Psychometric properties of the coordinate response measure corpus with various types of background interference[a)]

David A. Eddins
Department of Communication Sciences & Disorders, Global Center for Hearing & Speech Research, University of South Florida, 4202 East Fowler Avenue, PCD 1017, Tampa, Florida 33620
deddins@usf.edu

Chang Liu
Department of Communication Sciences and Disorders, University of Texas at Austin, 1 University Station A1100, Austin, Texas 78712
changliu@mail.utexas.edu

Abstract: The coordinate response measure (CRM) corpus has gained broad acceptance as a research tool for investigating speech intelligibility in background competition and has been widely used in studies of informational masking. The purpose of this study is to establish the psychometric characteristics of CRM target-word identification in various backgrounds with the goal of being able to determine when it is appropriate or not to use adaptive threshold procedures with the CRM corpus. Target-word identification performance based on adaptive tracking mapped directly onto the monotonic psychometric functions obtained for two-talker, four-talker, and cafeteria noise interferers.

PACS numbers: 43.71.Gv, 43.71.An [SGS]
Date Received: November 16, 2011 **Date Accepted:** December 29, 2011

1. Introduction

There are many factors that influence the ultimate choice of speech materials to be used in a listening experiment. These factors include but are not limited to acoustics, phonetics, semantics, syntax, predictability, stimulus duration, talker characteristics, and associated background interference. For standard test materials, the choice of materials may also be influenced by the availability of normative data, associated measurement methods, overall test efficiency, and known psychometric properties. The purpose of this report is to examine psychometric properties of the coordinate response measure (CRM) speech corpus (Bolia *et al.*, 2000) using various interfering sounds with a goal of determining if and when adaptive psychophysical methods are reasonable to use with the CRM corpus.

Sentences in the CRM corpus take the form "Ready [call sign] go to [color] [number] now." There are eight call signs (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, Tiger), four colors (blue, red, green, white), and eight numbers (1–8) yielding 256 different sentences recorded for eight different talkers giving a total of 2048 sentences in the corpus. It is most common to assess speech intelligibility by scoring the percent correct identification of the color and number associated with a given target call sign. The CRM has been championed for use in studies in which speech is masked by speech because the format of the speech materials allows the listener to lock onto a target phrase signified by its call sign even when competing sentences from

[a)]Portions of this research were presented at the 159th Meeting of the Acoustical Society of America, April, 2010 in Baltimore, MD.

the same corpus are presented simultaneously and from the same location. The corpus does not maintain specific phonetic or phonemic balance and does not permit an exhaustive analysis of confusion matrices. Nevertheless, its advantages include limited linguistic variation, limited predictability due to context, and limited vocabulary size. Because it is a closed-set format, it is easy to administer and score, and memory for items presented previously has minimal impact when those items are repeated. Finally, it is one of the few corpora suitable for repeated item presentation in studies with large sets of experimental conditions.

In most published studies, CRM target sentence identification has been measured using a method of constant stimuli where signal-to-noise ratio (SNR) is varied in fixed steps over a pre-defined range and the results are expressed as percent target identification as a function of SNRs, thereby yielding a psychometric function. However, it is possible that the utility of the CRM can be maximized if performance can be reliably measured using adaptive or "up-down" methods (e.g., Levitt, 1971; McMillan and Creelman, 1991), thereby reducing the number of trials in an experiment while focusing the majority of trials on a target percent correct level of performance. As noted by Levitt (1971), an essential assumption when using adaptive measurement methods is that the underlying psychometric function is monotonic with respect to the independent variable in the experiment (i.e., the signal-to-noise ratio). Informally, we and others have been criticized for using adaptive threshold methods for measuring performance using the CRM materials based on the fact that, in some masker conditions, the form of the function relating percent correct identification of color and number is non-monotonically related to changes in SNR ratio. It is important to note that in some masker conditions, CRM target identification monotonically increases with increasing SNR.

