Speech intelligibility and localization in a multi-source environment[a]

Monica L. Hawley,[b] Ruth Y. Litovsky, and H. Steven Colburn
Hearing Research Center and Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

Natural environments typically contain sound sources other than the source of interest that may interfere with the ability of listeners to extract information about the primary source. Studies of speech intelligibility and localization by normal-hearing listeners in the presence of competing speech are reported on in this work. One, two or three competing sentences [IEEE Trans. Audio Electroacoust. **17**(3), 225–246 (1969)] were presented from various locations in the horizontal plane in several spatial configurations relative to a target sentence. Target and competing sentences were spoken by the same male talker and at the same level. All experiments were conducted both in an actual sound field and in a virtual sound field. In the virtual sound field, both binaural and monaural conditions were tested. In the speech intelligibility experiment, there were significant improvements in performance when the target and competing sentences were spatially separated. Performance was similar in the actual sound-field and virtual sound-field binaural listening conditions for speech intelligibility. Although most of these improvements are evident monaurally when using the better ear, binaural listening was necessary for large improvements in some situations. In the localization experiment, target source identification was measured in a seven-alternative absolute identification paradigm with the same competing sentence configurations as for the speech study. Performance in the localization experiment was significantly better in the actual sound-field than in the virtual sound-field binaural listening conditions. Under binaural conditions, localization performance was very good, even in the presence of three competing sentences. Under monaural conditions, performance was much worse. For the localization experiment, there was no significant effect of the number or configuration of the competing sentences tested. For these experiments, the performance in the speech intelligibility experiment was not limited by localization ability. © *1999 Acoustical Society of America.* [S0001-4966(99)00606-2]

PACS numbers: 43.66.Pn, 43.71.Gv, 43.66.Qp [DWG]

INTRODUCTION

In everyday environments listeners are faced with complex arrays of signals arriving from multiple locations. The auditory system has a remarkable ability to separate out individual sources and to extract information from those sources. Historically, this ability has been referred to as the ''cocktail party effect'' (Cherry, 1953; Pollack and Pickett, 1958). More recently it has been described as a problem of ''sound source determination'' (cf. Yost, 1992, 1997) or ''sound source segregation'' (Bregman, 1990). Considerations of these phenomena lead to the fundamental question: how does the auditory system perform these functions in real-world daily environments, such as a crowded room, a subway station or a social gathering? This question not only is important for understanding how the normal auditory system functions, but has significant implications towards understanding how impaired auditory systems process complex signals in everyday environments.

Early studies on this topic (e.g., Cherry, 1953; Pollack and Pickett, 1958) employed a two-channel competition paradigm, whereby multiple speech signals were presented over headphones simultaneously to the two ears, such that each ear received completely different stimuli. These studies showed that under certain conditions listeners are able to ignore information presented to one ear and focus on the information presented to the other ear. In addition, intelligibility of speech was markedly worse if both signals were presented to a single ear. Based on these findings and other related work it was proposed that the ability to understand speech in noisy environments improves when the target speech and the competing sounds are spatially separated (e.g., Hirsh, 1950; Dirks and Wilson, 1969).

In order to test the hypothesis that spatial separation enhances speech intelligibility, one must manipulate the relative locations of the target and competing sources. Such manipulations have been conducted in several recent studies which employed virtual-acoustic simulation of environments under headphones (Bronkhorst and Plomp, 1992; Nilsson and Soli, 1994; Koehnke and Besing, 1996; Yost *et al.*, 1996; Peissig and Kollmeier, 1997), as well as actual stimuli in a sound-deadened room (Bronkhorst and Plomp, 1990; Yost *et al.*, 1996). Most studies measured speech intelligibility in the presence of masking noise with the target signal in front and with the direction of the interfering noise as a variable (e.g., Plomp and Mimpen, 1981; Peissig and Kollmeier, 1997). Not surprisingly, a consistent finding across these

studies is that intelligibility improves when the target and competing sources are spatially separated. However, little is known about the effect of spatial separation for targets located on the sides and with multiple competing sounds that are placed either on the same side or on opposite sides.

The nature of the interfering source(s) is another important consideration. In most studies the competing sources are not other speech tokens, but flat-spectrum noise (e.g., Dirks and Wilson, 1969), speech-spectrum-shaped noise (e.g., MacKeith and Coles, 1971; Plomp and Mimpen, 1979; Koehnke and Besing, 1996), multi-token babble (called speech-simulating noise) (Peissig and Kollmeier, 1997), or noise that is modulated with the envelope fluctuations of speech (Festen and Plomp, 1990; Bronkhorst and Plomp, 1992). A few studies have used speech signals as both targets and competing sounds that were either sentences (e.g., Jerger *et al.*, 1961; Plomp, 1976; Duquesnoy, 1983; Festen and Plomp, 1990; Peissig and Kollmeier, 1997) or single words (Yost *et al.*, 1996); however, typically there was only one competing sound. To the authors' knowledge, very few studies have been conducted with multiple competing sentences (Abouchacra *et al.*, 1997; Ericson and McKinley, 1997; Peissig and Kollmeier, 1997).

Previous studies have shown that speech intelligibility is markedly reduced as the number of competing sources is increased (Bronkhorst and Plomp, 1992; Yost *et al.*, 1996; Abouchacra *et al.*, 1997; Peissig and Kollmeier, 1997). In the present study, both the target and up to three competing sources were presented from several locations in the frontal hemifield in various configurations in an attempt to separate the effects of increasing the number of competing sources from the effect of spatial locations. Each sentence was played from a speaker at a common level so the changes in intelligibility or localizability of the target sound are consequences of the spatial locations of the target and competing sources. Presentation of target and competing sentences spoken by the same talker minimized nonspatial cues such as pitch differences between talkers (cf. Yost, 1997).

The ability of a person to tell where a sound is located in space may also be affected by other sounds in the environment. While many studies have been performed to measure a normal-hearing listener's ability to localize either natural sources (see Blauert, 1997 for review) or simulated sources (e.g., Wightman and Kistler, 1989b; Besing and Koehnke, 1995), relatively few have explored the effect of competing sources (Good, 1994; Wightman and Kistler, 1997b) or reverberation (Giguere and Abel, 1993; Besing and Koehnke, 1995) on localization accuracy. Most localization studies have used noise or click train stimuli; however, in our study the stimuli are sentences. Gilkey and Anderson (1995) showed that anechoic localization performance without competing sources was similar for speech sounds and click trains in the azimuthal dimension.

In the present study we measured both speech intelligibility and localization of speech signals in the presence of up to three competing sentences. In both experiments the location of the competing sources included locations that were close, intermediate and far relative to the target's location. Finally, the experiments were conducted under two conditions: (1) a sound-field condition in which sounds were presented from loudspeakers in a sound-deadened room without head movements and (2) a virtual condition in which sounds were prerecorded through the ears of KEMAR (Knowles Electronic Mannequin for Acoustical Research) from the same loudspeakers in the sound field and presented over headphones in a sound-booth.

I. METHODS

Each experiment (speech intelligibility and localization) was repeated with different subjects using the two methods of presentation (sound-field and virtual). The stimuli and recording techniques are common to both experiments.

A. Sound-field room

The sound-field room used for both virtual recording and sound-field testing is a large laboratory room (approximately 30 by 20 ft). Part of the room is separated by office partitions into a testing space approximately 9 by 13 ft. Egg-crate foam sheets line the walls and close the space above the partitions to the ceiling. The testing space was measured to have a reverberation time T_{60} (time required for level to decrease by 60 dB) of roughly 200 ms for wideband click stimuli. (The value of T_{60} was between 150 and 300 ms for one-third octave narrow-band noises centered between 250 and 4000 Hz.) The ambient noise level in the room was approximately 50 dBA. Seven loudspeakers were positioned along a 180° arc in the frontal hemifield at 30° increments at a distance of 5 ft from the listener, who was seated at a desk-chair. Head movement was minimized with a modified head/neck rest (Soft-2 Head Support, Whitmeyer Biomechanix). The computer was placed in a nearby IAC sound-booth to reduce fan noise. The experimenter's monitor and keyboard were on a desk in the same room, but outside the testing space.

