Speech intelligibility assessment in a helium environment. II. The speech intelligibility index

Lisa Lucks Mendel[a]
Center for Speech and Hearing Research, Department of Communicative Disorders, University of Mississippi, University, Mississippi 38677

Bruce W. Hamill
The Johns Hopkins University, Applied Physics Laboratory, Laurel, Maryland 20723-6099

James E. Hendrix
Oxford Signal Processing, Oxford, Mississippi 38655

Loring J. Crepeau and Jerry D. Pelton
Navy Experimental Diving Unit, Panama City, Florida 32407-7015

Matthew D. Miley
X^N Technologies, Cheney, Washington 99004

Elizabeth E. Kadlec
University of Georgia, Athens, Georgia 30602

(Received 13 September 1997; accepted for publication 29 May 1998)

The Speech Intelligibility Index (SII) was measured for Navy divers participating in two saturation deep dives and for a group of nondivers to test different communication systems and their components. These SIIs were validated using the Speech Perception in Noise (SPIN) test and the Griffiths version of the Modified Rhyme Test (GMRT). Our goal was to determine if either of these assessments was sensitive enough to provide an objective measure of speech intelligibility when speech was processed through different helmets and helium speech unscramblers (HSUs). Results indicated that SII values and percent intelligibility decreased incrementally as background noise level increased. SIIs were very reliable across the different groups of subjects indicating that the SII was a strong measurement for predicting speech intelligibility to compare linear system components such as helmets. The SII was not useful in measuring intelligibility through nonlinear devices such as HSUs. The speech intelligibility scores on the GMRT and SPIN tests were useful when the system component being compared had a large measurable difference, such as in helmet type. However, when the differences were more subtle, such as differences in HSUs, neither the SPIN nor the GMRT appeared sensitive enough to make such distinctions. These results have theoretical as well as practical value for measuring the quality and intelligibility of helium speech enhancement systems.

[S0001-4966(98)03109-9]

PACS numbers: 43.72.Kb, 43.72.Ew, 43.71.Gv, 43.70.Dn [JLH]

INTRODUCTION

Accurate assessment of speech intelligibility is important for determining the effectiveness of different communication systems used with Navy divers who are exposed to high noise levels in their work environments. Further, these divers work under hyperbaric conditions at depths of several hundred feet of seawater (FSW) in a helium–oxygen (heliox) environment. These conditions affect the acoustical properties of the speech they produce making it significantly different from typical discourse and difficult to understand.

Helium speech unscramblers (HSUs) attempt to convert the frequency spectra of helium speech into a frequency range that is closer to average conversational speech, making it more intelligible to listeners and thus improving communication among divers (Belcher and Andersen, 1983; Copel, 1966; Giordano *et al.*, 1973; Stover, 1967). While such devices provide significant improvement to intelligibility, the effects of noise, pressure, and helium still contribute significantly to poor speech understanding among divers. Thus refinements continue to be made in diver communication systems by using state-of-the-art technology to improve speech intelligibility. Our work has focused on evaluating the effectiveness of different speech intelligibility assessment tools for providing accurate estimations of speech understanding in order to test these system refinements.

Traditional speech recognition tests such as Griffiths (1967) version (GMRT) of the Modified Rhyme Test (House *et al.*, 1965) and the Speech Perception in Noise (SPIN) test (Kalikow *et al.*, 1977) have been used extensively for measuring helium speech intelligibility in Navy divers, but have not proven to be sensitive enough to measure small but noticeable intelligibility differences. Mendel *et al.* (1995) showed that GMRT and SPIN scores do not reflect divers' subjective impressions of the performance of different HSUs. In that investigation, the GMRT and SPIN produced almost

[a]Electronic mail: cdmendel@olemiss.edu

identical test scores for two types of HSU, suggesting that neither test was more sensitive than the other, at least in a dry chamber. Thus these traditional speech recognition assessment tools proved to be ineffective methods of determining accurate measures of speech understanding in a dry chamber helium environment.

The Speech Intelligibility Index (SII) is a different assessment tool that is based on the audibility of the speech signal and is calculated as the sum of a number of contiguous frequency bands that are each weighted according to their contribution to speech understanding (ANSI, 1995; Pavlovic, 1991). The SII is an updated version of the standard for the Articulation Index (AI, ANSI, 1970) which defines a method for computing an objective, physical measure that is closely correlated with the intelligibility of speech when measured by standard perceptually based speech intelligibility tests under a variety of adverse listening conditions. This objective SII measure appeared to have significant application to predicting speech understanding in a helium environment.

