

Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility

Jean C. Krause^{a)} and Louis D. Braida

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 16 February 2002; accepted for publication 31 July 2002)

Sentences spoken “clearly” (and slowly) are significantly more intelligible than those spoken “conversationally” for hearing-impaired listeners in a variety of backgrounds [Picheny, Durlach, and Braida, *J. Speech Hear. Res.* **28**, 96–103 (1985); Uchanski *et al.*, *J. Speech Hear. Res.* **39**, 494–509 (1996); Payton, Uchanski, and Braida, *J. Acoust. Soc. Am.* **95**, 1581–1592 (1994)]. However, it is unknown whether slower speaking rates are necessary for highly intelligible speech or whether an alternative form of clear speech exists at faster (i.e., normal) rates. To investigate this question, talkers with significant public speaking experience were asked to produce clear and conversational speech at slow, normal, and quick rates. A method for eliciting clear speech was introduced that ensured the clearest possible speech was obtained at each of these speaking rates. To probe for other highly intelligible speaking modes, talkers also recorded sentences in two other speaking modes: soft and loud. Intelligibility tests indicated that clear speech was the only speaking mode that provided a consistent intelligibility advantage over conversational speech. Moreover, the advantage of clear speech was extended to faster speaking rates than previously reported. These results suggest that clear speech has some inherent acoustic properties that contribute to its high intelligibility without altering rate. Identifying these acoustic properties could lead to improved signal-processing schemes for hearing aids. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1509432]

PACS numbers: 43.71.Bp, 43.71.Gv, 43.71.Ky [CWT]

I. INTRODUCTION

This report is concerned with clear speech, a speaking style that many talkers adopt in order to be understood more easily in difficult communication situations. Previous studies have demonstrated that this altered speaking style is significantly more intelligible (roughly 17 percentage points) than conversational speech for hearing-impaired listeners in a quiet background as well as for both normal-hearing and hearing-impaired listeners in noise and reverberation backgrounds (Picheny *et al.*, 1985; Uchanski *et al.*, 1996; Payton *et al.*, 1994). Furthermore, the intelligibility advantage is independent of listener, presentation level, and frequency-gain characteristic (Picheny *et al.*, 1985). These results suggest that signal-processing schemes that convert conversational speech to a sufficiently close approximation of clear speech could improve speech intelligibility for hearing aid users in many situations.

Before such signal-processing schemes can be developed, however, it is first necessary to determine the extent to which a reduction in speaking rate is responsible for the intelligibility improvement provided by clear speech, since the typical speaking rate for clear speech (100 words per minute) is roughly half that of conversational speech (Picheny, Durlach, and Braida, 1986). This question is particularly important for real-time hearing aid applications, since audio and visual signals must remain synchronized for maximum benefit to the listener. Although the role of speaking rate in highly intelligible speech is not well understood, some linguists hypothesize that clarity is independent of speaking

rate (Zwicky, 1972). If so, it should be possible to obtain clear speech and conversational speech at the same speaking rate. Therefore, this paper investigates: (1) whether alternative forms of clear speech can exist at normal speaking rates (or faster), and (2) whether clear speech (spoken slowly) has an intelligibility advantage over typical slow speech, without emphasis on clarity. In either case, it would follow that clear speech has some inherent acoustic properties, independent of rate, that account for its higher intelligibility. Furthermore, the subsequent task of identifying those properties would be greatly simplified, since clear and conversational speech at the same rate could be compared directly.

Most previous attempts to achieve clear speech at normal rates have focused on signal-processing techniques. For example, two studies investigated straightforward time-scale manipulations of clear and conversational speech, compressing clear speech to normal conversational speaking rates and expanding conversational speech to clear speaking rates. In the first of these studies (Picheny, Durlach, and Braida, 1989), the time scale of sentences was altered uniformly, and in a subsequent study (Uchanski *et al.*, 1996), a nonuniform time-scaling method was applied, altering phonetic segments within sentences to reflect the segmental-level durational differences previously measured between clear and conversational speech. Neither time-scaling procedure produced clear speech at normal speaking rates that was more intelligible than unprocessed conversational speech, though nonuniform time scaling was generally less harmful to intelligibility than uniform time scaling. Other work has examined the role of pauses in highly intelligible speech, because the reduced speaking rate found in clear speech is partly a result of more frequent and longer pauses (in conjunction with lengthened

^{a)}Electronic mail: jeanie@mit.edu

speech sounds) (Picheny *et al.*, 1986). One such study (Uchanski *et al.*, 1996) showed that key words excised from clear and conversational sentences have nearly the same intelligibility as the same words in sentence context, suggesting that differences in pause structure do not necessarily account for differences in intelligibility. This result is supported by another study (Choi, 1987) which found that artificial manipulations of pause structure do not substantially alter the intelligibility of clear and conversational speech; that is, adding pauses to conversational speech did not improve its intelligibility, and deleting pauses from clear speech did not decrease its intelligibility. Such manipulations did increase the effective speaking rate for clear speech, but not nearly enough to achieve normal speaking rates.

