COMPARISON OF TEMPORAL AND FREQUENTIAL METHODS OF SPEECH DATA BASE LABELLING

G Pérrenoux, N Vigouroux(1), L J Boe(2), G Caelen-Haumont(2), D Autesserre(3), D Fohr(4)

(1) CERFIA, (2) IGP, (3) IPA, (4) CRIN-INRIA

ABSTRACT

French laboratories of GRECO-PRC "Dialogue Homme-Machine" are involved in "Speech Input and Assessment Methodology", the Esprit Project SAM nr 2589.

For labelling of Speech Data bases, four of them are mainly concerned: CERFIA (Toulouse), IGP (Grenoble), IPA (Aix), CRIN (Nancy).

Based on knowledge cumulated under the auspices of the GRECO "Communication Parlée" for labelling of BDSON "Base de Données des Sons du Français" two approaches have been developed with different underlying theoretical conceptions and editing tools:

1) A frequential one based on an expertise of well known sonagramic display (CRIN) or spectral representation (CERFIA, IPA) delivered at the output of a hearing model including a 24 channel filter bank;

2) A temporal expertise of the signal (IGP) based on an articulatory interpretation of speech production.

The aim of this paper is to compare the two spectral approaches and to discuss them in relation to the temporal method.

The data used are the 23 sentences uttered by 2 French speakers (male and female) of the CD ROM EUROM 0. After presentation of criteria adopted by the different laboratories for center labelling and/or detecting of discontinuities between sounds, will be discussed for different categories of sounds (plosives, fricatives, sonorants, glides and vowels) consequences of the selected options on accuracy reliability, time consuming, possibility of automization, according to the aims (evaluation of speech recognition, synthesis, acoustical analysis) for fundamental research and/or applications.