The purposes of the studies presented here were to: (1) investigate the effectiveness of the GMRT and SPIN tests in assessing the intelligibility of divers' helium speech using different HSUs and helmets during two in-water saturation deep dives (Deep Dive 1995, DD 95, and Deep Dive 1996, DD 96) conducted in the Ocean Simulation Facility (OSF) at the Navy Experimental Diving Unit (NEDU) in Panama City, FL, and (2) determine the effectiveness of the SII in assessing speech understanding in dry and wet chambers during the same saturation deep dives.

I. SPEECH INTELLIGIBILITY INDEX MEASUREMENTS

A. Manikin testing

Initial measurements using two acoustic manikins were taken prior to the saturation deep dives. Digital audio tape (DAT) recordings from earlier deep dives conducted at NEDU were analyzed, and specific segments of these tapes were dubbed and edited for use during manikin and human subject testing. The SII measurement method we used required separate measurements of *speech without noise* and *noise without speech*. Therefore noise recorded on DAT from the helmet of a diver in the water at 850 FSW was used for the *noise without speech*, and speech segments taken from topside-to-diver communications of previous deep dives were used as the *speech without noise*.

A spectrum analysis of each of the measures of *speech without noise* and *noise without speech* was obtained using a Hewlett–Packard Dynamic Signal Analyzer (HP 35665A). Once the spectra and the SPLs of these segments were obtained at the desired overall levels, the data were written to ASCII files suitable for input to the SII critical band calculation program provided with the ANSI SII standard (ANSI, 1995).

B. Human subject testing

1. Subjects

In order to validate the SII calculations obtained on the manikins, human subject testing was conducted prior to the saturation deep dives. The goal of this testing was to determine how the SII calculations compared to typical speech recognition performance. A group of normally hearing nondivers ($N=14$; ages 18–44) and two groups of normally hearing divers (DD 95: $N=8$; ages 24–43; DD 96: $N=8$; ages 26–38) participated in the studies.

2. Procedure

The GMRT and the SPIN test lists were used in this study. A complete description of each test can be found in Mendel *et al.* (1995). Lists No. 8 of both the GMRT and SPIN tests were edited to sound like continuous discourse so they could be used as the *speech without noise* stimuli for the SII measurements.

a. Nondivers. All speech stimuli were presented at 68 dB SPL corresponding to the raised standard speech level. The *noise without speech* measurement consisted of seven different overall noise levels of the same ''helmet noise'' used in the manikin measurements (i.e., 50, 56, 62, 68, 74, 80, and 86 dB SPL). After a hearing screening, power spectral density (PSD) curves of the spectra of the edited (compressed) GMRT and SPIN lists and the spectra of each noise level were obtained for each subject and stored for later calculation of the SII. Each subject then listened to either three GMRT and SPIN lists at three different noise levels of four GMRT and SPIN lists at four different noise levels. To avoid learning effects, subjects were paired so that no subject heard the same randomized list pair from either test. Subjects provided written responses to the stimuli heard.

b. Divers. The same DAT tapes used with the nondivers were used to measure SIIs for the Navy divers participating in DD 95 and DD 96. Two identical equipment configurations were used to test two divers at a time in a dry living chamber (Alpha Chamber) with background noise reduced as much as possible. Each diver wore a headset on his better ear, and the DAT recordings of *speech without noise* and *noise without speech* were presented to that ear.

For DD 95, SIIs were measured at four different noise levels (i.e., 78, 84, 90, and 96 dB SPL); the chamber noise floor was 68 dB SPL. DAT recordings of the compressed GMRT and SPIN stimuli (*speech without noise*) were each presented at 90 dB SPL, and a PSD curve for each was measured and stored. For DD 96, SIIs were measured using noise levels at 106, 111, 116, and 121 dB SPL. Overall the noise levels were higher than those used in DD 95 because we were attempting to measure SII and speech intelligibility performance in more adverse listening conditions than had been measured previously. DAT recordings of the compressed GMRT and SPIN stimuli were each presented at 65 dB SPL, instead of 90 dB SPL, and a spectrum for each was measured and stored. Once the spectra and the SPLs of *speech without noise* and *noise without speech* were obtained and stored at the desired overall levels, SII calculations were made using the procedures described earlier.

