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

Some remarks on the average speech spectrum

Tarnóczy, T. and Fant, G.

journal: STL-QPSR
volume: 5
number: 4
year: 1964
pages: 013-014

KTH Computer Science and Communication

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

B. SOME REMARKS ON THE AVERAGE SPEECH SPECTRUM

T. Tarnóczy* and G. Fant

The purpose of our report is to present some recent data on the long-time average speech spectrum of Swedish, Hungarian, and German obtained by the "speaker-chorus" method and to attempt an interpretation of the results with respect to factors related to instrumentation, constant features of human speech, and language differences.

The "speech-chorus" or "random-chorus" method was described by Tarnóczy (4) several years ago. This method is based on a tape-recording of several speakers reading the same text and mixing them all on a single tape-record before the spectrographic analysis. This procedure randomizes the speech characteristics and minimizes the analysis efforts. A report on Hungarian and German speech spectra determined by this method was given at the IVth International Congress on Acoustics (5). Both groups of subjects contained 16 males and 16 females. These data are now to be compared to a recent study at the STL comprising 10 male and 10 female Swedish subjects. As in the earlier study the analysis was made with 1/3 octave filters. The text had a length of approximately one minute.

Significant differences appear in the region of 700-1500 c/s for the male speakers and 1000-2000 c/s for the female speakers, as may be seen in Figs. II-B-1 and II-B-2. German speakers have a relatively higher spectrum level in this octave than Swedes, and Hungarians have the lowest level. We consider this order to be related to language differences, e.g. in terms of the relative occurrence of vowels with a second formant in this frequency region.

The maxima and minima in the low frequency region are influenced by the harmonic structure. The first maximum reflects the average voice fundamental frequency, which is about one octave higher for females than for males. The spectrum level in the frequency range below the fundamental frequency is largely a function of the particular frequency of the fundamental and the tendency of a minimum at half an octave above the fundamental is likewise predictable. Because of the relatively narrow bandwidths of the analysis filters in this region and the tendency of the fundamental frequency to be distributed more often around the average pitch than at extreme frequency levels it may be concluded that the minimum simply reflects