B. Listeners

A total of 12 paid listeners (9 females and 3 males) 18–21 years old participated; all were native speakers of English with hearing thresholds at or below 15 dB HL between 250 and 8000 Hz. Three listeners were tested binaurally in the sound field (denoted S1, S2 and S3). Nine listeners were tested using the virtual stimuli: three listeners were tested binaurally (denoted as V1, V2 and V3) and six were tested monaurally, each listening with either their left or right ear (denoted as VL1, VL2, VL3, VR1, VR2, VR3). Each listener participated in both the intelligibility and the localization experiments under the same listening conditions. Listeners had no prior experience in any psychoacoustic experiments.

C. Stimuli

The speech tokens were sentences from the Harvard IEEE corpus (IEEE, 1969), which consists of 72 phonetically balanced lists of 10 sentences each. A subset of the sentences was recorded by each of two male talkers (45 lists by talker DA and 27 lists by talker CW). The spectrum of each talker
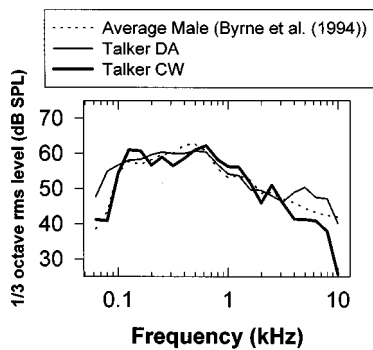
FIG. 1. Spectra for two male talkers in this study compared with average spectrum, normalized to 70 dB SPL overall level for each talker, for American English male talkers from Byrne *et al.* (1994).

is comparable to the average male talker from Byrne *et al.* (1994) as shown in Fig. 1. Each sentence was scaled to the same root-mean-square value, which corresponded to approximately 62 dBA when played from the front loudspeaker, and recorded by a single microphone (Bruel & Kjaer Type 4192) at approximately the position of the listener's head.

In the sound-field experiments, stimuli were played through loudspeakers (Radio Shack Optimus 7) controlled by a personal computer (486DX, Gateway 2000) using Tucker Davis Technologies (TDT) hardware with a dedicated channel for each loudspeaker. Each channel includes a digital-to-analog (D/A) converter (TDT DD3-8), a filter with cutoff frequency of 20 kHz (TDT FT5), an attenuator (TDT PA4) and a Tascam power amplifier (PA-20 MKII). Each speaker was equalized from 200 to 15 000 Hz to have a flat spectrum (± 1 dB) over this range by prefiltering each loudspeaker according to its impulse response measured in an anechoic chamber (cf. Kulkarni, 1997). For the sound-field experiments, when a target and competing sentence were presented from the same loudspeaker, the stimuli were digitally mixed prior to the D/A conversion.

For the virtual-source experiments, each of the 720 sentences was recorded when played from each of the seven positions in the room using the same loudspeaker. These recordings were made binaurally through a KEMAR mannequin (Etymotic ER-11 microphone system, with ear canal resonance removed) onto Digital Audio Tape (DAT). The appropriate combinations of sentences recorded from various positions were subsequently digitally mixed (using a Silicon Graphics Inc. Indigo) and recorded back onto a DAT tape. During testing, stimuli were played back (Sony DTC-8 or Sony DTC-700 DAT players) over headphones (Sennheiser HD 520 II) in an IAC sound-treated booth with no compensation for the headphone transfer functions.[1] Subjects were tested individually. For monaural testing, only one headphone was stimulated.

II. SPEECH INTELLIGIBILITY EXPERIMENT

In the speech intelligibility experiment, the subject's task was to identify the content of an unknown sentence presented from a known location that was fixed throughout a block of trials. Prior to each block, the listener was familiar-

ized with the location of the target sentence, the number of competitors to expect, and the content of each competitor, which were all printed on the answer sheet. However, the location of each competitor was not known and varied from trial to trial. Prior to each trial, a reference stimulus (500-ms, 250-Hz tone) was presented from the target loudspeaker. After each stimulus presentation, the listener was given approximately 16 s to write down the target sentence. *Post hoc* scoring consisted of tallying the number of key words out of five per sentence that had been incorrectly reported. As an example, the key words in the following sentence are italicized: The *streets* are *narrow* and *full* of *sharp turns*. Scoring was strict so that all errors of any kind were counted. An analysis of the results from the binaural virtual condition showed no significant difference between scoring strictly and allowing for minor errors including addition or deletion of suffixes (e.g., played/plays/play) and spelling mistakes (e.g., cloth/clothe) (Dunton and Jones, 1996). Listeners were allowed only one presentation of each target sentence, rather than being permitted an unlimited number of presentations as in the Yost *et al.* (1996) study.

Each sentence in the corpus was designated as either a ''target'' sentence or ''competing'' (nontarget) sentence based on its duration; all target sentences were shorter than every competing sentence. While sentences spoken by both talkers were used, in any given block the same talker spoke both the target and the competing sentences. Within a trial, all sentences had nearly synchronous onsets, but asynchronous offsets. The level of each competing sentence was equal to that for each target sentence; hence, the effective level of the overall competing sound increased as the number of competing sentences was increased.

A. Design

The two parameters of primary interest in this study were the number of competing sentences (one, two or three) and the relative locations of the target and competing sentences. The target sentence was played from one of three locations: left ($-90°$), front ($0°$) or right ($+90°$). Competing sentences were played simultaneously from various combinations of positions categorized as either *close*, *intermediate* or *far* for each target location as listed in Table I. Notice that multiple competitors always originate from separate locations. In the *close* configurations, one competing sentence was at the same position as the target. In the *intermediate* configurations, the nearest competing sentence was 30° to 90° from the target. In the *far* configurations, the nearest competing sentence was more than 90° from the target. The corresponding overall signal-to-noise ratios at the better monaural ear for the various competing sentence configurations are given in Table II.

A third parameter of interest was the effect of sound-field or virtual listening mode. A between-subjects design for listening mode was used in order to accommodate the large design of the experiment without repeating speech material. Therefore, the entire design was repeated for each subject in only one listening mode.

Each target-competing sentence configuration was re-

TABLE I. Categories for competing sentence configurations (− for left, + for right of midline) used in speech intelligibility experiment. Locations presented simultaneously are enclosed by parentheses. Δ_{min} is the minimum absolute separation between target location and location of any competing sentence.

Target location	No. of competing sources	Competing sentence configuration		
		Close $\Delta_{min}=0°$	Intermediate $30°\leqslant\Delta_{min}\leqslant90°$	Far $\Delta_{min}>90°$
−90°		−90°	−60°, −30°, 0°	+30°, +60°, +90°
0°	1	0°	−90°, −60°, −30°, +30°, +60°, +90°	
+90°		+90°	+60°, +30°, 0°	−90°, −60°, −30°
−90°		(−90° −60°), (−90° +90°)	(0° +30°)	(+60° +90°)
0°	2	(0° +30°)	(−90° −60°), (−90° +90°)	
+90°		(+60° +90°), (−90° +90°)	(0° +30°)	(−90° −60°)
−90°		(−90° −60° −30°), (−90° 0° +90°)	(−30° 0° +30°)	(+30° +60° +90°)
0°	3	(−30° 0° +30°), (−90° 0° +90°)	(−90° −60° −30°)	
+90°		(+30° +60° +90°), (−90° 0° +90°)	(−30° 0° +30°)	(−90° −60° −30°)

peated between 10 and 25 times (with different target and competing sentences), hence performance is based on scoring between 50 and 125 key words per condition. Different numbers of trials were used for each condition so that there was a similar number of blocks for each target location. Thirty-seven blocks were tested, each containing between 14 and 21 trials. Within a block, the location of the target and the number (and content) of the competing sentences remained constant while the location of the competing sentence(s) was randomized from trial to trial so that each competing sentence configuration was repeated at least twice per block. The order of the blocks was randomized and all listeners performed the experiment in the same order.[2] Due to an error in the preparation of the tape for virtual listening, two conditions (target at −90° or +90° and the competing sentences at −30°, 0° and 30°) were not randomized along with the other conditions. Therefore, for the virtual listening subjects only, these conditions were tested separately after the speech intelligibility and localization experiments were completed.

Listeners received the same, minimal training at the beginning of each testing session. First, a known sentence was presented twice from each of the seven possible locations from −90° to +90°, in order left to right. Then examples of listening to a known sentence in the presence of one, two and then three known competing sentences were played from se-

lected locations. Finally, listeners were given five practice trials for each of one, two and three competing sounds (from only a subset of the locations actually tested), which were not scored. The speech intelligibility experiment contained a total of 600 trials requiring approximately 6.5 h of testing per listener over multiple sessions.