After the SII measurements were obtained on the divers, speech intelligibility testing was conducted to determine how the SII calculations compared to typical speech recognition performance. Each diver heard one 50-item GMRT and one 50-item SPIN list at each noise level through the headset on his better ear. The lists were presented to each diver so that

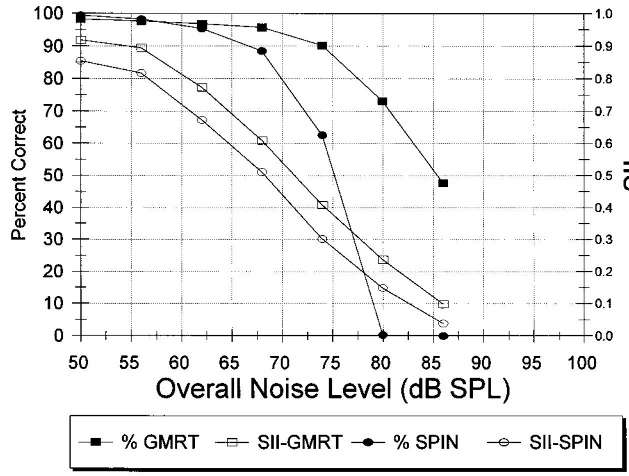

FIG. 1. Nondivers' mean percent correct GMRT and SPIN scores and Speech Intelligibility Index (SII) values for compressed GMRT and SPIN stimuli as a function of noise level for DD 95.

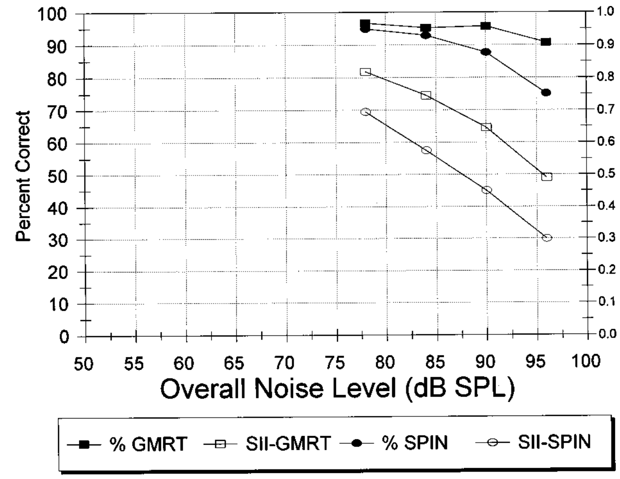

FIG. 2. Divers' mean percent correct GMRT and SPIN scores and Speech Intelligibility Index (SII) values for compressed GMRT and SPIN stimuli as a function of noise level for DD 95.

no subject heard the same randomized list pair from either test. Subjects provided written responses to the stimuli heard.

For DD 96, SIIs were also measured while divers were in the wet chamber. DAT recordings of the compressed GMRT and SPIN stimuli were presented to each diver immediately before he entered the water. A PSD measurement was then taken and stored on a diskette. The helmet was then placed on the diver and he entered the water. Once the diver was quiet and in the water, a spectrum measurement was taken of the background noise level in the helmet and stored on a diskette. This measure of actual helmet noise served as the *noise without speech* measurement needed for the SII calculation.

C. Results

1. Nondivers

Figure 1 plots the mean GMRT and SPIN percent correct scores and SII values as a function of noise level for the nondivers. GMRT and SPIN percent correct scores and SIIs remained high for the lower noise levels even though SII values decreased more rapidly than percent correct scores. GMRT intelligibility scores of 90%–98% were comparable to SII values ranging from 0.41 at 90% intelligibility to 0.92 at 98% intelligibility. SPIN intelligibility scores of 88%–99% were comparable to SIIs ranging from 0.51 at 88% to 0.85 at 99%. Both SII values decreased appropriately as the level of noise increased for both the GMRT and SPIN.

2. Divers

The average results of the SII calculations and intelligibility performance for the divers are plotted as a function of noise level in Fig. 2 (DD 95) and Fig. 3 (DD 96). As seen with the nondivers, SII values decreased appropriately as the level of noise increased for both the GMRT and SPIN stimuli; intelligibility and SII values generally were lower for the SPIN than for the GMRT. Overall GMRT percent correct performance was better than SPIN performance which may be a reflection of the open-set nature of the SPIN test versus the closed-set nature of the GMRT. Also, Figs. 2 and 3 show that SPIN performance drops more rapidly at the higher noise levels than does GMRT performance.

