Use of speech production repair strategies to improve diver communication.

L. LUCKS MENDEL[1], J. H. WALTON[2], B.W. HAMILL[3], J.D. PELTON[4]

[1]School of Audiology and Speech-Language Pathology, The University of Memphis, Memphis, TN 38105; [2]Center for Speech and Hearing Research,National Center for Physical Acoustics, Department of Communicative Disorders, University of Mississippi, University, MS 38677; [3]Institute for Advanced Science and Technology in Medicine, The Johns Hopkins University, Applied Physics Laboratory, Laurel, MD; [4]NEDU, Panama City, FL 32407-7015

Lucks Mendel L, Walton JH, Hamill BW, Pelton JD, Use of speech production repair strategies to improve diver communication. Undersea Hyperb Med 2003; 30(4): 313-320 - The purpose of this investigation was to determine if speech intelligibility improved when divers made specific modifications to their speaking patterns while in a hyperbaric helium-oxygen (heliox) environment. Divers were trained to produce a variety of sentences using speech with three types of alterations: (1) slowed rate, (2) increased loudness, and (3) a combination of slightly slowed rate, a minimal increase in loudness, increased pause time, and greater mouth opening (composite strategy). Both diver and non-diver listeners judged these sentences for intelligibility. In addition, acoustic analysis of the cues for the identification of voicing, place, and manner of articulation was conducted to determine if such cues might become more audible in the speech signal when repair strategies were used. Both perceptual and acoustic results showed the composite method to be the best for natural-sounding, intelligible speech. It had the effect of slowing rate and increasing loudness just enough to increase intelligibility without causing distortion. It was concluded that teaching divers to produce speech using this method would be a worthwhile investment for improving speech intelligibility.

repair strategies, divers, communication, intelligibility, speech perception

INTRODUCTION

Speech intelligibility is critical to the safety and communications of deep sea divers. U.S. Navy divers work under hyperbaric conditions at depths of several hundred feet of sea water (fsw) in a helium-oxygen (heliox) environment. These conditions affect the acoustical properties of the speech they produce making it different from typical discourse and consequently more difficult to understand. Despite these adverse environmental conditions, divers must be able to communicate clearly with other divers at depth and with topside personnel.

Several factors are responsible for the intelligibility problem that occurs under these deep-diving conditions: (a) noise from the flow of breathing gases through some older diving helmets (e.g., EX 14 MOD1) and the diver's own breathing activity produce a constant noise level in the range of 105-115 dB SPL, which masks the diver's speech; although active and passive forms of hearing protection are used to help minimize noise levels at the ear, these hearing protectors often fail to provide an adequate signal-to-noise ratio for adequate communication through headphones (1); (b) elevated atmospheric pressure attenuates the speech signal, resulting in speech sounds that are greatly altered in intensity; and (c) the gas which divers breathe at 850 fsw is composed of as much as 99% helium in the heliox mix (2), resulting in high-pitched, nasal, and distorted speech (commonly known as *helium speech*). These three major factors interact to produce speech that is highly unintelligible to the listener.

Many investigations into the intelligibility of helium speech were conducted from the early 1960's through the mid-1980's (3-7). Since then little has been published about these intelligibility issues. In addition, over the last 40 years several attempts have been made to improve the intelligibility of helium speech with electronic helium speech unscramblers (HSUs). Researchers designed HSUs, using a variety of techniques, to convert the frequency spectra of helium speech into more normal spectra to make the speech more intelligible to listeners and to improve communication (8-11). Despite great advances in HSU technology, the combination of noise, pressure, and helium still conspires to impair speech intelligibility because researchers have not been completely successful in altering the helium signal so that it resembles normal speech. In addition, any subjective intelligibility improvements that have been made have not always been objectively verifiable (1,12).

