Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Structural classification of Swedish phonemes

Fant, G.

journal: STL-QPSR
volume: 1
number: 2
year: 1960
pages: 010-015

F. AUTOMATIC IDENTIFICATION OF SOUND FEATURES

A few pilot studies devoted to the identification of nasal consonants in Swedish speech material have been started. No conclusive results may be reported yet. Measurements with the aid of anti-resonance circuitry did not support the view of a rather large first formant bandwidth of nasal sounds compared with other sounds such as [l] [j] [v] and vowels of a very low F_1.

C. Cederlund, B. Lindblom, J. Mårtony, A. Møller, S. Öhman

G. STRUCTURAL CLASSIFICATION OF SWEDISH PHONEMES

The following tabulation of Swedish phonemes in a distinctive feature code is in all essentials based on the system of Jakobson-Fant-Halle (1) but with the modifications owing to the general advance of the theory and the specific views held by the author. The particular solution for the vowel system is the same as that proposed in recent publications (2), but differs from that of an earlier study (3). The consonant system has not been published before.

The distinctive feature scheme primarily serves the purpose of linguistic theory but includes acoustic descriptions which theoretically could be regarded as an instruction for machine recognition of spoken text. In practice the acoustic definition of the features often involves stipulations concerning the differences between alternative phonemes. These abstractions may not be translated to identification rules without taking into account the specific range of qualities utilized by the particular speaker in a specific context.

Most of the distinctive features or rather "phonemic distinctions" are identical with elementary phonetic categories which are well established in linguistic theory. Any scheme for machine recognition of spoken items will to some extent rely on a classification in terms of these categories. On the other hand, it is clear that an optimal recognition process will differ from the traditional distinctive feature system in terms of the particular choice and definitions of features and the sequence of operations. Vowels could thus be identified directly from properly normalized formant frequencies. Independently of the purpose of the identification scheme, whether this be linguistic theory or machine recognition of speech, it should be recognized that distinctive features or the several cues which may underlie a distinction are not always static constituents of a single sound segment* but quite often involve several adjacent segments and dynamic relations within a sequence of segments. These are well established facts.**