B. Results

Performance was scored separately for each listener as the error rate, the percentage of key words that were incorrectly identified in each condition. The average and standard deviation across listeners for the two binaural testing groups (sound-field or virtual listening) are shown in Fig. 2. Results for the one, two and three competing sentence cases are plotted in separate panels. The virtual listening subjects (Vs) are denoted by open symbols and the sound-field listening subjects (Ss) are denoted by filled symbols. The symbol shape corresponds to the location of the target, which was either at −90° (circle), 0° (square) or +90° (triangle). The different symbols are offset slightly for clarity.

Figure 2 shows several important findings. First, for each fixed number of competing sentences, configurations with the largest errors are those in which the competing sentences are close to the target location (i.e., one competing source has the same location as the target). Second, the maxi-

TABLE II. Signal-to-noise ratio at the better ear for all configurations tested. Values were computed using the Shaw and Vaillancourt (1985) source-to-eardrum transfer functions. When there were multiple conditions contributing to the average, the average for all conditions in a category for a given number of competitors is given first and the value for each individual condition is given in parentheses in the same order as reported in Table I. Only the conditions on the left side are shown, the symmetric cases on the right side are equivalent.

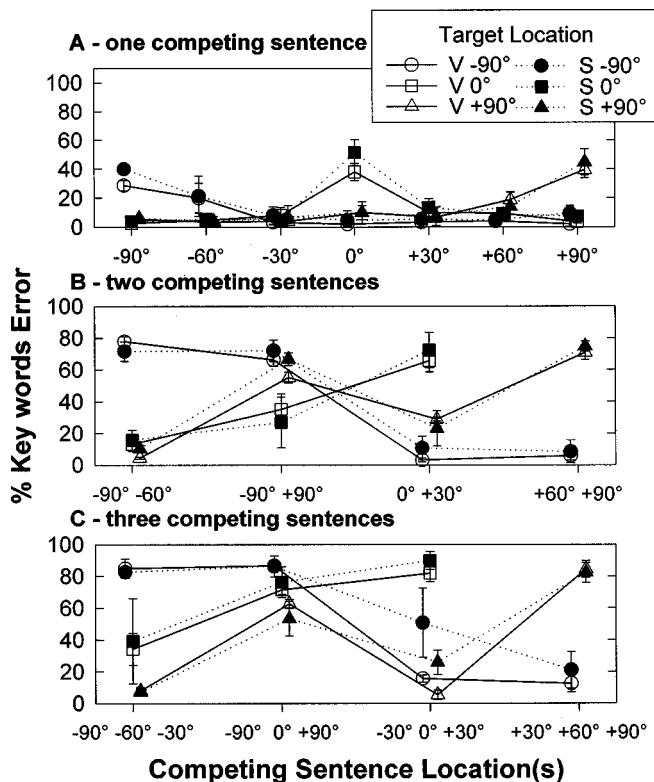| Target location	No. of competing sources	Competing sentence configuration		
		Close $\Delta_{min}=0°$	Intermediate $30°\leqslant\Delta_{min}\leqslant90°$	Far $\Delta_{min}>90°$
−90°		−2.5	−1.0 (−3.3, −3.0, +3.3)	+10.9 (+7.1, +12.6, +13.0)
0°	1	−0.7	+7.0 (+9.4, +8.6, +3.0)	
−90°		−5.7 (−7.6, −3.8)	−1.1	+8.1
0°	2	−5.1	+0.35 (+4.3, −3.6)	
−90°		−5.4 (−7.7, −3.2)	−4.6	+5.1
0°	3	−6.7 (−7.5, −5.8)	+1.8	

FIG. 2. Speech intelligibility error rates for binaural listening with one- (panel A), two- (panel B) and three- (panel C) competing sentences. Average values and standard deviation over three listeners in each group (virtual-listening and sound-field) are plotted for each competing sentence configuration. Virtual-listening results are shown using open symbols and sound-field listening results are shown with closed symbols. Target location is indicated by symbol: circles for −90°, squares for 0°, and triangles for +90°.

mum error rates observed generally increase as the number of competing sentences in the environment is increased (although in some cases, ceiling and floor effects lessen the difference). Third, similar error rates are observed for the sound-field and virtual listening conditions. These observations are consistent with analysis of variance (ANOVA) results. Specifically, a two-way mixed design ANOVA [binaural testing condition (sound-field or virtual listening) ×competing sentence location(s)] was performed for each target position (−90°, 0° or +90°) and number of competing sentences (one, two or three) separately (a total of nine ANOVAs). The binaural testing condition was a between-subjects factor and number of competing sentences was a within-subjects factor. For all analyses, significance was considered at a level of 0.05, using Scheffe's correction for multiple comparisons when necessary. For all nine ANOVAs, the results yielded a significant effect of competing sentence location. The configurations when the competing sentences were *close* (one competing sentence at the same location as the target) with respect to the target position yielded significantly more errors than the configurations which were *intermediate* or *far*. There was no significant effect of testing condition (sound-field versus virtual listening); however, an interaction was observed between the testing condition and the competing sentence location for the target at −90° and +90° and the three-competing sentences condition, with the

difference only for the competitor locations of (−30°, 0° and +30°). The error rates were lower when the condition was tested separately (virtual listening) than when tested randomized with other conditions (sound-field listening). Further studies are needed to determine if this difference is due to the particular competitor location or due to the (lack of) randomization.

The ANOVA analyses were repeated using an arcsin transformed version of the error rate which takes into account the limitation of the measurement range (Walker and Lev, 1953; used by Nabelek and Robinson, 1982). The only difference in the results was that there was a significant interaction between the testing condition and three-competing sentence location only for the target at +90° and no longer for the target at −90°.

A further validation of the equivalence of sound-field and virtual presentations (indicated by the ANOVA results) is the high correlation between the average error rate for the two binaural testing conditions ($r = 0.96$, $p < 0.01$); the regression line is not statistically different from the unity line [correlation coefficient: $t(41) = 0.82$, $p > 0.05$; intercept: $t(41) = 1.40$, $p > 0.05$]. Thus, speech intelligibility results are essentially equivalent when tested in the sound field or recorded in the sound field and played back over headphones.

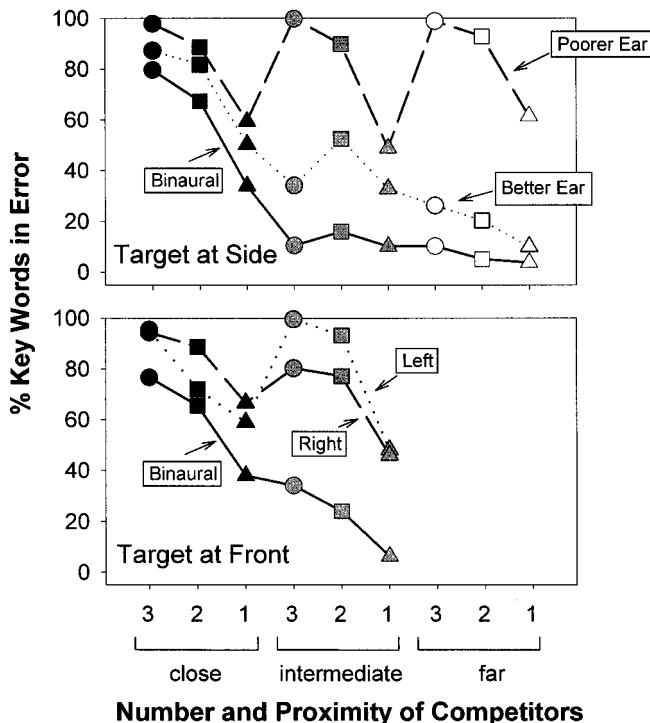Figure 3 shows performance for virtual conditions sepa-

FIG. 3. Speech intelligibility error rates for virtual-listening with the target at side (−90° and +90° averaged) and at front (0°). Average values are plotted for each group (binaural and each monaural ear) and each competing sentence configuration. Values for binaural virtual-listening are connected with solid lines, those for monaural left virtual-listening (or better monaural) are connected with dotted lines, and those for monaural right virtual-listening (or poorer monaural) are connected with dashed lines. Symbol shading denotes the proximity of the competitors to the target: *close* (black), *intermediate* (gray) or *far* (white) and the number of competing sentences is denoted by the symbol shape *three* (circle), *two* (square) or *one* (triangle).

rated into panels by target position: side (average of symmetrical conditions for target at −90° and +90°, top panel), front (0°, bottom panel). Each panel shows performance under binaural (connected by solid lines) and monaural (connected by dotted and dashed lines) conditions. The proximity of the competing sentences to the target location is denoted by the symbol shading with black for *close* (one competing sound in the same location as the target), gray for *intermediate* (closest competing sound 30° to 90° from the target location) and white for *far* (closest competing sound more than 90° from the target location) competing configurations. The symbol shape denotes the number of competing sentences: circle for three, square for two and triangle for one. The symmetrical conditions, in which the relative location of the target and competing sentences were mirror-images of each other, were averaged since the symmetrical conditions differed by an average of only 2.4% key word errors.