Figure 4 plots the dry chamber data for both nondivers and divers collected during DD 95 and DD 96. In this figure, GMRT and SPIN percent correct intelligibility scores are plotted as a function of the SII in three conditions: (a) nondivers tested at one atmosphere in 1994 (M/000; S/000); (b) divers tested in dry chambers at 850 FSW in DD 95 (M/850/95; S/850/95); and (c) divers tested in dry chambers at 850 FSW in DD 96 (M/850/96; S/850/96). All GMRT points are represented by rectangles and all SPIN points are ovals. Examination of this figure suggests that the data collected in DD 96 were quite similar to those collected in the laboratory situation and in DD 95 and that SII values decreased as percent intelligibility decreased. Note that the two unaveraged points on the curve reflect the GMRT and SPIN scores for the one diver who was exposed to the highest noise level. Even though these points are not averaged, they do fall appropriately on the curve and line up well with the other data.

Figures 5 and 6 show plots of the GMRT and SPIN

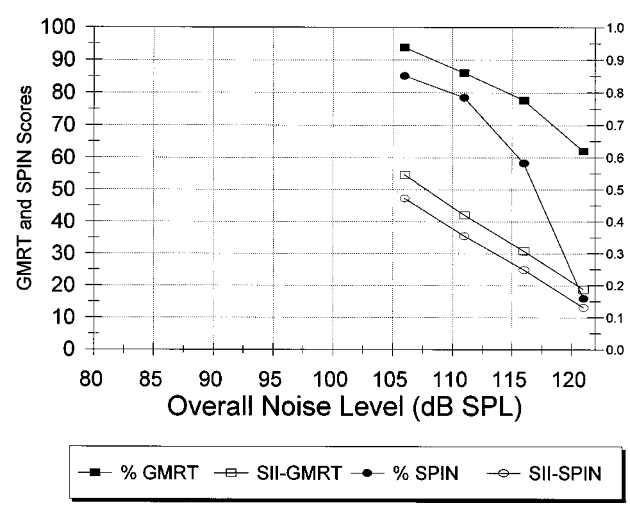

FIG. 3. Divers' mean percent correct GMRT and SPIN scores and Speech Intelligibility Index (SII) values for compressed GMRT and SPIN stimuli as a function of noise level in the dry chamber for DD 96.

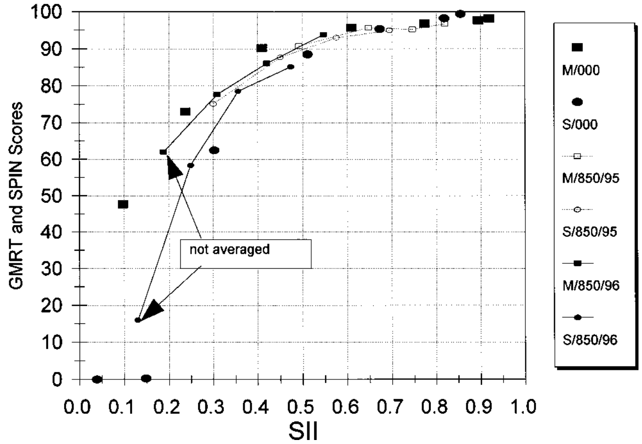

FIG. 4. GMRT and SPIN mean percent correct performance as a function of Speech Intelligibility Index (SII) in the dry chamber for three studies: (a) nondivers in laboratory environment, (b) divers in DD 95, and (c) divers in DD 96. (*M*/000=GMRT scores at one atmosphere; *S*/000=SPIN scores at one atmosphere; *M*/850/95=DD 95 GMRT scores at 850 FSW; *S*/850/95 =DD 95 SPIN scores at 850 FSW; *M*/850/96=DD 96 GMRT scores at 850 FSW; *S*/850/96=DD 96 SPIN scores at 850 FSW.)

intelligibility performance, respectively, as a function of SII for all dry and wet chamber data taken to date. The wet chamber data are shown as large open squares or ovals and are not averaged or connected because the noise levels in the water could not be controlled. Thus these data are not grouped into distinct noise levels as they were in the dry chamber measurements. The data points for the SPIN do not line up quite as well as they do for the GMRT, but there is still a clear pattern in the results. There was considerably more scatter in the SPIN wet chamber intelligibility data than in the GMRT intelligibility data which is probably a reflection of the open-set nature of the SPIN test. The SPIN scores also appeared to be more affected by the adverse noisy conditions in the water than the GMRT.

