Categorical Perception and Difference Limens in Helium-Oxygen Speech

Kolbjørn Slethei

Department of Linguistics and Phonetics
University of Bergen
N-5007 Bergen

ABSTRACTS

Speech distorted by helium in the breathing gas, as is the case in saturation diving at depths below 50 msw, is rendered unintelligible by an upward frequency shift of spectral components.

When responents label spectras of linear interpolations between /i:/ and /e:/ spoken in air, we get a categorical transition in the vowel continuum. Interpolation spectras from 54, 120 and 300 msw are being categorised with decreasing accuracy, while the respondents' ability to label appropriately increases from 300 to 500 msw.

The difference limens (DL) for F1 and F2 for the vowel /i:/ have been investigated for the same depths. DL for F1 remains relatively stable, with a rise from 300 to 500 msw. DL for F2 is raised from 0 to 300 msw, and lowered from 300 to 500 mws.

These findings will be discussed.

INTRODUCTION

In saturation diving the nitrogen and most of the oxygen in the breathing gas is replaced by helium. Table 1 lists typical compositions of breathing gases at various depths, in this case from an experimental dive in pressure chambers at The Norwegian Underwater Technology Center (NUTEC) in Bergen, Norway. (The small quantity of $N_2$ is a consequence of an unintentional contamination.)

| | Depth (msw): 0: | 54: | 120: | 300: | 500: |
|---|---|---|---|---|---|
| Pressure in atm.: | 1 | 6.4 | 13.0 | 31.0 | 51.0 |
| $O_2$ in %: | air | 8.6 | 4.1 | 1.6 | 0.9 |
| $N_2$ in %: | air | 0.1 | 3.4 | 1.6 | 0.0 |
| He in %: | air | 90.4 | 92.6 | 96.9 | 99.1 |

Table 1. Contents of breathing gases at various depths /1/.

Intelligibility can be described by the output from a Modified Rhyme Test (MRT), where intelligibility is the percentage of correct identifications in a multiple forced choice test, adjusted to correct for the potential guesswork involved.

Table 2 list typical MRT-scores for speech at depths between 0 and 500 msw.

| | Depth (msw): 0: | 100: | 200: | 300: | 400: | 500: |
|---|---|---|---|---|---|---|
| MRT-score: | 97 | 56 | 50 | 46 | 42 | 47 |

Table 2. Intelligibility as a function of depth. Data modified from Slethei /2/.

The decrease in MRT-score is mainly caused by the helium, but the increase in ambient pressure contributes to the effect /3/. (Rank-order correlation between MRT-score and the proportion of helium is -0.99, between MRT-score and ambient pressure -0.77, based on the data in Tables 1 and 2.)

The loss in intelligibility is small from 200 to 400 msw, and from 400 to 500 there is even an increase. This flattening and rise in MRT-score cannot be accounted for by changes in depth, ambient pressure or composition of the breathing gas. A somewhat more detailed analysis seems to be needed.

It should be borne in mind that the MRT-score may disguise differences that are pertinent to the understanding of real speech, because the vowel phoneme is the same for all words that are candidates for the respondents' best forced choice. This might suggest that studies of the perception of vowels could shed some more light into the auditory darkness at depths below 200 meters.

In order to approach some of the problems related to the perception of helium speech, we have made two studies; one deals with categorical perception of vowels in helium-oxygen speech (Part I), the other deals with how difference limens (DLs) behave in this breathing gas (Part II). DLs will be studied for F1 and F2 separately.

At the time of finishing this paper, both parts comprise data from 15 respondents with no prior experience with helium-oxygen speech. Both studies will be extended to 20 respondents.

METHOD.

Part I: Formant parameters (F1-F4, BW1-BW4) for the Norwegian vowels /i:/ and /e:/ spoken in air /4/ by one diver were used as end point values, and 18 linear interpolations were calculated. All the 20 vowels were synthesized by a LPC-based formant cascade synthesizer. The same procedure was repeated for the same vowels spoken in atmospheres for 54, 120, 300 and 500 msw.

Each of the 5 sets was headed by the /i:/-/e:/-pair 3 times to serve as anchoring points for the identification tasks. Each set was randomized individually. These 5 sets, together with pauses and some sinusoid control signals, were DA-converted directly onto analog audio tape.

The stimuli were presented to the respondents via earphones, and the respondents were asked to tick off their best identification as either /i:/ or /e:/ for each stimulus. The empirical material for Part I thus consists so far of 1500 individual and independent data points.

Part II. F1 for the vowel /i:/ spoken in air was varied with respect to frequency and bandwidth and pairs of vowel-like stimuli were produced. Stimulus pairs were organised to fit into an AX-paradigm, where F1 frequency for the X spectrum was varied from 2% to 12% above that of the A spectrum in a cumulative manner. $X_1$ has a first formant frequency 2% above A, $X_2$ has a first formant frequency 2% above that of $X_1$ and so on. $AX_1$ and $AX_2$ would then constitute two different pairs of vowel stimuli.

