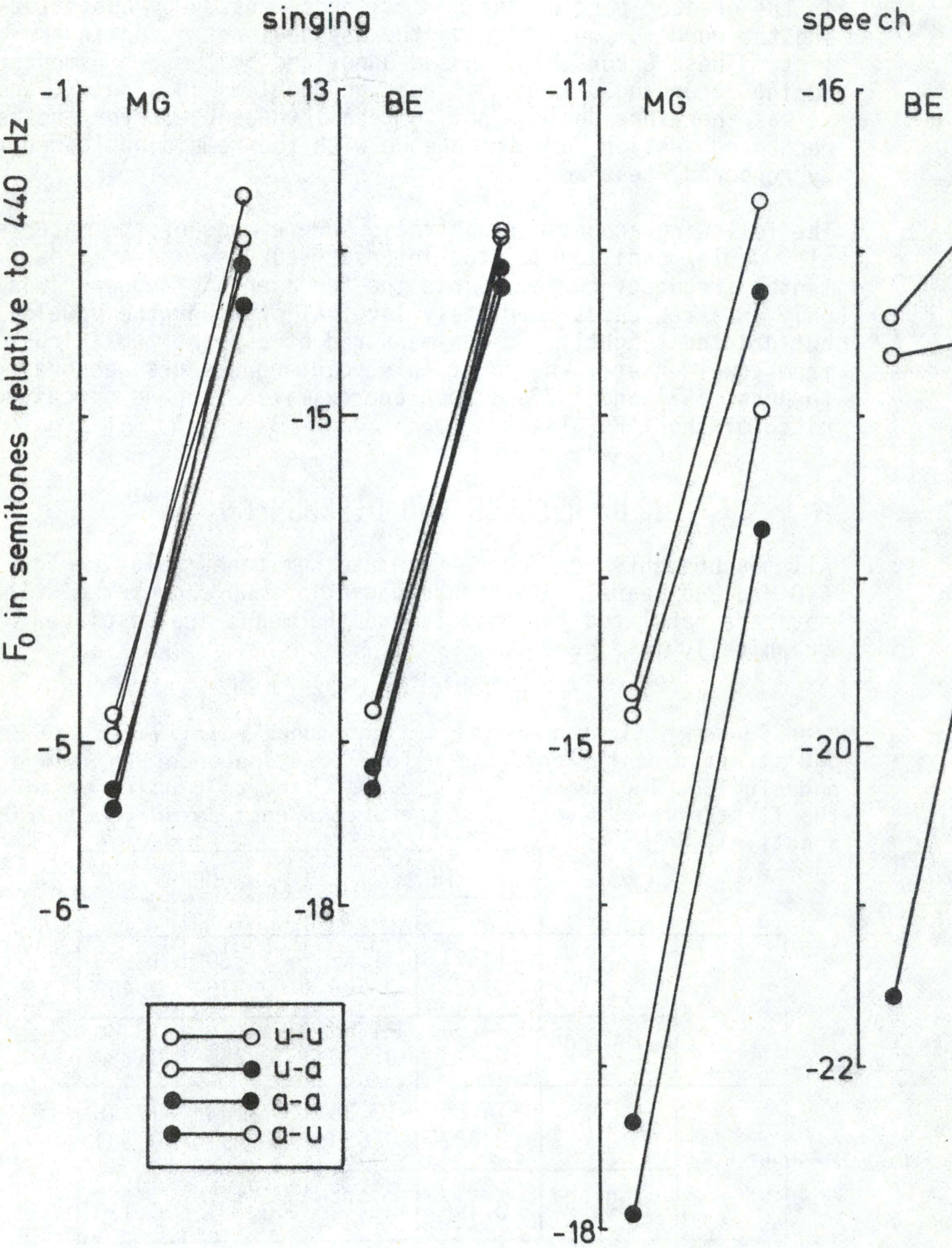


Figure 1

Mean fundamental frequencies in semitones relative to 440 Hz in singing and speech for the two subjects MG and BE. The two vowels in each test sequence are connected by lines.