In addition to attempts to obtain clear speech at normal speaking rates through signal-processing techniques, one preliminary experiment (Uchanski *et al.*, 1996) sought to elicit clear speech at normal speaking rates naturally. In that experiment, a professional talker attempted to produce clear speech at a variety of rates. Results of intelligibility tests suggested that the talker could not improve his intelligibility without slowing down. However, only one talker was examined, and talkers vary considerably in their ability to produce highly intelligible speech (Bradlow, Toretta, and Pisoni, 1996). Therefore, more work in this area must be completed before any conclusions regarding the existence of naturally produced clear speech at normal speaking rates are justified.

In order to increase the chances of finding at least one talker in this study who could produce a form of clear speech at normal speaking rates, much attention was given to screening talkers, and a method was developed to elicit the clearest possible speech from talkers at a given speaking rate by providing them with training and feedback on intelligibility. In addition, the possibility was also investigated that speaking modes other than clear speech could exist with comparable or even greater intelligibility advantages relative to conversational speech. Some naturally occurring speaking modes differ acoustically from conversational speech [e.g., mothers addressing infants produce more extreme vowels (Kuhl *et al.*, 1997)], but without intelligibility measurements, it is unknown whether these speaking modes also differ significantly from conversational speech in intelligibility. In this study, the intelligibility of loud and soft speech was evaluated relative to conversational speech presented at the same intensity. These two speaking modes were chosen because they occur frequently in natural speech, are relatively easy to elicit from talkers, and have acoustic effects on the speech spectrum that are independent of intensity (Licklider, Hawley, and Walking, 1955). Finally, this study also investigated whether clear speech (spoken slowly) is more intelligible than typical slow speech, without emphasis on clarity. Although other studies have examined acoustic properties of slow speech (Crystal and House, 1982, 1988; Han, 1966), none has considered intelligibility. Such an analysis is straightforward to conduct and lends insight into whether these studies are useful for understanding clear speech.

In this paper, an objective method is introduced for training talkers to produce the clearest possible speech at a variety of speaking rates. Methods are also described for

eliciting soft and loud speech. Intelligibility results are reported for all speaking modes, loud and soft as well as clear and conversational speech at slow, normal, and quick rates. The goal of the intelligibility tests was to determine whether normal-hearing listeners with simulated hearing losses, achieved by additive noise, could derive intelligibility benefits (relative to conversational speech) from one or more of the speaking modes elicited at a given speaking rate. If so, further examination of such highly intelligible speaking modes could lend insight into possible signal-processing approaches for hearing aids with the potential to improve speech clarity as well as audibility.

II. SPEECH ELICITATION METHODS

In order to improve the chances of obtaining clear speech at normal speaking rates (i.e., rates comparable to conversational speech) naturally, much attention was given to talker selection and training. Talkers were recruited from the New England area, but only talkers with significant public speaking experience (e.g., students or professionals in television or radio broadcasting, public speaking, or other communications disciplines) were considered, because their experience increased the likelihood that they could respond to training in relatively short amounts of time. Since the training was fairly intensive, however, it was not feasible to train all potential talkers. Therefore, talkers who responded with at least 2 years of public speaking experience were asked to participate in a preliminary screening to evaluate their intelligibility. A description of the 15 talkers who participated in the screening and their speaking experiences is summarized in Table I. Based on the results of the screening, the five participants with the highest potential for producing clear speech at a variety of speaking rates were then selected for training.

A. Screening of talkers

In order to obtain a form of clear speech from each talker with minimal training, the talkers were familiarized with the characteristics of previously obtained clear speech. The talkers listened to samples of both conversational and clear speech materials recorded for an earlier study of clear speech (Picheny *et al.*, 1985), and differences between the two speaking modes were discussed. The talkers were asked to mimic the clear speech that had been presented, and they were given feedback by the experimenter on both rate and clarity. The goal of obtaining clear speech at normal speaking rates was explained, but each talker was instructed not to increase speaking rate at the expense of clarity. Each talker was then given 1 hour to practice producing clear speech. At the conclusion of the practice period, each talker was asked to record a unique set of 100 sentences, 50 spoken clearly and 50 spoken conversationally.

1. Recording procedure and sentence materials

All recording sessions took place with the talker seated in a sound-treated room. A Sennheiser MD 421 cardioid microphone was positioned approximately 6 in. in front of the talker's mouth. The roll-off filter on the microphone was adjusted to the speech setting, and the microphone output

TABLE I. The following data were used to evaluate talkers during the preliminary intelligibility screening: (1) speaking experience; (2) intelligibility (I), based on % correct key-word scores; and (3) speaking rate (r) in words per minute, excluding pauses. Conversational and clear data are differentiated with subscripts. The change in intelligibility as a function of rate is represented by the slope of a line that would connect the two data points, calculated by $m = -(I_{\text{clear}} - I_{\text{conv}})/(r_{\text{clear}} - r_{\text{conv}})$. Individual results are listed here for the five talkers who were ultimately selected to participate in formal training (T1–T5). Results for the remaining talkers who participated in the screening (T6–T15) are reported in aggregate.

