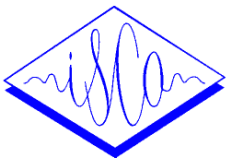

2nd International Conference on
Spoken Language Processing (ICSLP 92)
Banff, Alberta, Canada
October 12-16, 1992

THE PSH/DISPE HELIUM SPEECH CDROM

A. Marchal, C. Meunier* and P. Gavarry***

* Laboratoire Parole et Langage. URA 261 CNRS, Université de provence
13621, Aix-en-Provence, France.
** Institut National de Plongée Professionelle, Port de la Pointe Rouge
13008, Marseille, France.

10.21437/ICSLP.1992-412

ABSTRACT

Gas mixture and pressure modify the spectral characteristics of diver's speech. Additionnally, constraints imposed on jaw movements by wearing a facial mask affect the speech production process. The auditory feedback loop is equally concerned. Furthermore, underwater adverse working conditions are characterised by noise from different sources. As a result, divers' speech is poorly intelligible and communications between divers and surface control need to be enhanced. To this end, «voice unscramblers» are being used. However, the technological state of commercially available equipment is dated and the quality of speech remains insufficient.

To help with the design, testing and qualification of new communication devices, a bilingual (French-English) Database is set up. It consists of phonetically balanced lists of 200 words read by 17 divers under sea and in chambers at operationnal levels from the surface to 300m.These recordings have been edited, labelled and then stored on a CD-ROM.

The PSH/DISPE CDROM complies with the SAM EUROPEC standards (ESPRIT-SAM project n°2589). It consists of acoustic signal files (40 KHz, 16 bits) of hyperbaric speech and of associated files. These later ones provide informations about the speakers and the recording conditions.

INTRODUCTION

The use by divers of respiratory mixtures such as Heliox (helium and oxygen) has solved the physiological problem of the toxicity of gases under high pressure, thus allowing for deeper and longer dives. However, the synthetic atmosphere and the depth-generated pressure increase causes a decline in the intelligibility of the diver's speech [1]. Additional perturbations are caused by the noise of bubbles, ventilation of the chamber, lamination of gases, and deformations due to facial masks which are poorly suited to speech production.

Communication plays an important role in diving, in conditions of low visibility, difficult work, or a threatening environment, and above all, whenever an equipment breakdown would put the diver's life in danger. Communication between the underwater site, the diving bell, and the surface must be improved, not just for safety reasons but also for greater effectiveness, productivity, and profitability of subaquatic projects.

To answer to the need for better communication, the Speech and Language Laboratory of the French National Research Council in Aix-en-Provence and the Institute of Professional Diving in Marseille have collaborated in developing a bilingual French-English database of underwater and high-pressure speech. The Subaquatic and Hyperbaric Speech Database, abbreviated PSH/DISPE (Parole Subaquatique et Hyperbare/Divers' Speech) fulfills the twofold need of industry and research for a database of sounds that can be used to develop new hyperbaric speech unscrambling techniques and assess underwater vocal communication systems.

This article presents the PSH/DISPE database, focusing on its structure and processing environment.

1. DATABASE CONTENT

1.1. Corpus

The corpus consisted of:

(i) Words: four lists of 46 French words [2] and four lists of 50 English words [3], generated in compliance with M.R.T. test requirements.

(ii) Sentences: eight phonetically balanced sentences in French (table 1).

1. C'est un peu fade comme pâte, il faudrait un zeste de citron.
2. Emile a de la chance, car il fait rire les gens.
3. A pâques, nous irons à Cagnes ou dans le Var.
4. Marie a une bague qui va bien à son doigt.
5. Le chien de la meute est dans sa niche.
6. Il est bête comme ses pieds, mais il a du talent.
7. L'âne porte sur son dos un ballot de foin.
8. Tous les ans, on envoie mille tonnes de riz au Mali.

Table 1: Phonetically balanced French sentences.

(iii) A paragraph: one phonetically balanced paragraph in English («The Rainbow»).

1.2. Recordings

The corpus was pronounced twice by each of 17 speakers. Three recording conditions were set up:

(i) in a quiet environment before a dive (i.e. in air)

(ii) during a dive (i.e. under water)

(iii) in the chamber (i.e. during decompression).

In case of error (e.g. forgotten or poorly pronounced words, substitutions), the affected words were repronounced at the end of the recording session.

The recordings using a SONY DAT TCDD10 were done in simulated diving conditions at 60 m, 84 m, 100 m, 150 m, 180 m, 200 m, 250 m, and 300 m (chamber and pool).

2. DATA BASE ARCHITECTURE

The database complies with the SAM EUROPEC standards [4].

The EUROPEC system (devolopped at ICP, Grenoble) [5] automatically creates a record containing one word or sequence each time it detects an intensity level above a predefined cutoff.

Hyperbaric speech data has been put into Database format by re-recording using the EUROPEC software. This operation involves two steps:

(i) First, specification files must be set up containing information about the data in the database. This information concerns the speakers (see SPEAKERS.DBF in table 2), the corpus (CORPUS.DBF, *.TXT), and the recording conditions (see *.RCD in table 3).

D
-
SCD: DC
SNM: DUPUIS
SBN: Charles
SEX: M
DOB: 1959
HET: 1,90
WET: 89
NLN: Français
ACC: Standard
ETH:
EDL: Secondaire
SMK: non fumeur
PTH:

Table 2: The speaker files specify sex, age, height, spoken language, name, accent, education, smoking habits, and any speech-related disorders for each speaker.