As already noted in Fig. 2, the binaural listening condition results (symbols connected by solid lines in Fig. 3) show that the largest error rates for each target location are consistently seen for the *close* conditions, regardless of the number of competing sentences. The *intermediate* and *far* conditions show similar error rates. It is clear that the proximity of competing sentences is more influential on the error rate than is the number of competing sentences. Our results in the single competing source case are consistent with previous studies that have found a large decrease in threshold when the competing sound was separated 90° from the target location (e.g., Dirks and Wilson, 1969; MacKeith and Coles, 1971; Plomp and Mimpen, 1981; Bronkhorst and Plomp, 1992; Peissig and Kollmeier, 1997).

When listening monaurally, there is often a large difference in performance depending on which ear is being stimulated (dotted lines and dashed lines in Fig. 3) and the spatial relation of this ear to both the target and the competing sentence locations. The ''better monaural ear'' is defined as the ear that produces the lowest error rate (which also typically has the higher signal-to-noise ratio[3]) for a particular competing configuration. When the target and competing sounds arise from different spatial locations, the effective signal-to-noise ratio is usually higher at one ear than the other due to the acoustical shadow of the head.[4]

The performance in binaural conditions (Fig. 3) is always better than ''better monaural ear'' performance. Bronkhorst and Plomp (1992) found an average of 3 dB better performance (corresponds to roughly 30% fewer errors in our technique) for spatially separated competitors in binaural versus ''better monaural ear'' performance for environments including up to six competitors. Therefore, the lack of large improvement in many environments for the binaural listeners over the better ear monaural listeners is likely due to ceiling and floor limitations of our measurement technique. There are several situations, however, which show a large difference (maximal expected binaural advantage) between binaural and better monaural ear listening in *intermediate* configurations in which the target was at the front (0°) in the presence of two or three competing sentences. This suggests that in real-world situations (like the ones we are simulat-
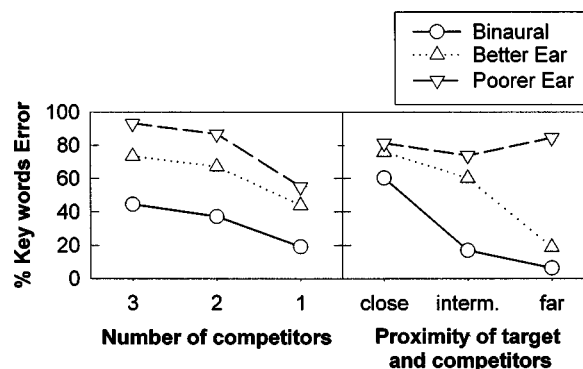

FIG. 4. Error rates in virtual-listening speech intelligibility experiments as a function of the number of competitors and the proximity of the competitors to the target location for binaural (circle), better monaural ear (up-pointing triangle) and poorer monaural ear (down-pointing triangle).

ing), the actual advantage of binaural listening is large only for particular, but typical situations.

C. Discussion

This study measured speech intelligibility in the presence of multiple competing sentences in which the competing sounds are varied in both their number and proximity to the target location. The major findings are: (1) results are similar in the sound-field and virtual listening conditions; (2) for binaural and better-monaural-ear conditions, the proximity of the target and competing sentences has more influence on the error rate than the number of competing sentences; (3) when the monaural ear is in an unfavorable position, then the number of competing sentences has more influence on the error rate than does proximity of target and competing sentences; (4) there is a large difference between binaural and best-monaural performance only for particular configurations.

There was no evidence that testing in a sound-field without head movements yields different results from testing in a virtual listening experiment, even though individualized ear recordings were not made. This view is consistent with early studies such as Dirks and Wilson (1969) and confirm that the virtual presentation of speech sounds is reasonable (e.g., Bronkhorst and Plomp, 1992; Koehnke and Besing, 1996) for research and clinic applications. Yost *et al.* (1996) showed a difference between sound-field and virtual conditions; however, the listeners in the sound-field were allowed head movements while virtual stimuli were presented independent of head position. Our results suggest that, even though the listeners did not hear the sources in their correct elevations (listeners V1 and V2 reported elevated sources), they were able to use the azimuthal separation of virtual sources to increase the intelligibility of the target.

Figure 4 summarizes the effect of number of competitors and proximity of competitors to the target location separately. For all listening modes, the effect of number of competitors is similar with decreases of 25%–38% as the number of competitors is decreased from three to one. The effect of proximity of target and competitors is only apparent for the binaural and better-monaural listening modes with fewer errors (changes of approximately 55%) when the competitor

locations are moved from close to far. No consistent change in error rates was seen for the poorer-monaural-ear condition as a function of proximity; all data points are near 100% error rate. Therefore, for the binaural and better-monaural-ear listening modes, the effect of proximity of target and competitors have more influence on the error rate than does the number of competitors. This finding is consistent with a previous study by Yost *et al.* (1996). Having both the target and competing sentences be tokens from the same talker may have emphasized this spatial aspect. The poorer-monaural-ear listening mode is more affected by the number of competitors simply because there is no effect of the proximity of the target and competitors. The listeners could not consistently use the spatial separation in the sources to better understand the target under monaural conditions (Yost *et al.*, 1996), which is perhaps related to poorer localization ability under monaural conditions (discussed in Sec. III C).

These data show that monaurally hearing listeners perform much worse than listeners with binaural hearing in many situations; however, they may perform nearly as well in specific conditions when the target is situated on the side of their good ear. Previous comparisons between binaural and monaural listening have generally been made for one competing sound (e.g., Dirks and Wilson, 1969; MacKeith and Coles, 1971) with the exception of Bronkhorst and Plomp (1992) who studied environments containing up to six competitors. While there appear to be small differences between binaural and better-monaural-ear conditions in the single competing sentence situation (triangles in Fig. 3), the largest advantages were observed for configurations with multiple competing sources in intermediate proximity. Our results further suggest that the large binaural advantages are also present in asymmetrical competing configurations when either two or three competing sentences are on the side of the nonfunctioning ear. Bronkhorst and Plomp (1992) determined that the binaural advantage was 2.4 to 4.3 dB depending on the number and location of competing sources. Therefore, the benefit of the binaural system is available regardless of the environment; however, it does not always improve intelligibility.

A similar pattern of binaural advantage (decrease in error rate from better monaural to binaural performance) for the various environments tested is obtained when the target and competitor location(s) are processed through a modified version[5] of the model by Zurek (1993) (Fig. 5). In this model, the intelligibility of the speech is predicted from an Articulation Index calculation based on the signal-to-noise ratio at the better monaural ear and binaural interaction in individual frequency bands weighted by their importance to understanding the speech. This analysis shows that although the signal-to-noise ratio is improved for binaural listening over better monaural listening conditions, the actual intelligibility is not always improved. This lack of consistent improvement is due to the ceiling and floor effects inherent to speech intelligibility.

