EUR-ACCOR:TIIE DESIGN OF A MULTICHANNEL DATABASE*

ACCOR: A.Marchal, W.Hardcastle, P.Hoole, E.Farnetani, A.Ni Chasaide, O.Schmidbauer, I.Galiana-Ronda, O.Engstrand, D.Recasens.

CNRS, Aix-en-Provence, University of Reading, Ludwig Maximilians Universität, München, CNR, Padova, University of Dublin, Siemens AG, München, Universidad Politecnica de Valencia, Stockholm Universitet, Universitat Autònoma de Barcelona.

ABSTRACT

The EUR-ACCOR database has been designed tó allow access and analysis of articulatory, aerodynamic and acoustic characteristics of seven European languages: French, English, German, Italian, Catalan, Swedish and Irish. The data are acquired using a PC-based multi-channel system developed in Reading (UK) and Aix (France), which enables simultaneous recording of up to 16 input channels. The design of the corpus, consisting of (1) structured VCV nonsense items, (2) real words matching the structure of (1), and (3) short sentences, illustrating the main connected speech processes in the different languages, is discussed with reference to the main scientific goals of the ACCOR project.

1.INTRODUCTION

The ACCOR project aims to provide a detailed description of the main articulatory and acoustic correlations in coarticulatory processes. A cross-language approach is adopted as a means of identifying, on the one hand, the major language-independent universal regularities of these processes (due to factors such as the mass, inertia and elasticity of the speech organs, the mechanical linkage between them and the neuromuscular complexities of the cranial nerve system) and also how these interact with language specific factors such as the phonological rules of the language.

In carrying out such an investigation, activities of the main physiological systems underlying speech production are investigated: the respiratory system (producing a flow of air), the laryngeal system (modifying the airflow by the valving mechanism of the vocal folds) and the complex system of supraglottal structures in the mouth and nose, such as the tongue, lips, jaw and soft palate which shape the vocal tract into different resonating cavities The ACCOR project aims at a detailed description of the complex coordination between these different structures and the resulting acoustic output. Specific articulatory processes in the seven languages will be examined with a view to determining how such processes differ according to the different phonological systems. Also, it will be possible to examine the functions of the different motor subsystems in the same speaker.

The cross-language nature of the project has meant considerable attention has had to be paid to the design of a suitable database. Three main scientific goals of the project were considered in the design:

[i] to allow the testing of several hypotheses regarding articulatory/phonological interactions in the different languages.

[ii] to investigate fundamental issues in speech production theory particularly those relating to coarticulatory processes.

[iii] to provide a unique resource for researchers in speech science and related fields.

2. DESIGN CONSIDERATIONS FOR DATABASE

2.1. Testing Articulatory/Phonological Interactions

One of the main design considerations for the database has been to provide a rich enough corpus to allow the testing of various different hypotheses regarding interactions between articulatory characteristics and phonological patterns in the different languages.

The seven languages under investigation differ considerably in their vowel and consonant inventories, and it seems reasonable to expect that the structure of such inventories may constrain the type and extent of coarticulatory influences. For example, comparable consonants in the different languages may differ in their coarticulatory patterns because of the presence of other phonemes in the language. Italian, for example, contrasts alveolar with palatal nasals, and alveolar with palatal laterals; the two alveolars may be more resistant to the coarticulatory effects of a "palatal" vowel such as /i/, than another language where the possibility of perceptual confusion does not exist. Similarly, French, with the presence of nasal vowels in its inventory may exhibit less coarticulatory nasality than other languages without nasal vowels, such as German. It may be possible to formulate a general "density" hypothesis; coarticulatory processes vary inversely with the number of elements in the phonological class.

The database should allow the testing of more general hypotheses also such as :

- do some languages show greater carry-over/anticipatory coarticulation than others?
- does the same subject show different anticipatory/carryover effects for different motor subsystems (lip, jaw, tongue, velum)?
- do all languages exhibit a tendency towards "instability" of alveolar stops in connected speech?
- do all sounds show the same degree of contextually induced variability, or is there an increasing scale, of the order, say, of /s,t,d,n,l/, valid for all languages?

The VCV context would seem most appropriate for examining coarticulatory effects, as it is possible to study a variety of transitional phenomena in such a sequence: vowel-to-vowel effects, VC and CV effects. This context therefore forms the basis of the corpus. Another requirement for testing the above hypotheses is the inclusion of a range of sounds involving the different motor subsystems under investigation as well as voiced/voiceless distinctions. Thus, labial sounds such as /u/, /p/, /b/ were included as well as oral lingual consonants differing in both place and manner of articulation(e.g. /t,d,k, ʃ ,s,z,tʃ,/) and nasals such as /n/.

2.2 Speech production theory

The database has been designed to allow the investigation of fundamental issues in speech production theory. One area of considerable theoretical significance, for example, is the question of inter-gestural timing for consonants and the corpus has been designed to allow for the investigation of this aspect. The /kl/ sequence seems a particulary useful item to use for studying timing relationships, such as for example, between the tongue body gesture for the /k/ and apical contact for the /l/. There is preliminary evidence that the timing between the body and tip of the tongue in this sequence is language specific and may be related to prosodic aspects such as rhythm, syllabic structure and patterns of stop aspiration and voicing. Another area of significance is a specification of both acoustic/articulatory relationships in complex sounds such as fricatives. This

area is currently under investigation as part of Workpackage 2 of the project.