* see p. v

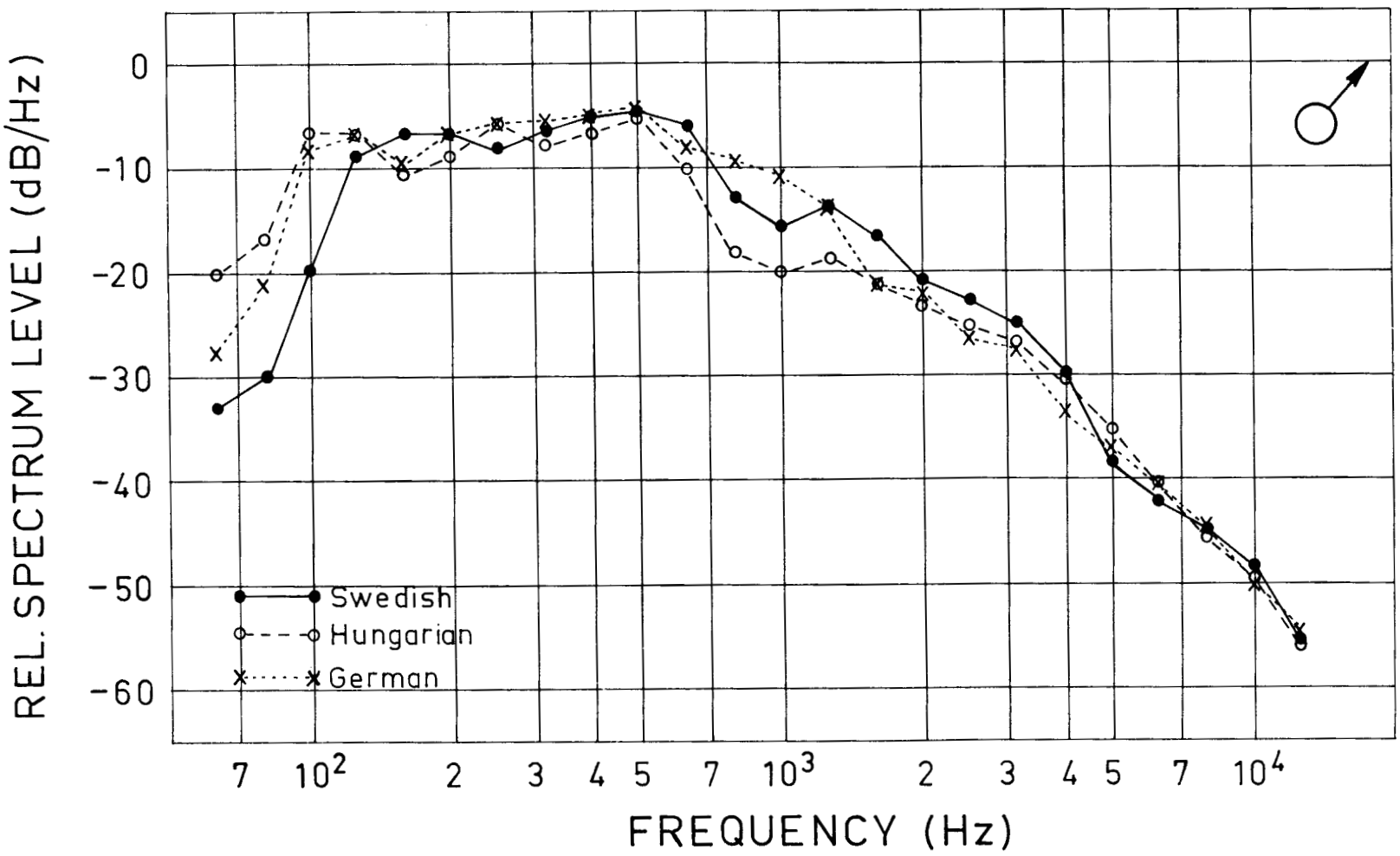

Fig. II-B-1. Average speech spectra of Swedish, Hungarian and German male subjects determined by the chorus method.

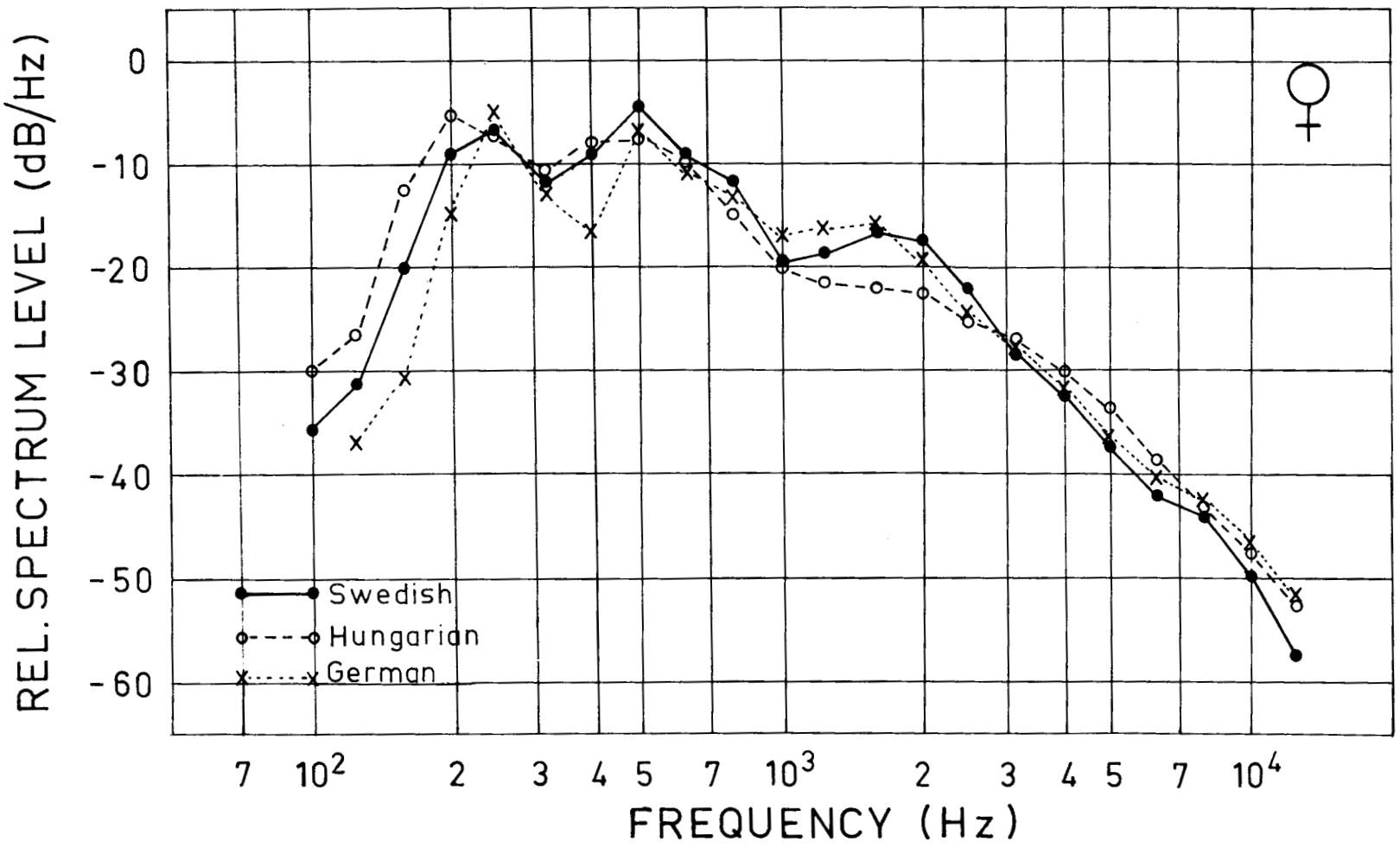

Fig. II-B-2. Average speech spectra of Swedish, Hungarian and German female subjects determined by the chorus method.

