EFFECT OF VOWEL QUALITY ON PITCH PERCEPTION

I.Raimo, O.Aaltonen, Å.Hellström* and E.Vilkman

Phonetics, University of Turku, Finland
***Psychology, Stockholm University, Sweden**

ABSTRACT

The intrinsic F0 (IF0) phenomenon was hypothesized to cause expectations of different pitches for different vowels. Listeners judged for pairs of synthetic vowels which members had the higher pitch. The judgments were clearly based on vowel quality; there were also heavy effects of the time-order. The results can be explained by vowel-specific expected F0. This supports the view that intrinsic F0 of vowels is centrally controlled.

1. INTRODUCTION

Many explanations have been given for the intrinsic F0 (IF0) of vowels: under comparable circumstances, the high vowels [u, i] are produced with a higher F0 than the low vowels [a, æ], cf. [7]. According to the acoustic coupling hypothesis, F0 is affected by vowel-specific changes in vocal tract acoustics. Mechanical coupling hypotheses suggest that IF0 depends on physiological interaction between the articulatory and the phonatory systems. From the results of our own acoustical and physiological experiments [8] we have concluded that none of these hypotheses is entirely satisfactory.

It has recently been suggested that IF0 is not merely a passive reflection of the biological characteristics of speech mechanisms, but centrally controlled [4]. This suggestion is supported by preserved IF0 in the esophageal speech of laryngectomized patients [7]. If IF0 is learned and automatized in language acquisition, then listeners may have different expectations for vowel pitches, which in turn may cause pitch perception to depend on vowel quality. The present study tested this hypothesis experimentally.

2. PROCEDURE

The Finnish low vowels [a] and [æ], and the high vowels [u] and [i] were synthesized using the cataract type synthesis. All vowels had the same input amplitude configurations and the following formant structures (Hz) and method-dependent relative amplitudes:

	F1	F2	F3	dB
a	700	1100	2500	+5-6
æ	650	1700	2500	+3-4
u	300	600	2500	+1-2
i	300	2250	2850	+0-1

The durations of all vowels were 23 cs. Five F0 levels (1-5) were used. For Level 1, F0 was 104 Hz initially, reaching 114 Hz after 9 cs and then declining to 84 Hz. Levels 2, 3, 4, and 5 deviated at all points from level 1 by +3, +6, +9, and +12 Hz. Levels 2 and 4 were used as the first members and all five levels as the second members of pairs. Thus the largest differences within the pairs were 9 Hz (more than a semitone). 1.1 s intervened between the members of each pair and 3.6 s between the pairs. All possible vowel pairs, 160 vowel pairs in all, were recorded in random order and presented to listeners who had to judge for each pair which vowel had the higher pitch or if they had equal pitch. For each vowel combination, 20% of the pairs had equal F0, in 40% the first vowel was higher, in 40% the second. There were two groups of Finnish-speaking listeners: 32 university language students (4 men and 28 women, mean age 22) and 66 members (29 men, 37 women, mean age 38) of a well regarded amateur symphonic choir, who were thought to be more than normally trained in discriminating vowel pitch.

3. RESULTS

In terms of correct judgments, the choir performed slightly better than the students: For the eight pairs in which equal-quality vowels were juxtaposed with the maximal (9 Hz) F0 differences, the choir made 64% and the students 51% correct judgments. Below, the percentages of selections as the higher are given for the vowels in each combination (mean over the two time orders). The percentages of "equal" judgments are given in parentheses. (For the students, each row represents 640, for the the choir, 1320 judgments.)

	Students	Choir
æ - u	79- 9 (12)	64-24 (13)
a - u	75- 8 (17)	56-27 (17)
i - u	68- 9 (23)	56-27 (18)
æ - i	49-24 (27)	45-37 (18)
æ - a	40-31 (29)	53-28 (19)
a - i	43-35 (22)	39-42 (19)

Thus, in both groups, [æ] was heard as higher and [u] as lower compared with any other vowel.

For all vowel combinations, the groups made the time-order dependent judgments:

	Students	Choir
V1-V2	28-41 (31)	30-45 (25)

Thus, the second vowel was heard as the higher more often than the first. This is called (see [2]) a negative time-order error (TOE). There were, however, clear differences between the vowels, as well as between the groups:

	Students	Choir
æ - æ	12-35 (54)	17-37 (46)
a - a	11-34 (55)	17-33 (50)
i - i	23-18 (59)	17-34 (49)
u - u	25- 8 (66)	23-26 (50)

Thus, for both groups, the higher in pitch a vowel was judged when compared with other vowels, the stronger was its tendency to be judged as higher when second in a pair and compared with itself. For [u] with the students, the negative TOE was reversed to positive.

For describing and explaining TOEs (which are found for many kinds of stimuli including tonal loudness and pitch), Hellström [2] developed a general model for stimulus comparisons. According to this model, the two pitches are not compared directly; their mean judged difference (as measured e.g. by D%, the difference between the percentages of "first higher" and "second higher" judgments) is proportional to the difference between two compounds. In the present case, each compound corresponds to one of the vowels in the pair, and is a weighted sum of its actual pitch, with relative weight *s*, and its expected pitch (its adaptation level, AL), with relative weight *1-s* (NB: *s* may be either >1 or <1).