2.3 Unique resource for speech science

Recently a large number of digitized acoustic databases have been compiled for various purposes, (e.g. ESPRIT/SAM EUROM.O [1]; DARPA/TIMIT [2]). The EUR-ACCOR database is unique in that it includes articulatory and aerodynamic (nasal and oral air flow) as well as acoustic data in a number of different languages. The requirements of the database are such that it is necessary to simultaneously investigate the activity of different motor subsystems of speech, the resulting movement trajectories, aerodynamic effects and the acoustic output. To achieve this aim, the multi-channel data acquisition systems, EDIT and PHYSIOLOGIA have been developed at Reading [3] and Aix [4] respectively.

3. CORPUS

The study of coarticulation and its possible implications for the linguistic and neuromotor programming processes would seem a priori to exclude the use of nonsense items. On the other hand, although undoubtedly desirable, the analysis of spontaneous speech is problematic in that it is difficult to control all the parameters which may be responsible for coarticulatory effects. The emphasis on cross-language comparison makes this problem even more acute. In the ACCOR project we decided the best compromise was to work in parallel on three different types of speech material; structured VCV nonsense items, real words matching the nonsense items in their VCV structure as closely as possible, and connected speech in the form of a series of short sentences.

In choosing the items for inclusion in the corpus we were constrained by a number of criteria. Firstly, all items, including the nonsense items, had to comply with the phonological rules of the language. For vowel sounds, it was desirable to cover a range of different tongue postures and different perceptual characteristics. The three "extreme" vowels /i/, /a/ and /u/ were chosen as these vowels exist as phonemes in each language, and their phonetic manifestations in the different languages are similar. For the consonants, a range of different places and manners of articulation were chosen. It was decided to focus on VCV sequences for the reasons stated above. This raised the problem of stress placement. Obviously, it would have been preferable if the stress placement were uniform for each language. Unfortunately, however, language specific stress placement rules precluded this possibility if the items were to be produced "naturally". In all cases, except French and Swedish, stress was placed on the first syllable. In French the more natural placement is on the second syllable and in Swedish there is equal stress on each syllable. Details of the three corpora are as follows:

3.1. Nonsense items

Vowels /i,a,u/ in isolation. VCV sequences, where C = /p,b,t,d,k,s,z,n,l, ʃ tʃ / and the sequences /kl,st/; V = /i/, /a/ (/ ə / when unstressed in English) and /u/. These items are phonotactically permissable in all the languages under investigation, with one or two exceptions, (e.g. /z/ and / ʃ / do not occur in Swedish).

3.2. Real words

These match the VCV nonsense sequences in section 3.1. as closely as possible. Thus nonsense item /iti/ is matched by English "meaty", /uti/ in Italian is matched by "muti" etc. It was not possible to obtain a complete set of matching real words for all the languages.

3.3. Sentences

A set of short sentences was constructed in each language to illustrate the main connected speech processes in that language. Thus in English, sentences such as "Fred can go, Susan can't go, and Linda is uncertain" have numerous examples of word final alveolar stops and nasals which many speakers would assimilate into the place of articulation of following velars when spoken naturally.

Systematic comparison of data from the three types of recordings seems to us indispensable to take account of the specific nature of the processes involved and to resolve the inherent theoretical problems in establishing a multi-lingual corpus such as this. Moreover it should enable us either to show clearly the limitations of a single corpus approach or to vindicate the exclusive use of nonsense items. We are finding, in the initial results, that there are interesting and predictable differences between nonsense and real words, at least on some articulatory measures.

The items in 3.1, 3.2 and 3.3 constitute a core corpus which is read 10 times by each of the 10 speakers in the seven languages. Of the 10 repetitions, five are carried out with the Electropalatograph (EPG), Laryngograph and audio recording (sampled at 20 KHz), and five with EPG, nasal volume velocity of air, oral volume velocity, (using a Rothenberg mask and pneumotach), Laryngograph and audio recording. The core corpus will constitute approximately 300 Mbyte per speaker. In addition to the core corpus, a number of additional items are added for each language to illustrate particular language-specific features, e.g. geminates in Italian, front rounded vowels in German, palatals in Catalan, palatalisation in Irish Gaelic etc. The EUR-ACCOR database is well on schedule; approximately 5 speakers in each language have already been recorded and the first phases of data processing is nearing completion.

REFERENCES

[1] A.J.Fourcin, G.Harland, W.Barry and V.Hazan (Eds), "Speech input-output assessment; multi-lingual methods and standards", Chichester: Ellis Horwood, 1989.

[2] DARPA/TIMIT "Acoustic phonetic continuous speech database". CD-ROM Produced in the USA by PDO, 1988.

[3] A.Trudgeon, C.Knight, W.Hardcastle, G.Calder, and F.Gibbon, "A multi-channel physiological data acquisition system based on an IBM PC, and its application to speech therapy", Proceedings of Speech 1988 (7th Fase Symposium, Edinburgh: Institute of Acoustics, pp 1093-1100, 1988.

[4] B.Teston and B.Galiando "Design and development of a workstation for speech production analysis", Proceedings of VERBA Conference, Rome, January, 1990.

ACKNOWLEDGEMENT

We acknowledge financial support from The European Economic Commission DGXIII under the auspices of ESPRIT II Basic Research Action.

*The paper is presented on behalf of the ACCOR consortium