Our research has focused on improving diver-to-diver communications. The present study, conducted at the Navy Experimental Diving Unit (NEDU) in Panama City, FL, was intended to provide important information on intelligibility that might be useful in improving communication among divers. The purpose of this study was to determine if speech intelligibility improves when divers make specific modifications to their speech while in a heliox environment. Divers used different compensatory speaking strategies so that listeners could judge the relative intelligibility of each. Results of this study offer an explanation of how this information can be used to improve speech intelligibility under adverse diving conditions.

METHODS

Subjects

Speakers. Nine male U.S. Navy-qualified divers, ages 25 to 32, participated in this study. All divers had normal hearing acuity as defined by Navy hearing requirements. Normal hearing was defined as pure-tone thresholds better than 20 dB HL for the octave frequencies from 500 to 8000 Hz. Three divers were identified as excellent speakers, three as adequate speakers, and three as poor speakers. From these groups, recordings from six of the divers were chosen by two of the authors (JHW and LLM) to be used in the listening procedure. The parameters used to categorize the speakers into their respective groups included: (a) clarity of speech production, (b) accuracy of producing the stimuli according to the specified repair strategy, and (c) appropriate use of rate, pauses, and loudness. Recordings from those six speakers who were judged to produce the best speech (i.e., had very precise articulation and a standard dialect, and were particularly adept at producing the repair strategies that are described below) were used in the perceptual analysis. The poor speakers were not included in the analysis because their utterances were inadequate in demonstrating the differences among the repair strategies investigated.

Listeners. Fifteen male and female non-divers (ages 18 to 38) and 13 male divers (ages 28 to 44) who did not participate as speakers served as listeners for this experiment. All listeners had normal hearing and middle ear function. The diver listeners were familiar with helium speech processed through an HSU; the non-diver listeners were completely unfamiliar with helium speech.

Stimuli and Speaking Strategies

Sentences from a published intelligibility test (13,14) along with sentences that divers frequently hear in a hyperbaric environment, were chosen as stimuli (see Appendix). At 0 fsw in air each diver read ten sentences, varying in length from 4 to 27 words, in his normal speech pattern. Then each diver was taught to read the sentences with three different "strategies" of speaking: (1) slowed rate, normal loudness; (2) increased loudness, normal rate; and (3) a composite of strategies. The composite strategy was a combination of slightly slowed rate, a minimal increase in loudness, increased pause time, and greater mouth opening. Imagery and demonstration were used to tailor the training for each diver's specific articulation pattern.

The divers' production of the repair strategies was monitored qualitatively by the experimenters, and necessary adjustments were communicated to the divers using descriptive terms and demonstration. The increased volume levels were also monitored via the VU meter on the recording system. It was difficult to control quantitatively the potential influences of increased loudness on fundamental frequency and rate, but when qualitative influences were noticed by the experimenters, the divers were instructed to repeat the stimulus in an attempt to control for those factors.

Recording of Perceptual Stimuli

Sentence stimuli were recorded to digital audiotape (DAT) using a Panasonic DAT Deck (Model SV-3700) and a Sony DAT recorder (Model PCM-2500 A/B). All stimuli were recorded at 0 fsw in air and 850 fsw in heliox in an Industrial Acoustics Company (IAC) sound-treated booth. Speech produced at 850 fsw was processed through an HSU and presented through the Tethered Diver Communication System (TDCS). Divers read the sentences using their normal speech and the three test speaking strategies: (1) slow rate, (2) increased loudness, and (3) composite.

At the completion of this phase of data collection it was determined that the *loud* strategy was too fatiguing to the divers and increased their work of breathing. Further, the resulting speech signal was significantly distorted when it was processed through the HSU. Perceptual comparisons, therefore, were made only among the remaining three speech patterns: normal speech, slow rate, and composite repair strategies.