Assuming identical, linear relations between F0 and pitch for all vowels in the small F0 range used, the expected F0 (AL) (in Hz relative to the mean F0, 106 Hz) and *s* values (up to a scale constant, *k*) could be estimated by multiple linear regression of *D%* for each pair on its F0 values (R was .954 for the students, .892 for the choir):

	Students			Choir		
	ks1	*ks2*	*AL*	*ks1*	*ks2*	*AL*
a	4.5	5.4	-40	3.1	5.2	-10
æ	4.0	5.7	-16	4.0	5.7	-13
i	3.2	5.2	-4	3.6	4.9	-13
u	0.8	3.6	+7	4.1	5.4	-2
Mean	3.1	5.0	-14	3.7	5.3	-9

For both groups and all vowels, the vowel's weight when first in a pair (*s1*) was higher than its weight when second (*s2*). For the students, AL (expected F0) was highest for **[u]** and lowest for **[a]**. For the choir, **[u]** had a higher AL than the other vowels.

4. DISCUSSION

The purpose of our study was to test if perception of vowel pitch depends on vowel quality, because in articulation pitch varies with quality. The result was clearly positive: other things being equal, the vowels **[æ]** and **[a]**, which have low IF0s, were heard as highest in pitch. The results cannot be explained by the amplitude differences, as we found no clear relation between amplitude and experienced or expected pitch. The distribution of energy in the vowel spectra might be of greater importance.

However, our results indicate that the most important factor for vowel-specific pitch is expected F0, which is higher for the high than the low vowels. By reference to Hellström's [2; 3] model, the different AL and *s* values explain both all vowel-specific pitch differences and the TOEs in our data. Besides, pitch discriminability (indicated by *ks*) in the student group was much poorer for **[u]** than for the other vowels. The results thus clearly support our hypothesis that because in articulation F0 varies between vowels, perceived vowel pitch depends on vowel quality.

It is interesting to note that in another recent study [9] the vowel **[u]** was produced with higher subglottal pressure than the other vowels **[i æ a]**. Thus the vowel **[u]** seemed to be different from other vowels also in terms of the physiology of speech production. These effects may have been emphasized by the especially dark quality (low F2) of Finnish **[u]**. The question whether our findings share a common basis, e.g. higher respiratory effort in production perceived as stress, remains open for speculation.

Our study also supports the view that IF0 is an inherent property of vowel prototypes in the brain; even trained singers cannot eliminate its effect on perceived pitch. In vowel pitch perception, then, the IF0 behaves somewhat like formants, which are not perceived separately, but as integrated characteristics of vowel quality. IF0, nevertheless, has no phonologically distinctive function in languages [5]. Our results are in accordance with those speech perception theories which maintain that speech perception is based on articulatory rather than acoustic parameters of speech sounds; see [6].

5. ACKNOWLEDGMENTS

We express our deepest gratitude to the Speech Transmission Laboratories of the Royal Institute of Technology in Stockholm and especially to Rolf Carlson and Björn Granström for their invaluable help in producing the synthetic vowels.

6. REFERENCES

[1] GANDOUR, J. and WEINBERG, B. (1980), "On the relationship between vowel height and fundamental frequency: evidence from esophageal speech", *Phonetica, 37,* 344-354.

[2] HELLSTRÖM, Å. (1985), "The time-order error and its relatives: mirrors of cognitive processes in comparing", *Psychological Bulletin, 97,* 35-61.

[3] HELLSTRÖM, Å. (1989), "What happens when we compare two successive stimuli?", in G. Ljunggren & S. Dornic (Eds.), *Psychophysics in Action,* Berlin: Springer.

[4] HONDA, K. & FUJIMURA, O. (forthcoming), "Phonological vs. biological explanations - intrinsic vowel pitch and phrasal declination", *Vocal Fold Physiology,* Proceedings of the 6th Vocal Fold Physiology Conference, Stockholm 1989.

[5] LEHISTE, I. (1970), *"Suprasegmentals"*, Cambridge: The MIT Press.

[6] LIBERMAN, A. & MATTINGLY, I. (1985), "The motor theory of speech perception revised", *Cognition, 21,* 1-36.

[7] SAPIR, S. (1989), "The intrinsic pitch of vowels: theoretical, physiological, and clinical considerations", *Journal of Voice, 3* , 44-51.

[8] VILKMAN, E., AALTONEN, O., LAINE, U. & RAIMO, I. (forthcoming), "Intrinsic pitch of vowels - a complicated problem with an obvious solution?", *Vocal Fold Physiology,* Proceedings of the 6th Vocal Fold Physiology Conference, Stockholm 1989.

[9] VILKMAN, E., RAIMO, I. & AALTONEN, O. (1991), "Is subglottal pressure a contributing factor to the intrinsic F0 phenomenon?", in this publication.