The differences between *u* and *a* in speech and singing are given in table II, which also shows the results of a series of t-tests applied to the data. In speech the differences are in all cases statistically significant and are comparable in magnitude to the findings of previous investigations of inherent Fo level variation in Danish (see Reinholt Petersen 1978, 1979, and 1980).

Table II

Differences in semitones between means of *u* and *a* in speech and singing. The levels of significance acheived are indicated by ++ (p<0.01) and + (p<0.05).

		1st/stressed vowel		2nd/unstressed vowel	
		$m^{u}_{a}mu$	$m^{u}_{a}ma$	mum^{u}_{a}	mam^{u}_{a}
MG	speech	2.60++	3.07++	0.52+	0.74++
	singing	0.44++	0.47++	0.42++	0.41++
BE	speech	-	3.93++	0.95++	-
	singing	0.50++	0.38++	0.34++	0.21+

In singing the differences are much smaller than in speech, but yet statistically significant in all cases as can be seen from table II. In comparing speech and singing it may be pointed out that the within-group variability (cf. table I) is considerably smaller in singing than in speech. The overall within-group variance for speech is 2.562 and for singing 0.043, and the difference between these variances is statistically significant (F=59.744 p<0.01).

These findings indicate that subjects, when necessary - as in singing - can minimize not only the random fundamental frequency variation but also the systematic Fo variation arising from differences in vowel height, i.e. they are able to counteract an effect on Fo which is assumed to be an automatic consequence of properties inherent in their speech production system. A similar tendency has been reported by Hombert 1976a, who found that the effect of consonant type on Fo in the following vowel seems to be smaller in a tone language (Yoruba) than the effect normally found in non-tonal languages.

Although smaller than in speech the existence of an effect of vowel height on the fundamental frequency in singing has been clearly demonstrated and, as appears from table III and figure 2, this effect gives rise to appreciable differences among the realizations of the intended musical interval of a minor third (i.e. 3 semitones) between the first and the second

Table III

Means and standard deviations in semitones for produced intervals in singing.

	MG	BE
a–u	3.37 0.284	3.47 0.245
a–a	3.04 0.116	2.95 0.266
u–u	2.90 0.245	3.30 0.120
u–a	2.56 0.211	2.75 0.304

vowel in the test sequences. As could be expected the smallest intervals produced are found in *u–a* sequences and the largest in *a–u* sequences. The differences between the two types of sequences approach one semitone (0.72 for MG, and 0.81 for BE). A one-way analysis of variance applied to all four types (*a–u, a–a, u–u, u–a*) showed a statistically significant effect on the size of the produced intervals (MG: $F=17.995$, $p<0.01$, BE: $F=22.756$, $p<0.01$).

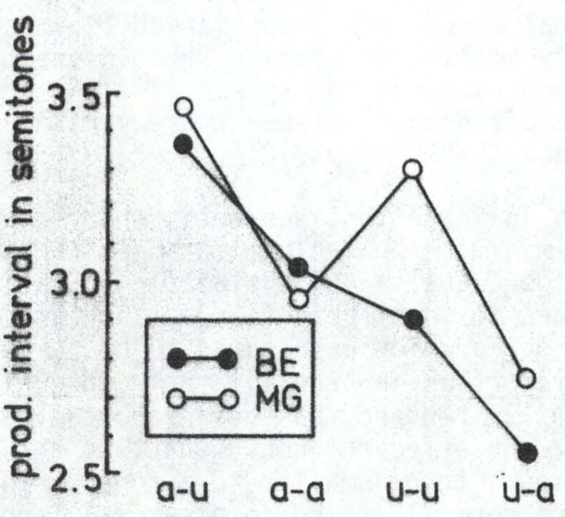


Figure 2

Mean produced intervals in semitones.

It is also seen from figure 2 and table III that in extreme cases (*a-u* for MG, and *u-a* for BE) the produced intervals deviate by nearly half a semitone from the theoretical value of 3 semitones of the intended minor third. Sundberg (1978, p. 65f) and Hagerman and Sundberg (1980, p. 40) report data on deviations from theoretical values when subjects adjusted successive synthetic tones to different musical intervals. The mean deviations were always below 0.1 semitones, and this implies that the deviations found in the present experiment are certainly above the threshold of detection.