SCD: start of conditions
RCC: 180PiscMN
VER: V3.0
VOL:
DIR:
SNB: 2
SBF: 01
SSB: 16
NCH: 1
LGG: 0
PCN: 1
SAM: 40000
MIN: micro_name char(10)
FLT: filter char(10)
PRO: ext_proc char(10)
ENV: Piscine
DPT: 180m
GAZ: Héliox
PPO: 570 mb
RO2: 3
RHE: 92,7
TGZ:
TAQ: 5
MSK: X-LITE
ECD: end of conditions

Table 3: An example of a recording conditions file.

(ii) Then the recordings *per se* are processed. This step involves defining a configuration file for each recording (sampling frequency, type of input, cutoffs for triggering the recording, etc.). Given the high level of noise (respiration, bubbles) of our recordings, it was necessary to segment manually the data and to introduce silent intervals between words. Each record in a given signal file (*.WFS) is associated with its graphic transcription in a corresponding «orthographic» file (spelling; *.WFO).

The PSH/DISPE files are classified according to the language spoken (French or English), the ambient pressure (atmospheric = 1 bar, or hyperbaric > 6 bars), the recording location (chamber or pool) and the speech production depth (see Fig. 2).

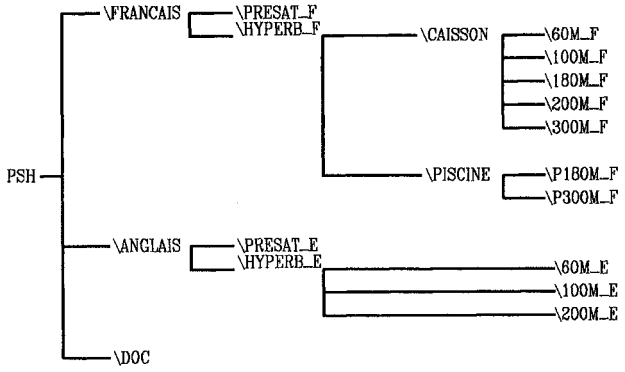

Fig. 2: Directory structure of the PSH/DISPE database.

A given corpus, speaker, sentence, word, or sequence of phonemes can be accessed under DOS by stating the corresponding file name in the lowest level sub-directory, or by running the GERSONS system, which uses the phonetic or orthographic transcription to access the desired sequence.

The file naming system complies with the recommandations of ESPRIT SAM project. The names of the orthographic files are identical to the names of the corresponding signal files except that the letter "S" in the extension is replaced by the letter "O" (see Fig. 3).

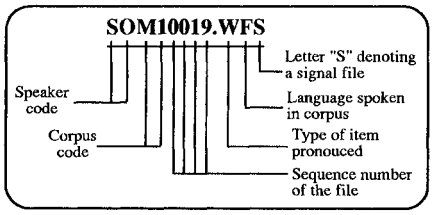

Fig. 3: Europec file naming recommandations.

Finally, each signal file is associated with a configuration file (data pertaining to sampling frequency, re-recording cutoffs, signal onset, etc.). The name of a configuration file is identical to the name of the corresponding signal file, except that the extension is CFG. For example, the name of the configuration file associated with the SOM10019.WFS signal file is SOM10019.CFG.

CONCLUSIONS AND PERSPECTIVES

Again, a database is not an end in itself. The objective is to make the database into a knowledge base.

The parameters that affect speech production in a hyperbaric environment are varied, and prior research has not dealt with them in a systematic manner.

In order to solve the vocal communication problem posed here, a systemic approach appears necessary. By this, we mean:

(i) The entire communication chain should be considered, from the time when the vocal emission is produced by the speaker until it is received by the listener.

(ii) An exhaustive list should be drawn up of all possible communication conditions between divers and the bell, among divers in the chamber, and between the chamber and the surface.

(iii) It appears crucial to examine the effects of each component of the communication chain. These include the microphones, the dead volume in the cavity created by the face mask, transmission lines (which often deteriorate under water), amplifiers, etc. In short, all communication equipment used at any point and at any work station must be considered.

In this sense, the PSH/DISPE fulfiles the need of both basic Research and proves a useful tool in the design and testing of new speech unscrambling techniques.

Acknowledgements. We would like to express our gratitude to M.H. Casanova who did most of the recordings at the Institute of Professional Diving (INPP) and at the French Navy Hyperbaric Center in Toulon (GISMER), and to the divers who volunteered to participate in the project. Our thanks go also to J.F. Sérignat and J. Zeilinger at the I.C.P. in Grenoble who trained us in the use of the SAM EUROPEC software. This research was funded by grants from CEP&M, Région Provence-Alpes-Côte-d'Azur, D.R.E.T., I.N.P.P., and the C.N.R.S.

REFERENCES

[1] H. Hollien and J.W. Hicks. "Helium/pressure effects in speech: updated initiatives for research." Proceedings of the IEEE Acoustic Communication Conference, Washington D.C. D4, pp.1-26, 1982.

[2] A. Marchal and C. Meunier. "Intelligibilité de la parole en milieu hyperbare et/ou subaquatique", Actes du congrés ORIA 91, Marseille, 1991.

[3] J.D. Griffiths. "Rhyming minimal contrasts: a simplified diagnostic articulation test." J. Acoust. Soc. Am. 42, pp.236-241, 1967.

[4] A.J. Fourcin, G. Harland, W. Barry and V. Hazan. Speech Input and Output Assessment. New-York: edited by Ellis Horwood, collection E. H. Books in Information Technology, 1989.

[5] J. Zeiliger and J.F. Sérignat. "EUROPEC V 4.0: a speech database recording software. User's guide." ESPRIT PROJECT 2589 Periodic Progress Report, Grenoble, 1990.