(ARIPUC 12, 1978)

FORM FACTORS FOR POWER SPECTRA OF VOWEL NUCLEI

Ulf J. Stålhammar[1]

Abstract: The spectrum of the vowel nucleus is, within the present framework, regarded in terms of 'Form Factor Elements'. One such element is reflected in the structure of formant clusters and may be expressed in terms of bandwidth. It is shown that in two-formant synthesis a vowel can change its identity from /i/ to /e/ (or /y/ to /ø/ with a decreased pole span) simply by an increase of the higher pole bandwidth. Prior to this discovery, and indicating its potential, a four-formant synthesis experiment was conducted. In this part of the investigation the possibility of constructing /e/ vowel spectrum envelopes containing either higher, equal or lower spectral centres of gravity, as compared to envelopes generating the auditory impression of /i/, is demonstrated. N.B. in the process of generating the power spectra for /i/ and /e/, the frequency parameters F_4, F_1 and F_0 were held constant. The spectrum balance was achieved either by means of only frequency adjustments of F2 and F3 or, with these frequencies "frozen", by amplitude modifications of F2, F3 and F4. The two experiments emanate from an empirically found paradoxical relationship for one female voice between such parameters as Tongue Height plus Fronting versus Second Formant Prime, (F'_2) versus Centre of Gravity for the Spectral Components above F_1, (denoted here as G'_2).

The psychoacoustic evidence obtained focuses the attention towards the development of an "excitation area" theory of perception based on form factors.

---

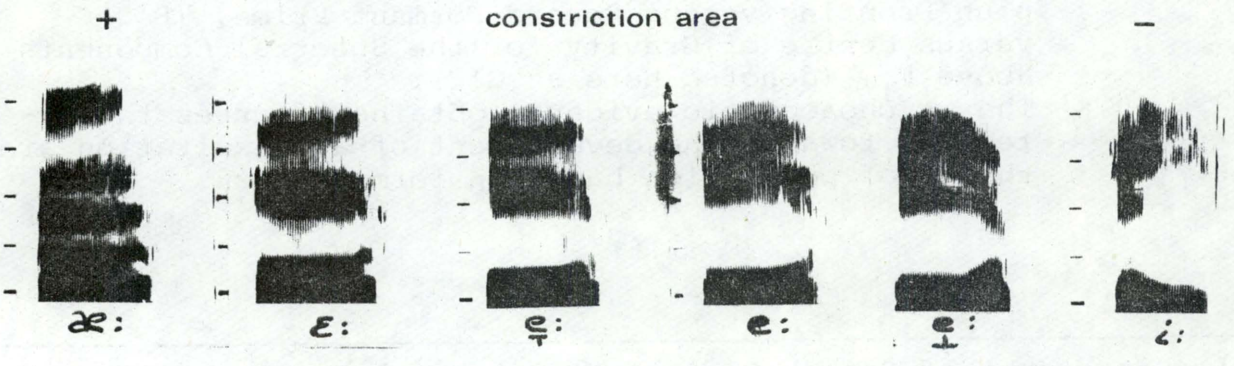
1) Department of Speech Communication, Royal Institute of Technology (KTH), S 100 44 Stockholm 70, Sweden.

Introduction

The work to be described forms part of a research program dealing with vowel nuclei in production, synthesis and perception. In the present paper the latter two issues are concerned. The paper provides some background philosophy to and a summary presentation of the talk given at the "Svensk-Dansk Fonetiker Seminar 1977" under the title: "'Formfaktorn' och dess betydelse för vokalperception" - "The 'Form Factor' and its Significance in Vowel Perception". The research is carried on in support of perception and automatic recognition studies. The aim is to investigate how modified spectra can nevertheless be perceptually identified as one and the same vowel, and ultimately to express each vowel in terms of one parameter only, if possible, and to use the same parameter for male and female voices.

Background

Space geometric properties of the human vowel tract allow the second formant, F_2, to be positively correlated with tongue height/fronting until the extreme prepalatal region is reached. Within this latter region the two parameters will be reversely correlated as demonstrated in the spectrogram sequence of figure 1. This fact focuses on problems referring to automatic speech recognition (ASR), decision logic and human perception.



1.1 1.2 1.3 1.4 1.5 1.6

Figure 1

Effect of increasing tongue height/fronting upon formant patterns. - Speaker US.

Figure 2 contains a vowel plot in terms of the F_1/F_2 spaces of adult males, adult females and five to eight-year-old children of both sexes (Stålhammar, 1971). Each data point represents the mean over ten subjects. Attention should be given to the relation between the vowels /i/ and /e/. The figure shows the relationship $F_{2i} < F_{2e}$ for all three speaker categories. Figure 3 demonstrates that the relation is manifested, again as a mean, also in different contexts. Context 1 represents vowels in isolation, context 2 vowels embedded in C-C environment and context 4 vowels in fluent speech (Stålhammar, Karlsson, Fant, 1973). Obviously, it is clear that the i/e relationship cannot be adequately described in terms of the F2 feature alone. This is a drawback for ASR systems having as vowel identifier the distance F_2-F_1. From a perceptual point of view, however, /i/ is perceived as [+high] in relation to /e/. An examination of figure 1.6 might provide the answer, the effect of the weak F2 amplitude due to its large

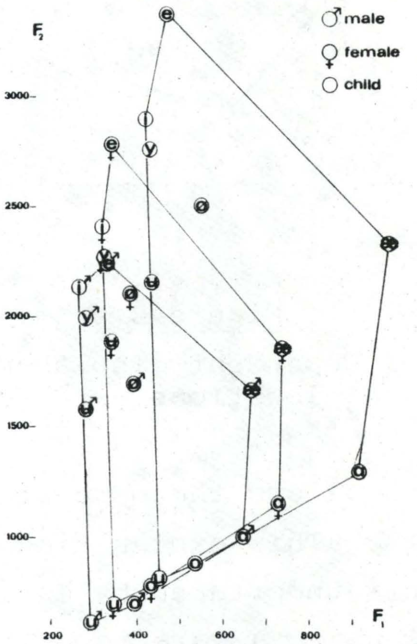


Figure 2

Sex dependent displacement of the F_1/F_2 locations.