Talker	Sex	Speaking experience	Yrs	I_{conv}	r_{conv}	I_{clear}	r_{clear}	m
T1	F	College television, radio, public speaking	5	42	307	48	169	0.04
T2	F	Professional speaker	5	42	213	70	97	0.24
T3	F	Broadcasting student	2	25	315	64	61	0.15
T4	F	Debate team	6	52	175	75	57	0.20
T5	M	Debate team	7	48	164	79	90	0.42
T6–T15	5 M, 5 F	Varied	2–5	20–45	160–285	34–64	55–198	0.10–0.34

was amplified using a Symetrix SX202 dual microphone pre-amplifier. The speech was recorded at a 48-kHz sampling rate, using a SONY 59ES digital audio tape deck. The files were then downsampled to 20 kHz and normalized for rms level. The sentence materials recorded were obtained from the corpus of nonsense sentences described by Picheny *et al.* (1985). These sentences have five to eight words each and provide no semantic context that could aid listeners in identifying key words, which are defined as all nouns, verbs, and adjectives (e.g., “The right cane could guard an edge.”).

2. Listeners and testing conditions

Two normal-hearing listeners (one male, one female) were obtained from MIT and the surrounding community. They were native speakers of English who possessed at least a high school education and had no prior experience with the task. Their hearing thresholds were no greater than 20 dB HL at frequencies between 250 and 4000 Hz.

Each talker’s speech was presented in a background of speech-shaped noise (Nilsson, Soli, and Sullivan, 1994) to listeners monaurally over TDH-39 headphones. The stimuli were stereo signals with speech on one channel and speech-shaped noise of the same rms level on the other channel. The speech was attenuated by 4 dB and added to the speech-shaped noise, and the resulting signal (SNR = -4 dB) was presented to the listeners from a PC through a DAL card. Listeners were given the opportunity to choose a comfortable listening level and to select which ear would receive the stimuli. They were also encouraged to switch the stimulus to the other ear when fatigued. Listeners responded by writing their answers on paper. They were given as much time as needed to respond, but were presented each sentence only once. Intelligibility scores for key words (nouns, verbs, and adjectives) were determined using the scoring rules described by Picheny *et al.* (1985). Under these testing conditions, normal-hearing listeners have been shown to receive benefits from clear speech that are consistent with hearing-impaired listeners in quiet (Payton *et al.*, 1994).

3. Evaluation and selection of talkers

Speaking rates in words per minute (wpm) and intelligibility scores for each talker, averaged across listener, are summarized in Table I. Since the goal of the screening was to

identify talkers who could not only produce clear speech but also demonstrate a high potential for producing clear speech at normal speaking rates, pauses (silent periods of 10 ms or longer) were excluded from the speaking rate calculation. For clear speech, this procedure provided a rough estimate of the minimum speaking rate that could be achieved by training talkers without sacrificing intelligibility, given that deleting pauses from clear speech does not reduce its intelligibility (Choi, 1987). It also provided an indication of articulation rates, which facilitated analysis of talkers’ abilities to produce clear speech at normal rates.

When speaking clearly, all talkers achieved some improvement in intelligibility over speaking conversationally, but none achieved this improvement without slowing down. Since the screening did not provide talkers with training, this result was not unexpected. However, the relationship between intelligibility and speaking rate was considered an important criterion for screening talkers’ potential to achieve clear speech at normal rates. In order to quantify this relationship, an equation for each talker was derived for the line passing through the two data points (conversational and clear) achieved by that talker during the screening, with intelligibility as a function of speaking rate. The equation was of the form $I = a - mr$, where I represents intelligibility, r represents speaking rate excluding pauses, and a and m are positive constants. The value of m satisfying this equation for each talker is also summarized in Table I.

Since it was unknown what talker characteristics were most likely to be associated with talkers who were capable of producing clear speech at normal rates, talkers with different characteristics were selected in order to improve the chances of finding at least one talker who could produce clear speech at normal speaking rates. The data for the five selected talkers are listed in Table I. T1 was selected for her ability to speak at a higher rate than most other talkers, both in conversational (307 wpm) and clear (169 wpm) modes. In addition, she exhibited an unusually low value of the slope parameter m . It was hoped this value of m could be increased with training. Talkers T2 and T4 were selected because their overall intelligibility in both modes was higher than most of the other talkers at similar speaking rates. Two talkers with this characteristic were selected because it was considered likely to be associated with an ability to speak clearly at

normal rates. T3 was selected because she had the greatest increase in intelligibility between conversational and clear speech. She also demonstrated the ability to change her speaking rate significantly, from 61 wpm in clear mode to 315 wpm in conversational mode. T5 was selected because his clear speech had the overall highest intelligibility at 79 percent as well as the highest value of m . These five talkers (four females, one male) participated in the formal training sessions described below.

B. Training and recording procedure

After participating in intensive training, talkers recorded clear and conversational speech from talkers at slow, normal, and quick speaking rates, resulting in six speaking styles (referred to by mode/rate): clear/slow, conv/slow, clear/normal, conv/normal, clear/quick, conv/quick. The novel training procedure, described below, included quantitative feedback on both speaking rate and intelligibility, ensuring that the clearest possible speech for a given speaking rate was obtained from talkers. Whether this speech could provide an intelligibility advantage over conversational speech of the same rate was evaluated with intelligibility tests described in the next section.