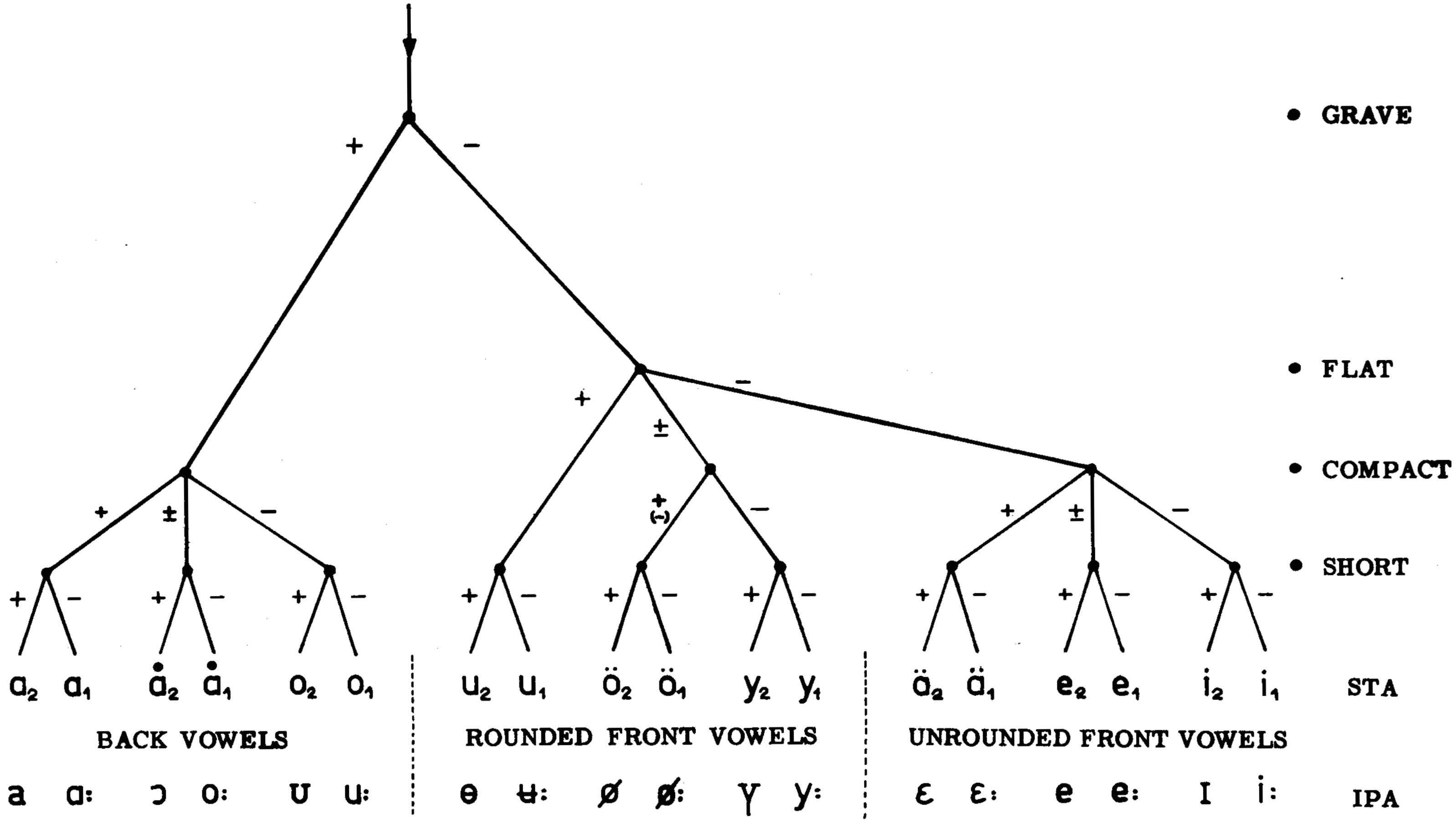
The Swedish vowel system comprises nine long vowel phonemes $/o_1/$ $/å_1/$ $/a_1/$ $/y_1/$ $/u_1/$ $/ö_1/$ $/i_1/$ $/e_1/$ $/ä_1/$ and nine short vowel phonemes $/o_2/$ $/å_2/$ $/a_2/$ $/y_2/$ $/u_2/$ $/ö_2/$ $/i_2/$ $/e_2/$ $/ä_2/$. These phonemic notations of the STA-alphabet conform with common Swedish orthography. Phonetic values of the basic allophones are indicated in Fig. I-9 and I-10.

The relation of $/ä_1/$ to $/e_1/$ or of $/e_1/$ to $/i_1/$ specified by the compactness feature is that of an open versus a close vowel, referring to the mouth cavity. The relation of $/a_1/$ to $/å_1/$ or $/å_1/$ to $/o_1/$ may either be identified by the compactness feature as in Fig. I-9 or with the relation of unrounded to rounded (flat) vowels, Fig. I-10. There are of course hybrid alternatives, such as the labeling of $/o_1/$ to $/å_1/$ as flat versus $/a_1/$ and opposing $/å_1/$ to $/o_1/$ in terms of compactness. This solution avoids the use of ± terms within the back vowels. The same relations hold for short vowels.

* The difference between the prosodic and inherent features are thus to some extent eliminated since both involve temporal relations. The prosodic features, however, generally operate over speech-wave units of a greater length than the inherent features.

** An attempt to construct a scheme of segment classification according to an inventory of narrow phonetic categories instead of phonemic distinctions has been undertaken in a recent article.(4)

SWEDISH VOWELS (SYSTEM A)



STA = Swedish Technical Alphabet IPA = International Phonetic Alphabet

Fig. I-9 Distinctive feature coding of Swedish vowels. Back vowels separated in terms of compactness.

