Evaluation of speech intelligibility with the coordinate response measure

Douglas S. Brungart

Citation: The Journal of the Acoustical Society of America **109**, 2276 (2001); doi: 10.1121/1.1357812
View online: https://doi.org/10.1121/1.1357812
View Table of Contents: https://asa.scitation.org/toc/jas/109/5
Published by the Acoustical Society of America

Evaluation of speech intelligibility with the coordinate response measure

Douglas S. Brungart[a)]
Air Force Research Laboratory, Human Effectiveness Directorate, Wright-Patterson AFB, Ohio 45433-7901

(Received 20 April 2000; revised 28 June 2000; accepted 23 January 2001)

The sentences in the coordinate response measure (CRM) corpus described in a recent letter to this journal [Bolia *et al.*, J. Acoust. Soc. Am. **107**, 1065–1066 (2000)] have been used to measure speech intelligibility as a function of signal-to-noise ratio with a speech-spectrum-shaped noise masker. The data from this experiment, along with those of an earlier experiment comparing intelligibility with the CRM and the well-known modified rhyme test (MRT), have also been used to estimate performance with the CRM as a function of the articulation index (AI) for a variety of different masking signals. The results provide a normative evaluation of the CRM for potential users of the CRM corpus, and can be used to compare the CRM with other measures of speech intelligibility. [DOI: 10.1121/1.1357812]

PACS numbers: 43.71.Gv, 43.72.Kb [DOS]

I. INTRODUCTION

In a recent letter to this journal, Bolia *et al.* (2000) described a publicly available speech corpus for speech intelligibility experiments based on the coordinate response measure (CRM) originally developed by Moore (1981). This speech corpus consists of sentences of the form ''Ready (call sign) go to (color) (number) now'' spoken by eight talkers with each of eight call signs, four colors (''blue,'' ''red,'' ''green,'' and ''white''), and eight numbers (1–8). Speech intelligibility can be assessed with the corpus from the percentage of correct number and color identifications for the utterances addressed to the listener's assigned call sign. The CRM was originally developed by researchers at the Air Force Research Laboratory (Moore, 1981) to provide an intelligibility test with greater relevance to military communications than the commonly used modified rhyme test (MRT) (House *et al.*, 1965). Moore measured speech intelligibility with both the CRM and the MRT in a variety of jamming conditions and found that performance with the CRM was less sensitive to interfering noise than the MRT, but that overall performance was highly correlated across the two intelligibility tests. However, little additional information about speech intelligibility with the CRM has been reported since this original publication. This short article presents a normative evaluation of speech intelligibility with the CRM speech corpus as a function of signal-to-noise ratio with a speech-spectrum-shaped masking noise. The results of this evaluation, along with the original data provided by Moore, are used to relate performance with the CRM to the well-known articulation index (Kryter, 1962). This provides a basis for comparing the CRM with other measures of speech intelligibility.

[a)]Electronic mail: douglas.brungart@wpafb.af.mil

II. CRM INTELLIGIBILITY EXPERIMENT

A. Procedure

Nine normal-hearing paid volunteers with previous experience in the CRM task (four males, five females) participated in the experiment. The listeners sat at a control computer in a sound-deadened room and heard diotic headphone presentations of the sentences from the CRM corpus. Each utterance was masked by a noise signal that was spectrally shaped (with a 512-point FIR filter) to match the average overall spectrum of the 2048 sentences in the CRM corpus[1] and rectangularly gated to the same length as the speech signal. The overall levels of the masking noises varied randomly from approximately 64 to 70 dB in 1-dB steps,[2] and the levels of the speech signals were scaled to produce signal-to-noise ratios (measured from the rms power of the speech and noise) ranging from −18 to 15 dB in 3-dB steps. Only sentences with the call sign ''Baron'' were used in the experiment, and the sentence presentations were balanced so that each listener heard exactly 120 sentences from each of the eight talkers in the corpus in random order. All other parameters of the sentence presentation, including color, number, and signal-to-noise ratio, were randomly selected (with replacement) on each trial. After each sentence presentation, the listeners identified the spoken color and number combination by moving the mouse pointer to the appropriate colored digit on the screen of the control computer. The trials were divided into 8 blocks of 120 trials, with 1–2 blocks of trials collected during each day of the experiment.