The advantage of binaural listening over better monaural ear listening is not the most relevant comparison, however, since the monaural listener does not have the luxury to choose the one ear that is "better" for a given situation. A
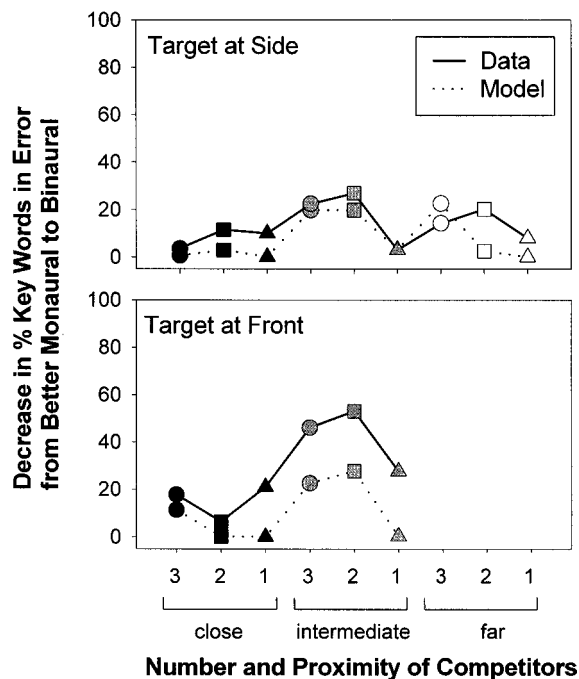

FIG. 5. Decrease in error rates for binaural listening over better monaural listening mode for speech intelligibility experiments as a function of the number of competitors (symbol shape) and the proximity of the competitors to the target location (symbol color). The data are connected by solid lines and the model prediction [modified version of the model by Zurek (1993)] for the same environments are connected by dotted lines.

truly monaural listener would be at a significant disadvantage some of the time. It is interesting to note that the configurations with the target sentence in front and the competing sounds toward the nonstimulated ear (environment with the largest binaural advantage) would be a common environment encountered by a unilaterally deaf person trying to utilize speech-reading.[6] The monaural listener has to either employ speech-reading or to turn his or her head such that the target is at the side and the competing sentences are in front; by reorienting his or her head, the listener would obtain a 40% decrease in error rate for monaural listening. MacKeith and Coles (1971) noted that "if the listener has only one usable ear he will more often be in a position of having either auditory advantage or visual advantage, but not both."

Our data generally support the claim by Peissig and Kollmeier (1997) that when the target is at 0°, performance is better if two competing sentences are on the same side, compared with conditions when they are distributed on both sides. Although the analysis of our data is limited by the fact that our results include floor effects for the comparable conditions that were tested, there was a small reduction in error from the condition of two competing sentences that were presented from either the left and right (−90° and +90°) to the condition where both were from the left (−90° and −60°).

In summary, the advantage of binaural listening in a realistic environment was shown to be dependent on specific details of the environment. When the speech was already highly intelligible monaurally, there was no improvement to be had by using binaural listening. When the speech intelligibility was poor monaurally, the improvement in signal-to-

3442   J. Acoust. Soc. Am., Vol. 105, No. 6, June 1999

Hawley *et al.*: Speech intelligibility and localization   3442

noise ratio by a few decibels was sometimes not enough to substantially improve the intelligibility, although in other environments those few decibels led to large improvements in intelligibility. Therefore, binaural hearing was sometimes able to drastically improve the intelligibility; however, the real benefit comes from having the better monaural ear always available.

III. LOCALIZATION EXPERIMENT

In the localization experiment, the listeners' task was to perform an absolute identification of source location for a known sentence in the presence of unknown competing sentence(s) presented from unknown locations. The number of competing sentences and the content of the target sentence were fixed throughout a block of trials. After each stimulus presentation, the listener was given approximately 6 s to write down the judged location of the target sentence, specified by the number of the location (an integer in the range 1 to 7).

None of the sentences used in the localization experiment had been previously used as targets in the speech intelligibility experiment (which required 600 target sentences out of 720 sentences in the IEEE corpus). The twelve shortest sentences from this pool of 120 sentences were designated as target sentences for this experiment, therefore guaranteeing that the target sentence was shorter than any of its competitors. The competing sentences were chosen from the remaining sentences in the pool and varied from trial to trial. While sentences spoken by both talkers were used, in any given block the same talker spoke both the target and the competing sentences. Within a trial, all sentences had nearly synchronous onsets, but asynchronous offsets. Each sentence was scaled to the same root-mean-square value, corresponding to approximately 62 dBA, the same as was used in the speech intelligibility experiment. Hence, the effective level of the overall competing sound increased as the number of competing sentences was increased.

A. Design

The parameters of interest in this study were the number (one, two or three) and configuration (relative locations) of competing sentences. The target sentence was presented randomly from one of the seven speakers ($-90°$ to $+90°$) simultaneously with competing sentences (one, two or three) in the same configurations as used in the intelligibility experiment (Table I). Note that the competing sentences were clustered on the left, in front, on the right, or distributed on both sides.

The total number of competing sentence configurations is 15 (7 with one competing sentence and 4 for each with two- and three-competing sentences). All conditions for each number of competing sentences were randomized and then tested in blocks of either 70 or 84 trials each with the content of the target sentence fixed. The number of trials in each block differed to give roughly equal number of blocks for each number of competing sentences. Each configuration was repeated about 15 times[7] for each of the seven target positions. The order of the blocks was randomized and all listeners performed the experiment in the same order. A total

of 1575 trials was tested per listener. In addition, the monaural (virtual-listening) subjects localized the target without competing sources.

Listeners were given minimal training, but no trial-by-trial feedback was given. At the beginning of each testing session, sample sentences were played from each location in order between $-90°$ to $+90°$, twice, and listeners were given several practice trials in which they localized a known sentence. There was no feedback during the training trials, and performance was not scored. The localization experiment required approximately 6 h of testing per listener over multiple sessions.

B. Results

To quantify the effect of competing sentence configuration on localization performance, several statistics were calculated for each competing sentence configuration: percent correct, root-mean-square error (rms error) and correlation coefficient r for the least-squares linear regression. The percent correct is a measure of how many correct identifications are made independent of the size of the errors on incorrect trials; the rms error is a measure of the average size of the errors between observed and perfect performance; and r^2, the proportion of variance accounted for by a linear regression, is a measure of the size of the deviations between the best fitting line through the data and the actual responses (Good and Gilkey, 1996). Random guessing would result in 14% correct, a rms error of $85°$, and a r^2 value of 0. The statistics of rms error and r^2 gave consistent information, therefore only rms error is considered further.

The percent correct and rms error for each listener are shown in Fig. 6. The data for each listener are grouped in clusters, with columns slightly offset according to the number of competing sentences. Results for individual subjects are arranged so that binaural listeners are on the left and monaural listeners are on the right. The symbol shape and shading denotes the number of competing sentences. Average and standard deviation are plotted for all conditions having the same number of competitors. Dark lines are plotted for localization with no competing sentences (only measured for the monaural virtual-listening subjects).

When listening binaurally, overall performance was quite good with 89% correct averaged across all listeners and all competing sentence configurations. Individual binaural listeners accurately localized the target sentence on at least 92% of the sound-field trials and 72% of the virtual-listening trials averaged over all competing sentence configurations. At least 95% of the responses for each binaural listener were within one speaker of the correct location. This corresponds to a rms error typically less than $30°$. Two of the three listeners in the virtual listening group (V1 and V2) reported that the sounds appeared to be presented from locations that matched the recorded azimuth, but elevated from the recorded locations. While listeners V1 and V2 have the lowest overall percent correct of the binaural listeners, the rms error observed is similar to the other binaural listeners.