Why, then, did the subjects not detect their deviations when listening to their own voices during recording? One explanation could be lack of skill on the part of the subjects. This is unlikely, however, considering their educational background and their experience as semi-professional performers, but, of course, the possibility cannot be precluded. Another explanation could be that they, listening to their own voices, and listeners in general, listening to solo-singing, perceptually compensate for the effect of vowel height on the fundamental frequency. A compensation of this kind has been found in speech by Hombert (1976b) and Rosenvold (1981).

In order to try out this explanation, a listening test was arranged in which a highly competent listener (a staff member of the Institute of Musicology) was asked to evaluate the intervals produced by the subjects.

By means of a sampling, editing and test generating program implemented on the PDP/8 computer of the Institute of Phonetics the test intervals were cut out of the carrier tune, stored and played back to the listener in the order in which they had occurred in the original recording. Each interval was repeated 3 times followed by a pause of four seconds for the listener to put down his response. The first and last four stimuli (i.e. the first and last randomizations) were regarded as dummies so that only those responses were analysed which correspond to intervals having been measured. Thus ten responses were given to each of the four different test sequences. The listener was instructed to evaluate the intervals on a seven-point scale ranging from fully acceptable (1) to not acceptable (7).

After the test the listener complained that he had found the task extremely difficult, and that he had not always felt certain that he had perceived the stimuli as music. The listener's difficulties can, perhaps, be explained by the fact that the stimuli were rather short and had been stripped of their musical context. Therefore the results of the test as presented below should be taken with some reservation.

It should be expected, under the hypothesis of compensation for inherent Fo level differences, that the location on the scale of the bulk of the responses would be independent of the heights of the vowels in the intervals responded to. If, on the contrary, the listener had made no perceptual compensation