Based on the work of Brungart and colleagues (e.g., Brungart, 2001a,b; Brungart and Simpson, 2007), the following summary can be stated. A single-talker interferer yields a non-monotonic psychometric function when the interferer is the same talker (Brungart 2001a,b; Brungart and Simpson, 2007), a different talker of the same sex (Brungart 2001a,b; Brungart and Simpson, 2007), or a different talker of the different sex relative to the target sentence (Brungart 2001a,b; Brungart and Simpson, 2007). Similarly, when the target sentence and a single-talker interferer are presented to the same ear, and a second interfering sentence is presented to the other ear, the resulting psychometric function is non-monotonic only if one of the interferers is the same talker as the target sentence (Brungart and Simpson, 2007). Finally, if one interfering sentence is spoken by a talker with the same sex as the target talker, and a second interfering sentence spoken by a talker of the same or different sex is presented at a different SNR as the first interferer, then a non-monotonic psychometric function may result (Brungart, 2001b). On the other hand, when the interferer presented to the target ear consists of two or more talkers (Brungart, 2001b), is speech-shaped noise (Brungart 2001a,b), or envelope-modulated noise (Brungart 2001a,b), monotonic psychometric functions have been reported.

In this study, we seek to establish whether or not it is feasible to measure CRM target identification performance in two-talker, four-talker, or cafeteria noise interferers using adaptive psychophysical methods.

2. Method

2.1 Listeners

Thirteen young, normal-hearing listeners ranging in age from 22 to 26 years participated in this study. All listeners were native speakers of American English and had pure-tone thresholds ≤ 15 dB HL at octave intervals between 250 and 8000 Hz (ANSI, 2010). All procedures were approved by the Institutional Review Board at the University of Texas Austin and all listeners consented to participate.

2.2 Stimuli

Target sentences from the coordinate response measure (CRM) corpus took the form "ready [call sign] go to [color] [number] now." The number 7 was excluded since its bisyllabic phonetic structure differed from numbers 1 to 6 and 8. Thus, there were a total of 224 combinations of eight call signs, four colors, and seven numbers. The CRM corpus includes eight different talkers. The stimuli for talker 2 (adult, male) were used here. Correct target identification was based on identification of both the number and color in the target sentence.

Three types of interfering sounds were used: two-talker CRM interferer, four-talker babble, and cafeteria noise. The two-talker CRM interferer was composed of two sentences from the CRM corpus, one spoken by a different male talker than the target sentence and one spoken by a female talker. The interfering sentences had different call-signs, numbers, and colors than the target sentence and were adjusted to have the same root-mean-square (rms) level prior to summation. The four-talker babble was taken from a commercially available recording distributed by Auditec of St. Louis and was composed of two male and two female talkers reading independent passages. The cafeteria noise was taken from the Nonsense Syllable Test of Resnick *et al.* (1975) also distributed by Auditec of St. Louis. For the two-talker CRM interferer, the onsets of the target and interfering sounds were simultaneous. For the four-talker babble and cafeteria noise, a three-second segment of interferer was selected randomly for each trial from a longer recording and the target sentence was temporally centered within the interferer segment. The overall level of the interfering sounds was fixed at 70 dB SPL and the target sentence level was varied to produce the required signal-to-noise ratio. Sound pressure levels were measured at the output of the insert earphones via a GRAS IEC 126 2-cm^3 coupler connected to a Larson-Davis (Model 2800) sound level meter (linear weighting) with a Larson-Davis (Model 2575) microphone. Stimuli were re-sampled to 44 828 Hz and were presented to the right ear via a calibrated ER-2 insert earphone. Stimulus presentation and response collection were controlled by TDT (Tucker-Davis Technologies, Alachua, Florida) SYKOFIZX software and a series of TDT hardware modules including an enhanced real-time processor (RP2.1), programmable attenuator (PA5), signal mixer (SM5), and headphone buffer (HB7).

2.3 Procedures

2.3.1 General procedures

Listeners were seated in a sound-treated booth (Tracoustics Model RS244) in front of a LCD screen and PC mouse. The LCD screen displayed a subject interface consisting of a seven-column, four-row matrix of buttons with labels corresponding to the CRM response set (seven digits and four colors). On each trial, following stimulus presentation, listeners responded by button press using the mouse. The listener was required to correctly identify the number and color of the target sentence for the item to be scored as a correct response.