Due to the large number of stimuli, a subset of items from the DAT recordings were used for the perceptual aspect of this experiment. The selected stimuli were dubbed to a master DAT, and then individual DATs with the stimuli randomized and paired for comparisons were made from the master for use during the testing procedure. For each of the six speakers, a sentence produced with one strategy was paired with the same sentence spoken by that speaker using a different strategy. This produced the stimulus set of 180 sentence pairs (6 speakers X 3 strategy pairings X 10 sentences). The DAT recordings of the stimuli were randomized to avoid an order effect. DAT recordings of helmet noise (gas flow noise from the EX14 MOD1 helmet) were added to the stimuli presented in all listening combinations to simulate conditions typical of the divers listening environment. The 180 comparisons were then presented to both groups of listeners (divers and non-divers) in random orders for perceptual judgments.

Listening procedure

It was not possible to test the non-diver and diver listeners in the same listening environment because the two groups were in two different geographic locations. However, we attempted to make the environments as similar as possible given the differences in equipment and testing locations. The sentence stimuli were routed from a Panasonic DAT Deck (Model SV-3700) and helmet noise was presented from a JVC (Model TD-W707) audiocassette deck to the non-diver listeners, who were tested individually in an IAC sound-treated booth. Both signals were then routed through a Beltone 2000 audiometer to the listeners' ears via supra-aural earphones. The diver listeners were evaluated in groups of 6 and 7 on the Medical Deck at NEDU. The speech and noise stimuli were routed from two Sony DAT decks (Model DTC-ZE700) through the TDCS in the control room to SetCom5 monaural headsets worn by the divers. The headsets were calibrated so that outputs varied no more than 0.9 dB.

All stimuli were presented binaurally to the non-diver listeners with the noise and the stimuli presented simultaneously at 70 dB SPL to both ears. Because the headsets used by the diver listeners were monaural, the stimuli and noise were presented at 70 dB SPL simultaneously to each listener's better ear. Listeners were asked to judge which sentence in each pair was more intelligible. Each of the 28 listeners made 180 judgments for a total of 5,040 responses.

Acoustic Analysis of Stimuli

An acoustic analysis was conducted on a subset of the sentence stimuli using the Computerized Speech Lab (CSL) by Kay Elemetrics, Model 4300B, Software Version 5.X. Stimuli were sampled at 51,200 Hz, filtered, and stored to disk for analysis. Time measurements and phoneme boundaries were made from information derived from a combination of the waveform and a spectrographic display. All frequency and intensity measurements were made from a fast Fourier transformation (FFT) which had a sampling rate of 51,200 Hz and a frame length set at 1024. When visual inspection of the voiced aspects of speech was needed to determine accurate time and formant measurements (with the narrowest bandwidth possible), the syllables were filtered and down-sampled to 25,600 Hz for speech obtained at 850 fsw and to 12,800 Hz for samples produced at 0 fsw.

RESULTS

Perceptual Analyses

In a three-way Analysis of Variance (ANOVA) conducted on the means of the three variables (listener group [diver, non-diver], repair strategy [normal, slow, composite], and speaker group [excellent, adequate]), no significant main effect was found for the listener group. Therefore, the data for the two listener groups were combined, and a two-way ANOVA was conducted on the variables of repair strategy and speaker group. This ANOVA revealed a significant main effect for repair strategy [$F(2,354)=106.65$, $p<0.001$] and a significant interaction between repair strategy and speaker group [$F(2,354)=24.40$, $p<0.001$]; there was no main effect for speaker group. Post hoc Tukey all pairwise comparisons revealed that listeners perceived the composite strategy as significantly more intelligible than either the slow strategy or the no training (normal speaking) condition ($p<0.05$) (Figure 1). There was no significant difference between normal speech and the slow repair strategy.

The significant two-way interaction between repair strategy and speaker group revealed that the different levels of speaker group (excellent versus adequate speakers) depended on which repair strategies were being compared. Figure 2 shows that excellent speakers were judged to be more intelligible than adequate speakers when the composite strategy was used ($p<0.05$). Adequate speakers were more intelligible than the excellent speakers for slow speech ($p<0.05$).

The adequate speakers were also more intelligible than excellent speakers when they used normal speech ($p<0.05$). Overall, these results indicate that slowing the rate of speech did not improve speech intelligibility. Use of the composite strategy, however, had a significant effect on intelligibility.