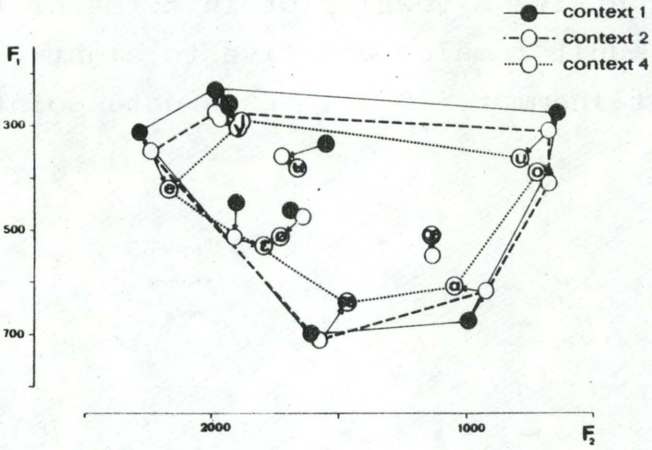


Figure 3

Context dependent displacement of the F_1/F_2 locations.

distance from F_3, which allows the perceptual focus to be oriented towards higher formants. This explanation is also supported by the matching experiments undertaken by Carlson, Fant, Granström (1975), in which test subjects were reported to match closer to F_3 than to F_2 for a synthetic /i/ vowel.

Consequently, the use of a new parameter (G'_2-F_1) in place of (F_2-F_1) should bring about an improvement from an ASR point of view; G'_2 represents the centre of gravity of spectral components above F_1. Under the described conditions which usually prevail in reality, a positive correlation between the parameters Tongue Height/Fronting and G'_2 is obtained.

While the newly defined parameter (G'_2-F_1) represents a definite improvement over the previously used (F_2-F_1) parameter, a contradictory case of a female voice has nevertheless been found.

Female case

Two vowel spectra originating from a female test subject, AKS, are shown in figure 4 in which the power spectra pertaining to the instants identified by the sampling arrows are presented. It is noticed that F_{1i}=285 Hz, F_{2i}=2600 Hz, F_{3i}=3670 Hz, F_{4i}= 5190 Hz; similarly, F_{1e}=305 Hz, F_{2e}=3030 Hz, F_{3e}=3700 Hz, F_{4e}=

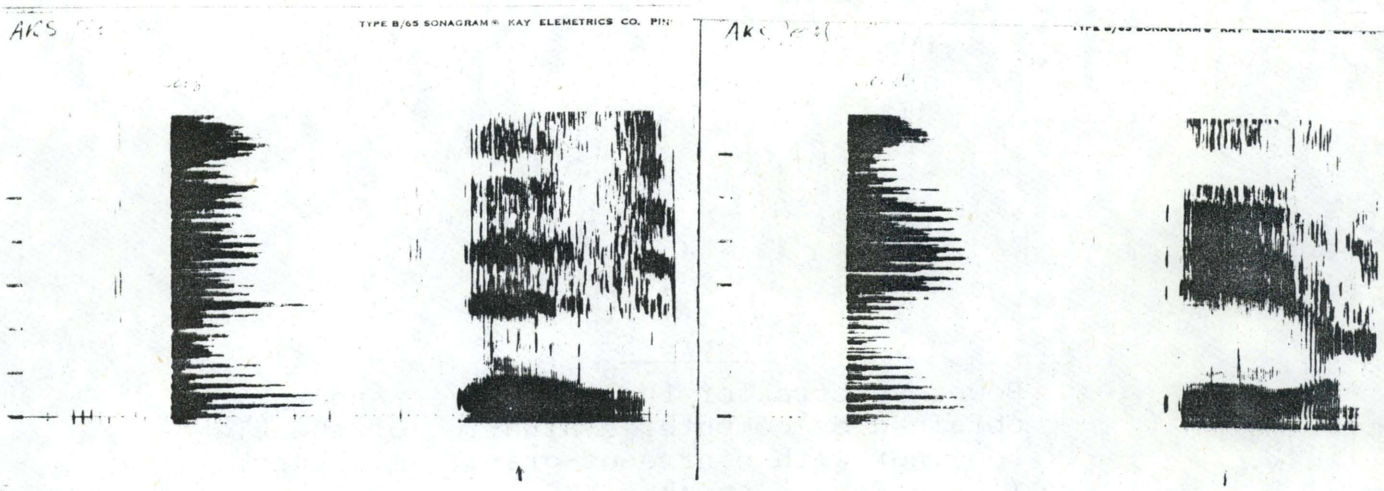


Figure 4

Visible speech and power spectrum cross-section patterns for vowels /i/ and /e/. - Female speaker AKS.

4795 Hz. Again it is observed that $F_{2i} < F_{2e}$ at the sampling locations, i.e. Δ = 430 Hz. More importantly, an unusually high amplitude of F_{2i} is observed. Based on visual examination it might be hypothesized that the G'_{2i} could be even lower than the G'_{2e} within the steady state portions of the two sounds. In order to scrutinize this hypothesis quantitatively, the frequency, $f_{G'}$, of the spectral centre of gravity was computed:

$$f_G = \frac{\sum f \cdot A_f}{\sum A_f} \quad ; \qquad (1)$$

where A_f is the linear amplitude of the component of frequency f.

The cross-section patterns presented in figure 5 for the two vowels /i/ and /e/ were obtained by means of a computer operating at a sampling rate of 100 Hz.

