Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

A filter bank speech spectrum analyzer

Liljencrants, J.

journal: STL-QPSR
volume: 6
number: 2
year: 1965
pages: 001-007

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

KTH Computer Science and Communication

I. INSTRUMENTATION FOR SPEECH ANALYSIS

A. A FILTER BANK SPEECH SPECTRUM ANALYZER*

J. Liljencrants

At the Speech Transmission Laboratory a flexible spectrum analyzer has been designed and constructed.

The input signal is analyzed with a bank of band-pass filters. These can be set in a number of different combinations of center frequencies and bandwidths. In this way it is possible to make a gross analysis for orientation as well as detailed analysis with a high resolution in either frequency or time.

The signals from the band-pass filters are rectified and smoothed in low-pass filters with a variable cutoff frequency. The channels are synchronously sampled and the signal is quantized on a logarithmic scale in 120 steps of 0.5 dB.

The scanning of the channels and the subsequent data handling employ digital techniques. In this way the analyzer is fairly independent of the speed of peripheral equipment which thus can be chosen according to its suitability for specific tasks.

The following is a short description of the main parts of the analyzer with examples of the mode of operation.

Signal source

The signal to be analyzed is normally recorded on one track of a conventional magnetic audio-recording tape. The other track contains a time-code signal giving a time scale that is fixed relative to the signal and which can be used to control the sampling. The time code is generated and received by a special unit within the analyzer.

In certain cases the signal should be repeated a number of times which can be accomplished either by cutting the tape into loops or by using a tape-scanner.

---

*Extended version of paper to be presented at the 5th International Congress on Acoustics, Liège, Belgium, Sept. 7-14, 1965.

Heterodyne system

The input signal is band-limited to 11 kc/s and modulated with a first carrier, the frequency of which can be varied between 27 and 36.5 kc/s. Part of the lower sideband is isolated with a band-pass filter, 18 to 28 kc/s, and is modulated with a second carrier fixed at 28 kc/s. The lower sideband of this modulation is in the audio-frequency band and is fed to the filter bank via a power amplifier.

The result of these modulations is thus that the input band is displaced upwards or downwards with an amount equal to the difference frequency between the two carriers.

The carrier frequencies are derived by means of frequency synthesis from appropriate partials of a set of square waves. These are taken from a binary divider chain counting a master clock. The synthesis of the first carrier frequency is controlled by a 5 bit binary signal determining the appropriate multiple of 0.5 kc/s of frequency shift.

Filter bank

The filter bank contains 51 channels. Each channel is built as a plug-in unit containing a band-pass filter, a linear full-wave rectifier, and a low-pass filter.

The low-pass filters have an approximate linear phase which implies that the transition from pass band to rejection band is not as sharp as for instance in a Butterworth filter. Instead the overshoot in the pulse response is very small at the same time as the duration of the response is comparatively small as related to the inverted value of the bandwidth. These properties are rather essential in the analysis of rapidly changing signals for the avoidance of masking effects. The filter is a mininum-phase network (no finite zeros) of the third degree and thus the attenuation will rise 18 dB/oct in the rejection band.

The band-pass filters are derived from a corresponding low-pass filter by means of the low-pass/band-pass frequency substitution $W = s + \omega_1^2/s$. The attenuation and transient properties are analogous to those of the low-pass filter.

The channels of the filter bank are divided into two groups called A and B. Group A comprises ten channels from 0 to 900 c/s with constant spacing and variable bandwidths. This group is not used together with the heterodyne system. Group B consists of the remaining 41 channels the first of which is always centered at 1 kc/s while the remainder covers a range above this frequency. There are 11 different combinations of frequency spacing and bandwidth of the B-group and 4 combinations of the A-group, as shown in Fig. I-A-1 and in Tables I-A-1 and I-A-2.

Table I-A-1

Group	Combi-nation	Δf	B	W(kc)
B	1	12.5	31.3	0.5
B	2	12.5	62.5	0.5
B	3	12.5	125	0.5
B	4	25	31.3	1
B	5	50	125	2
A, B	6	100	62.5	4
A, B	7	100	125	4
A, B	8	100	250	4
A, B	9	100	500	4
B	10	100/80 m	250	9.6
B	11	83 m	83 m	9

Filter spacing Δf, bandwidth B, and total analysis range W covered within the B-group. Filter group A is limited to the combinations 6-9.

The first nine combinations have constant spacings of center frequencies throughout the filter bank. The filter spacing in combinations 6 through 9 is 100 c/s which conforms with the fixed spacing of the A-group so that the B-group may be added without transposition.