B. Results

The results show that performance in the CRM is related to signal-to-noise ratio by an S-shaped curve typical of most measures of speech intelligibility in noise (Fig. 1). At signal-to-noise ratios (SNRs) of 0 dB or above, performance plateaus at about 98%-correct responses. At SNRs less than 0 dB, performance degrades rapidly, falling by about 10%/dB as the signal-to-masker ratio declines from −6 to −12 dB. At lower SNRs, performance again levels off as color iden-

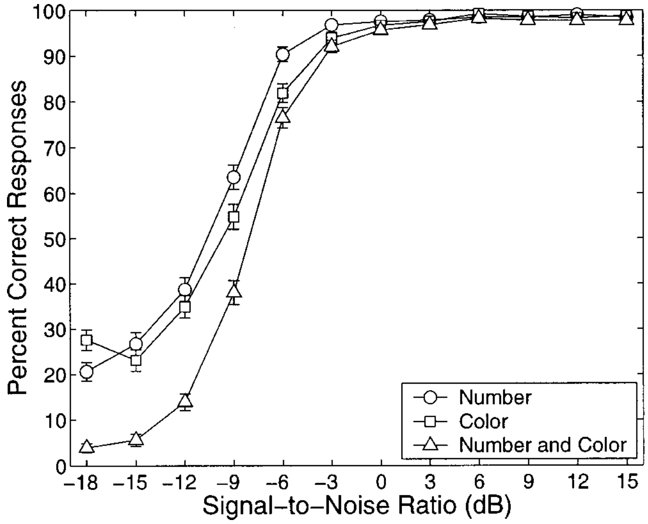

FIG. 1. Percentage of correct identifications of color, number, and both color and number as a function of signal-to-noise ratio with a speech-spectrum-shaped masker. The error bars have been calculated from the arcsine transformation and represent ±1.4 standard errors. Mean values with nonoverlapping error bars would be significantly different at the $p<0.05$ level in a one-tailed *t* test.

tifications reach chance performance (25%). There are two interesting features of the data. The first is that correct identifications of the color and number in each utterance were essentially independent events. The probability of correct overall (number and color) identifications differs from the product of the probabilities of correct number and color identifications by an average of only 1.1%. This implies that coarticulation played a relatively minor role in the intelligibility of the color–number pair.

The second interesting feature is that the percentage of correct number identifications was consistently higher than the percentage of correct color identifications, in spite of the larger vocabulary in the number category (eight numbers versus four colors). This somewhat counterintuitive result, which indicates a lack of phonetic balance in the four color coordinates used in the response set, can be explained in part by differences in intelligibility across the key words in the CRM vocabulary. The digits ''six'' and ''seven'' were correctly identified more frequently than the other color and number coordinates (Fig. 2). This result is probably related to the high-frequency content of the strident fricative /s/, which stands out against the primarily low-frequency (<5 kHz) speech-shaped masking noise. The second syllable in the word ''seven'' may also have contributed to its enhanced intelligibility. The distinctive diphthong /ai/ may account for the relatively large number of correct identifications of the key words ''five'' and ''white.'' The large differences in intelligibility across the CRM key words suggest that the phonetic features of the color and number words played a larger role in determining overall performance than the number of alternatives in the color and number sets.

Although there were some variations in overall performance across the talkers and listeners used in the experiment, these differences would have relatively little impact on the measurement of speech reception thresholds with the CRM. The threshold SNR values for 50%-correct identification of

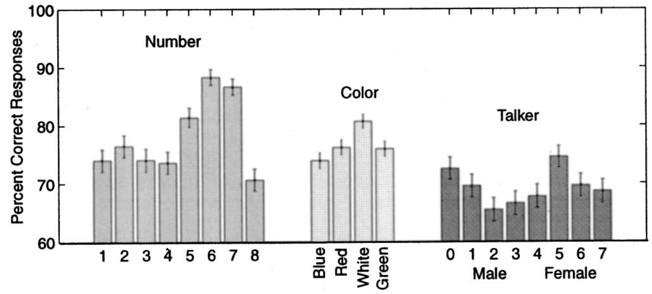

FIG. 2. Effects of number, color, and talker. The first set of data shows the percentage of correct number identifications as a function of the number in the target phrase. The second set shows the percentage of correct color identifications as a function of the color in the phrase. The third set shows the percentage of correct overall responses as a function of the talker (numbered 0–7 as in the CRM corpus). The error bars have been calculated from the arcsine transformation and represent ±1.4 standard errors. Mean values with nonoverlapping error bars would be significantly different at the $p<0.05$ level in a one-tailed *t* test.

both the color and number coordinates fell between −8.6 and −7.3 dB for seven of the eight talkers and seven of the nine listeners used in the experiment. The overall percentages of correct responses varied by about 10% across the different talkers used in the corpus (Fig. 2), and by about 10% across the different listeners used in the experiment.