Two separate evaluations of G'_2 were carried out using the frequency ranges 0.8-4.8 kHz (test 1) and 0.8-6.8 kHz (test 2), respectively for the evaluation of the spectral centre of gravity. Test 1 was undertaken mainly to focus on the contribution of F_2 and F_3 to the spectral centre of gravity, and test 2, with the high frequency limit increased to 6.8 kHz, to include the influence

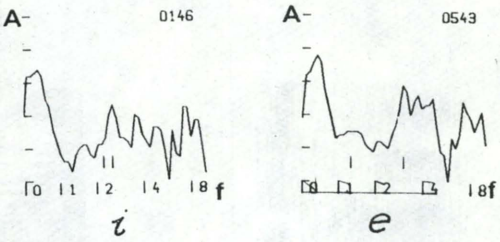


Figure 5

Power spectra for the vowels /i/ and /e/, obtained by computer evaluation of the signal together with centre-of-gravity markings, ('). - Female speaker AKS.

of higher spectral components upon the spectral centre of gravity. The following numerical values of G'_2 were obtained:

Test 1: $G'_{2i} = 2137$ Hz

$G'_{2e} = 2687$ Hz

$\Delta = G'_{2i} - G'_{2e} \approx -500$ Hz

Test 2: $G'_{2i} = 2375$ Hz

$G'_{2e} = 2825$ Hz

$\Delta = G'_{2i} - G'_{2e} \approx -500$ Hz

Although for test 2 the upper cut-off frequency is increased by 2 kHz the increase in G'_2 is about ten times smaller, i.e. only approximately 200 Hz; Δ remains essentially constant. Although usually /i/ is associated with a higher spectral centre of gravity, $G'_{2i} < G'_{2e}$ in this case.

Comparison with previous equivalent formant methods

An F'_2 calculation was carried out using the 1975 formula due to Fant (Carlson, Fant, Granström 1975):

$$F'_2 = \frac{F_2 + c(F_3F_4)^{1/2}}{1 + c} \qquad (2)$$

$$c = \left(\frac{F_1}{500}\right)^2 \cdot \left(\frac{F_2-F_1}{F_4-F_3}\right)^4 \cdot \left(\frac{F_3-F_2}{F_3-F_1}\right)^2$$

The following numerical values were obtained:

$$F'_{2i} = 2860 \text{ Hz}$$

$$F'_{2e} = 3450 \text{ Hz}$$

$$\Delta = F'_{2i} - F'_{2e} \approx -0.6 \text{ kHz}$$

The formula was later revised as follows (Bladon, Fant 1978):

$$F'_2 = \frac{F_2 + c^2(F_3F_4)^{1/2}}{1 + c^2} \qquad (3)$$

$$c = K(f)\frac{A_{34}}{A_2}$$

where A_{34} is the vocal tract transfer function in the valley between F3 and F4 at the frequency $F_{34} = (F_3F_4)^{1/2}$ and A_2 is the transfer function at the second formant peak, F2. The factor K(f) in the weighting function is intended to include the additional preemphasis originating from source, radiation and higher pole corrections and in addition a correction for differences in equal loudness levels.

With the new version the following numerical values were obtained:

$$F'_{2i} = 2725 \text{ Hz}$$

$$F'_{2e} = 3460 \text{ Hz}$$

$$\Delta = F'_{2i} - F'_{2e} \approx -0.7 \text{ kHz}$$

N.B. again with the /i/ as the lower counterpart.

Although the previous formulæ are adequate to represent most vowels, the contradictory example indicates that some additional modifications will be required to cover all vowels.

It should be observed here that for speaker AKS, given the F_n data presented above, F'_{2i} = 2860 Hz and F'_{2e} = 3450 Hz. In order to transform the F'_{2i} value to coincide with the F'_{2e} value by means of shifting only one formant at a time, the F_2 has to be shifted up to 3422 Hz, the F_3 up to 4031 Hz, or the F_4 down to 3283 Hz. It is understood that such a transformation of F_4 is not permissible if the resulting transformed value is smaller than F_3. F'_{2e} cannot in this case be reached by a shift in F_1 only. Similarly, in a CVC environment, for the observed female formant set (Stålhammar et al. 1973) for the vowel /I/, (F_1 350 Hz, F_2 2600 Hz, F_3 3075 Hz, F_4 4000 Hz) and the male set, (F_1 325 Hz, F_2 2315 Hz, F_3 2915 Hz, F_4 3400 Hz) it is noticed that each individual female formant occupies a higher frequency position, yet the female F'_2 = 2910 Hz and the male F'_2 = 3035 Hz, i.e. $F'_{2fe} < F'_{2ma}$. In order to transform the F'_{2fe} value to coincide with the F'_{2ma}, F_1 has to be shifted up to 505 Hz, F_2 to 3030 Hz, F_3 to 3140 Hz, or F_4 has to be lowered to 3845 Hz. An increase of F_4 results in a downshift of F'_2.

Since it is now possible to state that neither the (F_2-F_1) parameter nor the (G'_2-F_1) parameter nor the (F'_2-F_1) parameter cover all possible cases, there must be another, as yet unknown, factor of importance. A preliminary attempt to determine the nature of such a factor will be done by means of (1) a four-formant synthesis experiment and (2) by means of a two-formant synthesis experiment. However, for the sake of completeness, the rôle of formant transitions should first be considered.