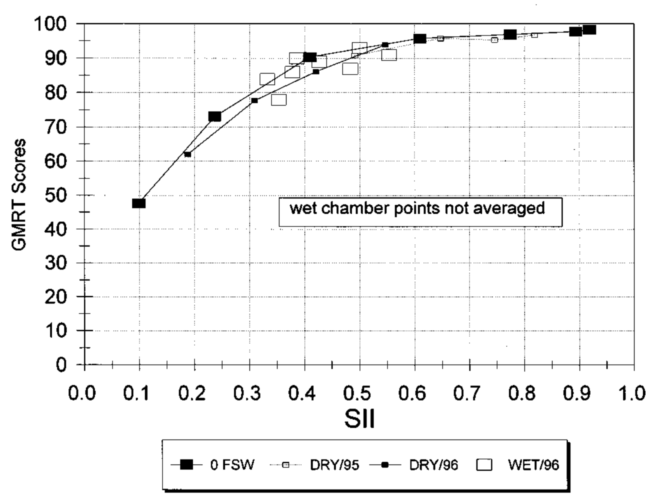

FIG. 5. GMRT mean percent correct performance as a function of Speech Intelligibility Index (SII) in the dry and wet chambers for three studies: (a) nondivers in laboratory environment, (b) divers in DD 95, and (c) divers in DD 96. (O FSW=GMRT scores at one atmosphere; DRY/95=DD 95 GMRT scores at 850 FSW; DRY/96=DD 96 GMRT scores at 850 FSW; WET/96=DD 96 GMRT scores at 850 FSW.)

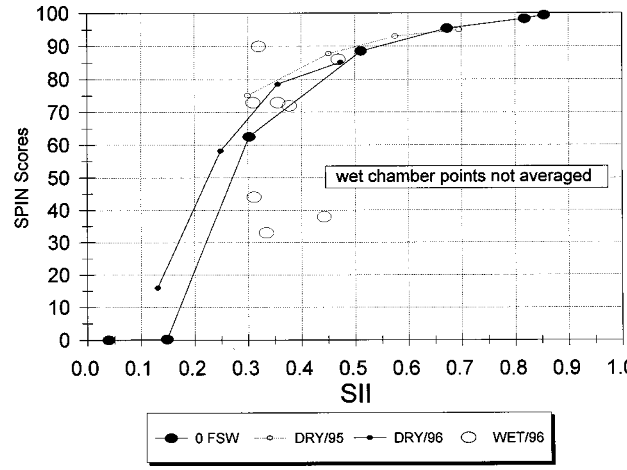

FIG. 6. SPIN mean percent correct performance as a function of Speech Intelligibility Index (SII) in the dry and wet chambers for three studies: (a) nondivers in laboratory environment, (b) divers in DD 95, and (c) divers in DD 96. (O FSW=SPIN scores at one atmosphere (laboratory environment); DRY/95=DD 95 SPIN scores at 850 FSW; DRY/96=DD 96 SPIN scores at 850; WET/96=DD 96 SPIN scores at 850 FSW.)

II. IN-WATER GMRT AND SPIN TESTING AT 850 FSW

A. Subjects

The same two groups of Navy divers from DD 95 and DD 96 served as subjects during the in-water test phase.

B. Stimuli and instrumentation

A tethered diver communications system (TDCS) was used to enable divers to communicate with each other and support personnel in the OSF control room. Two different helmets were used during all dives: a standard helmet and a modified helmet which had no supply valve and was reported to be much quieter than the standard helmet. Also, two different HSUs were employed during the dives. HSU1 used a 16-bit processor which performed sampling, compression, and reconstruction in the time domain using a speech signal processing algorithm and A/D converter (Dildy, 1992). HSU2 used a high-speed, floating point digital signal processor to provide improved synchronization over the entire dynamic range of the signal, smooth transitions in the waveforms, and overlapping of waveforms (Hendrix, 1996).

The same GMRT and SPIN lists used in the dry chamber SII measurement and validation were used in this portion of the experiment. DAT recordings and helium live-voice presentations of eight lists each of the GMRT and SPIN tests were used. The DAT recordings of the stimuli were presented via a Panasonic DAT deck (model SV-3700).

C. Procedure

During DD 95 and DD 96, all divers were tested in pairs while in the water. Each diver heard four lists of stimuli: one DAT-recorded SPIN and GMRT list presented from topside (topside–diver) and a randomization of both lists presented live-voice (using helium speech) from the paired diver (diver–diver). In all cases, the diver heard the stimuli from a particular list before he read them to his paired diver to avoid

TABLE I. Overall mean percent intelligibility scores and standard deviations (s.d.s) for the GMRT and SPIN in the wet chamber across condition (topside–diver versus diver–diver). All GMRT scores were corrected to account for the closed-set nature of the test. The formula used to correct the scores was 2 ×[No. right−(No. wrong/4)]. Starred values (*) indicate significant differences.

