3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

DIVERS' SPEECH: VARIABLE ENCODING STRATEGIES

A. Marchal and C. Meunier

Laboratoire "Parole et Langage", URA 261 CNRS
29, av. Robert Schuman
13621 Aix-en-Provence
FRANCE

10.21437/Eurospeech.1993-120

ABSTRACT

In subaquatic and hyperbaric communication, several causes of distorsion have been identified: physical effects related to gas density and increased ambiant pressure. The present study investigates an other source of variation, i.e. the speaker himself who tends to adopt also specific strategies to compensate for the loss of intelligibility.

Keywords: Helium Speech, Production, adverse environment, variability.

1. INTRODUCTION

Underwater operations are currently conducted at very deep levels . A number of factors adversely affect the speech production process. These include the density of respiratory gas mixtures, increased ambiant pressure, constraints imposed by wearing a facial mask, perturbations of the auditory feedback loop... not to mention the hostile underwater environment itself. In addition, it has also been reported that divers may under certain circumstances experience stress. Consequently, diver's speech is poorly intelligible.

An acoustic analysis of the following parameters has been carried out for 5 situations and 2 divers: F1, F2.

As expected, we observe differences which are related to the physical parameters of the dives; but more interestingly, differences which would reflect different individual divers' responses to a given situation. Divers adopt in fact variable encoding strategies to overcome the problems caused by the "adverse" environment and to enhance their speech intelligibility. The effectiveness of the various strategies which have been used are discussed against the results of intelligibility tests.

2. THE PSH/DISPE CDROM

To help with the design, testing and qualification of new communication devices, a bilingual (French-English) hyperbaric speech database has been set up: The PSH/DISPE CDROM. It consists of lists of words and short sentences recorded by numerous divers at various depths from 0 to 300 meters and for different situations.

(i) Words: four lists of 46 French words [1] and four lists of 50 English words [2], generated in compliance with M.R.T. test requirements.

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

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

This database complies with the SAM standards and is available on a CDROM.

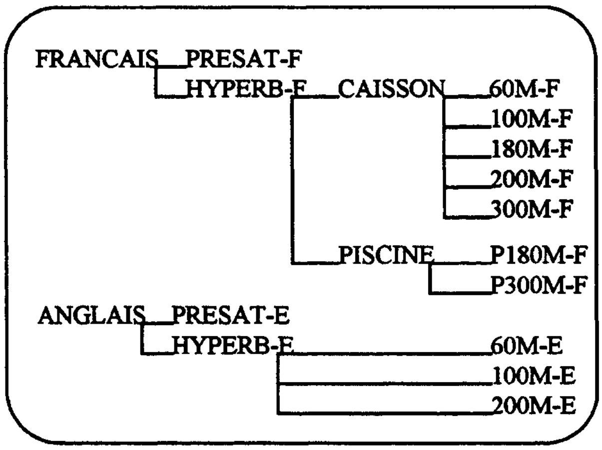

Figure 1: this is a hierachical table which shows the content and the directory access to the PSH/DISPE database For this study: PRESAT-F, 180M-F, 300M-F, P180M-F, P300M-F.

3. ACOUSTIC ANALYSIS

For the purpose of this study, the database management system GERSONS was used [3] to extract of words containing the vowel /i/: mille - pile - qui - fiche - mig - bile - gui - quiche.

The speakers were recorded in two diving conditions: in a pressure chamber and in simulated dives at the following dephts: 180 and 300 meters. We also consider 5 types of communications:

- 0 meter (atmospheric pressure, air)
- 180 meters, heliox in chamber
- 180 meters, heliox in pool
- 300 meters, heliox in chamber
- 300 meters, heliox in pool

The speakers under study were selected due to their recognised lack of regional accent: speaker codes are: GD and SO.

PTS software [4] was used to compute F1 and F2 values.these values were manually measured at the center of the stable part of the vowel by a traineb phonetician.

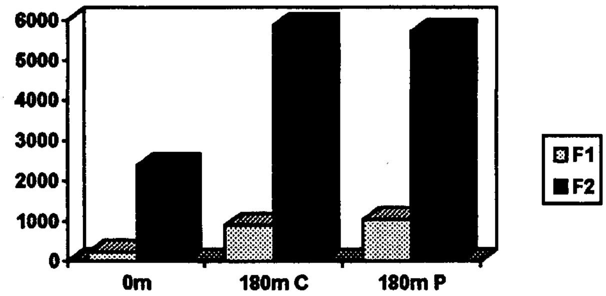

Figure 2: F1 and F2 values in the three different situations for speaker GD. We see an important increase of the both formants between 0 and 180m.