The Vowel - a Spectral Chameleon

The low F_1 vowels in Swedish, /i/, /y/, /ʉ/ and /u/, are mostly characterized by pronounced F_n transitions when uttered in isolation (see e.g. the /i/ and /ʉ/ vowels in figure 6 and Stålhammar, Karlsson 1972). These vowels tend to be diphthongized towards a target of extreme tongue-palate closure for /i/ and /y/ and a target of labial closure following /ʉ/ and /u/. In order

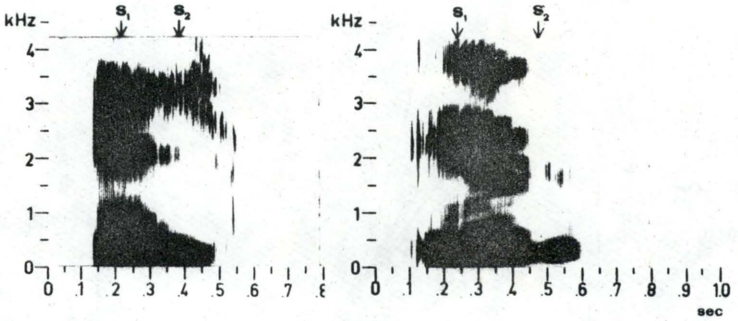


Figure 6

Spectra of an /i/ vowel, left, and an /ʉ/ vowel, right, together with indications of sample 1, (S_1), and sample 2, (S_2), locations. - Speaker US.

to obtain a more adequate representation of F_n for this category of vowels, the location of two samples, S_1 and S_2, are defined. S_1 is located at approximately 20% of the total duration from voice onset and the location of S_2 is derived from an articulatory target as follows: for /i/ and /y/ it is where the F_3 transition changes direction from positive to negative as in figure 6 or at other target criteria, and in /ʉ/ and /u/ where the F_2 reaches steady state. The values of F_1, F_2 and F_3 at the instant S_1 are redefined as being equal to zero; thus the endpoints which correspond to S_2 reflect the difference, ΔF_n, between the final target value (at S_2) and the initial value (at S_1) of the F_n's. In figure 7 these ΔF's are plotted for F_1, F_2 and F_3. The encircled digits indicate test subjects. A close examination of the plot reveals that the F-pattern shows great variability as a function of the two sampling locations. For the vowel /i/ the F_1 transition is negative and simultaneously associated with positive F_2 and F_3 transitions for subjects 3 and 7. Similarly, for subjects 0 and 4, F_1, F_2 and F_3 all show negative transitions. For subjects 5 and 8, F_1 is negative, F_2 constant and F_3 positive. For subject 2, F_1 is constant, F_2, F_3 negative. For subject 1, F_1 is also constant, F_2 negative; however, F_3 is positive. For subjects 6 and 9, F_1 and F_2 are negative, while F_3 is positive.

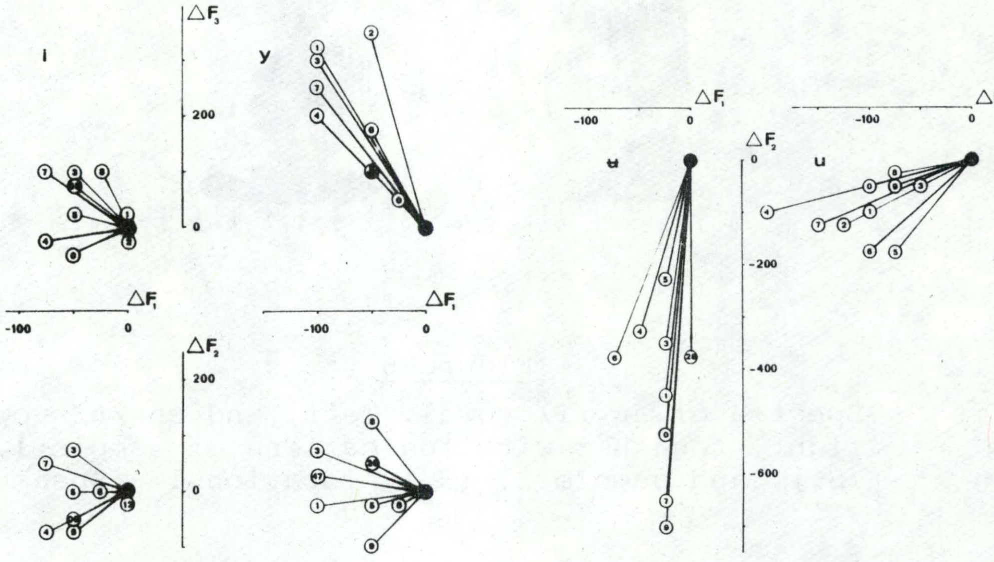


Figure 7

ΔF's for formant transitions between the instants S_1 and S_2 for various vowels. ● = normalized location for S_1. o = ΔF's, obtained for various subjects at S_2.

For the vowel /y/ all subjects display negative F_1 transitions combined with positive F_3 transitions, while F_2 shows a less consistent pattern. The vowels /ʉ/ and /u/ have clear negative transitions for both the lower formants. The negative $F_{2\text{ʉ}}$ transition for subject 9 is about 700 Hz.

In view of the inconsistent transition patterns for F_1, F_2 and F_3, it is imperative to search for more consistent parameters such as, e.g., the form factors described in this paper.