2.3.2 Method of constant stimuli

The SNRs were manipulated in 3-dB steps from −15 to +6 dB for the four-talker babble and cafeteria noise, and from −9 to +12 dB for the two-talker CRM interferer. For a given condition, each of the 28 target sentences (4 colors × 7 numbers) was present twice at each level, resulting in a total of 56 trials for each condition. On each trial, the call sign of the target sentence was randomly selected from the eight choices. Thus, percent correct speech identification was based on 56 trials for each condition and each listener. For a given interferer type, the order of SNR conditions was randomized.

2.3.3 Adaptive up-down tracking

Target identification thresholds were estimated for the three types of interferer using a single interval, 28-alternative forced-choice procedure with either a two-down, one-up

or a one-down, two-up tracking algorithm, estimating points on the psychometric function corresponding to 70.7 and 29.3% correct (Levitt, 1971). The final threshold for each condition was calculated by averaging thresholds for two blocks of 56 trials (i.e., each target stimulus was presented twice). Threshold was estimated for an additional block if thresholds for the first two blocks differed by more than 2 dB. This occurred on 17 of 78 conditions. Overall, the order of the two measurement methods and three interferer types was randomized for each listener.

3. Results

Figure 1 shows the psychometric functions obtained for each of the three interferer types in panels A (two-talker), B (four-talker), and C (cafeteria noise) and the fitted psychometric functions based on the average fitting parameters across the 13 subjects (see details of panel D below). Target sentence identification, in terms of percent correct identification, is plotted as a function of SNR. The bold function in panels A, B, and C represents the percent correct identification at each of the eight SNRs averaged across the 13 listeners. Comparison across the three panels reveals that the psychometric functions for the four-talker babble and cafeteria noise are similar to each other and shifted to the left of the psychometric function for two-talker babble by about 6 dB. Psychometric functions for individual listeners resemble the classic sigmoidal shape and therefore the functions for each individual and each condition were fit with a model representing the logistic distribution as illustrated in Eq. (1)

$$p(c) = \alpha + \left((1.0 - \alpha) \Big/ \left(1.0 + e^{-(x-x_0)/b}\right)\right), \tag{1}$$

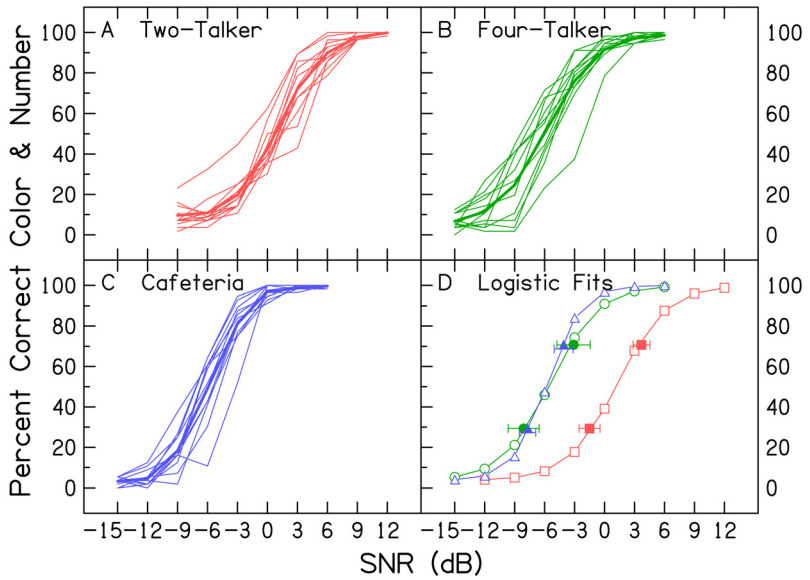

Fig. 1. (Color online) Psychometric functions and adaptive thresholds for target identification using the CRM and three types of interfering sounds. Psychometric functions are shown for each of the 13 listeners in panels A (two-talker CRM inteferer), B (four-talker babble), and C (cafeteria noise). Thick lines represent the function obtained by averaging percent correct scores at each SNR value across the 13 listeners. Open symbols connected by lines in panel D show the averaged psychometric functions derived from the individual fitting functions obtained using Eq. (1) for each of the 13 listeners, computing the average value of the growth constant (b) and midpoint (X_0), and solving for the correct SNR values. Filled symbols with error bars indicate the mean and standard deviation of the 70.7 and 29.3% correct thresholds across the 13 listeners estimated from the adaptive up-down method.