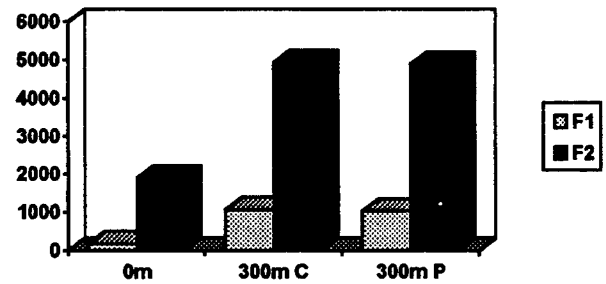

Figure 3: F1 and F2 values in the three different situations for speaker SO. We see an important increase of the both formants between 0 and 300m.

4. RESULTS

4.1. Descriptive statistics

Table 1: General results (means, standard deviations, etc) of the analysis of speaker GD formants.

GD speaker	F1 0m	F2 0m	F1 180m C	F2 180m C	F1 180m P	F2 180m P
Mean	226.125	2382.000	905.500	5855.000	1037.500	5699.500
Std. Dev.	34.676	36.348	35.813	59.792	46.865	229.182
Std. Error	12.260	12.851	12.662	21.140	16.569	81.028
Count	8	8	8	8	8	8
Minimum	201.000	2348.000	872.000	5771.000	1006.000	5302.000
Maximum	268.000	2416.000	939.000	5973.000	1140.000	6075.000
# Missing	0	0	0	0	0	0

Table 2: General results (means, standard deviations, etc) of the analysis of speaker SO formants.

SO speaker	F1 0m	F2 0m	F1 180m C	F2 180m C	F1 180m P	F2 180m P
Mean	167.500	1895.750	1064.750	4932.500	1022.750	4890.625
Std. Dev.	35.813	59.389	66.706	107.439	47.376	75.441
Std. Error	12.662	20.997	23.584	37.985	16.750	26.673
Count	8	8	8	8	8	8
Minimum	134.000	1812.000	1006.000	4765.000	939.000	4832.000
Maximum	201.000	2013.000	1208.000	5100.000	1073.000	5033.000
# Missing	0	0	0	0	0	0

The results indicate, as might have been anticipated, a large increase in formant values from 0 to 180m (fig. 2) and from 0 to 300m (fig. 3).As reported by the literature [5], this increase is due to the combined effect of the gas density and the increased ambiant pressure.

4.2. F1/F2 ratio

We calculated the increase in formant values to give an idea of the non linearity of increase between the two formants:

GD speaker:

- F1 (180m C) = **4** * F1 (0m)
- F1 (180m P) = **4.59** * F1 (0m)

- F2 (180m C) = **2.46** * F2 (0m)
- F2 (180m P) = **2.39** * F2 (0m)

SO speaker:

- F1 (300m C) = **6.37** * F1 (0m)
- F1 (300m P) = **6.11** * F1 (0m)

- F2 (300m C) = **2.60** * F2 (0m)
- F2 (300m P) = **2.58** * F2 (0m)

We can note an increase of F1 proportionnally more important for SO than for GD.

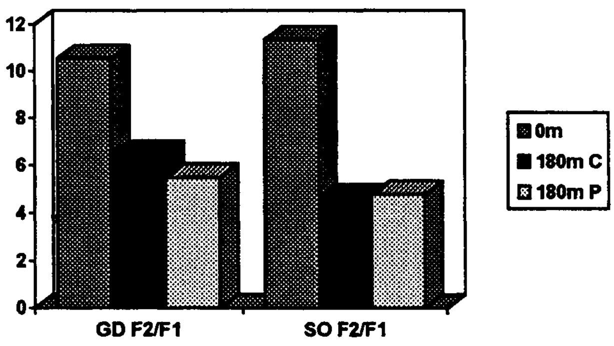

Figure 4: F2 is divided by F1. We can see the ratio for the five situations and for the two speakers (GD and SO). The ratio between F1 and F2 is more important at 0m than at 180 and 300m. This significate that F1 increases proportionnally more than F2.

