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

Speech synthesis. Instrumentation for parametric synthesis (OVE II)

Fant, G. and Mártony, J.

journal: STL-QPSR
volume: 3
number: 2
year: 1962
pages: 018-019

KTH Computer Science and Communication

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

II. SPEECH SYNTHESIS

A. SPEECH SYNTHESIS

1. Instrumentation for parametric synthesis (OVE II)

Our formant coded speech synthesizer OVE II is organized as shown in the block diagram of Fig. II-1. There are three main synthesis filters which are operated independently labelled the F N and K systems. Each of these is designed as a cascade (series) connection of elementary zero and pole functions. These systems are fed from a voice source and a noise source through source amplitude modulating gates A_O A_N A_H A_C. The gate A_O determines the amplitude of voicing for the vocalic filter system and A_N is the corresponding amplitude of voice for the nasal system N. The notation A_H pertains to the noise amplitude for the vocalic filter system F and A_C is the amplitude of noise for the fricative filter K. The outputs of the F- N- and K-systems are added in a mixer stage.

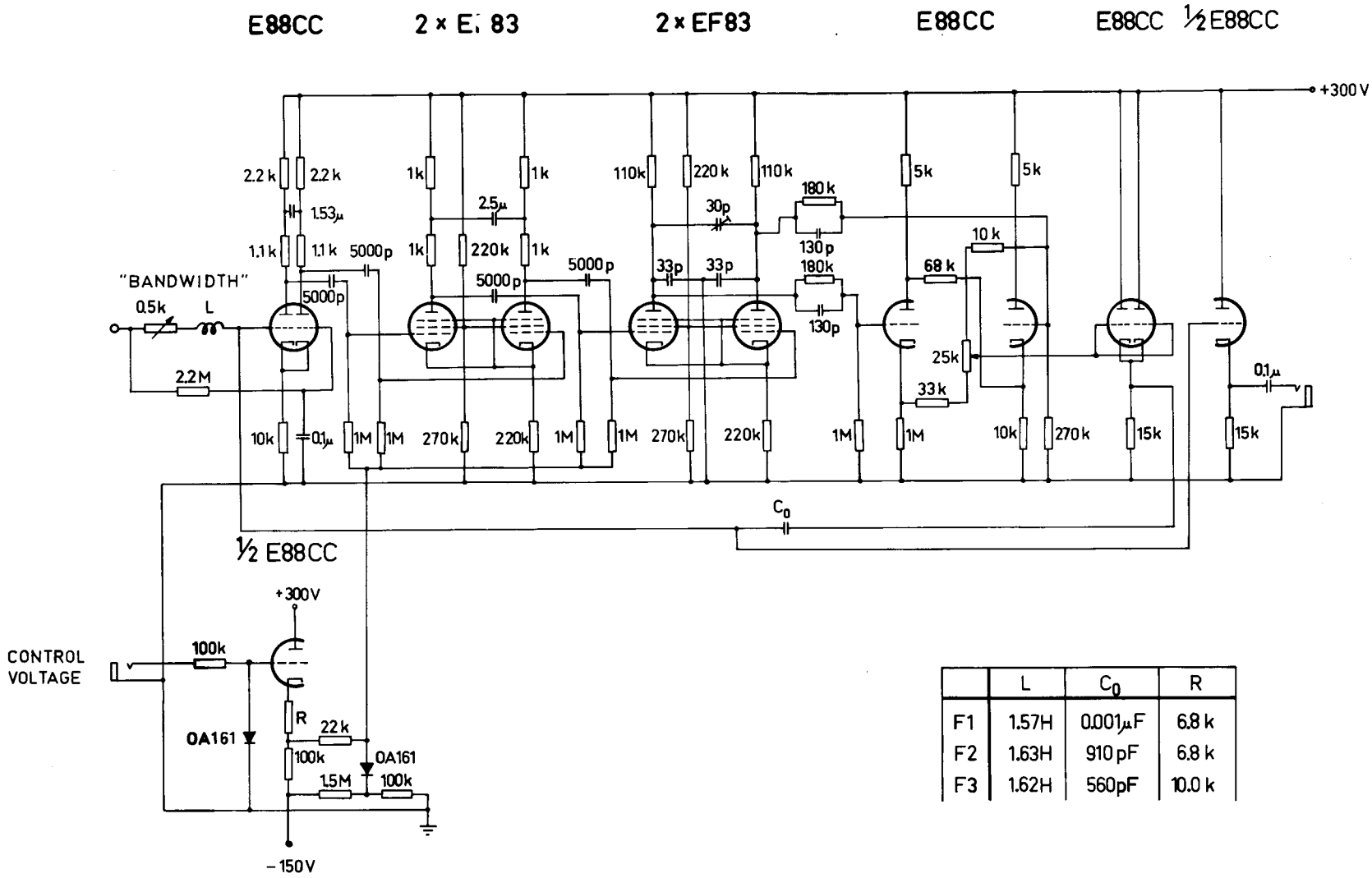
A mechanical function generator provides means for voltage control of 12 parameters during 3 seconds of speech. Normally we utilize 11 of these, namely the voice fundamental frequency F_O, the four gating functions mentioned above and further 3 variable F-formants, 2 variable K-formants, and one variable K-anti-resonance. These control voltages are smoothed through low-pass filters the time constants of which may be varied in octave steps from 2.5 msec to 40 msec.