Dive	Condition	Helmet	GMRT mean	GMRT s.d.	SPIN mean	SPIN s.d.
DD 95	topside–diver	NA	78*	12.6	66*	29.5
	diver–diver	NA	32	16.5	27	19.3
DD 96	topside–diver	NA	84*	6.3	63*	21.6
	diver–diver	NA	41	16.3	36	19.7
	topside–diver	standard $n=8$	81	4.9	47	16.4
		modified $n=8$	87*	6.4	81*	9.0
	diver–diver	standard ($n=8$)	34	11.4	27	16.3
		HSU 1 ($n=4$)	31	13.3	33	21.6
		HSU 2 ($n=4$)	38	10.1	21	7.9
		Modified ($n=8$)	48	13.1	46	18.7
		HSU 1 ($n=4$)	50	18.1	38	12.3
		HSU 2 ($n=4$)	46	8.1	54	22.5

learning effects within a listener. All stimuli were presented at comfortable listening levels, and all diver responses were written on answer sheets.

D. Results

1. In-water GMRT and SPIN intelligibility scores

The overall mean percent correct scores and standard deviations for DD 95 and DD 96 on the GMRT and SPIN for the two listening conditions (taped, topside–diver and live-voice, diver–diver) are provided in Table I. All GMRT scores were corrected to account for the closed-set nature of the test using the formula: 2×[No. right-(No. wrong/4)]. Analyses of variance (ANOVAs) conducted on the means of the two intelligibility scores for each diver's GMRT and SPIN tests in the water revealed a significant main effect favoring topside–diver presentation of test lists over diver–diver presentation [DD 95: $F(1,7)=77.29$, $p<0.0001$; DD 96: $F(1,6)=55.03$, $p<0.0005$], and a significant main effect of overall higher test scores for the GMRT than for the SPIN in both conditions [DD 95: $F(1,7)=39.82$, $p<0.0005$; DD 96: $F(1,6)=16.62$, $p<0.01$]. No interaction effects were observed.

An item analysis of the divers' responses to the stimuli on the SPIN and GMRT showed that divers made considerably more errors on the SPIN than on the GMRT. More SPIN errors were made on low-predictability (noncontextual) items than on high-predictability items.

The data in Table I also show DD 96 pooled GMRT and SPIN mean percent intelligibility scores and standard deviations across both helmet types along with a breakdown of the scores as a function of helmet type. An ANOVA conducted on the means of the two intelligibility scores for each diver's GMRT and SPIN tests in the water revealed a significant main effect for helmet [standard versus modified, $F(1,6)=8.12$, $p<0.05$] revealing higher test scores measured in the modified helmet compared to the standard helmet. A significant two-way interaction between tests and helmets, $F(1,6)=6.93$, $p<0.05$, revealed that GMRT scores obtained using the modified helmet were higher than SPIN scores in the standard helmet. The significant two-way interaction between conditions and tests, $F(1,6)=27.96$, $p<0.005$, indicated that diver–diver scores on both tests were lower than topside–diver SPIN scores, which, in turn, were lower than topside–diver GMRT scores. There was also a significant three-way interaction.

Also shown in Table I are the mean percent intelligibility scores and standard deviations for the GMRT and SPIN as a function of the type of HSU used during the diver–diver presentations in the water during DD 96. An ANOVA conducted on these GMRT and SPIN test scores revealed no significant effects across HSUs for either the modified or the standard helmet.

2. Comparison of SIIs and intelligibility scores for DD 96

Correlations were computed to examine the relationship between the GMRT and SPIN percent intelligibility scores and their corresponding SII values obtained during DD 96. Four general relationships were examined: (a) GMRT score and the SII value using the GMRT stimuli (SII-GMRT), (b) SPIN score and the SII value using the SPIN stimuli (SII-SPIN), (c) GMRT and SPIN scores, and (d) SII-GMRT and SII-SPIN. Each subject heard two different GMRT and SPIN lists, and correlations were computed in two ways: (1) on the overall means of the scores by test list and (2) on the individual lists heard by divers as a function of helmet. The correlations are shown in Table II.

The correlations measured were considerably higher for the modified helmet than for the standard helmet, which is consistent with the main effect for helmets found in the ANOVA. This finding indicates that, except for the GMRT-SPIN percent correct correlation, intelligibility test data and

TABLE II. Correlations examining the relationship between speech intelligibility scores and Speech Intelligibility Index (SII) values. Correlations were computed using scores by individual test lists per subject and means of scores for two test lists per subject.