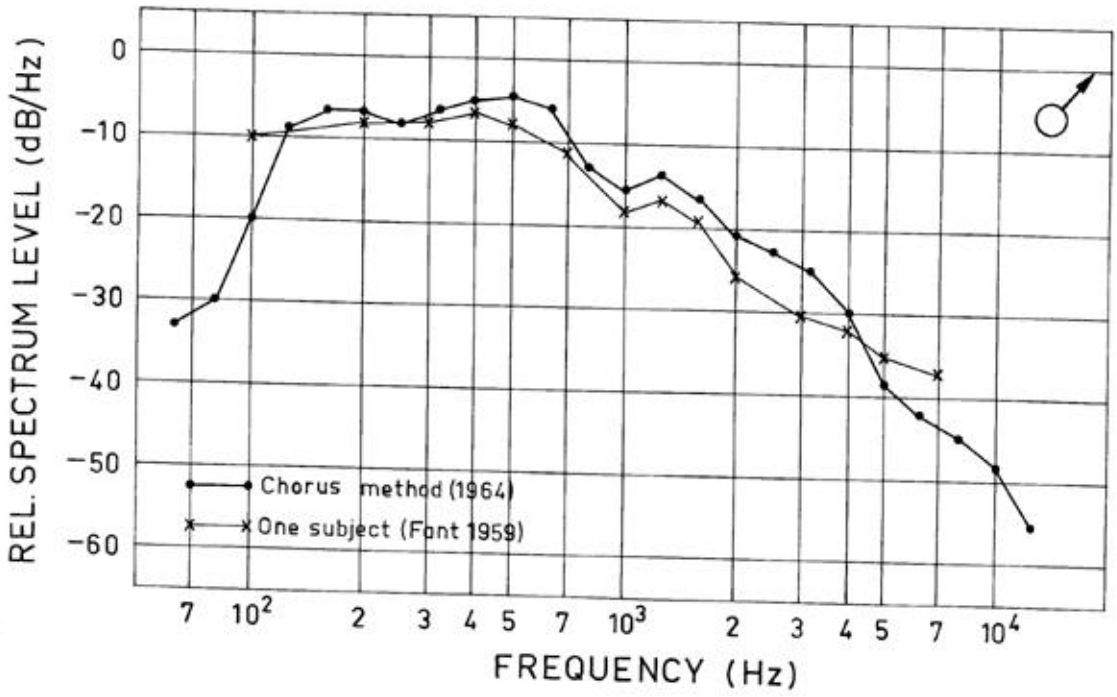

Fig. II-B-3. Average speech spectra of Swedish male subjects determined by the chorus method in 1964 compared with data of a single male subject from Fant, 1959[3].

the relative lack of energy at frequencies of 1.5 times the fundamental as in narrow band "harmonic analysis" of an isolated voiced sound.

There is one spectrum feature which appears to be independent of language and sex. This is the tendency of a spectrum minimum at about 900 c/s in male speech and at about 1000 c/s in female speech. The only exception is the German male group, where this effect could be counteracted by the specific language structure. In Fig. II-B-3 our Swedish male data are compared with an earlier measurement by Fant [2] of a single speaker. It can also be traced in most of the original data of Dunn and White [1] for English. In the recent years' studies of voice source characteristics at the STL by means of spectrum matching and inverse filtering experiments, we have observed the same kind of spectral minimum as a typical characteristic of the voice source of isolated vowel samples [3].

The most reasonable explanation of the 1000-c/s minimum in long-time average spectra is that it reflects a source feature rather than a statistical property of the distribution of the second formant frequency of vocalic sounds.

References:

(1) Dunn, H.K. White, S. D.: "Statistical measurements on conversational speech", J.Acoust.Soc.Am. 11 (1939), pp. 278-288.

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

(3) Mártony, J.: "On the vowel source spectrum", STL-QPSR-1/1964, pp. 3-4.

(4) Tarnóczy, T.: Folia Phoniat. 8 (1956), pp. 65-70; J.Acoust.Soc. Am. 28 (1956), pp. 1270-1275; Acta Phys.Hung. 7 (1957), pp. 87-105; Acustica 8 (1958), pp. 392-395.

(5) Tarnóczy, T.: "Sprachliche Verschiedenheiten erhalten durch Sprechchormethode", Paper G27, Proc. of the IVth Internat. Congr. on Acoustics, Copenhagen 1962 (Copenhagen 1963).