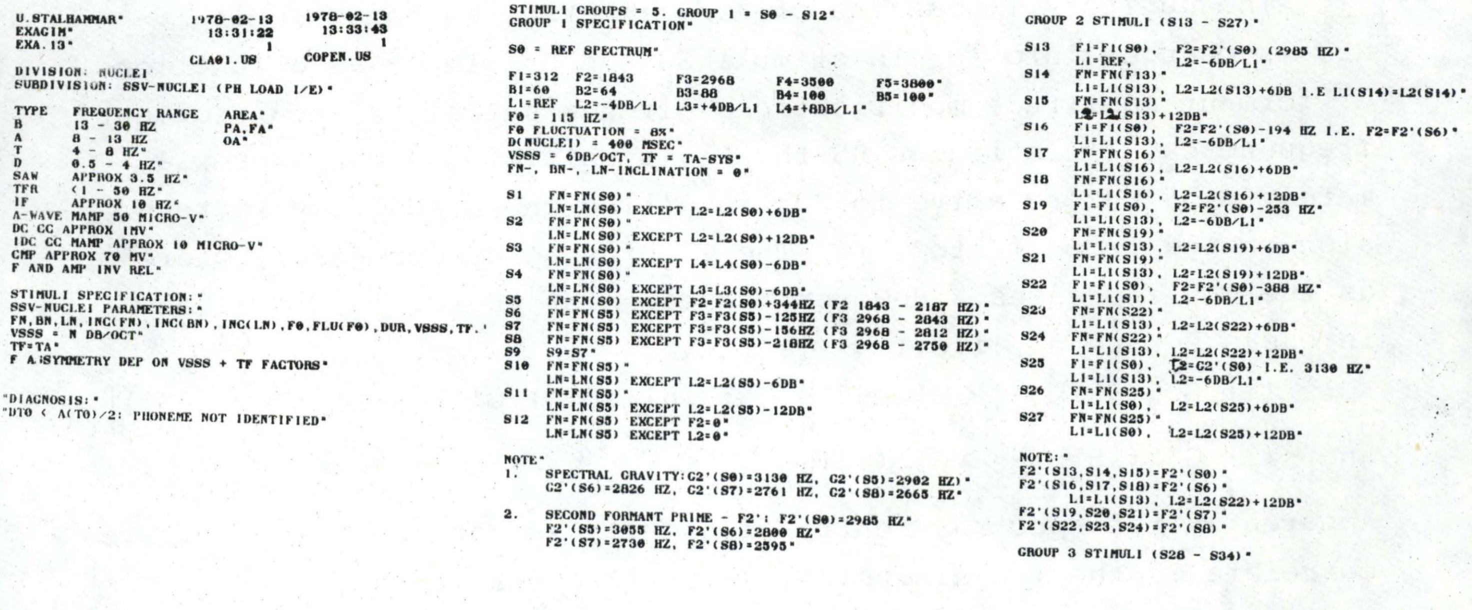
In order to eliminate the possibility of any spurious effects on the present study, the signal obtained from speaker AKS was examined in terms of the following questions:

1. Are the vowels produced by speaker AKS adequate and correct in all respects?
2. What about the higher formant amplitude decreases in the later part of the /i/ vowel and the negative higher formant transitions in the later part of the /e/ vowel?

With respect to question (1) a panel of 30 listeners judged the utterances of the speaker as being entirely proper and showing no spurious effects. With respect to question (2) the problem of amplitude decreases/transitions was avoided by means of retaining only the steady state portions of the vowels, i.e. approximately 50% of the total durations. After the truncations were made, the steady states were repeated to reestablish the original durations. According to the same panel of listeners the reconstituted signals retained a sufficient phonemic identity (/i/ and /e/, respectively).

Four-formant case

In order to retain control over formant amplitudes, the four-formant stimuli were produced by a computer simulation of parallel-formant synthesis. Input data forming the reference construct, SO, are presented in figure 8. The data are typical of a male vowel /i/ with a slight modification in F_1 so as to obtain an intermediate value between F_{1i} and F_{1e}. This weighted value was adopted



```
U.STALHAMMAR"              1978-02-13    1978-02-13
EXAGIM"                      13:31:22      13:33:43
EXA.13"                             1             1
                             CLA01.US      COPEN.US
DIVISION: NUCLEI
SUBDIVISION: SSV-NUCLEI (PH LOAD I/E)"

TYPE   FREQUENCY RANGE    AREA"
B      13 - 30 HZ         PA,FA"
A      8 - 13 HZ          OA"
T      4 - 8 HZ"
D      0.5 - 4 HZ"
SAW    APPROX 3.5 HZ"
TFR    <1 - 50 HZ"
IF     APPROX 10 HZ"
A-WAVE MAMP 50 MICRO-V"
DC CC APPROX 1MV"
IDC CC MAMP APPROX 10 MICRO-V"
CMP APPROX 70 MV"
F AND AMP INV REL"

STIMULI SPECIFICATION:"
SSV-NUCLEI PARAMETERS:"
FN,BN,LN,INC(FN),INC(BN),INC(LN),F0,FLU(F0),DUR,VSSS,TF.'
VSSS = N DB/OCT"
TF=TA"
F A;SYMMETRY DEP ON VSSS + TF FACTORS"

"DIAGNOSIS:"
"DT0 < A(T0)/2: PHONEME NOT IDENTIFIED"
```

```
STIMULI GROUPS = 5. GROUP 1 = S0 - S12"
GROUP 1 SPECIFICATION"

S0 = REF SPECTRUM"

F1=312  F2=1843     F3=2968     F4=3500     F5=3800"
B1=60   B2=64       B3=88       B4=100      B5=100"
L1=REF  L2=-4DB/L1  L3=+4DB/L1  L4=+8DB/L1"
F0 = 115 HZ"
F0 FLUCTUATION = 8%"
D(NUCLEI) = 400 MSEC"
VSSS = 6DB/OCT, TF = TA-SYS"
FN-, BN-, LN-INCLINATION = 0"

S1    FN=FN(S0)"
      LN=LN(S0) EXCEPT L2=L2(S0)+6DB"
S2    FN=FN(S0)"
      LN=LN(S0) EXCEPT L2=L2(S0)+12DB"
S3    FN=FN(S0)"
      LN=LN(S0) EXCEPT L4=L4(S0)-6DB"
S4    FN=FN(S0)"
      LN=LN(S0) EXCEPT L3=L3(S0)-6DB"
S5    FN=FN(S0) EXCEPT F2=F2(S0)+344HZ (F2 1843 - 2187 HZ)"
S6    FN=FN(S5) EXCEPT F3=F3(S5)-125HZ (F3 2968 - 2843 HZ)"
S7    FN=FN(S5) EXCEPT F3=F3(S5)-156HZ (F3 2968 - 2812 HZ)"
S8    FN=FN(S5) EXCEPT F3=F3(S5)-218HZ (F3 2968 - 2750 HZ)"
S9    S9=S7"
S10   FN=FN(S5)"
      LN=LN(S5) EXCEPT L2=L2(S5)-6DB"
S11   FN=FN(S5)"
      LN=LN(S5) EXCEPT L2=L2(S5)-12DB"
S12   FN=FN(S5) EXCEPT F2=0"
      LN=LN(S5) EXCEPT L2=0"

NOTE"
1.    SPECTRAL GRAVITY:G2'(S0)=3130 HZ, G2'(S5)=2902 HZ)"
      G2'(S6)=2826 HZ, G2'(S7)=2761 HZ, G2'(S8)=2665 HZ"

2.    SECOND FORMANT PRIME - F2': F2'(S0)=2985 HZ"
      F2'(S5)=3055 HZ, F2'(S6)=2800 HZ"
      F2'(S7)=2730 HZ, F2'(S8)=2595"
```

```
GROUP 2 STIMULI (S13 - S27)"

S13   F1=F1(S0),   F2=F2'(S0) (2985 HZ)"
      L1=REF,      L2=-6DB/L1"
S14   FN=FN(F13)"
      L1=L1(S13), L2=L2(S13)+6DB I.E L1(S14)=L2(S14)"
S15   FN=FN(S13)"
      L2=L2(S13)+12DB"
S16   F1=F1(S0),   F2=F2'(S0)-194 HZ I.E. F2=F2'(S6)"
      L1=L1(S13), L2=-6DB/L1"
S17   FN=FN(S16)"
      L1=L1(S16), L2=L2(S16)+6DB"
S18   FN=FN(S16)"
      L1=L1(S16), L2=L2(S16)+12DB"
S19   F1=F1(S0),   F2=F2'(S0)-253 HZ"
      L1=L1(S13), L2=-6DB/L1"
S20   FN=FN(S19)"
      L1=L1(S13), L2=L2(S19)+6DB"
S21   FN=FN(S19)"
      L1=L1(S13), L2=L2(S19)+12DB"
S22   F1=F1(S0),   F2=F2'(S0)-388 HZ"
      L1=L1(S1),   L2=-6DB/L1"
S23   FN=FN(S22)"
      L1=L1(S13), L2=L2(S22)+6DB"
S24   FN=FN(S22)"
      L1=L1(S13), L2=L2(S22)+12DB"
S25   F1=F1(S0),   F2=G2'(S0) I.E. 3130 HZ"
      L1=L1(S13), L2=-6DB/L1"
S26   FN=FN(S25)"
      L1=L1(S0),   L2=L2(S25)+6DB"
S27   FN=FN(S25)"
      L1=L1(S0),   L2=L2(S25)+12DB"

NOTE:"
F2'(S13,S14,S15)=F2'(S0)"
F2'(S16,S17,S18)=F2'(S6)"
      L1=L1(S13), L2=L2(S22)+12DB"
F2'(S19,S20,S21)=F2'(S7)"
F2'(S22,S23,S24)=F2'(S8)"

GROUP 3 STIMULI (S28 - S34)"
```