Correlations by test list	Overall (df=14)	Standard helmet (df=6)	Modified helmet (df=6)
GMRT×SII-GMRT	0.602[b]	0.373	0.668[d]
SPIN×SII-SPIN	0.007	−0.086	0.324
GMRT×SPIN	0.708[a]	0.511	0.770[c]
SII-GMRT×SII-SPIN	0.261	0.103	0.616
Correlations by mean scores	**Overall (df=6)**	**Standard helmet (df=2)**	**Modified helmet (df=2)**
GMRT×SII-GMRT	0.706[c]	0.474	0.833
SPIN×SII-SPIN	0.076	0.076	0.331
GMRT×SPIN	0.755[c]	0.529	0.906[d]
SII-GMRT×SII-SPIN	0.252	−0.161	0.624

Significance levels: [a]$p<0.01$; [b]$p<0.02$; [c]$p<0.05$; [d]$p<0.10$.

SIIs collected from divers wearing the standard helmet do not correlate as well as might be expected. One interesting finding was that the SPIN percent score correlates fairly well with the GMRT percent score, but does not correlate with the SII-SPIN.

III. DISCUSSION

A. SII assessments

The combined dry chamber data from our laboratory work, DD 95, and DD 96 shown in Fig. 4 reveal that SII values and percent intelligibility for the GMRT and SPIN stimuli decreased incrementally as the background noise level increased. These findings agree with earlier studies by Kryter (1962a, b) on the original Articulation Index, which produces calculations of speech audibility similar to those of the SII. Low SII values agreed with low GMRT and SPIN intelligibility scores, suggesting that little speech information was available to the listener when the background noise level was high. Likewise, higher SII values corresponded to higher percent correct scores when the noise levels were low, indicating that more speech information was available, contributing to enhanced intelligibility. Despite differences in the environments where we employed the SII for the nondiver and diver portions of DD 95 and DD 96, there was a very close match at all points along the curves. The data collected in DD 96 contributed to the lower portion of the curve showing performance in the most adverse noise conditions. The similarity of these results across the three different experiments and varying conditions lends support to the strength of the SII measurement for predicting speech intelligibility.

A consistent finding observed across all studies we performed using both the GMRT and SPIN has shown that SPIN scores were generally lower than the GMRT scores. This difference is possibly a result of the difference in response format across the two tests. The closed-set nature of the GMRT allows for a guessing floor, while the open-set format of the SPIN makes responses more difficult. However, even when GMRT scores were corrected to compensate for its closed-set nature, this disparity in scores still existed. Also, the majority of SPIN errors were on the low-predictability stimulus items where context could not be used to help determine an appropriate response.

Overall, the SII data collected in DD 96 confirm our findings from DD 95 suggesting that it can be used as a speech intelligibility assessment tool to compare linear system components in a diving environment. Ultimately, we wanted to be able to use the SII as an objective tool for speech intelligibility assessment that could test different communication systems, including HSUs, without the use of Navy divers.

Unfortunately, we have recently ascertained that the SII may not to be an appropriate tool for intelligibility measurements of speech processed through inherently nonlinear devices, such as HSUs. Fundamentally, the SII only measures the audibility of a signal above the background noise level, and its calculations are based on long-term average spectra of normal, intelligible speech. The speech coming into an HSU is distorted, that is, not within the normal frequency range of speech, and it is often unintelligible. Because of the nature of the signal processing performed by HSU technology, it is not appropriate to use the SII to compare such devices since the SII will only provide information about how well noise is suppressed through the HSU, not how intelligible the speech is. However, based on our findings during DD 95 and DD 96, the SII remains a promising tool for testing other diver communication system components, especially helmets, within a helium environment.

B. GMRT and SPIN assessments

The purpose of the in-water intelligibility testing using the SPIN and GMRT was to expand our previous findings by using these tests in a noisy, in-water environment with well controlled variables. We wanted to determine whether either test was sufficiently sensitive to accurately measure small changes in the communication systems when in a high level of background noise. The significant main effects observed for test condition (i.e., topside–diver scores higher than diver–diver scores) and for type of test (i.e., GMRT scores higher than SPIN scores) were found across both studies. Interestingly, this significant difference in test scores for the GMRT and SPIN was not found during DD 92 (Mendel *et al.*, 1995) which measured intelligibility only in a dry chamber. Most likely, the results obtained during DD 92 in the dry chamber were ceiling effects which contributed significantly to the inability of either the GMRT or SPIN tests to be more sensitive than the other in a less adverse environment. Thus these findings suggest that the GMRT may be more useful under noisier conditions such as when used in the water.