where $\alpha = 0.0357$ (chance performance for 1/28), b is the growth constant (i.e., steepness factor) of the function, X_0 is the midpoint (~51.5% correct) of the function, and x is the SNR value. The individual psychometric functions were fit using the Marquardt–Levenberg algorithm (Marquardt, 1963) implemented in SIGMAPLOT® v10.0. The algorithm searched for parameters that minimized the sum of the squared differences between the behavioral and predicted data using an iterative process. The RMS error associated with the fitting functions, when averaged across all listeners and SNR values, was 4.9% (two-talker), 4.7% (4-talker), and 3.1% (cafeteria noise). A chi-square test indicated significant goodness of fit for all three interference types ($p > 0.05$). Based on the fitted functions, SNR values corresponding to 70.7% and 29.3% correct were estimated, corresponding to the points on the psychometric function targeted by the two adaptive rules discussed below.

The fitting parameters growth constant (b) and midpoint (X_0), obtained from the logistic fits to the psychometric functions for each individual and each condition, were analyzed using separate one-factor (interferer type) repeated-measures ANOVAs with each fitting parameter as the dependent variable. There was a significant effect of interference type on the growth constant of the psychometric function ($F_{2,\ 24} = 13.92$, $p < 0.05$). Tukey *post hoc* tests showed that the growth constant for the cafeteria noise ($b = 1.68$ dB corresponding to the maximum slope at 13.9% per dB) was significantly steeper than for the two-talker ($b = 2.45$ dB corresponding to the maximum slope at 9.7% per dB) and four-talker ($b = 2.39$ dB corresponding to the maximum slope at 9.9% per dB) babble (both $p < 0.05$) with no significant difference between the two babble types ($p > 0.05$). Similarly, there was a significant effect of interference type on the midpoint (X_0) of the psychometric function ($F_{2,\ 24} = 289.81$, $p < 0.05$). Tukey *post hoc* tests showed that the midpoint for the two-talker CRM interferer ($X_0 = -1.3$ dB) was significantly greater than for the four-talker babble ($X_0 = -5.4$ dB) and the cafeteria noise ($X_0 = -5.7$ dB) (both $p < 0.05$) with no significant difference between the four-talker babble and cafeteria noise ($p > 0.05$).

The two adaptive procedures produced SNR values corresponding to either 70.7 (two-down, one-up) or 29.3 (one-down, two-up) percent correct. These values are shown in Fig. 1, panel D as solid symbols represented the average thresholds across the 13 listeners and the associated standard deviations. These values then were compared with estimates of the 70.7 and 29.3% correct points derived from the individual fitted functions using equation 1. Separate two-factor (method x interferer type) repeated-measured analyses of variance (ANOVA), with the threshold as the dependent variable, were conducted for the 70.7 and 29.3% correct data. Results indicated no significant difference in threshold between the two methods (70.7% threshold: $F_{1,12} = 3.28$, $p > 0.05$; 29.3% threshold: $F_{1,12} = 0.11$, $p > 0.05$), while there was a significant effect of interferer type for both thresholds (70.7% threshold: $F_{2,24} = 312.69$, $p < 0.05$; 29.3% threshold: $F_{2,24} = 199.63$, $p < 0.05$). No significant interaction effects were found for either thresholds (70.7% threshold: $F_{2,24} = 0.058$, $p > 0.05$; 29.3% threshold: $F_{2,24} = 0.90$, $p > 0.05$). Tukey *post hoc* tests indicated that thresholds were significantly higher for the two-talker CRM interferer than the four-talker babble and cafeteria noise (all $p < 0.05$), while threshold at 70.7% was significantly higher for the four-talker babble than the cafeteria noise ($p < 0.05$) with no significant difference in thresholds at 29.3% between the four-talker babble and cafeteria noise ($p > 0.05$). Visual inspection of the data in panel D of Fig. 1 shows the close correspondence between the adaptive thresholds and the six corresponding points on the psychometric functions averaged over the thirteen listeners. Figure 2 shows the relationship between the individual 70.7% and 29.3% thresholds from the two methods. The correspondence between the two methods is reflected by the degree to which data points are located near the diagonal.