In combination 10 the spacing is constant 100 c/s over the first 800 c/s then increasing and the bandwidth is constant 250 c/s. If the input signal is transposed upwards 1 kc/s the B-group will approximate a Koenig scale. As an alternative the transposition can be disposed with and an A-8 selection be combined with B-10. We then generate a frequency scale that is linear up to 1.8 kc/s and then gradually compressed.

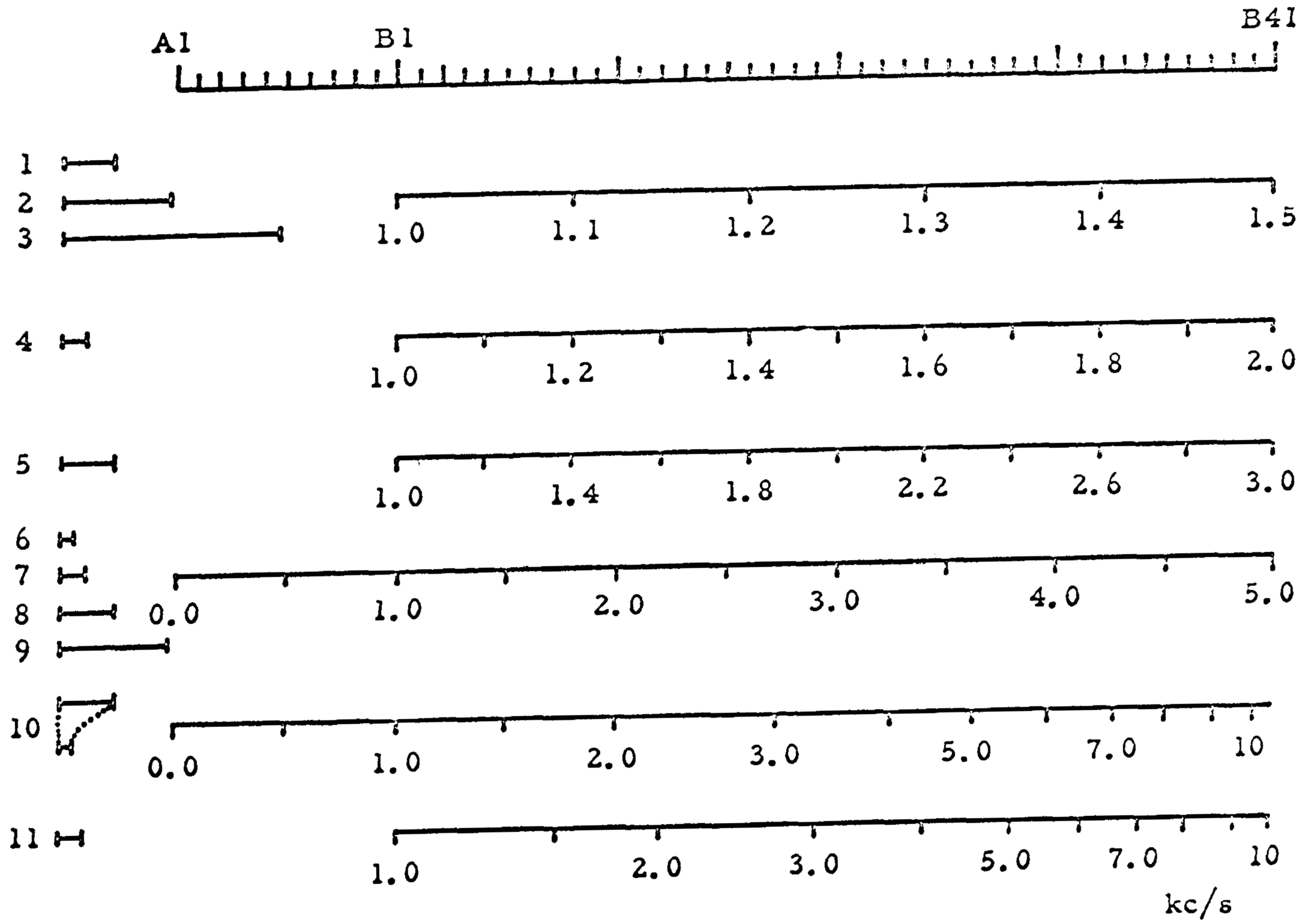


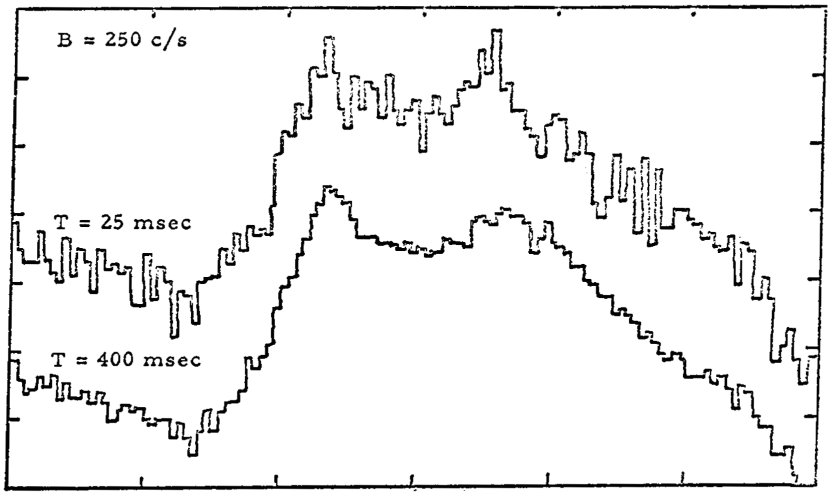
Fig. I-A-1. The top scale indicates the channel numbers while the other scales show corresponding center frequencies of the band-pass filters for the 11 possible alternatives as indicated. The bars at left show the filter bandwidth.

The effective filter frequencies may be shifted upwards or downwards in 500 c/s increments using the heterodyne system.

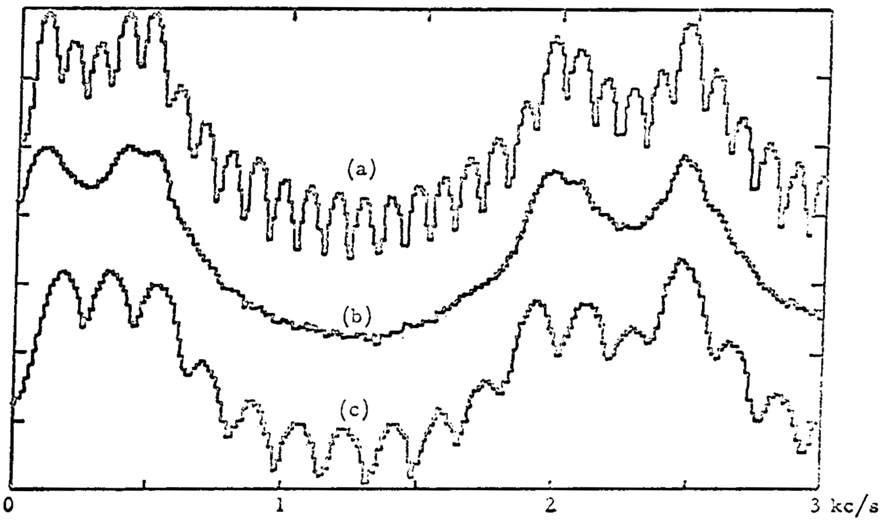
Channel number		Combination number				
		10		11		
physical	logical	f	Δf	f	Δf	B
B 1	10	1000	100	1000	60	58
B 2	11	1100	100	1060	63	61
B 3	12	1200	100	1123	66	65
B 4	13	1300	100	1189	71	69
B 5	14	1400	100	1260	75	73
B 6	15	1500	100	1335	79	77
