SOME IMPLICATIONS AND APPLICATIONS OF USING DIGITAL SOUND SPECTRA

ANTTI IIVONEN

The spectral properties of sounds are usually described (successfully) by means of direct measurement of certain acoustic variables (e.g., formant and antiformant frequencies, bandwidths, and intensities; e.g., Fant 1968). There are, however, some reasons supporting the need for a measurement method which would be independent of the acoustic variables in the ordinary sense.[1]

In spite of the varying fundamental frequency, the possibility of comparing different spectra with each other remains, when the shape of the spectral envelope is characterized by means of amplitude values which are measured at constant points on the frequency scale. Thus the envelope is defined by means of a series of numbers which indicate the amplitude at selected measurement points. The description so obtained is called a DIGITAL SPECTRUM. The method described can be applied along with (or before) the variable-dependent procedure. I have discussed the method on other occasions (first in an unpublished thesis at the University of Helsinki [1967]; later e.g., 1969 and 1970). The following points of view can be taken as comments on the method.

The method of using digital spectra is independent of the source of speech; all kinds of spectra can be treated with the same principle (e.g., voiced and whispered sounds). Computer applications of different kinds are possible. The problems and limitations of the method concern the domain of applications: ability of the method to describe the original signal, details of the measurement procedure, dynamics of the spectrum, distortions caused by the research equipment, the need for normalizing the amplitude relationships of the digital spectra, the relationship between the digital

---

[1] E.g., the following reasons can be mentioned: (1) the acoustic variables often have an *a priori* nature: one knows in advance what one should find; (2) the acoustic variables are primarily defined on the basis of resonancy; (3) in several cases the spectral peaks (formants) cannot be separated or labelled by the investigator (for difficulties of this kind see Lindblom 1962, Ladefoged 1967:81-86, Ungeheuer 1968:173; and (4) the formant structure of female voices is often not so clear as that of male voices, and this makes the measurement of formants more difficult. The goal of investigation should determine whether or not there is need to indicate the acoustic variables after the application of the method.

spectrum and the auditive response of the hearer. Many of the problems are common to other spectrometric methods as well.

Spectral comparisons of different kinds are possible. One possible application is to group together digital spectra which are acoustic representations (signals) for the same linguistic entity occurring in the same context, and which are produced by the same speaker. Indicating the minimum, maximum, and mean amplitude values at each measurement point, and plotting these statistical values (also the standard deviation can be indicated) on a coordination system, a numerical and graphic description of the acoustic variation area and the average distribution of the spectral energy in repetitions of the same linguistic unit can be presented (Figures 1 and 2).
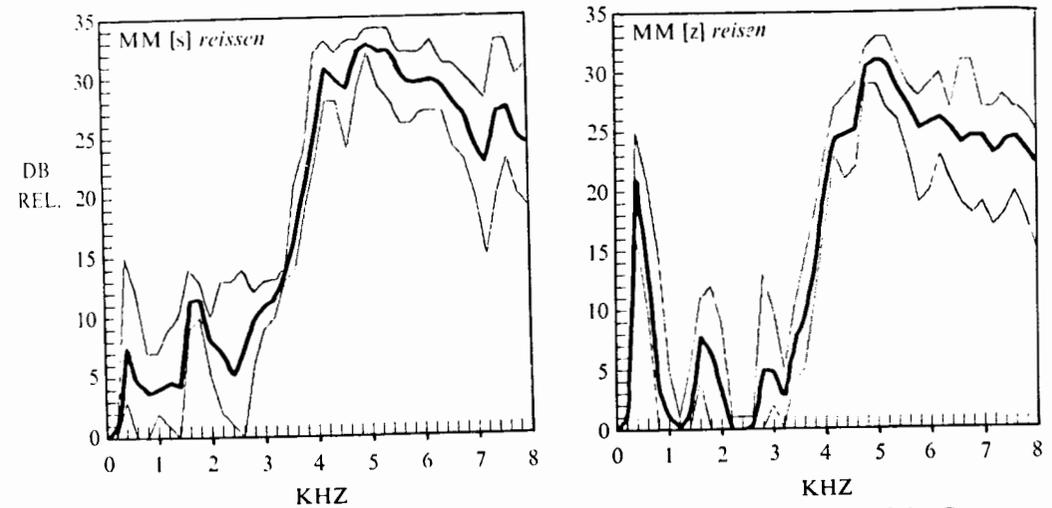