III. RELATIONSHIP BETWEEN THE CRM AND THE ARTICULATION INDEX

The results in Fig. 1 show how the percentages of correct responses in the CRM vary with signal-to-noise ratio for a speech-shaped noise masker. In order to compare the results to other measures of speech intelligibility, it is helpful to relate these results to the articulation index (AI). The AI, which estimates the intelligibility of speech from the spectral properties of the speech and the masking noise, has been shown to accurately predict performance in a variety of phonetically balanced intelligibility tests across a wide range of different listening environments (Kryter, 1969). In Fig. 3, the overall identification results of the CRM test shown in Fig. 1 have been plotted as a function of the AI for each of the 12 SNRs used in the experiment. The articulation indices were calculated using the 20-band method described by Kryter (1962), with the assumption that the overall rms noise level was 70 dB across all of the trials.

The performance data plotted in Fig. 3 provide a guideline for estimating performance with the CRM as a function of the AI, but they should be interpreted with some caution. The CRM has a very restricted vocabulary, and its color and number response words are not phonetically balanced. It is therefore problematic to assume that CRM performance with a speech-shaped noise masker will be the same as performance with all other types of masking signals that produce the same AI value. The original CRM data collected by Moore (1981) provide some insight into the performance of the CRM with different types of masking sounds. Moore collected data using both the CRM and MRT methods with live talkers and a panel of ten listeners wearing full military flight gear, including oxygen masks and flight helmets. He tested five different jammers (an FM tone, an FM drifting

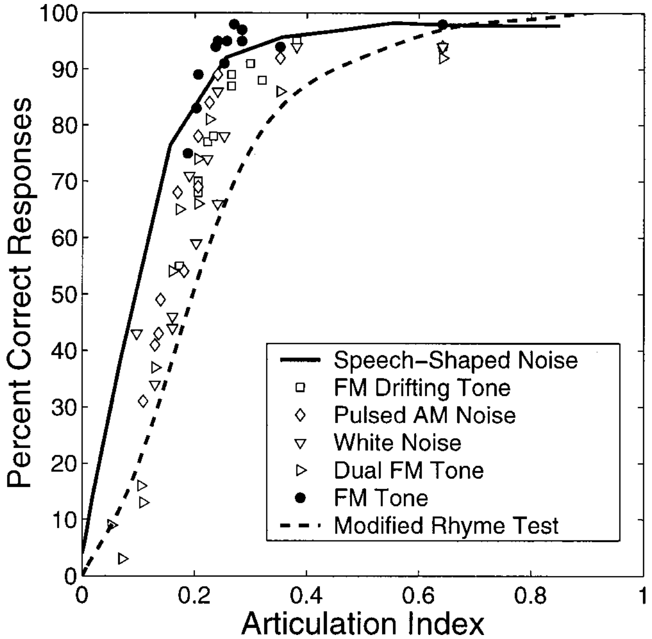

FIG. 3. Performance in the CRM as a function of the articulation index. The solid line represents the percentage of correct color and number identifications in the presence of a speech-shaped noise masker as a function of the articulation index, which was calculated by the 20-band method (Kryter, 1969). Each symbol in the plot represents a single jammer type, jammer-to-noise ratio, and ambient noise level from Moore's 1981 study comparing the CRM to the MRT. The performance curve for the MRT as a function of AI that was used to transform Moore's data is also shown (Kryter, 1969). See the text for details.