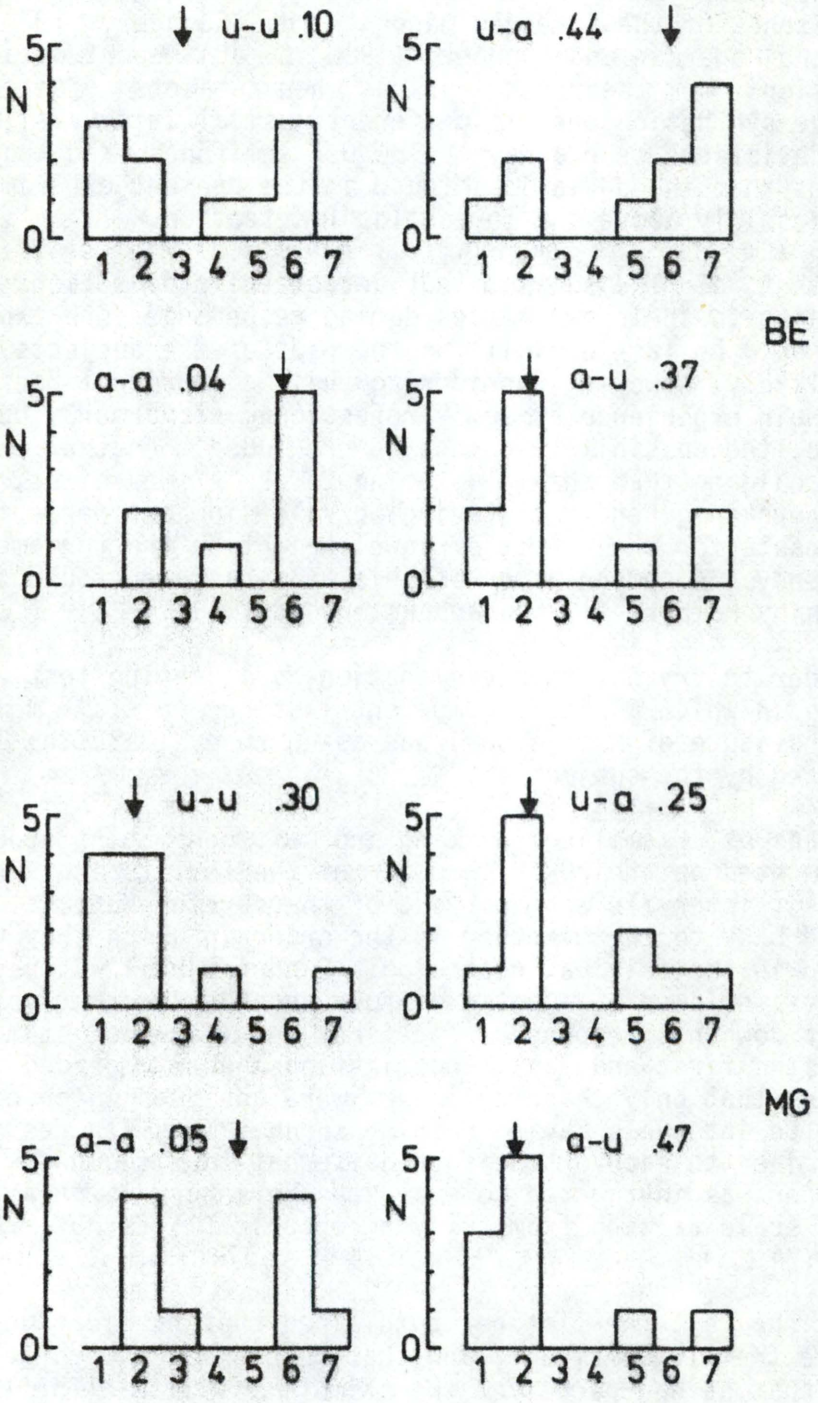


Figure 3

Distributions of responses on a seven-point scale to each of the four test sequences. 1 is "fully acceptable" and 7 is "unacceptable". In each panel in the figure the median of the response distribution is indicated by an arrow and the size of the deviation from the theoretical value of 3 semitones is stated.

for the effect of vowel height on Fo, the location of the responses on the scale should be expected to vary as a function of the deviation of the produced intervals from the theoretical value of 3 semitones.

The distribution of responses to each of the four types of test interval is presented in figure 3. The arrows in the figure indicate the medians of the response distributions, and for the sake of comparison the measured deviations are given. The response patterns do not seem to reveal any obvious relationship between deviations from theoretical values and the location of the central tendencies of the responses. MG has large deviations in *u-u*, *u-a*, and *a-u* intervals, but all these intervals have received responses at the "acceptable end" of the scale, whereas the *a-a* interval, which shows the smallest deviation, received the greatest number of responses at the "unacceptable end" of the scale. For BE similar discrepancies are revealed between response patterns and size of the deviations of the produced intervals from theoretical values. These findings speak in favour of the hypothesis that listeners perceptually compensate for inherent Fo level differences in solo-singing.

In order to obtain a quantitative evaluation of the question of a relation between produced intervals and response distributions the data for each subject were submitted to a χ^2-test. The χ^2-test, however, requires the expected frequency in each cell to be 5 or more. Since this requirement could not be met by the arrangement of the present data in four stimulus conditions, seven response categories and 40 responses, it was neccessary to regroup the data as follows: The responses were dichotomized into one group comprising the categories 1 through 3, and one comprising the categories 4 through 7. In regrouping the stimulus conditions the two test sequences having the smallest deviations from the theoretical value of 3 semitones were combined into one group, and the two having the largest into another. The resulting 2x2 matrices are shown in table IV which also gives the values obtained for χ^2 together with the probabilities of the occurrence of these values under H_0, i.e. under the hypothesis of no association between stimulus and response.

Table IV

Matrices of dichotomized stimulus categories versus dichotomized response categories. For further explanation see text.