This procedure was carried out for all depths for F1 and F2.

The test material consisted of 800 vowel pairs. They were presented to the respondents via earphones, and the respondents were asked to determine whether A and X were identical or different in quality by ticking off appropriate boxes on a response sheet. The empirical material for Part II consists of 12000 individual and independent data points.

RESULTS.

Part I. Table 3 presents the results for Part I. Stimulus No 1 is the end point /i:/ and No 20 is the end point /e:/.

| Stim. No: | Number of respondents identifying S as /i:/ for each depth: 0: | 54: | 120: | 300: | 500: |
|---|---|---|---|---|---|
| 1 | 15* | 15* | 7 | 5 | 11 |
| 2 | 15* | 12* | 4 | 11 | 11 |
| 3 | 15* | 13* | 4 | 8 | 13 |
| 4 | 15* | 11 | 3 | 5 | 10 |
| 5 | 15* | 13* | 3 | 7 | 9 |
| 6 | 15* | 11 | 3 | 7 | 8 |
| 7 | 10 | 8 | 3 | 6 | 3 |
| 8 | 14* | 13* | 7 | 7 | 8 |
| 9 | 6 | 10 | 4 | 5 | 7 |
| 10 | 2* | 10 | 1* | 9 | 4 |
| 11 | 1* | 2* | 3* | 7 | 5 |
| 12 | 1* | 3* | 4 | 5 | 3* |
| 13 | 2* | 3* | 3* | 3* | 1* |
| 14 | 0* | 1* | 2* | 11 | 5 |
| 15 | 0* | 2* | 4 | 10 | 4 |
| 16 | 0* | 1* | 2* | 7 | 5 |
| 17 | 0* | 1* | 5 | 3* | 3* |
| 18 | 0* | 0* | 1* | 4 | 1* |
| 19 | 0* | 0* | 1* | 7 | 1* |
| 20 | 0* | 0* | 3* | 7 | 4 |

Table 3. Number of respondents identifying stimuli 1-20 (/i:/-/e:/) as /i:/ for the depths 0 ,54, 120, 300 and 500 msw. Significant identifications as N* .

In Table 3 we have indicated which of the identifications were statistically significant according to the binomial distribution (alpha=0.05). A non-directional hypothesis was considered for depth = 0, and a directional hypothesis for depth > 0. In the stimulus continuum, the turnover point (between S 7 and S 8) was used as dividing point for directing the null-hypothesis towards /i:/ or /e:/.

From Table 3 we can calculate the number of statistically significant identifications for each depth. The results from this calculus are presented in Figure 1.

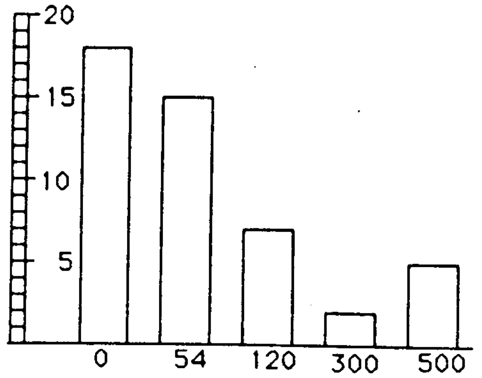


Figure 1. Number of statistically significant identifications in vowel stimulus continuum per depth.

If we disregard 0 msw and consider the cases where a stimulus is correctly identified as either /i:/ or /e:/, i.e. when stimulus number is either < 8 or stimulus number is > 7, we find that only 4 out of 28 stimuli have been identified correctly as /i:/, while 25 out of 52 have been correctly identified as /e:/. Testing these proportions against an expected equal proportion, we find that the difference is statistically significant. (Chi-square goodness of fit, with expected equal proportions.)

Part II. In an AX-paradigm, the DL is defined as the minimally detectable difference between A and X. For 15 respondents, we can reject a hypothesis that a threshold has been detected erraneously if 11 out of 15 responents agree that A and X are different. (Binomial distribution, non-directional, alpha = 0.05.)

Table 4 lists the DLs as the mean percentage of difference between A and X, and the typical formant frequency when the difference has been detected by 11 respondents. The typical formant frequency is the mean of the A and X values.

| Formant: | Depth: | DL: | SD: | Typical frequency: |
|---|---|---|---|---|
| F1: | 0 | 10.7 | * | 366 |
| | 54 | 11.2 | 1.2 | 786 |
| | 120 | 10.4 | * | 1225 |
| | 300 | 10.4 | 3.1 | 1597 |
| | 500 | 15.0 | 4.6 | 1917 |
| F2: | | | | |
| | 0 | 9.9 | 2.7 | 2451 |
| | 54 | 12.7 | 3.9 | 4823 |
| | 120 | 14.9 | 3.2 | 5237 |
| | 300 | 22.0 | 6.8 | 5974 |
| | 500 | 18.8 | 5.4 | 6185 |

Table 4. DL and standard deviations as a function of depth. (* indicate insufficient data.) Typical formant frequencies when DL is detected.