SWEDISH VOWELS (SYSTEM B)

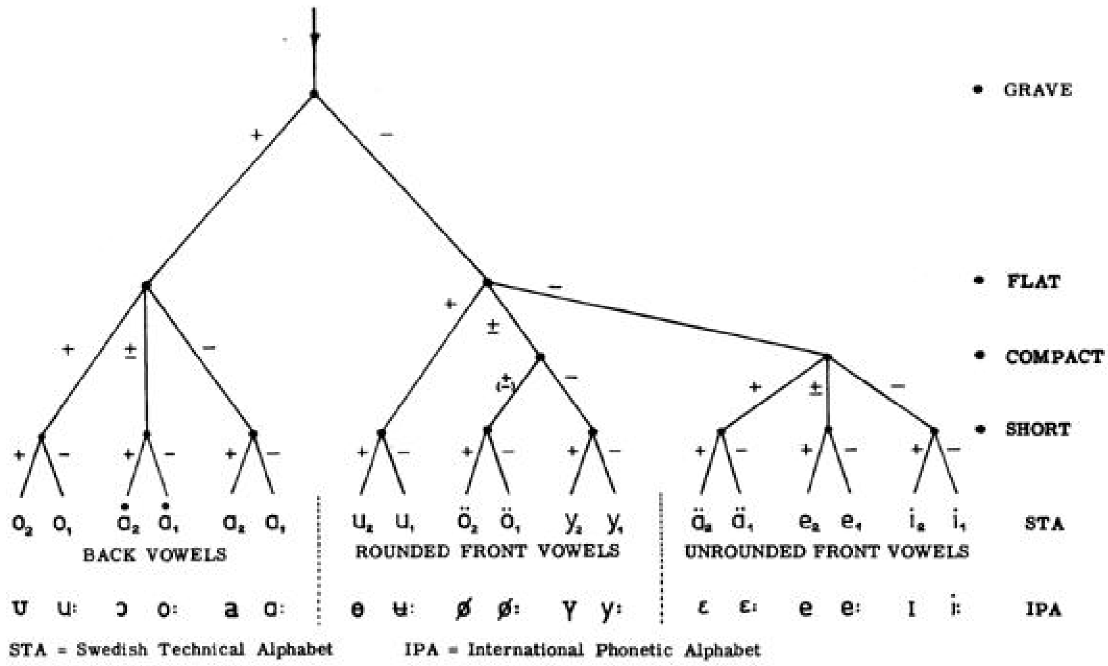


Fig. I-10 **Distinctive feature coding of Swedish vowels. Back vowels separated in terms of flatness.**