1. Regulating speaking rate

Speaking rates were specific to individual talkers and were regulated at the sentence level in order to make the task as natural for talkers as possible. Moreover, regulating sentence rate rather than word or syllable rate allowed the talker freedom to determine the duration of individual speech segments and words, thus maximizing the chance of obtaining natural clear speech at normal and quick rates. After an initial determination of speaking rates, talkers were required to retain their individual slow, normal, and quick rates throughout all training and recording sessions. The appropriate speaking rate was communicated to the talker by presenting over headphones the output of a metronome that was set to the desired sentence rate. Talkers could take as much time as desired between sentences, but were required to fit each sentence between two consecutive metronome clicks. Most talkers adjusted easily to monitoring their speaking rate in this way and did not appear to find the procedure difficult or unnatural.

Individual speaking rates were initially determined by the average rates achieved by each talker when asked to do the following: (1) slow—produce 100 clear sentences, with no constraints on rate, at the conclusion of training; (2) normal—read 200 sentences at a rate appropriate for normal conversation; (3) quick—read 50 sentences as rapidly as possible. Clear speech was used to establish the slow rate for two reasons. First, imposing no rate constraints on the production of clear speech was similar to methods used for eliciting clear (slow) speech in the screening and in previous studies (Picheny *et al.*, 1985; Uchanski *et al.*, 1996) and ensured that highly intelligible speech would be elicited. Second, this method allowed for elicitation of conv/slow speech at rates comparable to clear (i.e., clear/slow) speech, which allowed for a direct comparison of the intelligibility of clear and conversational modes at the same rate. Although speak-

ing rates were not uniform across talkers, Fig. 1 shows that there was very little overlap in the distributions of slow, normal, and quick rates across talkers. Only T2, who varied her rate relatively little, chose a quick rate (193 wpm) that overlapped with the remaining four talkers' normal rates, which averaged 187 wpm and ranged from 171 wpm (T5) to 198 wpm (T3).

2. Regulating intelligibility

The procedure for regulating intelligibility required the talker to repeat a sentence with increased emphasis on articulation until it was perceived correctly by a listener. This procedure was derived from an existing method for eliciting clear speech with syllables (Chen, 1980) in which a talker repeated a syllable until the listener perceived it correctly in the presence of masking noise. While such a method could be used with nonsense sentences, its disadvantage is that repetition of sentences increases their intelligibility (Uchanski *et al.*, 1996). To avoid the intelligibility benefit of repetition, four normal-hearing listeners were employed successively to provide the talker with feedback on intelligibility. The talker's speech was distorted by multiplicative noise (Schroeder, 1968) and presented to each of the listeners, in turn, monaurally over headphones. Multiplicative noise was used because it maintains a constant SNR, thus preventing the talkers from increasing intelligibility simply by speaking more loudly. At the beginning of each session, the SNR was set to 0 dB, and it was decreased in increments of 0.2 dB until the listeners received on average no more than one key word correctly from the talker's first utterance of the sentence. Throughout the experiment, the SNR was decreased in increments of 0.2 dB as the talker's intelligibility improved.

The talker and listeners were seated separately and did not have visual contact. Each listener could hear the talker only when directly addressed and presented with a sentence. The designated listener then responded verbally with the sentence heard. The talker and the experimenter could both hear the listener's response, and they could communicate freely with each other throughout the session. The experimenter provided instruction, reminding the talker to adhere to the timing cues provided by the metronome (for clear/normal and clear/quick modes) and pointing out patterns of mistakes among the listeners. The experimenter also served as a judge of the listener's responses and decided whether or not the talker should repeat a sentence. The listener's response was regarded as correct if more than half of the key words were correctly identified. If the response was incorrect, the talker repeated the sentence to the next listener; if the response was correct, the talker presented a new sentence to the next listener. The four listeners were not given feedback on whether or not the response was correct. A sentence was not repeated additional times after it had been presented to all four listeners. If three sentences in a row were presented to all four listeners without a correct response, the SNR was increased by 0.2 dB.

This procedure was used for both training and recording. During training sessions, talkers were encouraged to experiment with different speaking strategies and allowed to practice as much as desired. After speaking strategies had been