4. Discussion and conclusions

Target identification for the CRM corpus in the presence of two-talker CRM interferer, four-talker babble, and cafeteria noise results in monotonically increasing percent

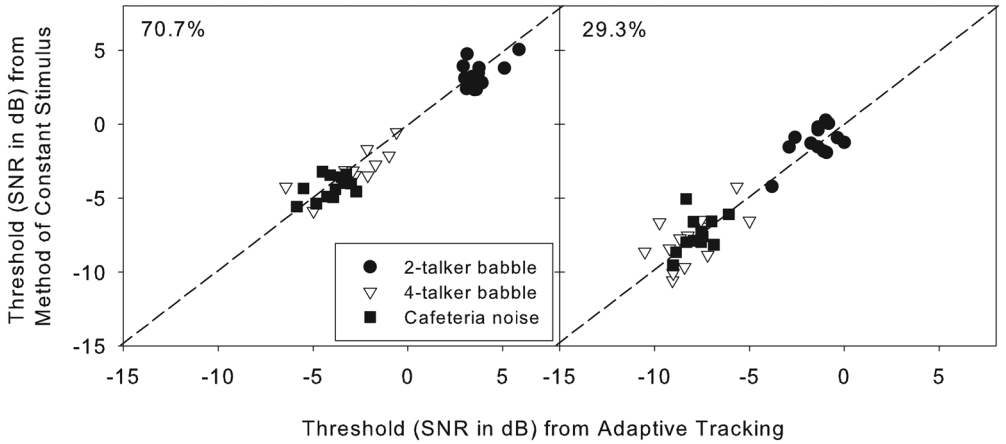

Fig. 2. The individual 70.7% (left panel) and 29.3% (right panel) thresholds (SNR in dB) from the method of constant stimuli as a function of thresholds (SNR in dB) from the adaptive tracking method for the three types of interferers. The diagonal dashed line in each panel indicates a perfect match between the two sets of thresholds. The correlations between the thresholds of the two methods across the three maskers were significant (Pearson $r = 0.971$ for the 70.7% thresholds and $r = 0.936$ for the 29.3% thresholds; $p < 0.05$).

correct performance as a function of signal-to-noise ratio. Thus, for these three maskers, the monotonicity assumption (Levitt, 1971; McMillan and Creelman, 1991) required for the use of adaptive up-down threshold procedures is met. Furthermore, the adaptive thresholds measured here are in close agreement with corresponding points on the psychometric function estimated for the same stimuli using the classic method of constant stimuli. The data for the two-talker CRM interferer are consistent with those reported by Brungart *et al.* (2001; see their Fig. 1, middle panel, filled triangles). Performance on the CRM in the presence of the four-talker babble and cafeteria noise used here have not been reported previously, but the data for four-talker babble used here (Auditec of St. Louis) are similar to the data for three-talker CRM interferer reported by Brungart *et al.* (2001; see their Fig. 1, lower panel, right pointing and upward pointing triangles). In the context of sounds previously used with CRM reviewed in the Introduction, it is clear that many types of interferer meet the monotonicity criterion for use with adaptive up-down methods. Obvious exceptions include single-talker interferers, maskers with temporal envelopes modeled after single-talker speech, and two-talker interferers in which one interfering talker is substantially lower in level than the other. The inherent properties of the CRM corpus combined with adaptive up-down methods make this corpus ideally suited for use in large-scale, parametric experimental designs requiring speech identification.

Acknowledgments

Work supported in part by NIH NIA Grant No. P01 AG009524.

References and links

ANSI (**2010**). S3.21-2010, *Methods for Manual Pure-tone Threshold Audiometry* (American National Standards Institute, New York).

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

Brungart, D. S. (**2001a**). "Evaluation of speech intelligibility with the coordinate response measure," J. Acoust. Soc. Am. **109**, 2276–2279.

Brungart, D. S. (**2001b**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2007**). "Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **122**, 1724–1734.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**(2), Suppl. 2, 467–77.
Marquardt, D. W. (**1963**). "An algorithm for least-squares estimation of nonlinear parameters," J. Soc. Indust. Appl. Math. **11**, 431–441.
McMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, New York).
Resnick, J. B., Dubno, J. R., Hoffnung, S., and Levitt, H. (**1975**). "Phoneme errors on a nonsense syllable test," J. Acoust. Soc. Am. **58**, Suppl. 1, 114.