B 7	16	1600	100	1414	84	82
B 8	17	1700	100	1498	89	87
B 9	18	1800	103	1587	95	92
B10	19	1903	108	1682	100	97
B11	20	2011	115	1782	106	103
B12	21	2126	121	1888	112	109
B13	22	2247	128	2000	119	116
B14	23	2375	136	2119	126	122
B15	24	2511	143	2245	133	130
B16	25	2654	151	2378	142	137
B17	26	2805	160	2520	150	146
B18	27	2965	169	2670	158	154
B19	28	3134	179	2828	169	163
B20	29	3313	189	2997	178	173
B21	30	3502	199	3175	189	183
B22	31	3701	211	3364	200	194
B23	32	3912	223	3564	212	206
B24	33	4135	236	3776	224	218
B25	34	4371	249	4000	238	231
B26	35	4620	264	4238	252	245
B27	36	4884	278	4490	267	259
B28	37	5162	295	4757	283	275
B29	38	5457	311	5040	299	291
B30	39	5768	329	5339	318	308
B31	40	6097	347	5657	336	327
B32	41	6444	368	5993	357	346
B33	42	6812	388	6350	377	367
B34	43	7200	411	6727	400	389
B35	44	7611	433	7127	424	412
B36	45	8044	459	7551	449	436
B37	46	8503	485	8000	476	462
B38	47	8988	512	8476	502	490
B39	48	9500	542	8978	536	519
B40	49	10042	573	9514	565	550
B41	50	10615		10079		582

Table I-A-2. Center frequencies and spacings of combination 10 (bandwidth B = 250 c/s), and center frequencies, spacings, and bandwidths of combination 11.

