

## A STOCHASTIC MODEL OF PHONEMIC PATTERNS IN SPOKEN ITALIAN

U. Bortolini<sup>\*</sup> - F. Degan<sup>\*</sup> - C. Minnaja<sup>\*</sup> - L.G. Paccagnella<sup>\*,\*</sup>,  
 Centro di Studio per le Ricerche di Fonetica (C.N.R.), Padova, Italy  
<sup>\*</sup>Istituto di Matematica Applicata, Università, Padova, Italy

Many situations in automatic speech recognition/understanding require decisions which have often to be made on the basis of incomplete or uncertain information. Stochastic modeling is a flexible general method for handling such situations. It consists of employing a specific probabilistic model which helps in uncertainty or incompleteness of the information. In this paper a specific class of stochastic model is discussed - models based on the theory of Markov processes.

The order of the source has been fixed on the basis of the average length of phonetic words. Transitional probabilities are estimated from the occurrences of sequences of two and three phonemes. Such statistics are evaluated from a corpus of spoken Italian, consisting of 49.533 phonemes, derived from 7.667 phonetic sequences.

The construction of the model is very simple, and it is based on drawing out random numbers. The sequences generated by our model point out phonotactic restrictions which are peculiar of the Italian language. The phonetic and syllabic structure of the sequences obtained reflect with good approximation structures which are the most frequent in the natural language.

First some statistics are exhibited and the proprieties of the general model are discussed; then we considered some examples of situations in automatic speech analysis in which such a model can be applied.

#### References

- Baker, Y.K. (1975): "Stochastic modeling for automatic speech understanding" in Speech Recognition, Reddy D.R. (ed.), 521-542, New York: Academic Press.
- Bortolini, U., Degan, F., Minnaja, C., Paccagnella L., Zilli, G. (1978): "Statistics of spoken Italian" in Proc. Ling. 12.