## **SPEECH INTELLIGIBILITY, SPATIAL UNMASKING, AND REALISM IN REVERBERANT SPATIAL AUDITORY DISPLAYS**

Boston University Hearing Research Center, Departments of Cognitive and Neural Systems and Biomedical Engineering

*Barbara Shinn-Cunningham*

ear a ee diffeece (ILD), adec a ae, e and effective dection [10, 18, 37, 40]. Previous results examining how well subjects localize in rooms (coduced a deae-sed cassroom boadbad  $T_{60} = 650$  ms) show that direction accuracy is only that accuracy is only that  $T_{60} = 650$  ms show that direction  $T_{60} = 650$  ms show that direction  $T_{60} = 650$  ms show that direction  $T_{60} = 650$  ms show that direct modestly degraded (mean localization errors  $e_{\mu\nu}$ ,  $e_{\mu}$  are increased by roughly 25%), but a cenepce  $\mathcal{M}$  of  $\mathcal{N}$  is significantly e <sup>n</sup>aced (b and de f an ode) compared to in a ec. c ace  $[39, 40]$ .

In addition of the location and length space of the control and length spatial perception, echoes and reverse the temporal modulation and  $r$  $\texttt{lead}$  a ed. In partial, echoes and reverberations and reverse  $t$ ed e ar ar ear out amplitude modulations,  $a \cdot a \cdot a \cdot A \cdot e$ ,  $d \cdot a \cdot f e$  e ce [41-46].

## **4. SPEECH IN REALISTIC ROOMS**

 $F_{\text{A}}$ <sub>r</sub> a speed-ae speech signal, such a speech signal, such as  $\mathbb{Z}$  decatures acoustic features are converted by the relative energy content at each frequency of  $\mathbb{R}^n$  . The isothermore  $\mathbb{R}^n$ constant over the voltons of  $\alpha$  other the voltons of  $\alpha$ speech sounds, information is convexed that it changes in ee ee ead ferec; e, ro fe information  $\mathbf{A}_1$  a speech signal is convexed by temporal is  $\mathbf{a}_1$  and  $\mathbf{a}_2$ modulations in the energy of the speech signal  $\mathbf{a}$ a each fe<sub>e</sub>urenc [47-49].

Beca<sup>n</sup>e eche and element can edice the e temporal modulations, echoes and reverberant can degrade speech intelligibility in some acoustic environments.  $H$  ee, for most ordinary (i.e., relatively small rooms), the  $\epsilon$  temporal extending  $\epsilon$  extending  $\epsilon$  and  $\epsilon$  extending is shortly in compared to the modulations in speech, and only delivered to the modest speech and only modest  $\mathbf{d}$ perce<sup>ptu</sup>a degada<sub>tion</sub>s agree, a least at least at the ear receiving the early t more intense direct sound (e.g., see [42]). Of course, the severity of the effects of echoes and reverberation on the effects of echoes and reverberation on the effects o signals at the ears varies with the location of the source relative to the checause to direct and environmental varies and example. decadd a<sup>n</sup>ce.

The effector are demonstrated in Figure 1, which plots are a same fa eec<sup>k</sup> sa each se eff ear anecho space (in black) suberimposed over the signal that would reach the ear in a normal (moderate-sized) classroom (plotted) classroom (plotted in a  $f$  and  $f$  are a and a denote a distance of 90. the  $t$  (in the  $t$ ,  $t$  and a edge ard plane containing the ears).  $\texttt{T}$  ee $\texttt{e}$  results were generated by measuring the head-related by measuring the head-related impulse responses (HRIR) in the classical policies  $\mathfrak{e}^{\bullet}$  classroom using a  $m$ axim $n$ -length sequence technique, then processing range techniques  $r$ speech aneft waveforms the either pseudo-anechoic HRIRs (in ceche and eele and reverberation  $\mathfrak{e}_1$ time windowing),  $e^a e^b e^b a^b$  HRIRs (in which both the decade e be<sup>4</sup>a o de <sup>re</sup>e cuded).

Results of that echoes and reverberation have the large effect on the total signal signal at the source is the source in the source is a the source is in the source is in the source is in the source is in the source in the source is in the source in the source is in the s  $a^2$ 90 $\ldots$  equaled and the efter  $a$  signal is considered. In the cases,  $\bullet$  energy is defined and  $\bullet$  can energy is relatively low, lee ca.e. i.e. u.e.  $\frac{1}{2}$  to large equality and reversion.