Fig.:s II-2, II-3, II-4, and II-5 provide detail information on the design of the main building blocks, including the formant generators, the anti-resonance circuit, and the amplitude modulated gates.

The F-filter is intended for the synthesis of vowellike sounds and comprises the voltage controlled units F1 F2 F3 and the fixed position, manually controllable units F4 and F5' and KH.

The voltage-frequency calibration of F1 F2 and F3 is linear which is a requirement for their sharing a common field on the function generator. With normal condenser settings in the resonance circuits the ranges are as follows:

F1 140 c/s ⟶ 1000 c/s
F2 500 c/s ⟶ 3000 c/s
F3 1100 c/s ⟶ 4500 c/s

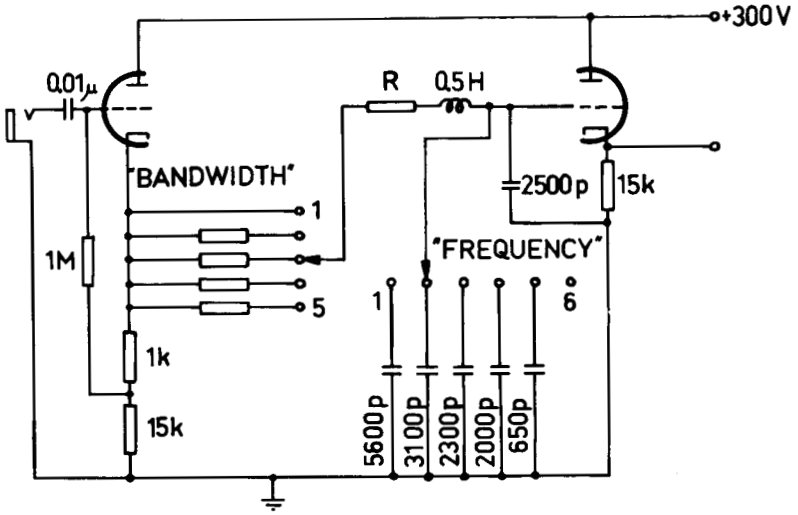


	L	C_0	R
F1	1.57H	0.001µF	6.8 k
F2	1.63H	910 pF	6.8 k
F3	1.62H	560pF	10.0 k

Fig. II-2. **Circuit diagram of the voltage controlled pole circuit in OVE II.**

F4

E 88CC



BANDWIDTH

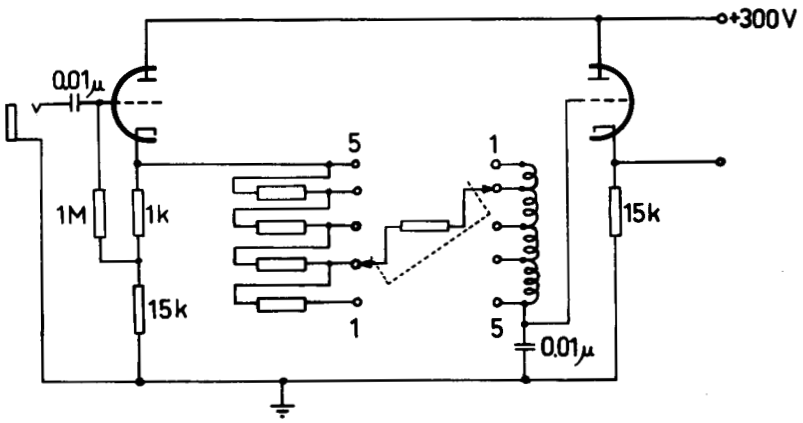
POS.	B_4 c/s	R_{tot} Ω
1	100	314
2	140	440
3	200	628
4	280	880
5	400	12500

FREQUENCY

POS.	F_4 c/s
1	2500
2	3000
3	3250
4	3500
5	4000
6	4500

F5'

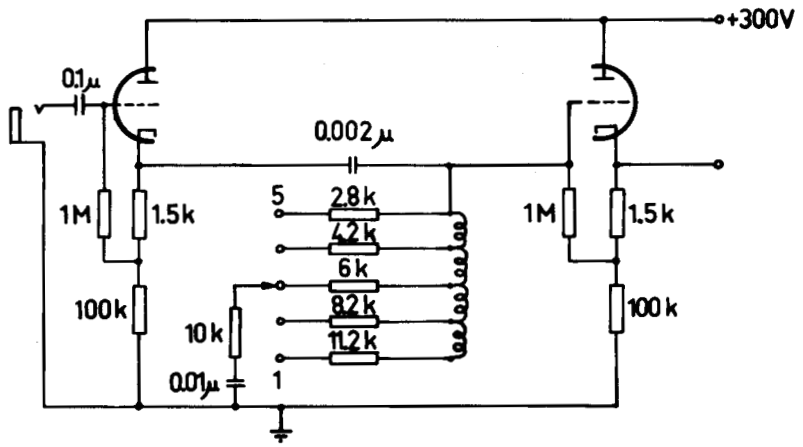
E 88CC



POS.	F'_5 c/s	L H	R_{tot} Ω
1	3000	2.83	533
2	3500	2.08	457
3	4000	1.60	400
4	4500	1.26	350
5	5000	1.03	320