Figure 8

Specification of the 4-formant, SO-S12, and the 2-formant, S13-S27, stimuli.

so as to make the stimuli more sensitive to subsequent adjustments of the higher formant frequencies. Based on this reference, adjustments have been made according to figure 8; the resulting preemphasized (+6 dB/octave within 0.2-5 kHz) computer generated envelopes are displayed in figure 9. In addition, loudness envelopes of the same stimuli produced by a HP 8051 Loudness Analyzer are shown. The Analyzer performs a continuous 1/3 octave level analysis of the noise to be measured and computes, by the method indicated by Zwicker (ISO recommendation 532), the loudness S in sones$_G$. In figure 8 FN represents any formant, BN the formant bandwidth, and LN the formant amplitude. 30 subjects judged the quality of the stimuli.

Evaluation

Stimuli S0 and S1 showed a definite codability with /i/, S2 a high degree of codability with /i/, and S3 and S4 were decoded as /e/. Stimuli S5-S9 were classified as /e/. Stimulus S10 was ambiguous, and S11 and S12 showed a definite codability with /i/.

Comments

The auditory impression of /i/, inherent in the stimuli S0-S2, is changed into /e/ in stimuli S3 and S4 simply as a function of formant amplitude modifications given a set of "frozen" formant frequencies. In stimulus S5 the F_2 is shifted upwards approximately 0.35 kHz relative to S0, resulting in an auditory impression change from /i/ to /e/! The G'_2 frequency decreases, whereas the F'_2 increases. However, cases exist where $G'_{2e} > G'_{2i}$ as in cases 2. and 3., i.e.:

(1) $G'_2(S5)\big|_{/e/} < G'_2(S0)\big|_{/i/}$; while (2) $G'_2(S5)\big|_{/e/} > G'_2(S1)\big|_{/i/}$ and (3) $G'_2(S5)\big|_{/e/} > G'_2(S2)\big|_{/i/}$

Apparently the spectral centre of gravity is not the only decisive correlate in the i/e dimension. Since the upshift of F_2 in S5 results in a perceptual shift from /i/ to /e/ relative to S0, it is evident that the positive correlation between F_2 and the