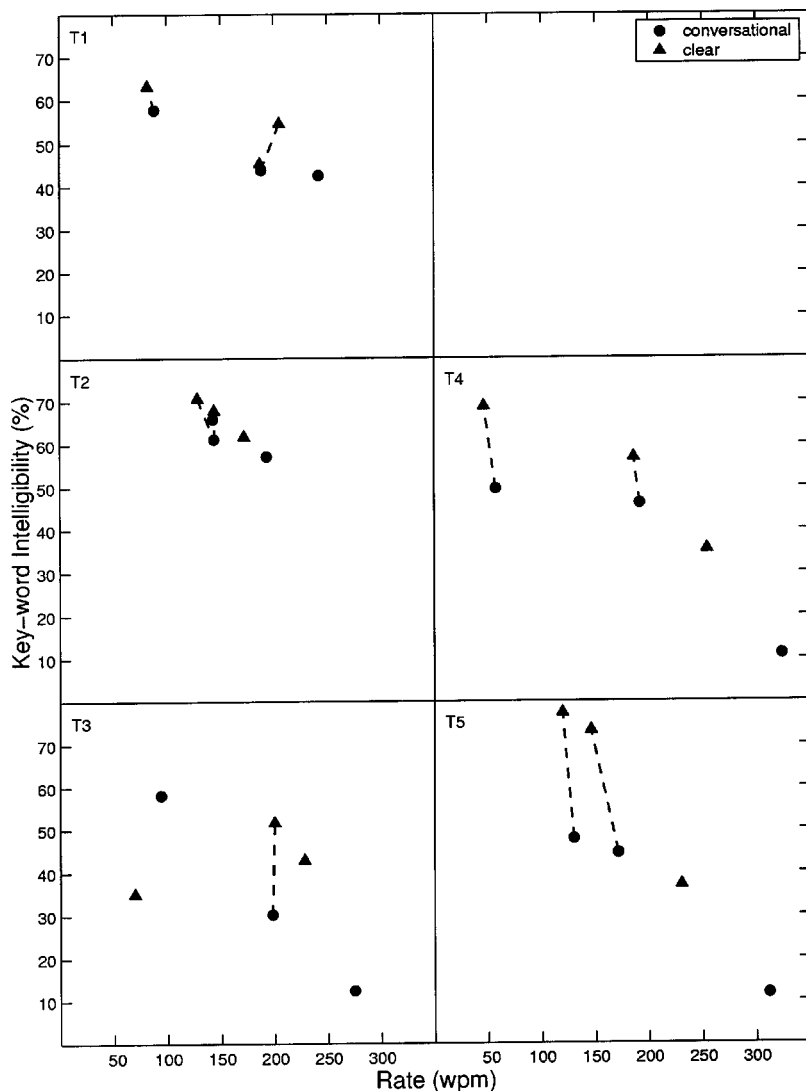


FIG. 1. Average key-word scores for each talker, averaged over listener, versus speaking rate. Dashed lines represent instances where intelligibility was improved without a significant change (no more than 25 wpm) in speaking rate. Two dashed lines emanate from T2's conv/normal data point, because both clear/slow and clear/normal speech provided an intelligibility advantage without a significant change in rate.

explored, training of each condition (clear/slow, clear/normal, clear/quick) was considered complete when the SNR had reached a constant value (within ± 0.2 dB). After a short break, the condition was then recorded.

3. Eliciting other speaking modes

Soft/normal and loud/normal speech were elicited from talkers with the aid of a Realistic digital sound-level meter, located approximately 2 1/2 ft. from the talker's mouth and set to measure the maximum sound pressure level (A-weighted, FAST-acting average: updated every 0.25 s) in a sentence. The talker was instructed to read ten sentences in a conversational manner, and the largest and smallest sound levels were noted. For soft/normal speech, each talker was instructed to speak at sound levels at least 15 dB below the largest level measured for that talker during conversational speech. The level on the meter was reported to the talker after each sentence was read, and the sentence was repeated if necessary. Loud/normal speech was elicited in a similar manner, except the talker was instructed to exceed the smallest level measured during conversational speech by at least 15 dB. It should be noted that although these modes were elicited at various intensities, all sentences were normalized

to the same rms level for the purposes of intelligibility tests. Thus, the question addressed in the intelligibility tests was whether these speaking modes provide an intelligibility advantage over conversational speech of the same intensity.

4. Recording sessions

Each talker recorded 700 nonsense sentences over four recording sessions, 2 to 3 h in length. The 600 sentences discussed here consisted of six unique sentence lists, each containing 50 sentences. Every sentence list was recorded in two speaking modes: L1—soft/normal and conv/normal; L2—loud/normal and conv/normal; L3—clear/normal and conv/normal; L4—clear/slow and conv/normal; L5—clear/quick and conv/quick; L6—clear/slow and conv/slow). One mode was the test condition, and the other was always conversational, to establish baseline intelligibility scores. In most cases, the two speaking modes were recorded at the same speaking rate in order to facilitate intelligibility comparisons without speaking rate as a factor. One list (L4), however, was recorded at two different rates: once in conversational mode at the normal rate and once in clear mode at the slow rate. The recording procedure was identical to that used for the screening, except that (1) the amplifier output

was recorded directly to a PC disk, using a DAL card with a 20-kHz sampling rate, and (2) pauses were excluded for normalization purposes in order to ensure that the rms level reflected the level of the speech only, since pause durations varied greatly across the three speaking rates. Sentences were checked for errors, and in a few cases, mispronounced words were noted so that responses could be graded accordingly during intelligibility tests.

III. INTELLIGIBILITY TESTS

Eight normal-hearing listeners (four males, four females; aged 18 to 29 years) participated in the final intelligibility tests. These listeners met the same requirements for normal-hearing listeners as those who participated in the screening. None had participated in the screening nor had any prior familiarity with the task. The eight listeners were divided into two separate testing groups.

Listeners participated in 16 2-h sessions over the course of approximately 8 weeks. In each session, listeners were tested on 4–5 sentence lists and were given a 5-min break after the presentation of each list and were also encouraged to rest briefly as necessary. In addition, a 10-min break was given near the halfway point of each session. The stimuli were prepared and presented in the same manner as in the screening, except that the SNR was set to -1.8 dB. This SNR was chosen to avoid ceiling and floor effects on all conditions, which varied considerably in difficulty. It was determined through informal listening tests that were administered to normal-hearing volunteers from the laboratory.