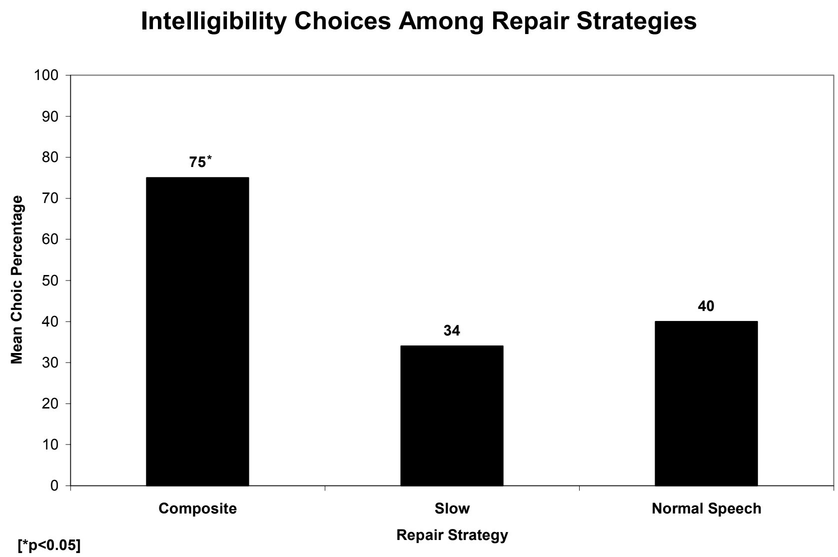

Figure 1. Mean perceptual choice percentages among all divers and non-diver listeners across the three repair strategies.

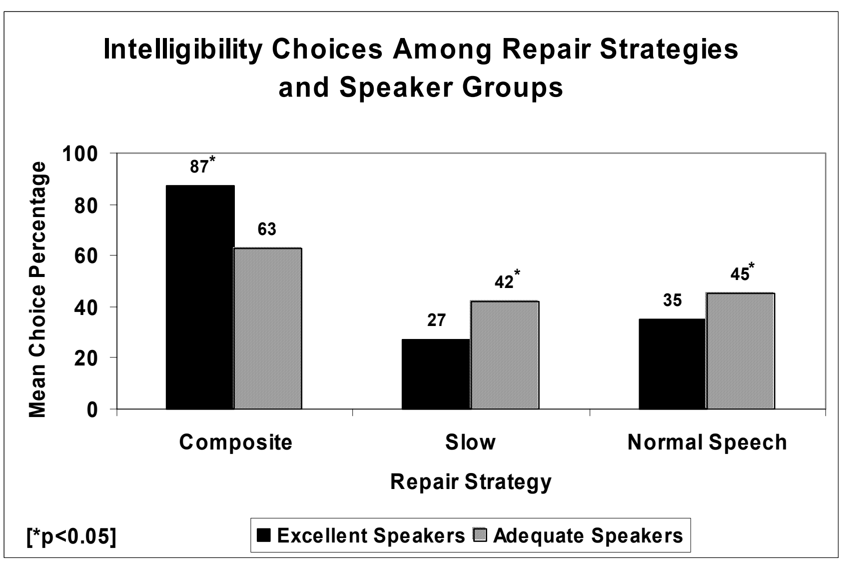

Figure 2. Mean perceptual choice percentages among all diver and non-diver listeners across the three repair strategies and two speaker groups.