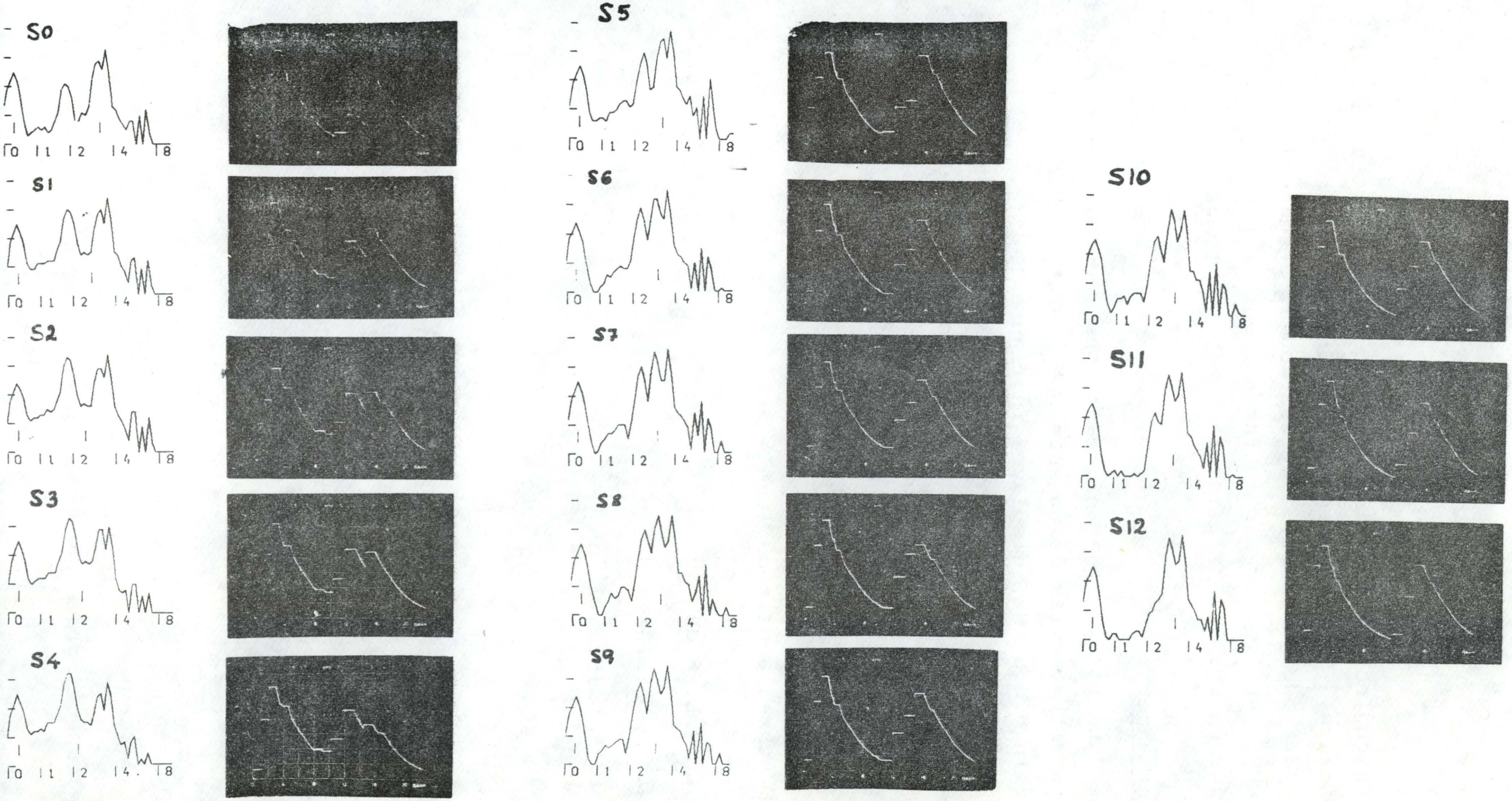


Figure 9

Left: Computer generated spectrum envelopes of the 4-formant synthetic stimuli, SO-S12.
Right: Loudness envelopes of the same stimuli produced by the HP 8051 Loudness Analyzer.

feature [+high] no longer exists when F_2 clusters with F_3 and higher formants. This is also evident in the samples S6-S8 where the /e/ vowel identity is preserved at a successive decrease of F_3. An L_2 decrease in the sequence S10 to S12 relative to S5 brings us back to /i/ (in S12 L_2 is set to zero).

Two-formant case

Input data forming the 2-formant constructs are to be found in figure 8 with the resulting envelopes, preemphasized +6 dB/octave within 0.2-5 kHz, displayed in figure 10.

The 2-formant reference construct S13 is generated with the same F_1 as F_1 of the 4-formant reference construct S0, while the upper formants are reduced to one pole equal to F'_2(S0). Similarly, the higher pole frequency, F_2, of stimulus S16 is derived from F'_2 of the 4-formant construct S6; F_2(S19) = F'_2(S7); F_2(S22) = F'_2(S8). However, F_2(S25) = G'_2(S0). Furthermore, for each F_2 three amplitude levels are used, e.g.:

S13 $L_2 = L_1 - 6$ dB
S14 $L_2 = L_2$(S13)+6 dB
S15 $L_2 = L_2$(S13)+12 dB

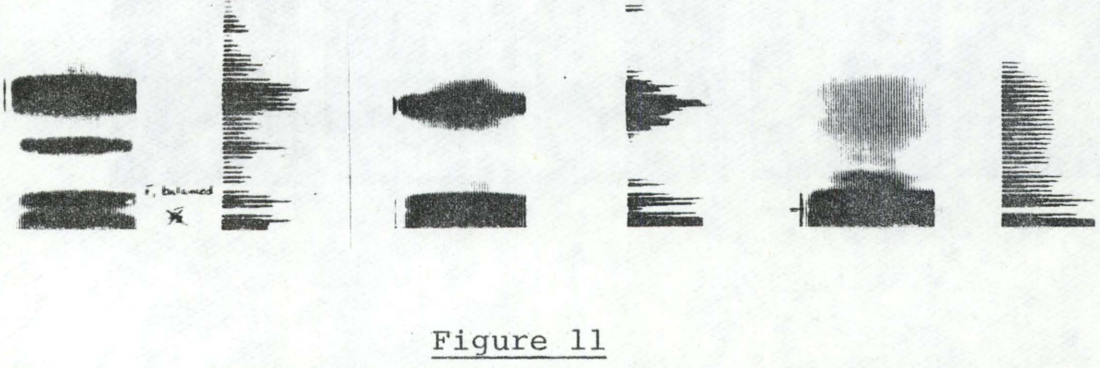


Figure 11

Broadband spectrograms and power spectra for stimuli S0, S13 and S30.