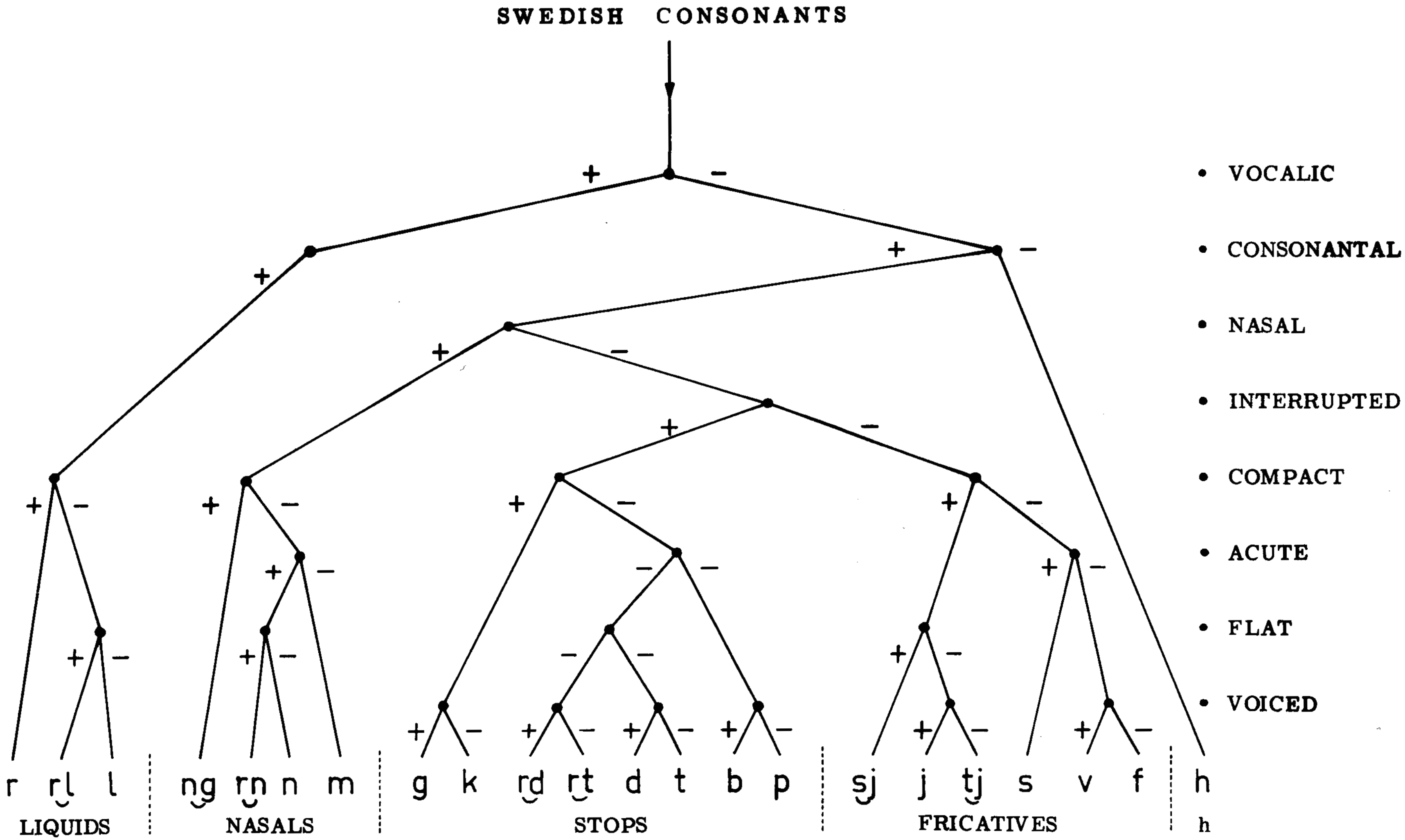


Fig. I-11 Distinctive feature coding of Swedish consonants.

The relations within the rounded front vowels /y_1/ /u_1/ /$ö_1$/ and /y_2/ /u_2/ /$ö_2$/ are more complex. The phoneme /$ö_1$/ is definitely compact compared with /y_1/ and /$ö_1$/ is compact compared with /y_2/ but the phonemes /u_1/ and /u_2/ cannot be consistently specified by this feature. Thus /u_2/ is generally a more compact vowel than /$ö_2$/ whereas /u_1/ is less compact than /$ö_1$/. In general, however, the relation of /u_1/ to /$ö_1$/ and /y_1/ or of /u_2/ to /$ö_2$/ and /y_2/ is the same as that of /y_1/ and /$ö_1$/ compared with /i_1/ /e_1/ and /$ä_1$/. This is the motivation for the classification of /$ö_1$/ and /y_1/ and similarly also /$ö_2$/ and /y_2/ as ± flat.