Acoustic Correlates to Perceptual Results

Paired two-sample t-tests for unequal variances were used to compare instances of an acoustic characteristic produced by a speaker using two strategies: slow rate and composite. The composite repair strategy produced a significantly more intense signal than the slow repair strategy. This difference was demonstrated when the energy of the entire sentence was compared

[t (46) = 6.06, $p < .05$] and when comparisons were made for the nasals /n/ and /m/ [t (11) = 2.74, $p < .01$], for comparisons of the plosive /p/ [t (11) = 4.03, $p < .01$], for the fricative /s/ [t (11) = 1.93, $p < .05$], and for the glides /w/ and /j/ [t (8) = 4.93, $p < .01$]. Although the tendency for liquids was for the composite strategy to be greater in intensity than the slow condition, a low number of comparisons (only four) and excessive variability due to the phonetic context prevented the differences for /r/ and /l/ from reaching significance.

Although the intent of the slow strategy was to slow the speech of the diver to give the listener more time to process the auditory information, the composite strategy actually produced significantly slower speech [t (46) = 1.91, $p < .05$]. Because speakers vary in their strategies for slowing speech (some prolong the vowels while others increase pause time), the duration of individual phonemes was not considered a valid measure of the listeners' ability to process speech and was, therefore, not measured.

Eight individual phonemes were measured for frequency range and for the frequency of the resonance with the greatest amplitude. Measurements were made from an FFT either as an average signal throughout the entire span of the phoneme or at the point of cursor placement. Each variable was compared in a paired t-test statistic with the number of comparisons varying from 6 to 24 as a function of the frequency of occurrence of the individual phoneme in the sample. Five out of eight phonemes had significantly higher frequency ranges for the composite strategy: for the nasals /m/ and /n/ [t (11) = 2.95, $p < .01$], for the plosive /p/ [t (11) = 2.99, $p < .01$], and for the glides /w/ and /j/ [t (8) = 2.84, $p < .05$]. The frequency of the resonance with the greatest amplitude was higher in the composite strategy for only the /p/ phoneme [t (11) = 2.34, $p < .05$] with all other phonemes failing to reach significance.

DISCUSSION

This research was designed to investigate whether divers could be taught to alter their manner of speaking to compensate for the loss of energy at 850 fsw and whether those alterations improved their intelligibility. The perceptual and acoustic findings observed in this experiment clearly indicate that the composite speech production repair strategy is the best to use in a diving environment. Both listener groups preferred the composite strategy over the slow strategy and the normal speech condition for perceived intelligibility. For the composite repair strategy, the best speakers were judged more intelligible than the adequate speakers. However, for the slow repair strategy and for normal speech, the adequate speakers were judged more intelligible than the excellent speakers. It may be that the adequate speakers were able to produce the slow strategy condition more naturally than the excellent speakers thus resulting in higher perceived intelligibility.

The use of speech production repair strategies was found to be quite beneficial to speech understanding. By altering their method of speech, divers improved the acoustic characteristics of helium speech and thus made it more intelligible. For example, the frequencies produced with a relatively closed vocal tract using normal speech were attenuated the most at 850 fsw. But by increasing the amount of mouth opening, increasing intraoral pressure, and increasing vocal effort, speakers can retain more of the critical acoustic information in speech that is produced under adverse diving conditions.

Acoustic analyses of the strategies revealed that the composite strategy provided more of the cues necessary for accurate speech perception. The increase in vocal intensity and the higher frequency ranges observed in the composite compared to the slow condition suggest that more acoustic information was available in the former. These findings suggest that the greater the acoustic information available, the more the speech is preferred by these listeners.

CONCLUSION

The perceptual and acoustical findings measured in the repair strategy investigations clearly indicated that the composite speech production repair strategy was the best in a heliox diving environment. Listeners perceived speech produced by this strategy as the most intelligible and it provided the most acoustic information (e.g., increases in vocal intensity and higher peak resonances and frequency ranges) which is necessary for accurate speech perception. Therefore, using the composite repair strategy together with continued improvements to the HSU in both quiet and noise, divers/speakers should produce more intelligible speech in a heliox diving environment.