tone, a white-noise modulated AM tone, a pulsed AM noise, and dual FM swept tones) at six different jammer-to-signal ratios both in quiet and in the presence of a 105-dB jet-engine-shaped ambient noise. Although Moore did not calculate the AI directly, it is possible to derive a rough estimate of the AI in each of these 60 listening conditions by comparing the MRT performance measured by Moore to the MRT versus AI curve provided by Kryter (dashed line in Fig. 3). These estimates can then be used to plot CRM performance as a function of AI in each listening condition (symbols in Fig. 3). The overall shape of the curve defined by these data points is similar to the response curve measured with the speech-shaped noise masker, but has a steeper slope and is shifted to the right of the speech-shaped noise curve. This shift reflects the additional difficulty introduced into Moore's experiment by requiring the listener to correctly identify one of three call signs (''Ringo,'' ''Laker,'' or ''Baron'') in each stimulus presentation and respond only to those phrases containing the call sign assigned to that listener prior to the experiment. The data from four of Moore's jamming conditions (open symbols) are clustered into a reasonably well-defined response curve, which is remarkable when one considers that these CRM performance estimates are sensitive to errors in both the determination of MRT performance and in the determination of CRM performance in each listening condition. However, the data from the FM jammer (filled symbols) consistently indicate better performance than the other four types of jamming signals (open symbols) under listening conditions with similar AI values. Although Moore does not provide a detailed description of the FM jammer, it is clearly an example of a masking signal that interferes substantially less with the phonetically unbalanced CRM vocabulary than with the phonetically balanced word list used in the MRT. Based on these results, it appears that the relationship between CRM performance and AI shown in Fig. 3 is reasonably stable for many types of masking sounds, but that performance can deviate substantially from this curve for certain specific types of masking signals. It is therefore not appropriate to use the CRM to determine the AI of speech in the presence of an arbitrary masking signal.

It is, nevertheless, useful to use the AI performance curves as a basis for comparing the CRM to other types of speech-perception tests. For example, the AI performance curve for the MRT shown in Fig. 3 allows a direct comparison between the CRM and MRT. For all the different types of masking signals tested, it is clear that the CRM is substantially more sensitive to changes in AI than the MRT at AI values less than or equal to 0.25, and substantially less sensitive than the MRT at AI values greater than 0.25.

IV. APPLICATIONS OF THE CRM

While the limited vocabulary of the CRM prevents its use as a replacement for comprehensive, phonetically balanced intelligibility measures such as the MRT, the CRM does have advantages over other speech-intelligibility tests that justify its use in some special testing situations. One advantage demonstrated by these results is the sensitivity of the CRM to small intelligibility changes in extremely difficult listening environments (characterized by AI values less than 0.25). This feature makes the CRM attractive for testing intelligibility in very noisy environments. It also makes the CRM an excellent threshold test for determining how powerful a jamming signal must be to render a communications channel inoperative. Another advantage of the CRM is its intrinsic portability across different languages. Since all languages contain words for colors and numbers, the CRM can provide a rough functional measure of intelligibility without deriving a phonetically balanced word list for each language tested. A third advantage of the CRM corpus is its applicability to speech-intelligibility testing with multiple simultaneous talkers. The call signs in the corpus allow the experimenter to designate the target phrase without relying on differences in its location, onset time, or talker characteristics to distinguish it from the simultaneous masking phrases. The listeners are simply instructed to respond with the color and number coordinates spoken in the phrase addressed to their assigned call signs. A final advantage of the CRM is the relative ease of setting up and running speech-intelligibility experiments with the publicly available CRM corpus (Bolia *et al.*, 2000). Because the listener is always limited to the same 32 possible responses, every trial of the CRM can be evaluated with exactly the same response list. This makes it much easier to collect and process intelligibility data with the CRM than with sentence-based tests or phonetically balanced tests such as the MRT. Thus, although the CRM should not be viewed as a comprehensive measure of speech intelligibility, it does appear to be representative of other speech perception tests and it should be considered when a rapid measure of functional intelligibility is desired.

ACKNOWLEDGMENTS

The author would like to thank Mark Ericson, Todd Nelson, Bob Bolia, Brian Simpson, Kim Scott, Tom Moore, and Howard Rothman for their helpful comments in the preparation of this manuscript.

[1]Note that the speech materials in the CRM corpus have been low-pass filtered with an 8-kHz cutoff frequency.

[2]The listeners were allowed to choose a comfortable listening level for the stimuli in the experiment.

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

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

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

Kryter, K. C. (**1969**). ANSI S3.5-1969, ''American National Standards Methods for Calculation of the Articulation Index'' (American National Standards Institute, New York).

Moore, T. (**1981**). ''Voice communication jamming research,'' in AGARD Conference Proceedings 331: Aural Communication in Aviation, pp. 2:1–2:6, Neuilly-Sur-Seine, France.