Monaural virtual-listening subjects performed much poorer than the binaural listeners. The monaural listeners had an overall average accuracy of 27%, with individual listen-
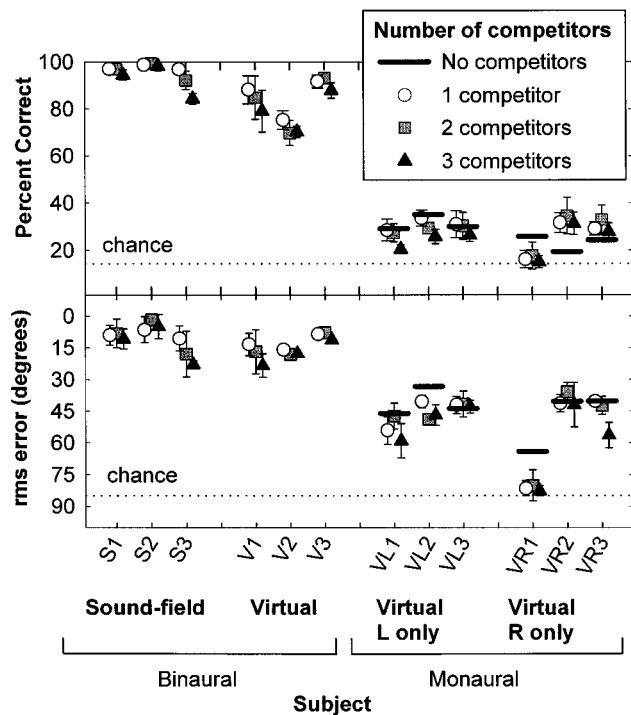
FIG. 6. Results for localization experiments for individual listeners. The percent correct and the root-mean-square (rms) error are plotted for each competing sentence configuration. The symbol and shading denoting the number of competing sentences (white for one, gray for two, black for three). Dark lines denote performance for localization in quiet for monaural listeners.
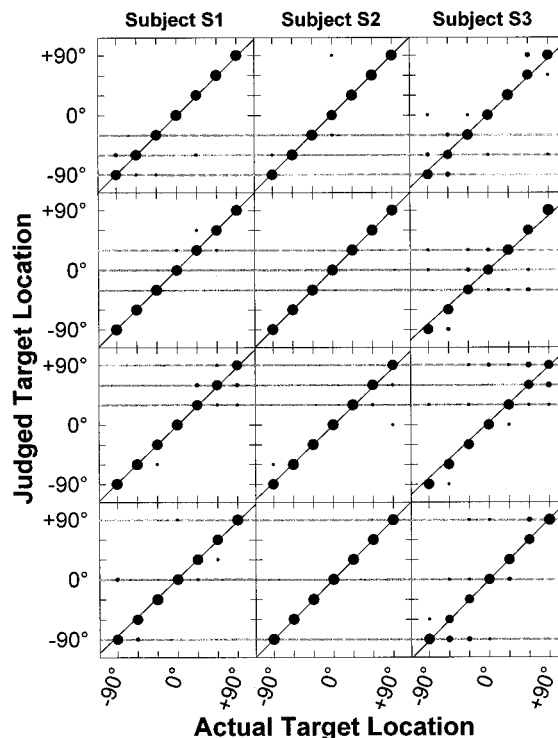

FIG. 7. Confusion matrix of localization responses for the binaural sound-field listening group for three-competing-sentence conditions. Separate listeners are shown in each column. The location of each competing sentence is shown by the gray horizontal lines (first row: −90°, −60° and −30°; second row: −30°, 0° and +30°; third row: +30°, +60° and +90°; fourth row: −90°, 0° and +90°). The size of each circle is proportional to the number of responses judged to that location. The black line is the least square error linear regression for each panel.

ers' averages in the range of 17%–31%. (Note that chance performance is 14% since there are only seven possible responses.) The overall rms error was 52°, with individual listeners' averages ranging from 40° to 81°. Localization ability in quiet was not consistently improved over localization ability in the presence of competitors for these listeners.

A two-way mixed design ANOVA [binaural listening condition (sound-field or virtual-listening)×competing sentence configuration (all 15 configurations including one, two or three competitors)] yielded a significant effect of listening condition ($p<0.001$), but not of competing sentence configuration ($p>0.41$) or for the interaction between these factors ($p>0.98$) when evaluated using either percent correct or rms error. The binaural listening condition was a between-subjects factor and the competing sentence configuration was a within-subjects factor. The average percent correct was 96% and 83% for the sound-field and virtual-listening groups, respectively. The average rms error was 10° and 14° for the sound-field and virtual-listening groups, respectively. Although the rms error differs significantly between the groups, the difference in rms error is quite small; average errors are much smaller than the separation between the speakers.

The response distributions of the binaural listening groups for the three-competing-sentence configurations are shown in Figs. 7 and 8 for the sound-field and virtual-listening subjects, respectively. The three-competing-sentence configurations are shown since these configurations typically show the poorest performance observed for the subjects. In these figures, results for a single subject are given in

each column. Each row of panels shows a different three-competing-sentence configuration with the location of each competing sentence denoted by a horizontal gray line. Each panel shows the distribution of judged target locations for each target location. The size of the circle indicates the relative number of responses for each judged location. Perfect performance corresponds to all judgments along the diagonal. The black line shows the best linear fit for the data in each panel separately.

The pattern of errors for individual subjects highlights the individual differences observed. All three binaural sound-field listeners (Fig. 7) localized the target well, even in the presence of three equal-level competing sentences. The large circles are concentrated along the diagonal and the best linear fit is very close to the diagonal. However, an analysis of the errors observed for the three-competing-sentence configurations shows that while the total error rates are low for all listeners, they occasionally mislocalized by more than 30° from the target location to the location of a competing sentence (S1: 1.3%, S2: 0%, S3: 5.5% of total trials). These infrequent large errors for subject S3 explains the lower percent correct and larger rms error for this subject in these three-competing-sentence configurations as compared with the other sound-field listeners. The virtual-listening subjects (Fig. 8) tend to have more errors that are near the target than was observed for the sound-field listeners. An analysis of the errors observed for the three-competing-sentence configura-

tions shows that, while the total percent correct tends to be lower than was observed for the sound-field listeners, the percentage of errors which were more than 30° and at the location of a competing sentence was not increased (V1: 4.0%, V2: 0.5%, V3: 0.2% of total trials).

Responses for representative monaural (virtual-listening) subjects for the three-competing sentence configurations are shown in Fig. 9 for the poorest performing subject (VR1) and two typical subjects (VL3 and VR2). The three-competing-sentence conditions plotted are the same as shown for the binaural subjects in Figs. 7 and 8. All six monaural-listening subjects showed poorer performance than the binaural subjects did. Responses are no longer tightly clustered near the diagonal, and the best fitting line is often very different from a diagonal line. Subject VR1 shows no relation between his judgments and the actual target location, with correlation coefficients near zero. The other two listeners shown (and the other three not shown) tended to choose the correct side much of the time, but with little differentiation between locations.

C. Discussion

The major findings of the localization study were: (1) under binaural conditions, localization performance is excellent (average of 90% correct and 12° rms error) for all configurations that were studied; (2) under monaural conditions, localization performance is poor for all configurations that were studied; (3) localization in the sound-field gives significantly smaller rms error than localization under virtual lis-

FIG. 9. Confusion matrix of localization responses for representative listeners in the monaural virtual-listening group for the three-competing-sentence conditions. The left column shows listener with accuracy close to chance. The middle and right columns show typical performance for the other five listeners. See Fig. 7 for explanation.

tening conditions; and (4) some subjects showed large infrequent mislocalizations to the location of competing sentences as the environment increased in complexity.

The binaural listeners were able to localize the target sound well even in the presence of competitors for all configurations tested. The stimuli were whole sentences (about 3 s long) presented at a level where the target is detectable even when presented with three competing sentences, although not always intelligible (Hawley and Colburn, 1997). This is consistent with the recent study by Yost *et al.* (1996) in which words that were correctly identified were localized correctly at least 80% of the time.