The decrease in DL for F2 from 300 to 500 msw becomes more apparent when presented in graphical form in Figure 2.

DISCUSSION

Part I. The labelling tasks performed in Part I clearly demonstrate that vowels simulating helium speech spoken in isolation differ in difficulty as objects for labelling. The Modifyed Rhyme Test disguises this difference. This calls for developing descriptive techniques which combine the reliability of the MRT with an ability to exploit variation within the linguistic material. MRT is only able to

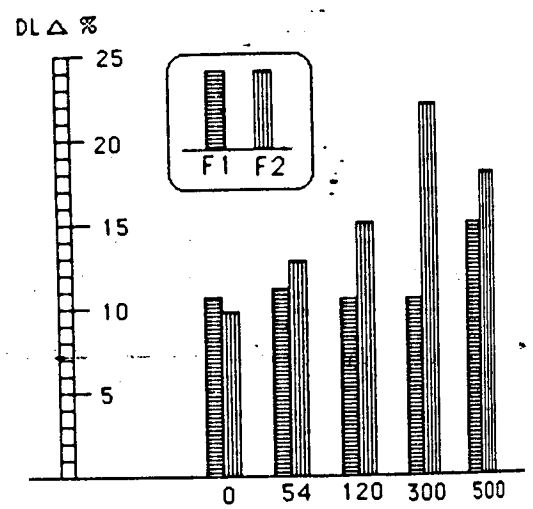


Figure 2. Difference limens for F1 and F2 per depth.

differentiate between those combinations of VC- and CV-structures that are included in the finite set of response words.

Algorithms aiming at reconstructing the speech signal as it would have been in air at 1 atmosphere, take the physical properties of the breathing gas into consideration. This is of course necessary, but auditory and perceptual aspects are being neglected. The upward shift in frequencies causes the formants to be moved out of the auditory region where pitch resolution is optimal for speech perception. For the vowels /i:/ and /e:/ this means that the second formant gradually loses its importance as cue for identification, until the first formant has been enabled to take its place. This is the most probable explanation for the effect shown in Table 4 and in Figure 1.

Part II: Flanagan's findings /5/ of DL for vowel formant frequencies in the region of 3-5% have recently been questioned by Ghitza and Goldstein /6/, who report DLs in the region above 12%. Our findings are largely in accordance with those of Ghitza and Goldstein, with an increase for F1 in the region above 1.5 kHz.

The considerable increase in DL for F2 from 0 to 300 mws (Table 5 and Figure 2) is only to be expected. The decrease from 300 to 500 mws is unexpected. Although the standard deviations are uncomfortably large, it is worth while asking why there is such a decrease.

The same explanation may suffice here. The frequency of F1 increases in relative importance as a cue to detecting the DL for F2 as the F2 frequency is moved out of the region of optimal pitch resolution.



CONCLUSIONS

Developing instruments for improving the efficiency of communication system is a demanding task, where the knowledge and skills from various professions may contribute.

There is a considerable room for improving the methods and techniques which describe the efficiency of such systems.

Knowledge from the fields of auditory and perceptual phonetics may contribute to the development of instrumentation for communication systems.

ACKNOWLEDGEMENTS

The Norwegian Underwater Technology Center (NUTEC) has given valuable assistance in the research reported here.

Instituut voor Perceptie Onderzoek (IPO) in Eindhoven generously allowed my to use their computer facilities to produce the speech synthesis.

REFERENCES

1. Belcher, E.O. and Hatlestad,S. (1983): Formant frequencies, bandwidths and Qs in helium speech. Journal of the Acoustical Society of America, 72(2): 428-432.
2. Slethei,K. (1984): Perception of Speech in a Hyperbaric Helium-Oxygen Atmosphere. Van den Broecke, M.P.R. and Cohen,A. (eds.): Proceedings of the Tenth International Congress of Phonetic Sciences, 438-442. Foris Publications.
3. Fant,G. (1980): Tal i vätgas- och heliumblandingar vid höga tryck - en teoretisk studie. Internal Report, Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm.
4. Belcher,E.O. and Hatlestad,S.(1982): Analysis of isolated vowels in helium speech. NUTEC Report No 26/82, Norwegian Underwater Technology Center, Bergen.
5. Flanagan,J.L. (1955): A difference limen for vowel formant frequencies. Journal of the Acoustical Society of America, 27:1223-1225.
6. Ghitza, O. and Goldstein,J.L. (1986): Scalar LPC Quantization Based on Formant JND's. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-34:697-708.