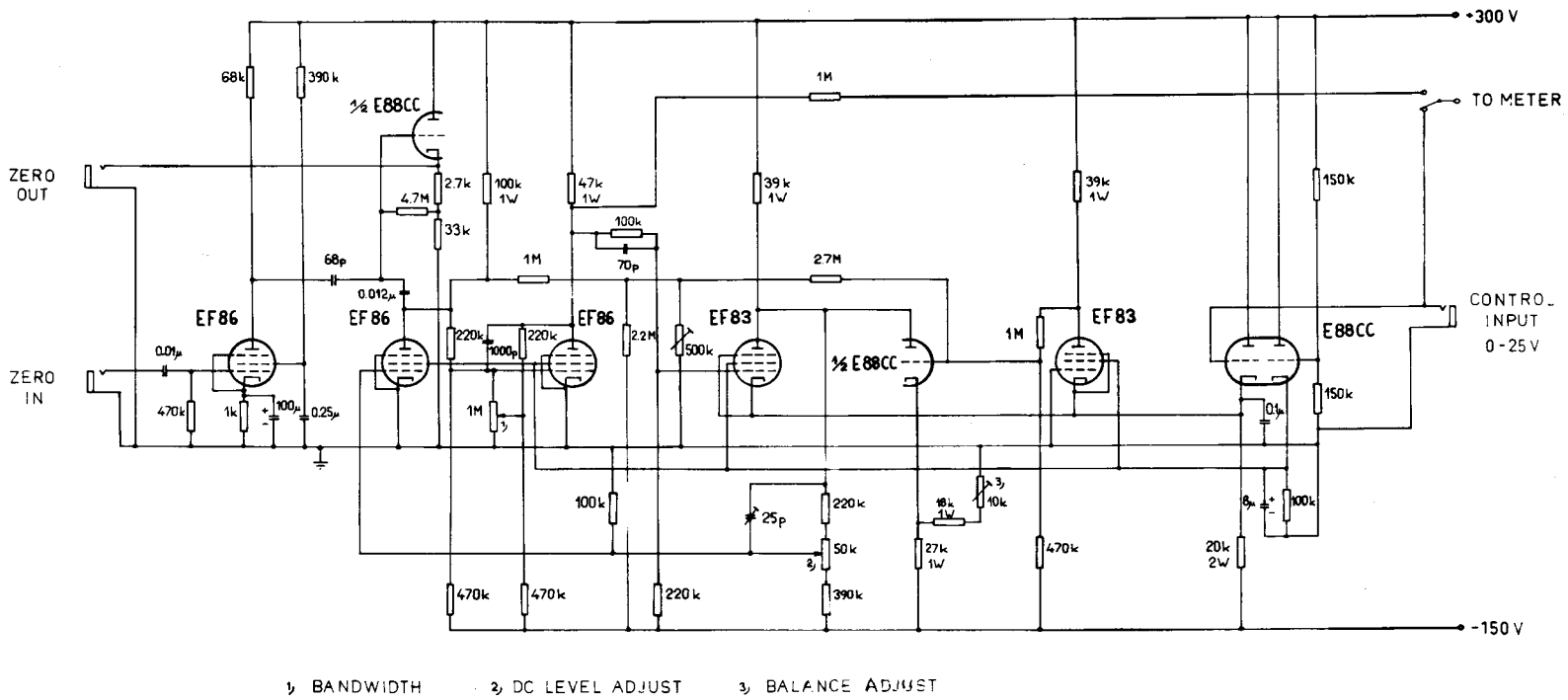
KH

ECC 83



POS.	f c/s	L H
1	3000	1.10
2	3500	0.81
3	4000	0.62
4	4500	0.49
5	5000	0.40

Fig. II-3. **Circuit diagrams of manually controlled poles F4 and F5' and higher pole correction KH in OVE II.**



1) BANDWIDTH 2) DC LEVEL ADJUST 3) BALANCE ADJUST

Fig. II-4. Circuit diagram of a voltage controlled zero circuit for use in OVE II.