The observation that localization is better under binaural than monaural conditions is expected (e.g., Middlebrooks and Green, 1991; Slattery and Middlebrooks, 1994; Wightman and Kistler, 1997a), since the localization accuracy in the azimuthal plane is thought to be dominated by the availability of interaural timing information (e.g., Wightman and Kistler, 1992). Monaural listeners do not have interaural timing information and therefore must rely on overall level and spectral differences. Five of six monaural-listening subjects in this study were able to judge the side of the target much of the time presumably based on the level of the target. Since the target was always played at the same level, the level cue would be reliable. The listeners in this study had no experience listening monaurally prior to this study and therefore may not be used to relying on spectral cues. Since the subjects were listening to someone else's (KEMAR) transfer functions, training with feedback may have allowed the subjects to learn these features and improve their performance.
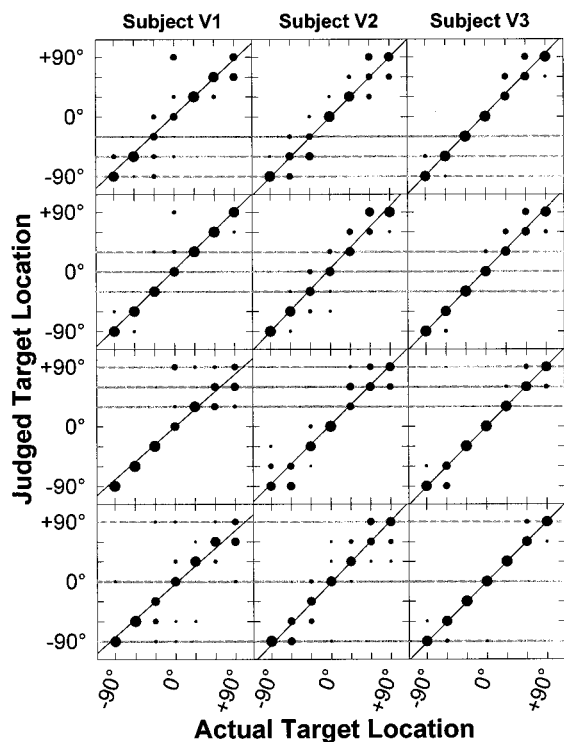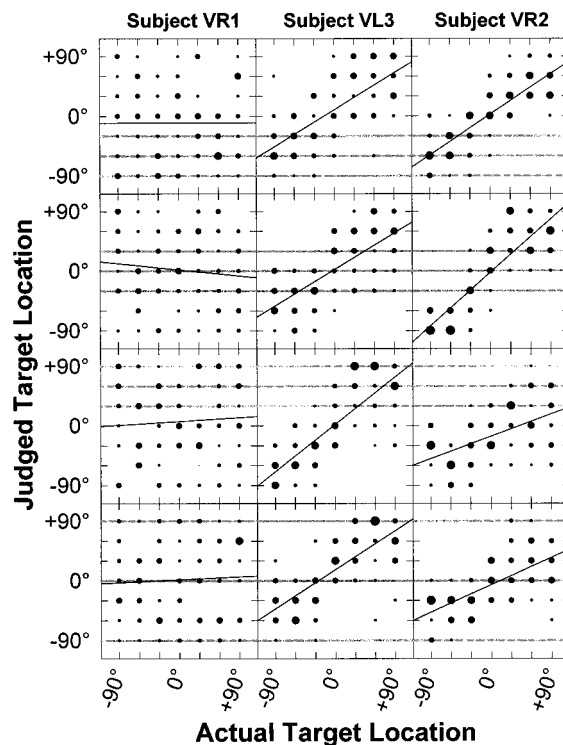

FIG. 8. Confusion matrix of localization responses for the binaural virtual-listening group for the three-competing-sentence conditions. See Fig. 7 for explanation.

Intersubject differences in listeners' ability to localize sounds under monaural conditions has also been seen by Slattery and Middlebrooks (1994) where some experienced monaural listeners performed better than and others performed comparably to nonexperienced monaural listeners; however, their performance was always poorer than binaural listeners.

Results suggest that listeners' abilities to make absolute identification judgments in the horizontal plane based on virtual sources did not appear to be greatly influenced by the distortions in perceived elevations experienced by subjects V1 and V2. Listeners appear to have categorized their localization judgments in a similar way regardless of the accuracy of the perceived elevation, yielding similar results to those of Besing and Koehnke (1995) who employed comparable virtual-listening methods. The frequent small errors and perceived elevation of sources reported by some of the listeners in the virtual listening group could be due to the use of KEMAR recordings and headphone presentation rather than more precise simulations (e.g., Wightman and Kistler, 1989a, 1989b; Kulkarni, 1997). While features of the head-related transfer functions (HRTFs) differ across listeners, particularly at higher frequencies (e.g., Wightman and Kistler, 1989a; Pralong and Carlile, 1996), the use of individualized functions have been shown to be less critical for azimuthal localization (e.g., Wenzel *et al.*, 1993).

Our results show a small, yet statistically significant decrease in performance between the sound-field and the virtual-listening subjects. A direct comparison between performance in the sound-field and virtual-listening conditions (e.g., Wightman and Kistler, 1989a; Wenzel *et al.*, 1993) cannot be made since our groups were composed of different listeners. Even for the virtual listener (subject V2) who showed the poorest overall performance (72% correct), the rms error is less than the difference between adjacent speaker locations.

Only two of the binaural listening subjects (listeners S3 and V1) showed increased difficulty localizing sounds as the number of competing sounds increased. For these listeners, the average percent correct decreased by 9.1% and 12.6% and the average rms error increased by 10.0° and 12.4° from the one- to the three-competing sentence environments. For the other four listeners, the average change from the one- to the three-competing-sentence environments was less than 5% correct and less than 3° rms error. Listeners S1 and V1 also showed an increased likelihood of making large mislocalizations to the location of a competing sentence, particularly in the three-competing-sentence environment (4.0% and 5.5% of the total trials). Since all sentences were spoken by the same talker, these listeners (S1 and V1) may have been confused as to which sentence was currently the target sentence and instead localized a competing sentence, although the experimental design intended to minimize this problem by having the target sentence written on the answer sheet in front of the listener and remain consistent for an entire block of trials.

IV. INTELLIGIBILITY AND LOCALIZATION IN THE SAME LISTENERS

Intersubject and presentation mode differences noted in the localization experiment did not translate into differences in speech intelligibility performance. Although there were significant differences in localization performance between the binaural groups listening in the sound-field and virtual environments, there were no significant differences between the groups in their intelligibility performance. Results from other experiments (Hawley and Colburn, 1997) show that these sentence stimuli can be localized in quiet at a level 20 dB below the level required for 50% correct key word intelligibility. Since the level of the target stimulus was well above this criterion in all cases, the localization performance for both the sound-field and virtual listening conditions is likely to be close to the asymptote of maximum performance for the subject. It is also likely that the separation of the speakers in the speech intelligibility was too large to show an effect of poorer localization of the stimuli on the intelligibility results. The intersubject differences (not tested statistically) among the monaural listeners is more striking than was seen for the binaural listeners with subject VR1 showing almost no localization ability at all and the other monaural listeners showing better than chance performance. However, the intelligibility performance of all monaural listeners was similar. Therefore, accurate localization of the target source is not necessary for intelligible speech under monaural conditions.

V. CONCLUSIONS

(1) In comparisons between sound-field or virtual listening results, there was no statistically significant difference for the intelligibility experiment and there were small, but significant, differences for the localization experiment.

(2) The proximity of the competing sentences to the target location was more influential than the number of sources on the intelligibility of sentences under binaural and better-monaural-ear listening conditions. Under poorer-monaural-ear listening conditions, the number of competing sounds was more influential on the observed intelligibility than was proximity of the target and competing sources.

(3) In environments with equal level target and competitor(s), large differences in intelligibility for binaural over better-monaural-ear listening were observed only for particular multiple-competing-sentence configurations. A model which predicts speech intelligibility in various environments confirmed this finding.

(4) Localizability of a clearly audible target is quite robust for sentences under binaural conditions even in the presence of three competing sounds regardless of their locations; however, there are occasional mislocalizations to the location of a competing sound.

(5) Monaural localization is much poorer than binaural localization as expected; however, performance does not typically degrade further even in the presence of three competing sounds. Listeners are generally able to only

judge the correct side that the sound was played from, even in quiet environments.

ACKNOWLEDGMENTS

[1]Kulkarni (1997) showed that the transfer function for the same headphone and listener varies with each placement. This is mentioned only to point out another source of variability in the simulation.

[2]It is unlikely that testing all subjects in the same order corrupted the results since the testing took place over six sessions and measurements for each condition were made during each session.

[3]While the ear with the higher signal-to-noise ratio will also typically give the lower error rate, there can be a discrepancy since the signal-to-noise ratio can improve without improving the intelligibility of the speech. This occurs if the signal-to-noise ratio improvement is due to frequency regions that are not important for speech. For all of the conditions tested in this study, the ear with the lower error rate also has the higher signal-to-noise ratio.

[4]One ear is always more favorable than the other ear in a single competitor environment if the spatial location of the target and competitor are different. However, when there are multiple competitors with symmetrically placed competitors and target from the front, there can be spatial separation without a more favorable ear.

[5]The original model is valid for a single steady noise competitor in anechoic space. Since the competitors are speech instead of noise, to get comparable intelligibility rates between the model and the actual performance the level of the target needed to be decreased by 5 dB. This amount is similar to the 6–8 dB lower thresholds obtained by Festen and Plomp (1990) for a speech competitor than for a noise competitor. To extend the model to multiple competitors, the amount of effective binaural interaction was assumed to be limited by the closest spatial competitor and equal to the binaural interaction as if that were the only competing source.