In DD 96, we were able to isolate the variable of helmet type and HSU to determine if either the SPIN or GMRT could measure differences in such system components. The significant ANOVA findings suggest that when the system component being compared has a large measurable difference, such as in helmet type, the SPIN and GMRT can be used as accurate speech intelligibility assessment tools. However, when the differences are more subtle, such as with differences in HSUs, neither the SPIN nor GMRT appears sensitive enough to make such distinctions. The two types of HSU used in this study were either too similar for differences to be measured, or the subtle differences in the HSUs may not have affected intelligibility enough for the GMRT or SPIN to assess those differences.

Stimulus context continued to enhance intelligibility in the SPIN test, as the lowest percent of error occurred among the high-predictability items that had semantic context imbedded in them. Also, although the GMRT scores were significantly higher than the SPIN scores, the overall mean scores were quite low, especially in the diver–diver condition. These low overall scores provide more justification for our assertion that these tests lack the sensitivity needed to measure small but noticeable intelligibility differences. Therefore the findings from the in-water testing generally agree with our earlier work and suggest that the traditional speech intelligibility tests we have studied (GMRT and SPIN) are inadequate tools for determining the true effectiveness of varied aspects of diver communication systems under adverse listening conditions.

IV. CONCLUSION

The results from Deep Dives 1995 and 1996 along with our laboratory work suggest that the SII holds some promise as a speech intelligibility assessment tool in a helium environment. The robust relationship between SII values and speech intelligibility performance measured here lends strong justification for the use of SIIs for some speech intelligibility assessments. The results of the studies presented here suggest that the GMRT may be more useful than the SPIN in assessing speech intelligibility in a noisy helium environment like the in-water conditions. It should be noted that the GMRT is a tool that is designed to measure the intelligibility of speech content through all the communication system components whereas the SII can only measure the noise masking effects on speech. Nonetheless, SII measurements may be preferred for some components since they can be performed using a manikin and would alleviate the need and expense of testing divers under difficult in-water conditions.

ACKNOWLEDGMENTS

The authors thank the Navy Experimental Diving Unit in Panama City, Florida for the use of their facilities and personnel during the collection of the data obtained from the Navy divers. Special appreciation is expressed to Jeff McDonell and Larry Gibbs for their work on the SII data collection. We also thank the anonymous reviewers for their constructive comments on earlier versions of this manuscript. This work was supported by the Naval Sea Systems Command.

ANSI (**1970**). ANSI S3.5-1970, ''American National Standard Methods for the Calculation of the Articulation Index'' (American National Standards Institute, New York).

ANSI (**1995**). ANSI S3.5-1995, ''American National Standard Methods for the Calculation of the Speech Intelligibility Index-Draft V4.0'' (American National Standards Institute, New York).

Belcher, E. O., and Andersen, K. (**1983**). ''Helium speech enhancement by frequency-domain processing,'' in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Boston, MA (IEEE, New York), pp. 1160–1163.

Copel, M. (**1966**). ''Helium voice unscrambling,'' IEEE Trans. Audio Electroacoust. **AU-14**, 122–126.

Dildy, C. (**1992**). *A Helium Speech Unscrambler—Version 2.0* (Panama City, Florida; Copyright-All Rights Reserved).

Giordano, T. A., Rothman, H. B., and Hollien, H. (**1973**). ''Helium speech unscramblers—A critical review of the state of the art,'' IEEE Trans. Audio Electroacoust. **AU-21**, 436–444.

Griffiths, J. D. (**1967**). ''Rhyming minimal contrasts: A simplified diagnostic articulation test,'' J. Acoust. Soc. Am. **42**, 236–241.

Hendrix, J. E. (**1996**). Personal communication.

House, A. S., Williams, C. E., Hecker, M. H. L., and Kryter, K. D. (**1965**). ''Articulation testing methods: Consonantal differentiation with a closed response set,'' J. Acoust. Soc. Am. **37**, 158–166.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Kryter, K. D. (**1962a**). ''Methods for the calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Kryter, K. D. (**1962b**). ''Validation of the articulation index,'' J. Acoust. Soc. Am. **34**, 1698–1702.

Mendel, L. L., Hamill, B. W., Crepeau, L. J., and Fallon, E. (**1995**). ''Speech intelligibility assessment in a helium environment,'' J. Acoust. Soc. Am. **97**, 628–636.

Pavlovic, C. V. (**1991**). ''Speech recognition and five articulation indexes,'' Hearing Instrum. **42** (4), 20–23.

Stover, W. R. (**1967**). ''Technique for correcting helium speech distortion,'' J. Acoust. Soc. Am. **41**, 70–74.