## **5. SPATIAL UNMASKING OF SPEECH IN ROOMS**

Eceadeebeaticae degadations both diectional early  $a^{\bullet}d$  eechoese.  $b^{\bullet}$  intelligiblishing  $b^{\bullet}$ a<sup>n</sup>d eebea<sup>t</sup>ion a deade ebeef faa  $s = s$  of target and masker sources on speech intelligibility.  $e^A$  is become  $\cdot$ 

In  $de$  , e.a., e., e., e., e., e., e., e., e.,  $de$  and reverberation frece aatrical spatial a conducted under head phones. Target and masker signals were simulated at different locations using the pseudoarechoic and reverberant HRIRs used to process the signal antishown in F $\uparrow$ , e $\uparrow$ 1.  $\uparrow$  The masker was a steady-state noise, which a always simulated at a position directly in front of  $\alpha$ the listener and a cell find cm. The  $a^{\bullet}$  energy  $a^{\bullet}$  arget signals were nonselle se sense sind at one of the three distances  $(0.15, 1, 2)$  and diections  $(0.a d.90.)$  for which HRIR et eared, ead  $\blacksquare$  different and the sixted target/masker a a configurations.

F<sub>or</sub> each  $\alpha$  are a figuration, subject the energy of  $\alpha$ while listening binaurally, while left ear, and with only the left early  $\mathbf{a}_1$  density on  $\mathbf{a}_2$ only the right ear, the  $d$  direct analysis of the advantages of the advantag of binaural processions  $\Gamma$  , red finding room conditions were reduced by a conditions were  $\Gamma$ e ed  $f'$  each and configuration and ear condition, one  $\Lambda$  and anechoic space and one simulations  $e$ e $e$ be $a$  coditions.

 $\mathbf{F}_{\mathbf{z}}$  reach condition, speech reception thresholds were easted adaptic by a single education  $50\%$  of the target level until  $50\%$  of the targe the sentence of the sentence and the sentence words and a sentence was a sentence was a sentence was a sentence easured found to the same final thresholds. For  $\Box$ , all-ea<sub>ch</sub> thec completed each extension of  $\Lambda$ Figure  $2$  plots the range range range range  $r$  and  $r$  thresholds of the direct-sound portion of the target  $f$  of the direct-sound level of the direct-sound level of the direct-sound level of the direction of the directio

 $m = \frac{1}{2}$  at the  $\frac{1}{2}$  speeches the  $\frac{1}{2}$  speeches the  $\frac{1}{2}$  speeches the  $\frac{1}{2}$  $Te$   $\qquad \qquad$   $\qquad \qquad$   $\qquad$   $\qquad \qquad$   $\qquad$   $\qquad$ 

Oea, eae ferse enaac four bect  $C_1$  are the better-ear (right ear; dotted ed line) and binaural (solid line) results, the data show that diectia e a<sup>n</sup>atitifa e a<sup>n</sup>d a e eads to ana processing advances of 3-5 dB in both anechoic and reverberant simulations. Thus, the interaction of  $e$  interaction of  $f$ te tare ad a e and e heat earth sificant decrease in the effective energy of binaural processing.

We are ad are are  $\blacktriangle$  e are direction in anechoic face, there is no significally figure is no significant or consistent or consistent or consistent or c difference across efft are achieved with the left early alone, rear alone,  $\epsilon$  becomes binaural. Here, in the elberant simulations, there is a distinct binaural processing advance is a contract in the target is a different diace a eac.