Within the consonant system, Fig. I-11, the alveolar phonemes /r͜l/ /r͜n/ /r͜t/ /r͜d/ are opposed to the pure dentals /l/ /n/ /t/ and /d/ in terms of the flatness feature. From a distributional point of view, however, the alveolars may be regarded as the realization of a phoneme /r/ plus a following dental phoneme. This is also the case for an /r͜s/ as opposed to /s/ with the complication that [r͜s] stands in complimentary distribution to a quite similar sound labeled [s͜j] which is not the result of a fusion between an [r] and a dental. Thus /r͜s/ is the same phoneme as /s͜j/. The phoneme /s͜j/ is acoustically flat (lower frequency of main formant) both in relation to the compact (palatal) /t͜j/ and the non-compact acute /s/.

The following is a condensed summary of acoustic correlates of phonemic distinctions with special reference to the Swedish phoneme system*.

1. The vowel system

(1) Acute. An acute vowel has a higher F_2-F_1 than a corresponding non-acute (grave) vowel.

(2) Flat. A flat vowel has a lower sum (with possible weighting) $F_1+F_2+F_3$ than corresponding non-flat (plain) vowels.

(3) Compact. A compact vowel has a higher F_1 than a corresponding non-compact (diffuse) vowel.

(4) Short. A short vowel within the Swedish vowel system has a shorter duration and generally a spectrum with a higher F_1 and a more neutral formant pattern than a corresponding long vowel of the same context.

* The articulatory correlates are well established except for some specific details of the Swedish vowel system. For a general discussion of articulatory correlates, see earlier publications, e.g., G. Fant "Acoustic Theory of Speech Production".(5)

The long vowel is combined with a short consonant and vice versa. The short/long distinction is in all essentials identical with the lax/tense distinction described by Jakobson-Fant-Halle [1].

2. The consonant system

(1) Vocalic. The F-pattern formants (F1 F2 F3) of frequencies F_1 F_2 F_3 respectively are more apparent and the overall intensity is higher in a vocalic than in a non-vocalic phoneme.

(2) Consonantal. Consonantal phonemes are acoustically characterized by sound segments fulfilling one or both of the following conditions:

a. The main energy is confined to other formants than F1 and F2.

b. Low second formant intensity F2 and generally a low first formant frequency position F_1 compared with adjacent sound segment.

These aspects of the consonantal feature appear as a temporal contrast effect either in the type of spectrum or in terms of a rapid transition of the formant frequencies and the sound intensity.

There is an appreciable overlap in the definition of the consonantal feature and the non-vocalic feature which parallels the classification of most consonants as being both consonantal and non-vocalic. The phoneme /h/ is classified non-consonantal because of the lack of contrast of formant frequencies relative to an adjacent vowel but it is non-vocalic because of the lower intensity, especially in the first formant range. The liquids /l/ and /r/ have more vocalic formant patterns than other consonants and may thus be coded as +vocalic and +consonantal.

The syllable

The nucleus of a common syllable is a single vowel or a diphthong, the speech-wave correlate of which is a vocalic non-consonantal sound segment. An adjacent sound segment belonging to the same syllable is always consonantal and produces together with the vowel a temporal contrast owing to its lower intensity* or spectrum of a non-vocalic type. On account of the +vocalic feature liquids may be said

*It is probable that a proper frequency selective pre-emphasis in the intensity measuring procedure, favoring the frequency range below 3000 c/s, would invariably provide a larger intensity of a vowel than an adjacent consonant, e.g., fricative. Results from experiments on voicing detectors support this view.

to rank next after vowels in terms of "syllabicity". In a consonant cluster liquids always occur next to the vowel in conformity with the idea of a successive decay of syllabicity away from the syllable nucleus. A liquid /l/ or /r/ out of contact with a vowel would thus in itself constitute a syllabic nucleus. The /r/ possesses a greater syllabicity than /l/ on account of the greater compactness.

In unstressed syllables the vowel is of lower intensity, shorter duration, and possesses a formant pattern closer to that of a neutral vowel than in stressed syllables. The contrast between the vowel and adjacent consonants is reduced.