Although the groups met at different times, both groups heard the same lists in the same order. It was not necessary to randomize the order of presentation of the lists, because learning effects over the course of an experiment are minimal with these speech materials (Picheny *et al.*, 1985). However, because each list was presented in two different modes in this experiment (conversational, for reference, and a test condition), learning effects within a list were a possibility. In order to minimize these learning effects, the conversational condition was presented prior to the test condition in roughly half of the cases and after the test condition in the remaining half. In addition, the amount of time between the first and second presentations of a each list was always at least 2 weeks.

IV. RESULTS

Intelligibility scores, averaged across talkers and listeners, and speaking rates were calculated as in the screening. To simplify discussion, the data discussed here are also averaged across all presentations of a given condition. Although the number of lists presented in each condition varied, the lists have been shown to be of equal difficulty (Rosengard, 2000). Therefore, this averaging method should provide the best estimates of the intelligibility of various conditions.

The clear/slow speaking style was most intelligible at 63 percent key words correct, followed in order of decreasing intelligibility by clear/normal (59%), loud/normal (53%), conv/slow (51%), clear/quick (46%), conv/normal (45%),

conv/quick (27%), soft/normal (26%) modes. The 18-percentage-point advantage for clear/slow (63%) relative to conv/normal speech (45%) is consistent with previous studies (Picheny *et al.*, 1985; Uchanski *et al.*, 1996). Moreover, a 14-point advantage was obtained for clear/normal (59%) relative to conv/normal speech, thus extending the benefit of clear speech to normal speaking rates. In addition, this advantage was comparable to the 12-point advantage obtained for clear/slow (63%) relative to conv/slow (51%) speech. Although a 19-point advantage was also obtained for clear/quick (46%) relative to conv/quick speech (27%), this finding does not extend the benefit of clear speech to quick rates, because the speaking rate for clear/quick speech (218 wpm) was significantly slower than the speaking rate for conv/quick speech (269 wpm). Of the speaking modes other than clear speech that were evaluated, soft/normal speech was less intelligible on average than conv/normal speech. On average, loud/normal speech (53%) was more intelligible than conv/normal speech (45%), but the advantage was less for loud/normal speech (8 points) than for clear/normal speech (14 points). Moreover, not all talkers achieved an intelligibility advantage with loud/normal speech. Thus, at normal rates, none of the alternative speaking modes tested provided as large or as consistent of an intelligibility advantage over conversational speech as clear speech.

An analysis of variance was performed on the key-word scores for the data specific to clear and conversational speech (conv/slow, conv/normal, conv/quick, clear/slow, clear/normal, clear/quick), after an arcsine transformation ($\arcsin \sqrt{I_j/100}$) to equalize the variances. The analysis of variance used a standard factorial model (nonrepeated measures) and analyzed four factors, assuming one random factor, listener, and three fixed factors: talker, mode, and rate. Although many factors and interactions had statistically significant F -ratios ($p < 0.01$), only five of them accounted for substantial portions of the variance: rate (20%), talker (19%), mode (13%), listener (7%), and rate \times talker (10%). For the purposes of this study, it is most important to note that a substantial portion of the variance is accounted for by the mode factor alone, and that a relatively small portion of the total variance (5% or less) is accounted for by each of the interactions with mode that were significant. In fact, with the exception of mode \times talker and mode \times rate \times talker, each statistically significant interaction with mode accounted for less than 2% of the variance, including mode \times listener and mode \times rate as well as some higher-order interaction terms. Therefore, to a first-order approximation, it can be concluded that the intelligibility benefit obtained by speaking clearly (mode) is statistically significant, and the size of the benefit is largely independent of listener, talker, and speaking rate. In other words, a comparable intelligibility benefit is obtained for a majority of talkers, independent of their overall intelligibility, for a majority of listeners, independent of their overall skill at the listening task, and for a majority of speaking rates (at least slow and normal), independent of the overall decline in intelligibility that comes with increased rate.

Of the interactions with mode that were statistically significant, the two terms accounting for the most variance were mode \times talker (5%) and mode \times rate \times talker (5%). These in-

TABLE II. Slopes $m1$ and $m2$ for each talker in this study, where $m1 = (I_{\text{normal}} - I_{\text{slow}}) / (r_{\text{normal}} - r_{\text{slow}})$ and $m2 = (I_{\text{quick}} - I_{\text{normal}}) / (r_{\text{quick}} - r_{\text{normal}})$. I represents intelligibility (%), r represents speaking rate (wpm), excluding pauses, and subscripts indicate nominal speaking rate. Asterisks indicate slopes in this study that could not be calculated because of data points that were eliminated due to the failure of the talker to achieve the desired rate or intelligibility. Comparison of average slopes (calculated from intelligibility and rate values averaged across talker) in this study with those obtained for previous attempts to elicit clear speech at normal rates shows that the present study's speech elicitation methods (talker selection and training) extended the advantage of clear speech to normal rates by increasing the value of $m1_{\text{clear}}$ to that of $m1_{\text{conv}}$.