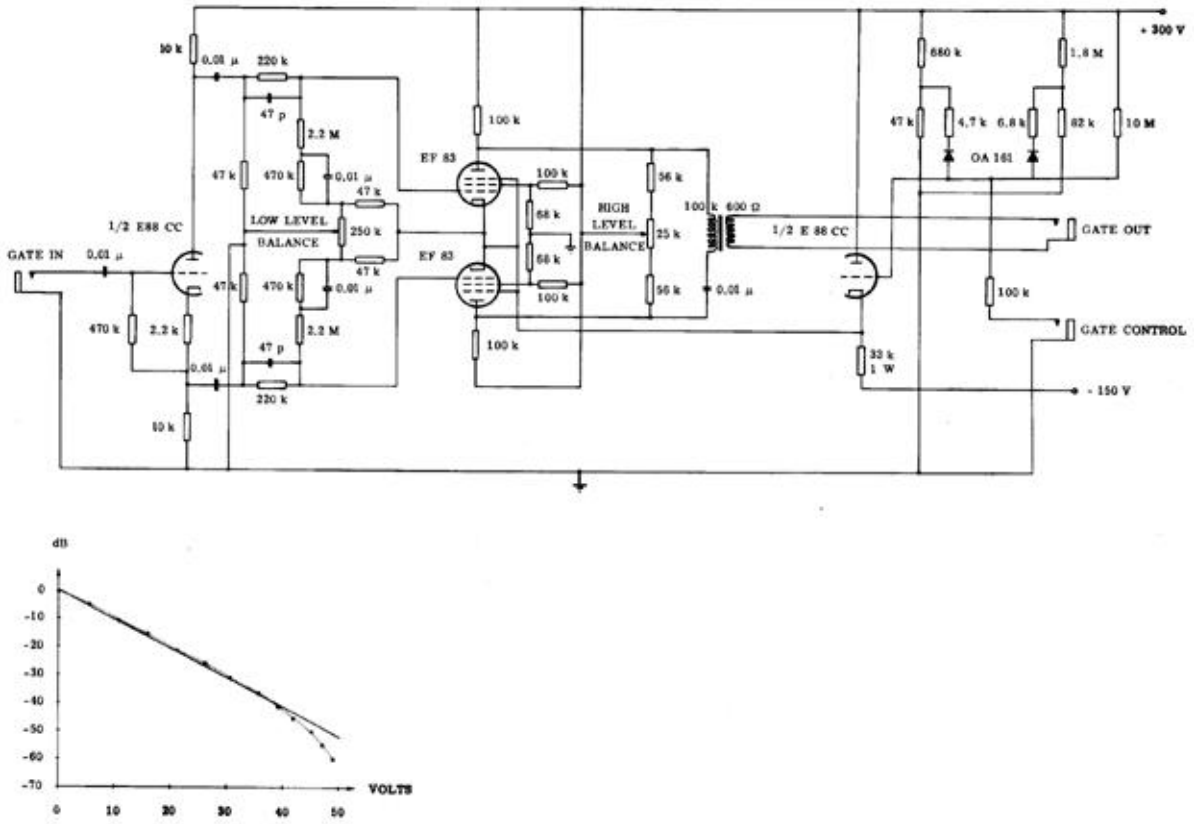


Fig. II-5. **Voltage controlled gate for OVE II providing linear variation on a dB-scale of signal level.**

The alternative settings for F4 and for the higher pole correction circuits F5' and KH (see ref. (1)) are as follows:

F4	2500 c/s	3000 c/s	3250 c/s	3500 c/s	4000 c/s
F5'	3.0 kc/s	3.5 kc/s	4.0 kc/s	4.5 kc/s	5.0 kc/s
KH	3.0 kc/s	3.5 kc/s	4.0 kc/s	4.5 kc/s	5.0 kc/s

Equivalent vocal tract length.

The bandwidth may be tuned in by hand for each formant. Unless otherwise specified we select the values:

B_1 = 70 c/s B_2 = 80 c/s B_3 = 100 c/s
B_4 = 140 c/s B_5 = 400 c/s

The second filter unit N, is intended for nasal sounds and is not controlled from the function generator except by means of the nasal gate which connects or disconnects the nasal filter to the common mixer. This unit has 4 pole circuits:

N1, N2, N3, N4, and a zero circuit N0

Frequencies and bandwidths are manually controllable over wide ranges. Typical settings for simulating a nasal consonant are:

[m] N_1 = 225 c/s N_2 = 700 c/s N_3 = 1200 c/s N_4 = 2200 c/s
N_5 = 2700 c/s N0 = 800 c/s

[n] N_1 = 225 c/s N_2 = 1100 c/s N_3 = 2100 c/s N_4 = 2500 c/s
N_5 = 2700 c/s N0 = 1500 c/s

The third filter unit, K, is intended for synthesis of non-vowellike noise sounds, i.e. unvoiced fricatives, affricates, and short fricative sound segments within the burst of a stop sound. This filter comprises 2 poles and one zero which are voltage controlled parameters. These 2 resonances and the anti-resonance cover the following frequency ranges:

K_1 1000 - 6500 c/s
K_2 2500 - 10000 c/s
K_0 1000 - 6000 c/s

The bandwidths are preset before synthesis to values of the order of 200-400 c/s.