BE

	response cat. 1-3	response cat. 4-7	
large dev. (*a-u*)+(*u-a*)	10	10	20
small dev. (*a-a*)+(*u-u*)	7	13	20
	17	23	40

$\chi^2=0.921$

$0.40>p>0.30$

MG

	response cat. 1-3	response cat. 4-7	
large dev. (*u-u*)+(*a-u*)	16	4	20
small dev. (*a-a*)+(*u-a*)	11	9	20
	27	13	40

$\chi^2=2.849$

$0.10>p>0.05$

For BE the effect of deviation size on the response distributions are clearly non-significant. For MG, however, there is a significant effect at the 10 percent level, but it should be noted that the effect goes in the opposite direction of what is to be expected, large deviations having been judged acceptable in more cases than small, and small deviations having been judged unacceptable in more cases than large ones.

Thus, if it can be assumed that the stimuli have been listened to as music (cf. the complaints of the listener mentioned above), it is justified to conclude that the effect of vowel height on Fo in solo-singing, the existence of which has been clearly demonstrated, and which gives rise to gross distortions of the intended musical intervals, seems to be perceptually compensated for by listeners to solo-singing and by singers listening to their own voices.

IV. NOTE

1. This represents the more distinct pronunciation used in singing. In speech the rendering is [ˈsd̥awl̩sʌnə ··· fʌˈg̊ʰɒːd̥əs].

REFERENCES

Di Cristo, A. and Chafcouloff, M. 1977: "Les faits microprosodiques du français: voyelles, consonnes, coarticulation", *Actes des 8èmes Journées d'Etudes du Groupe de la Communication Parlée,* Aix-en-Provence, p. 147-158

Hagerman, B. and Sundberg, J. 1980: "Fundamental frequency adjustment in barbershop singing", *Speech Transm. Lab., Quart. Prog. and Status Rep., Royal Inst. Techn. Stockholm 1,* p. 28-42

Hirst, D.J., Di Cristo, A. and Nishinuma, Y. 1979: "The estimation of intrinsic Fo: a comparative study", *Proc. 9th Int. Congr. Phon. Sc., vol. 1,* p. 381

Hombert, J.-M. 1976a: "Consonant types, vowel height and tone in Yoruba", *Working Papers in Phonetics, UCLA 33,* p. 40-54

Hombert, J.-M. 1976b: "Development of tones from vowel height", *Working Papers in Phonetics, UCLA 33,* p. 55-66

Johansson, I. 1976: *Inherenta grundtonsfrekvenser hos svenska vokaler och deras inflytande på satsintonationen, Stadsmål i övre Norrland 9,* (Umeå University)

Ladefoged, P. 1964: *A phonetic study of West African languages, an auditory-instrumental survey,* (Cambridge)

Lehiste, I. 1970: *Suprasegmentals,* (M.I.T. Press)

Lehiste, I. and Peterson, G.E. 1961: "Some basic considerations in the analysis of intonation", *J. Acoust. Soc. Am. 33*, p. 419-425

Neweklowsky, G. 1975: "Specifische Dauer und specifische Tonhöhe der Vokale", *Phonetica 32*, p. 38-60

Petersen, N. Reinholt 1978: "Intrinsic fundamental frequency of Danish vowels", *J. Phonetics 6*, p. 177-189

Petersen, N. Reinholt 1979: "Variation in inherent Fo level differences between vowels as a function of position in the utterance and in the stress group", *Ann. Rep. Inst. Phon. Univ. Cph. 13*, p. 27-57

Petersen, N. Reinholt 1980: "Coarticulation of inherent fundamental frequency levels between syllables", *Ann. Rep. Inst. Phon. Univ. Cph. 14*, p. 317-354

Peterson, G.E. and Barney, H.L. 1952: "Control methods used in a study of the vowels", *J. Acoust. Soc. Am. 24*, p. 118-127

Rosenvold, E. 1981: "The role of intrinsic Fo and duration in the perception of stress", *Ann. Rep. Inst. Phon. Univ. Cph. 15*, p. 147-166

Rossi, M. 1971: "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole", *Phonetica 23*, p. 1-33

Rossi, M. 1978: "La perception des glissandos descendants dans les contours prosodiques", *Phonetica 35*, p. 11-40

Sundberg, J. 1978: *Musikens ljudlära*, (Stockholm)