Fig. 1. Spectral variation area and average energy distribution of ten realizations of the German phonemes /s/ and /z/ occurring in the words *reissen* resp. *reisen* produced by the subject MM (male) in the sentence frame 'Nein, ich sagte ———.'. In all instances of the investigation procedure the volume regulations were kept unchanged for all utterances. The measurement (performed manually) was based on broad band sections produced with the Sound Sona-Graph model 6061-B. Localization of the sections at the temporal mid point of the sibilants. Computing of the minimum, maximum, and mean values at the 40 measurement points by means of Elliott 803. Graphic representation by means at Elliott Digital Calcomp Plotter.

Considering the auditory response, the frequency scale may be presented according to the mel or logarithmic scale. Some other operations in order to simulate the auditive response may also be applicable (see Iivonen 1970:90-104).

Coarticulation effects and individual properties of the spectrum (cf. Figures 1 and 2) can be studied.

Because of the articulatory instability it may be recommendable to use repetitions of the same linguistic unit. This may also reduce the effects of random distortions caused by the research equipment. Especially troublesome are the spurious components in the spectrum (see Lindblom 1962; Hakala and Savolainen 1966; Ungeheuer 1968:131).
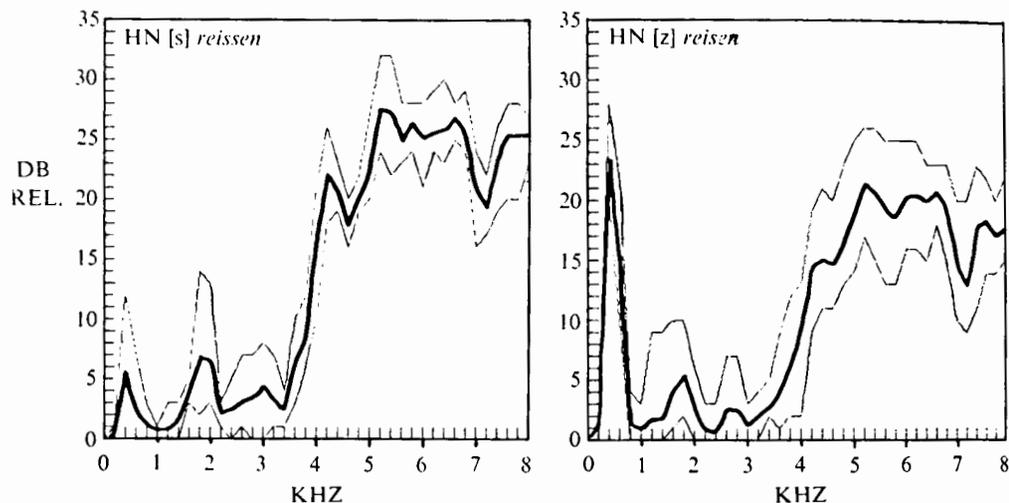
Fig. 2. Subject HN (male). Otherwise the same arrangements as in Figure 1. Comparisons concerning individual spectral differencies can be made with Figure 1.

The method is also applicable in defining the average spectral energy distribution of speech in similar sense as Tarnóczy has suggested (*Sprechchormethode*; 1962 and 1971) or in calculating the long time average speech spectrum (see Blomberg and Elenius 1970). Applications may also be found in speaker and speech recognition. It may also be possible that the speech synthesis can be based on the principle of digitalization in form of digital resonancies.

In measurement (performed manually) the sections produced with the Sound Sona-Graph can be used. In this case a digital spectrum represents a short time interval which corresponds with the time constant of the section filter of the Sona-Graph (cf. Lindblom 1962:197, Fant 1968:186). Concerning resonants (including voiced vowels, liquids, and nasals), the value of the average fundamental frequency of the subject can be used as the constant interval between the measurement points on the frequency scale. Otherwise intervals e.g., 120, 200 or 250 cps, may be applicable.

Dynamics of the section (Kay Elemetric Sound Sona-Graph model 6061-B) contains 35 dB.[2] A wider scale would make it possible that also weaker spectra and the weaker components of spectrum could be taken into account without using some auxiliary operations.

A need for normalizing the digital spectra occurs e.g., when the spectra produced by the same speaker in different occasions (when the relative speech volume can be different) should be compared with each other. One measurement point can be chosen as the reference point.