(3) Nasal. The formant structure of nasal sound segments has a reduced second formant intensity and possesses the typical qualities of the nasal murmur.

(4) Interrupted. Rapid onset or checking of the sound intensity combined with rapid formant transitions.

(5) Compact. The spectral energy of the sound segments of compact phonemes is concentrated to a more central location versus the main pitch of the immediately adjacent vocalic sound segments than non-compact (diffuse) phonemes*.

(6) Acute. Greater intensity of high-frequency formants than in non-acute (grave sounds).

(7) Flat. A shift down in the frequency location of formants retaining the general shape of the spectrum.

(8) Voiced. In Swedish consonants the voiced/voiceless opposition occurs in complimentary distribution with the lax/tense opposition. The common denominator is the relative lack or the shortness of duration of unvoiced segments of the speech wave. Examples are the burst and occlusion phases of voiced stops and the unvoiced segment of voiced (lax) continuants which are shorter in voiced than in unvoiced sounds.

* One suggested definition of main pitch is the second formant of a synthetic two-formant sound perceptually matching the quality of the sound segment. In first approximation this main pitch coincides with F_2 but is closer to F_3 or F_4 for high front vowels. In voiced stops and nasal consonants the front transitions carry a great part of the perceptually effective cues. The extent to which these cues may be included in the formulation above is not quite clear. In general the formant transitions and the consonantal sound segment constitute a compound stimulus.

Swedish consonants

		r	rl	l	ng	rn	n	m	g	k	rd	rt	d	t	b	p	sj =rs	j	tj	s	v	f	h
I	Vocalic	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
II	Consonantal	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−
III	Nasal				+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	
IV	Interrupted								+	+	+	+	+	+	+	+	−	−	−	−	−	−	
V	Compact	+	−	−	+	−	−	−	+	+	−	−	−	−	−	−		+	+	−	−	−	
VI	Acute					+	+	−			+	+	+	+	−	−				+	−	−	
VII	Flat		+	−		+	−				+	+	−	−			+	−	−	−			
VIII	Voiced								+	−	+	−	+	−	+	−		+	−		+	−	

Swedish vowels

		a_2a_1	$å_2å_1$	$o_2{}^{\cap}{}_1$	$ä_2ä_1$	e_2e_1	i_2i_1	$ö_2ö_1$	y_2y_1	u_2u_1
1	Acute	− −	− −	− −	+ +	+ +	+ +	+ +	+ +	+ +
2	Flat	(−)(−)	(±)(±)	(+)(+)	− −	− −	− −	± ±	± ±	+ +
3	Compact	+ +	± ±	− −	+ +	± ±	− −	+ +	− −	− −
4	Short	+ −	+ −	+ −	+ −	+ −	+ −	+ −	+ −	+ −

G. Fant

(1) Jakobson, R., Fant, C.G.M., Halle, M.: "Preliminaries to speech analysis. The distinctive features and their correlates", M.I.T. Acoustics Lab. Techn. Rep. No. 13 (1952) 3rd printing.

(2) Fant, G.: "Modern instruments and methods for acoustic studies of speech", Proc. of the VIII International Congress of Linguists, Oslo University Press, Oslo 1958, p. 282-358; also publ. in Acta Polytechnica Scandinavica, Ph 1 (246/1958).

——— : "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics No. 1, Vol. 15, 3-108 (1959).

(3) Fant, G.: "Phonetic and phonemic basis for the transcription of Swedish word material", Acta Oto-Laryngologica, Suppl. 116, 83-93 (1954).

(4) Fant, G.: "Descriptive analysis of the acoustic aspects of speech", paper presented at the 1960 Summer Symposia at Burg Wartenstein, Austria, "Comparative Aspects of Human Communication", September 4-10, 1960.

(5) Fant, G.: "Acoustic Theory of Speech Production", Mouton & Co., 's-Gravenhage 1960, 323 pp.