When  $c_1$  and  $c_2$  and  $c_3$  and  $c_4$  and  $c_5$  the conditions in which  $\alpha$  and  $\alpha$  to the  $\alpha$ right, comparisons between the left (worse) and right (better) eare results show very large differences in monaural  $\mathbf{e}$  f<sup>a</sup> are Given the  $\mathbf{a}$  data are not arred, the difference primarily reflects the large interaction of the large interaction of the large interaction of the l difference (ILD). It a concern  $\epsilon$  T is  $e$ and to the  $\det f$  the lister  $\in$  [50-52];  $\det f$  are ILD decreases with diate, eading to corresponding decrements in the entriest of  $\mathbf{d}$ difference in the left and  $r$ <sup>1</sup> early and a<sup>4</sup> right direction of the left and  $r$ distance. Comparing antechoral reverberant results, Figure 2 shows that the addition of echoes a direction of  $\mathbf{c}_1$  $t_{\rm e}$  decrease the differences in left- and  $r_{\rm e}$  and  $r_{\rm e}$  and rearmonaure d, e ec<sup>h</sup>a a e fae hid ace (where  $t$ e reverberation has the largest impact in  $t$  is the largest in  $t$  is the largest induced in ee<sup>r</sup> afefeced ae ee<sup>r r</sup>efiae a de<sub>tri</sub>mental to understanding the target, the effect is easily the target is easily the target is easily the target e $\uparrow$ a ed. Sec foall, enefeced at energy is (at least rafirst-order approximation) roughly equal at the two ears of for all conditions, whereas the direct sound ILD in the target sound ILD in the target  $\epsilon$ is quere for  $\Gamma$  ,  $e$  and to the right of the right of the right of the listener. The eche a die ebeautiste dit a e a a less activit would intellect intelligibility for the acoustical  $\mathcal{A}_1$  e.e., where  $\lambda$ re is very little the little tenergy reaching the little  $\alpha$  reaching the line  $\alpha$ space. Overall, then, the echoes and reverberation tend to

*Proceedings of the 2002 International Conference on Auditory Display, Kyoto, Japan, July 2-5, 2002*

## **8. REFERENCES**

1. We<sub>c</sub>e, E.M., *Localization in virtual acoustic displays*. P<sub>r</sub>e e ce, 1992. **1**(1): . 80-107.

- 34. A<sub>r</sub>b<sub>a</sub>, T.L. and J. Kidd, Gerald, *Evidence for spatial tuning in informational masking using the probe* $signal$  method.  $J \cup_{v \in A} f \cdot e$  Acoustical Society of A e. ca, 1999: . <sup>A</sup> b ed.
- 35. Brungart, D.S., *Informational and energetic masking effects in the perception of two simultaneous talkers'.*  $J_{\text{max}}$  a fee Acoustical Society of America, 2001. **109**(3): . 1101-1109.
- 36. Kidd, G., et al., *Reducing informational masking by sound segregation.*  $J \cup_{k} a$  fie Acoustical Society of A e ca, 1994. **95**(6): **A** 3475-3480.
- 37. Shinn-Cunningham, B.G., J.G. Desloge, and N. Kopco. *Empirical and modeled acoustic transfer functions in a simple room: Effects of distance and direction. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. 2001. Ne Pfa., Ne Y.
- 38. Sa a<sub>,</sub> e., S., *Auditory Localization of Nearby Sources* in Anechoic and Reverberant Environments, *Cognitive and Neural Systems.* 2000, B<sub>ost</sub>on, B<sub>oston</sub>, MA.  $\begin{array}{c} \nU & e \\
39.5 & \text{A}_{\text{O}}\n\end{array}$ : B
- 39. Shinn-Cunningham, B.G. *Learning reverberation: Implications for spatial auditory displays. International Conference on Auditory Displays*. 2000.  $A$  a  $a$ , GA.
- 40. S<sub>hinn</sub>-Cunningham, B.G. *Localizing sound in rooms*. *ACM/SIGGRAPH and Eurographics Campfire: Acoustic Rendering for Virtual Environments.* 2001. S<sub>nowbi</sub>d,  $Ua$ .
- 41. H<sub>u</sub>a, T. a d H.J.M. See e e , A review of the MTF *concept in room acoustics and its use for estimating* speech intelligibility in auditoria.  $J \cup a$  for Ac<sub>t</sub> ca S c e f A e ca, 1985. **77**(3):  $.1069 - 1077$ 1077.
- 42.  $H \t a$ , T., H.J.M. See e e, a d R. P *Predicting speech intelligibility in rooms from the modulation transfer function I. General room acoustics.* Ac ca, 1980. **46**: 60-72.
- 43. Nomura, H., H. Miaa, and T. Houtana, *Speechintelligibility and subjective MTF under diotic and dichotic listening conditions in r;*