[6]A proposed rehabilitative strategy for monaural hearing loss is the Contralateral Routing of Signals (CROS) hearing aid configuration (e.g., Harford and Barry, 1965) in which the sound that would have been received at the deaf ear is presented to the good ear through an open ear mold. This strategy was simulated by adding the left and right signals and presenting the sum to the right ear. An additional subject tested in this configuration and performance, relative to having only the right ear stimulus, was comparable to the performance of the better monaural ear only when the target was in front. This resulted in improvements only in the single competitor case (since the ears were not much different in the two- and three-competitor environments). Performance was much worse in conditions in which the right ear was the better monaural ear (target on the same side as the good ear) and not improved when the right ear was the poorer monaural ear. Therefore, limited benefit would be expected from this strategy in multi-source environments.

[7]Due to an error in the making of the tape for this experiment, an unequal number of repetitions were performed for individual target-competitor pairs. The range of repetitions was 8 to 23.

Abouchacra, K. M., Tran, T. V., Besing, J. M., and Koehnke, J. D. (**1997**). ''Performance on a selective listening task as a function of stimuli presentation mode,'' Association for Research in Otolaryngology Abstracts, 53.

Besing, J. M., and Koehnke, J. (**1995**). ''A test of virtual auditory localization,'' Ear Hear. **16**(2), 220–229.

Blauert, J. (**1997**). *Spatial Hearing*, 2nd ed. (MIT, Cambridge, MA).

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bronkhorst, A. W., and Plomp, R. (**1990**). ''A clinical test for the assessment of binaural speech perception in noise,'' Audiology **29**, 275–285.

Bronkhorst, A. W., and Plomp, R. (**1992**). ''Effect of multiple speech-like maskers on binaural speech recognition in normal and impaired hearing,'' J. Acoust. Soc. Am. **92**, 3132–3139.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Nasser Kotby, M., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). ''An international comparison of long-term average speech,'' J. Acoust. Soc. Am. **96**, 2108–2126.

Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech, with one and with two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Dirks, D. D., and Wilson, R. H. (**1969**). ''The effect of spatially separated sound sources on speech intelligibility,'' J. Speech Hear. Res. **12**, 5–38.

Dunton, L. B., and Jones, J. K. (**1996**). ''Intelligibility and localization of speech in a complex multi-talker environment: Listeners with normal hearing and hearing impairments,'' senior project, Biomedical Engineering, Boston University, Boston, MA.

Duquesnoy, A. J. (**1983**). ''Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons,'' J. Acoust. Soc. Am. **74**, 739–743.

Ericson, M. A., and McKinley, R. L. (**1997**). ''The intelligibility of multiple talkers separated spatially in noise,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 701–724.

Festen, J. M., and Plomp, R. (**1990**). ''Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**, 1725–1736.

Giguere, C., and Abel, S. M. (**1993**). ''Sound localization: Effects of reverberation time, speaker array, stimulus frequency, and stimulus rise/decay,'' J. Acoust. Soc. Am. **94**, 769–776.

Gilkey, R. H., and Anderson, T. R. (**1995**). ''The accuracy of absolute localization judgments for speech stimuli,'' Journal of Vestibular Research **5**(6), 487–497.

Good, M., and Gilkey, R. H. (**1996**). ''Sound localization in noise: The effect of signal-to-noise ratio,'' J. Acoust. Soc. Am. **99**, 1108–1117.

Good, M. D. (**1994**). ''The influence of noise on auditory localization in the free field.'' M.S. thesis, Psychology, Wright State University, Dayton, OH.

Harford, E., and Barry, J. (**1965**). ''A Rehabilitative Approach to the Problem of Unilateral Hearing Impairment: The Contralateral Routing of Signals (CROS),'' Journal of Speech and Hearing Disorders **30**, 121–138.

Hawley, M. L., and Colburn, H. S. (**1997**). ''Speech intelligibility and localization in a complex environment: relationship to binaural hearing,'' Second Biennial Hearing Aid Research and Development Conference, Bethesda, MD, p. 60.

Hirsh, I. J. (**1950**). ''The relation between localization and intelligibility,'' J. Acoust. Soc. Am. **22**, 196–200.

IEEE (**1969**). ''IEEE Recommended Practice for Speech Quality Measurements,'' IEEE Trans. Audio Electroacoust. **17**(3), 225–246.

Jerger, J., Carhart, R., and Dirks, D. (**1961**). ''Binaural hearing aids and speech intelligibility,'' J. Speech Hear. Res. **4**(2), 137–148.

Koehnke, J., and Besing, J. M. (**1996**). ''A procedure for testing speech intelligibility in a virtual listening environment,'' Ear Hear. **17**(3), 211–217.

Kulkarni, A. (**1997**). ''Sound localization in real and virtual acoustical environments,'' Ph.D. dissertation, Biomedical Engineering, Boston University, Boston, MA.

MacKeith, N. W., and Coles, R. R. A. (**1971**). ''Binaural advantages in hearing of speech,'' J. Laryngol. Otol. **85**, 213–232.

Middlebrooks, J. C., and Green, D. M. (**1991**). ''Sound localization by human listeners,'' Annu. Rev. Psychol. **42**, 135–159.

Nabelek, A. K., and Robinson, P. K. (**1982**). ''Monaural and binaural speech perception in reverberation for listeners of various ages,'' J. Acoust. Soc. Am. **71**, 1242–1248.

Nilsson, M. J., and Soli, S. D. (**1994**). ''Norms for a headphone simulation of the hearing in noise test: comparison of physical and simulated spatial separation of sound sources,'' J. Acoust. Soc. Am. **95**, 2994.

Peissig, J., and Kollmeier, B. (**1997**). ''Directivity of binaural noise reduc-

tion in spatial multiple noise-source arrangements for normal and impaired listeners,'' J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R. (**1976**). ''Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound wave (Speech or Noise),'' Acustica **31**, 200–211.

Plomp, R., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Plomp, R., and Mimpen, A. M. (**1981**). ''Effect of the orientation of the speaker's head and the azimuth on a noise source on the speech reception threshold for sentences,'' Acustica **48**, 325–328.

Pollack, I., and Pickett, J. M. (**1958**). ''Stereophonic listening and speech intelligibility against voice babbles,'' J. Acoust. Soc. Am. **30**, 131–133.

Pralong, D., and Carlile, S. (**1996**). ''The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,'' J. Acoust. Soc. Am. **100**, 3785–3793.

Shaw, E. A. G., and Vaillancourt, M. M. (**1985**). ''Transformation of sound pressure level from the free field to the eardrum in numerical form,'' J. Acoust. Soc. Am. **78**, 1120–1123.

Slattery, 3rd W. H. I., and Middlebrooks, J. C. (**1994**). ''Monaural sound localization: Acute versus chronic unilateral impairment,'' Hearing Res. **75**, 38–46.

Walker, H. M., and Lev, J. (**1953**). *Statistical Inference* (Holt, New York).

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). ''Localization using nonindividualized head-related transfer functions,'' J. Acoust. Soc. Am. **94**, 111–123.

Wightman, F. L., and Kistler, D. J. (**1989a**). ''Headphone simulation of free-field listening. I: Stimulus synthesis,'' J. Acoust. Soc. Am. **85**, 858–867.

Wightman, F. L., and Kistler, D. J. (**1989b**). ''Headphone simulation of free-field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878.

Wightman, F. L., and Kistler, D. J. (**1992**). ''The dominant role of low-frequency interaural time differences in sound localization,'' J. Acoust. Soc. Am. **91**, 1648–1661.

Wightman, F. L., and Kistler, D. J. (**1997a**). ''Monaural sound localization revisited,'' J. Acoust. Soc. Am. **101**, 1050–1063.

Wightman, F. L., and Kistler, D. J. (**1997b**). ''Sound localization in the presence of multiple distracters,'' J. Acoust. Soc. Am. **101**, A3105.

Yost, W. A. (**1992**). ''Auditory perception and sound source discrimination,'' Current Directions in Psychological Science **1**(6), 179–184.

Yost, W. A. (**1997**). ''The Cocktail Party Problem: Forty Years Later,'' in Binaural and Spatial Hearing in Real and Virtual Environments, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–347.

Yost, W. A., Dye, R. H. Jr., and Sheft, S. (**1996**). ''A Simulated ''Cocktail Party'' with Up to Three Sound Sources,'' Percept. Psychophys. **58**(7), 1026–1036.

Zurek, P. M. (**1993**). ''Binaural advantages and directional effects in speech intelligibility,'' in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Boston), pp. 255–276.