It is interesting to note that this increase does not affect linearly F1 and F2 (fig. 4). This is a larger increase in F1 than in F2.This effect has been linked mainly to the increase in pressure, whereas the change in gas density modifies in the same proportion both formants [6] [7].

4.3. Speaker strategy

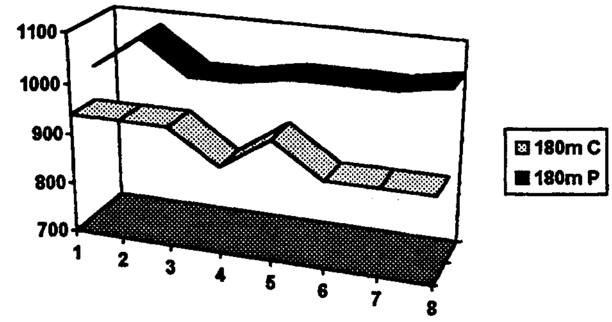

Figure 5: Speaker GD: values for F1 at the same depth (180m) but in two different situations (chamber C and pool P). The values are presented for the 8 words. This speaker significantly increase F1 from chamber to pool.

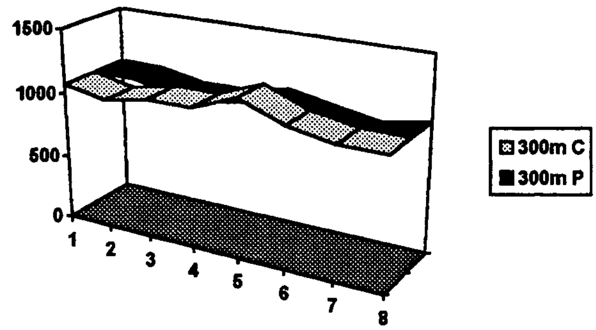

Figure 6: Speaker SO: values for F1 at the same depth (300m) but in two different situations (chamber C and pool P). The values are presented for the 8 words. This speaker does not change significantly F1 from chamber to pool.

We observe that one diver modifies significantly his speech production when wearing a facial mask (fig. 5). Is this a delibarate strategy or is it only the manifestation of constraints imposed by the mask poorly suited to this diver's facial morphology?

The changes from normal speech to hyperbaric speech are not identical from the speaker to the other. If the physical factors alone could explain the transformation, we would have expected the same effects to produce the same consequences. This is definitely not the case. This means that the speakers use in fact variable strategies to encode and to produce speech in this type of adverse environment.

This observation corresponds somewhat to what can also be found for speech produced in a noisy environment (Lombard effect).

CONCLUSION

The implication for diver's speech unscrambling and the enhancement of diver's speech intelligibility are numerous. Two directions for research are proposed: either to adapt the correction algorithm to a given speaker, or to train all speakers to cancel this "variable" effect. Further research is needed to assess the faisability of any of these approaches. From a more general pont of view, the investigation of speech produced in adverse conditions should contribute to a better understanding of the interaction between cognitive and sensory processes in speech production.

REFERENCES

[1]: Marchal, A., Casanova, M.H., Gavarry, P. and Avon, M. (1990). "Dispe: a diver's speech data-base", Third Australian International Conference on Speech Science and Technology, Melbourne, 452-457.

[2]: Griffiths, J.D. (1967). "Rhyming minimal contrasts: a simplified diagnostic articulation test," 42, 236-241.

[3]: Foulard, P., Sérignat, J. F. (1991) "GERSONS: Système de gestion de la Base de Données des Sons du français. Manuel utilisateur," GRECO-PRC "Communication Homme-Machine", Rapport interne, I.C.P. (Grenoble).

[4]: Caeroux, J. C., and Dolmazon, J. M. (1990). "PTS Software V 4.21. User's Manual," ESPRIT PROJECT 2589 Periodic Progress Report, june, (Grenoble).

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

[6]: Crestel, J. (1987). *Contribution à l'amélioration de l'intelligibilité de la communication orale en plongée hyperbare: modélisation de la source glottique*, Thèse de Doctorat de l'Université de Rennes I.

[7]: Guitton, M. (1987). *Contribution à l'amélioration de l'intelligibilité de la communication orale en plongée hyperbare: modélisation du conduit vocal*, Thèse de Doctorat de l'Université de Rennes I.