Talker/Study	Listener(s)	Scaling	$m1_{\text{conv}}$	$m2_{\text{conv}}$	$m1_{\text{clear}}$	$m2_{\text{clear}}$
T1	Normal	Natural+training	-0.14	-0.02	-0.07	*
T2	Normal	Natural+training	*	-0.08	-0.18	-0.21
T3	Normal	Natural+training	-0.27	-0.23	*	-0.31
T4	Normal	Natural+training	-0.03	-0.27	-0.09	-0.32
T5	Normal	Natural+training	-0.08	-0.23	-0.15	-0.43
AVG (present study)	Normal	Natural+training	-0.10	-0.20	-0.10	-0.33
Picheny <i>et al.</i> (1989)	Impaired	Uniform	-0.30	...
Uchanski <i>et al.</i> (1996)	Impaired	Nonuniform	-0.25	...
Uchanski <i>et al.</i> (1996)	Normal	Natural	-0.25	-0.29
Uchanski <i>et al.</i> (1996)	Impaired	Natural	-0.21	-0.28

teractions are shown in Fig. 1. For both T4 and T5, conv/slow speech was comparable in intelligibility to conv/normal speech, and the clear speaking mode at these speaking rates was much more intelligible than the conversational mode (average key-word scores were 18 points higher for T4 and 28 points higher for T5). Trends for the other three talkers are less clear. T3 failed to produce highly intelligible speech at the slow speaking rate, although her clear/normal and clear/quick styles were more intelligible than her conversational speech at similar speaking rates. T2 varied her speaking rate the least of all the talkers, and reported having difficulty adhering to the metronome. Her intelligibility drops off quickly at speaking rates above 150 wpm. T1 reported that she preferred speaking quickly, which may partly explain her achieving a higher key-word score for clear/quick speech than for clear/normal speech.

Instances where talkers achieved an intelligibility benefit (at least 5 percentage points) without a substantial change in speaking rate (no more than 25 wpm) are indicated in Fig. 1 with dashed lines. At normal rates, all talkers met this criterion, producing a form of clear speech that was more intelligible than conv/normal speech produced at nearly the same speaking rate (T1 met this criterion with clear/quick speech, which was produced within 25 wpm of her conv/normal speech). In addition, clear/slow speech was also more intelligible than conv/slow speech for all talkers except T3, according to this criterion. Although all talkers obtained an increase in intelligibility at quick rates by speaking clearly, none of them did so without also speaking more slowly. Additional training may have been necessary, or it may not be possible to produce clear speech at such high rates. Regardless, it should be noted that the clear/quick speech obtained for some of the talkers (particularly, T1 and T3) seems likely to be more intelligible than conversational speech at comparable rates, if such conversational speech had been obtained in this study (based on an interpolation of the conversational speech scores that were obtained for each talker). If so, it appears that the clear speech benefit can also extend to faster than normal speaking rates.

The clear speech benefit achieved by talkers in this study

was also analyzed by comparing two slope values across speaking modes: $m1$, the decrease in intelligibility from slow to normal rates, and $m2$, the decrease in intelligibility from normal to quick rates. These slopes, shown in Table II, were calculated for each talker and speaking mode from the data pictured in Fig. 1, with the exception of three data points, which were omitted because they failed to exhibit a change in rate or an improvement in intelligibility: T1's clear/normal data (no intelligibility improvement over conv/normal), T2's conv/slow data (not significantly slower than conv/normal), and T3's clear/slow data (no intelligibility improvement over conv/slow). In the first case, T1's clear/quick data point was substituted for the clear/normal omission, because it provided an intelligibility improvement over conv/normal speech and was within 25 wpm of her normal rate. In addition, average slope values were calculated from mean intelligibility and rate values averaged across talker, for this study as well as for previous attempts to elicit clear speech at normal rates. As shown in Table II, the average slope $m1_{\text{clear}}$ in this study was greater than $m1_{\text{clear}}$ for all previous studies and equal to $m1_{\text{conv}}$ (-0.10), indicating that the clear slope has been flattened (relative to previous studies) such that the size of the intelligibility benefit of clear modes over conversational modes remains roughly equal for slow and normal speaking rates. At faster rates, however, the average slope $m2_{\text{clear}}$ in this study became considerably steeper, dropping off more quickly than $m2_{\text{conv}}$ as well as all $m2_{\text{clear}}$ values obtained from previous studies. The point where these $m2$ lines converge may represent a physical limit on articulation resulting from physiological constraints at very high speaking rates.

V. DISCUSSION

The results of this study show that (1) at a given speaking rate, neither of the additional speaking modes examined provided an intelligibility benefit as large as that of clear speech; and (2) with proper training of talkers, the benefits of clear speech can be extended to faster speaking rates than those previously reported. Specifically, a form of clear

speech was obtained at slow (roughly 100 wpm) and normal (roughly 200 wpm) rates. Because the intelligibility advantage of clear/slow over conv/slow speech was comparable to that of clear/normal over conv/normal for nearly all talkers and listeners, it was also shown that over this range of speaking rates (slow through normal), the relative intelligibility benefit of clear speech is largely independent of rate, talker, and listener.

While the intelligibility advantage of clear speech did not extend to quick rates, as shown by the fact that $m2_{\text{clear}} < m2_{\text{conv}}$ for all talkers, it could still exist at faster than normal rates. Assuming physiological constraints on articulation, the intelligibility of clear speech at very high speaking rates must decrease more rapidly than the intelligibility of conversational speech in order to compensate for its higher intelligibility. Consequently, clear speech cannot maintain an intelligibility advantage above a certain “cutoff” speaking rate. Yet, with the limited number of rates that were examined in this study, only a lower bound (normal rates of roughly 200 wpm) on the cutoff rate was established. At least some of the talkers (T1 and T3) appear likely to have exceeded this lower bound, producing clear/quick speech that could be more intelligible than conversational speech at comparable rates, if such conversational speech had been elicited. Moreover, since the training provided in this study increased the cutoff rate beyond those previously reported, it is possible that additional training could increase the cutoff rate even further.