Fig. I-A-2. Examples of group analysis with spectrum sections as registered on the X/Y plotter. 10 dB per division of the vertical axis.



Spectra of a synthetic /s/ (OVE II) analyzed using filter combination 6. The influence of the integration time constant on the typical random variations in spectral level is apparent.



A synthetic vowel with constant formant frequencies and bandwidths analyzed with filter combinations 2 and 3.
(a) Voice fundamental F_o = 100 c/s, analysis bandwidth B = 63 c/s, the harmonics are resolved.
(b) B = 125 c/s will give a smooth spectrum envelope. The first two formant peaks are distorted due to absence of harmonics at the formant frequencies.
(c) F_o = 175 c/s and harmonics reappear in the picture. Problems of measuring formant frequency and level are obvious.

Finally in combination 11 without transposition the filter frequencies are distributed on a logarithmic scale covering the decade from 1 to 10 kc/s while the bandwidths are equal to the spacings. The effective frequency range of this decade may be changed by playing back the input tape at an increased or reduced speed. In case the input signal is transposed +1 kc/s combination 11 will provide a technical mel scale. This is defined by Fant (ref. [1], p. 76) as

$$m_t = 1000 \cdot \frac{\log(1 + f/1000)}{\log 2}$$

where m_t is technical mels and f frequency.

It should be pointed out that center frequencies are consistently defined in terms of an arithmetic mean.

The band-pass filters are followed by linear rectifiers and low-pass filters. These are realized as active RC-networks. The integration time is variable from 1.6 to 400 msec corresponding to cutoff frequencies from 250 to 1 c/s.

<u>Digitizing system</u>

Each of the channel filter units is followed by a simple analog to digital converter in order to allow synchronous sampling of all channels. The signal is quantized on a logarithmic basis into 120 levels 0.5 dB apart corresponding to 7 bits. The converters are based on the exponential decay of an RC-circuit and have been described in detail earlier [2].

The converters also serve as a digital memory for one spectrum section which can be scanned at an arbitrary speed before the next sampling.

Most functions of the analyzer are electronically controlled from a number of control units.

The program sequence for the heterodyne, the sampling, the output format, and other functions can be chosen with the aid of a number of patchboard plugs.

The sampling is initiated using the time code of the input tape-recording. The code signal consists of short clock pulses with the repetition frequency 640 c/s. Each 100 msec a pulse group of opposite polarity is added giving information of the absolute location on the tape within the code cycle which is about 25 seconds in length. The time-code receiver will give a warning if the input signal does not satisfy a parity check and other conventions.

Output equipment

Output data can be presented visually for monitoring and demonstration either on a digital display or reconverted to analog form on a CRT.

Graphic registration is made on a regular X/Y plotter for measuring purposes or photographically from a CRT for visible speech purposes.

To facilitate data processing by computer the data may be recorded in IBM compatible format with a digital magnetic tape-recorder.

Example of analysis procedure

Since each of the main parts is so flexible it is possible to select an optimal working mode for each specific project. The following may be a typical routine.

First the time code is recorded along with the input signal. Analysis is then made with one of the filter combinations covering the whole important part of the audio band. The analyzer is scanned periodically at a high rate and the result is recorded as visible speech along with a suitable representation of the time code. Alternatively the time code is presented as an oscillogram together with a duplex oscillogram of the speech. From these data the areas of interest may be located and labeled according to the time code as a reference for the following sampling of amplitude-frequency sections.

If spectrum sections of high frequency resolution are wanted we use the so called group analysis method. The analysis filters are for instance set to a frequency spacing of 25 c/s. The pertinent part of the filter bank will then cover the band from 1 to 2 kc/s. The input signal is transposed 1 kc/s upwards.

When the time code receiver finds the desired point in time a first set of measurements are taken representing the input band from 0 to 1 kc/s. While these are registered on the X/Y plotter the heterodyne is reset for no transposition. The input signal is then repeated and the next set of samples will now represent the range from 1 to 2 kc/s. The process is repeated till the desired range has been covered. The whole procedure is carried out automatically and the result is recorded as a coherent picture on the plotter.

References:

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

(2) Liljencrants, J.: "51-channel analyzer for spectrum sampling (The Sadistolog - an analog level to digital converter)", STL-QPSR 1/1962, pp. 3-5.

(3) Liljencrants, J. and Rengman, U.: "The 51-channel spectrum analyzer - a status report", STL-QPSR 4/1963, pp. 1-6.