The slower rate, clearer articulation, and slightly increased loudness components of the composite repair strategy are the primary contributors to clearer, more intelligible speech. However, it is possible that additional variables could have influenced these results. It should be noted that these modifications to speech production can result in greater work of breathing. Nonetheless, divers must be able to communicate through speech while in a heliox diving environment. Therefore, it is recommended that divers be trained to adjust their production of speech independent of increases in vocal intensity and rate. Fine adjustments in the production of speech sounds (e.g., opening the mouth wider) can be enacted against a background of overall adjustment in amplitude, allowing divers to maintain habitual levels of vocal intensity. These modifications in speech production together with continued improvements in the HSUs used during diving should improve diver-to-diver communications.

ACKNOWLEDGMENTS

The authors are indebted to the personnel at the U.S. Navy Experimental Diving Unit, Panama City, Florida, for technical support and use of their facilities. Thanks also go to the divers who participated as listeners and to the graduate students at the University of Mississippi who participated in data entry. This work was supported by the Naval Sea Systems Command.

REFERENCES

1. Mendel LL, Hamill BW, Crepeau LJ, Fallon E. Speech intelligibility assessment in a helium environment. *Journal of the Acoustical Society of America* 1995; 97:628-636.
2. U. S. Navy Diving Manual. NAVSEA 0994-LP-001-9020, Panama City, FL: Navy Experimental Diving Unit, 1999.
3. Brubaker RS, Wurst JW. Spectrographic analysis of divers speech during decompression. *Journal of the Acoustical Society of America* 1968; 43:798-802.
4. Sergeant RL. Speech during respiration of a mixture of helium and oxygen. *Aerospace Medicine* 1963; 34:826-828.
5. Morrow CT. Speech in deep-submergence atmospheres. *Journal of the Acoustical Society of America* 1971; 50(3, Part 1):715-728.
6. Holywell K, Harvey G. Helium speech. *Journal of the Acoustical Society of America* 1964; 36:210-211.
7. Fant G, Lindquist J. Pressure and gas mixture effects on divers speech. Speech Transmission Laboratory Quarterly Progress and Status Report, STL-QPSR-1, 1968.

8. Jack MA, Duncan G. The helium speech effect and electronic techniques for enhancing intelligibility in a helium-oxygen environment. Radio and Electronic Engineer 1982; 52(5):211-223.
9. Giordano TA, Rothman HB, Hollien H. Helium speech unscramblers--A critical review of the state of the art. IEEE Transactions on Audio and Electroacoustics 1973; AU-21(5):436-444.
10. Copel M. Helium voice unscrambling. IEEE Transactions on Audio and Electroacoustics 1966; AU-14(3):122-126.
11. Belcher EO, Andersen K. Helium speech enhancement by frequency-domain processing. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing 1983; New York: IEEE:1160-1163.
12. Mendel LL, Hamill BW, Hendrix JE, Crepeau LJ, Pelton JD, Miley MD, Kadlec EE. Speech intelligibility assessment in a helium environment. II: The speech intelligibility index. *Journal of the Acoustical Society of America* 1998; 104:1609-1615.
13. Voiers WD. Diagnostic acceptability measure for speech communication systems. Record of the International Conference on Acoustics, Speech, and Signal Processing 1977; New York: IEEE:204-207.
14. Dynastat, Inc. Diagnostic acceptability measure (DAM): A method for measuring the acceptability of speech over communication systems. Specification DAM-IIC, Dynastat, Inc., Austin, TX, 1995.

APPENDIX

Modified Diagnostic Acceptability Measure (DAM) Sentences:

1. Papa gave us apples and pears.
2. I patted the pad.
3. Betty saw Pat on the boat.
4. Paul was lost in the pond.

Diver Phrases:

5. The number four blade has a curl on the leading edge.
6. You have a ten-minute decompression stop at 20 feet.
7. Green diver, hook up the down haul cable.
8. Red diver, topside, how do you hear me?

Sentences:

9. It is a well-known fact that the sportsmen of the world are divided into two very distinct groups: the breakers and the fixers.
10. They have a Swiss Army knife in their pocket, backed up by a key chain which is attached to a little set of screwdrivers, including a Phillips.