Subjective comments from talkers regarding the training procedure indicated that the listener feedback provided during training was very helpful for developing clear speech. In particular, one talker noted that trends in listener responses raised his awareness of common phoneme confusions. He reported that this information was useful in deciding which phonemes to emphasize. Other talkers expressed interest in listening to speech distorted by multiplicative noise in order to gain information on how to speak more clearly. This request suggests that some talkers believe they have natural strategies for speaking clearly in difficult communication situations. Moreover, these strategies may differ depending on the nature of the distortion.

However, even if different types of clear speech exist, depending on the nature of the distortion, the clear speech obtained in this experiment warrants further study because it provided an intelligibility benefit to listeners with simulated hearing loss, without an accompanying reduction in speaking rate. Acoustical analyses of the differences between clear/slow and conv/slow speech and between clear/normal and conv/normal speech should help identify which characteristics of clear speech that have been reported previously (Picheny *et al.*, 1986; Cutler and Butterfield, 1990, 1991) contribute to its high intelligibility without altering rate. In addition, analyzing the intelligibility of clear speech at several rates between 200 and 300 wpm, elicited with varying amounts of training, should help determine the maximum cutoff rate for achieving a sizable clear speech benefit. Such

findings could ultimately lead to the development of signal-processing approaches for hearing aids that convert conversational speech to a close approximation of clear speech over as wide a range of speaking rates as possible.

ACKNOWLEDGMENTS

The authors are grateful to the many participants in this study, particularly the talkers for their dedication and enthusiasm during training. In addition, we thank Rosalie M. Uchanski for many helpful technical discussions. Financial support for this work was provided by a grant from the National Institute on Deafness and Other Communication Disorders (NIH Grant Number 5 R01 DC 00117) and a National Defense Science and Engineering Graduate fellowship from the Office of Naval Research.

- Bradlow, A., Toretta, G., and Pisoni, D. (1996). “Intelligibility of normal speech I. Global and fine-grained acoustic-phonetic talker characteristics,” *Speech Commun.* **20**, 225–272.
- Chen, F. R. (1980). “Acoustic characteristics and intelligibility of clear and conversational speech,” Master’s project, Mass. Inst. Tech., Cambridge, MA.
- Choi, S. (1987). “The effect of pauses on the intelligibility of sentences,” Bachelor’s project, Mass. Inst. Tech., Cambridge, MA.
- Crystal, T. H., and House, A. S. (1982). “Segmental durations in connected-speech signals: Preliminary results,” *J. Acoust. Soc. Am.* **72**, 705–716.
- Crystal, T. H., and House, A. S. (1988). “Segmental durations in connected-speech signals: Current results,” *J. Acoust. Soc. Am.* **83**, 1553–1573.
- Cutler, A., and Butterfield, S. (1990). “Durational cues to word boundaries in clear speech,” *Speech Commun.* **9**, 485–495.
- Cutler, A., and Butterfield, S. (1991). “Word boundary cues in clear speech: A supplementary report,” *Speech Commun.* **10**, 335–353.
- Han, M. S. (1966). “Acoustic-phonetic study on speech tempo,” *Stud. Sounds* **12**, 70–83.
- Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U., and Lacerda, F. (1997). “Cross-language analysis of phonetic units in language addressed to infants,” *Science* **277**, 684–686.
- Licklider, J., Hawley, M., and Walking, R. (1955). “Influences of variations in speech intensity and other factors upon the speech spectrum,” *J. Acoust. Soc. Am.* **27**, 207.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). “Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise,” *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Payton, K. L., Uchanski, R. M., and Braid, L. D. (1994). “Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing,” *J. Acoust. Soc. Am.* **95**, 1581–1592.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1985). “Speaking clearly for the hard of hearing. I. Intelligibility differences between clear and conversational speech,” *J. Speech Hear. Res.* **28**, 96–103.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1986). “Speaking clearly for the hard of hearing. II. Acoustic characteristics of clear and conversational speech,” *J. Speech Hear. Res.* **29**, 434–446.
- Picheny, M. A., Durlach, N. I., and Braid, L. D. (1989). “Speaking clearly for the hard of hearing. III. An attempt to determine the contribution of speaking rate to differences in intelligibility between clear and conversational speech,” *J. Speech Hear. Res.* **32**, 600–603.
- Rosengard, P. (2000). Personal correspondence *re*: Poster. International Hearing Aid Conference, Lake Tahoe.
- Schroeder, M. R. (1968). “Reference signal for signal quality studies,” *J. Acoust. Soc. Am.* **44**, 1735–1736.
- Uchanski, R. M., Choi, S., Braid, L. D., Reed, C. M., and Durlach, N. I. (1996). “Speaking clearly for the hard of hearing. IV. Further studies of the role of speaking rate,” *J. Speech Hear. Res.* **39**, 494–509.
- Zwicky, A. M. (1972). “On causal speech,” *Papers from the Eighth Regional Meeting* (Chicago Linguistic Society, Chicago), pp. 607–615.