Considering an automatic procedure the localization of the sample point desired

---

[2] The same dynamic scale is used in Halle 1959 and recently in Liiv and Remmel 1970. By means of the pre-amplification the dynamic range of the spectrum can be increased (high-shape amplification; see also Liiv and Remmel 1970:8-9). The scale used by Tarnóczy 1971 contains 60-80 dB.

on the time axis is a considerable problem. Some visual aid (e.g., cathode-ray oscilloscope display or spectrogram) may be used. Some preliminary attempts towards an automatic procedure have been made at the University of Oulu.[3] Another localization problem concerns the relevant signal points (cues) on the time axis of speech, but this problem concerns the acoustic phonetics in general.

University of Oulu
Oulu, Finland

BIBLIOGRAPHY

Bell, C.G., H. Fujisaki, J.M. Heinz, K.N. Stevens, and A.S. House
1961 "Reduction of Speech Spectra by Analysis-by-Synthesis Techniques", *Journal of the Acoustical Society of America* 33:1725.
Blomberg, M. and K. Elenius
1970 "Statistical Analysis of Speech Signals", *STL-QPSR* 4.8 (Stockholm).
Denes, P.
1962 "The Use of Computers for Research in Phonetics", in *Proceedings of the IVth International Congress of Phonetic Sciences*, Helsinki, 1961 (The Hague, Mouton) p. 149.
Fant, G.
1968 "Analysis and Synthesis of Speech Processes", in *Manual of Phonetics*, B. Malmberg, ed. (Amsterdam) p. 173.
Fujimura, O.
1962 "Analysis of Nasal Consonants", *Journal of the Acoustical Society of America* 34:1865.
Hakala, E. and H. Savolainen
1966 "Messprobleme bei der Analyse von Sonagrammen", *Phonetica* 14:91.
Halle, M.
1959 *The Sound Pattern of Russian* ('s-Gravenhage, Mouton) (2nd edition, 1971).
Iivonen, A.
1969 "Akustische und apparative Variation bei linguistisch identischen Einheiten", *Phonetica* 20:131.
1970 "Experimente zur Erklärung der spektralen Variationen deutscher Phonemrealisationen", *Commentationes Humanarum Litterarum* 45 (Helsinki).
Ladefoged, P.
1967 *Three Areas of Experimental Phonetics* (Oxford University Press).
Liiv, G. and M. Remmel
1970 *On Acoustic Distinctions in the Estonian Vowel System* (= *Academy of Sciences of the Estonian S.S.R., Soviet Fenno-Ugric Studies VI*).
Liljencrants, J.
1968 "A Filter Bank Speech Spectrum Analyzer", *Speech Transmission Laboratory Technical Report* (Stockholm).

---

[3] I would like to put the comments presented here in connection with the efforts in which the frequency dimension of the speech is divided with a bank of filters in bands, and the output of each filter is submitted (after rectifying and smoothing) to an analog-to-digital converter (Bell *et al.* 1961, Denes 1962, Fujimura 1962:1868: "digital spectrogram"; Blomberg and Elenius 1970) and with some other technical solutions which try to digitalize the speech signal using e.g., a computer (Liljencrants 1968; 1/3 octave filter real-time analyzer of Brüel & Kjaer; Ubiquitous Spectrum Analyzer of Federal Scientific Corporation; Real-time Analyzers of General Radio). Theoretically the principle of the digitalization of the time-frequency-spectrum is suggested by Meyer-Eppler 1959 (see also 1969:27: "Gabor-Matrix").

Lindblom, B.
  1962 "Accuracy and Limitations of Sona-Graph Measurements", in *Proceedings of the IVth International Congress of Phonetic Sciences*, Helsinki, 1961 (The Hague, Mouton) p. 188.
Meyer-Eppler, W.
  1959 *Grundlagen und Anwendungen der Informationstheorie* (Berlin, Göttingen, Heidelberg), (2nd edition, 1969).
Tarnóczy, T.
  1962 "Über das individuelle Sprachspektrum", in *Proceedings of the IVth International Congress of Phonetic Sciences*, Helsinki, 1961 (The Hague, Mouton) p. 259.
  1971 "Das durchschnittliche Energie-Spektrum der Sprache (für sechs Sprachen)", *Acustica* 24:57.
Ungeheuer, G.
  1968 "Systematische Signaldestruktion als Methode der psychoakustischen Phonetik", *Phonetica* 18:129.