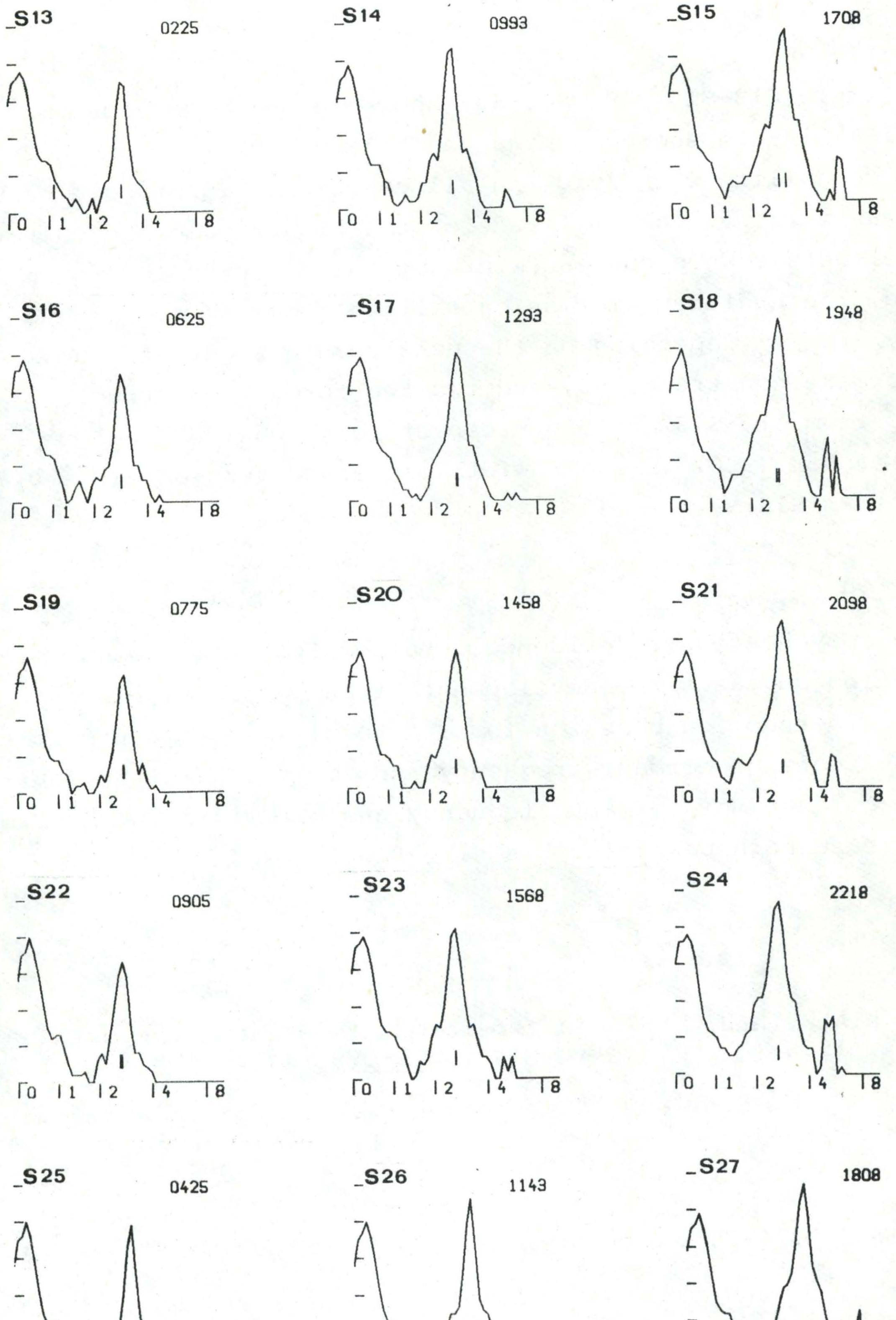


Figure 10

Computer generated spectrum envelopes of the 2-formant synthetic stimuli, S13-S27

Evaluation

Stimuli S13-S18 plus S25-S27 showed a definite codability with /i/, S19-S21 showed a high codability with /i/, and S22-S24 a high codability with /y/. In all cases the higher pole bandwidth was narrow. The amplitude of the higher pole had only minor effects. When the bandwidth of the higher pole was increased, the auditory impression shifted from /i/ to /e/ and from /y/ to /ø/ depending on the pole span. Figure 11 shows a four-formant construct, S0, and two two-formant constructs derived from it, S13 and S30, respectively, where $F_2(S13) = F'_2(S0)$ and $F_2(S30)$ is a bandwidth increased version, ($B = 0.75$ kHz), of $F_2(S13)$.

Conclusion

The psychoacoustic evidence obtained from the present investigation suggests the development of an "excitation area" theory of perception based on form factors. The combined integrated effects of formant frequency and amplitude in the four-formant case and of formant frequency and bandwidth in the two-formant case both point in this direction.

References

Bladon, R.A.W. and G. Fant 1978: "A two-formant model and the cardinal vowels", *STL-QPSR* 1, p. 1-8

Carlson, R., G. Fant and B. Granström 1975: "Two-formant models, pitch and vowel perception", pp. 55-82 in Fant, G. and Tatham, M.A.A. (eds.) *Auditory Analysis and Perception of Speech*, London

Stålhammar, U. 1971: Internal Report, STL-RIT, Stockholm

Stålhammar, U. and I. Karlsson 1972: "A Phonetic Approach to ASR", *1972 Conf. on Speech Communication and Processing* AFCRL Paper C10. IEEE cat.no. 72 CHO 596-7 AE, New York

Stålhammar, U., I. Karlsson and G. Fant 1973: "Contextual effects on vowel nuclei", *STL-QPSR* 4, p. 1-18