

- Lunner, T. (2003). "Cognitive function in relation to hearing aid use" *Int. J. Audiol.*, **42**, Suppl 1, pp. 49-58.
- Mishra, S., Rudner, M., Lunner, T., and Rönnerberg, J. (2010). "Speech understanding and cognitive spare capacity" in *Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing*. 2nd International Symposium on Auditory and Audiological Research. Elsinore, Denmark. Edited by J.M. Buchholz, T. Dau, J. Christensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. (The Danavox Jubilee Foundation, Copenhagen), pp. 305-313.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., and Wager, T.D. (2000). "The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis" *Cogn. Psychol.*, **41**, pp. 49-100.
- Pichora-Fuller, M.K. (2007). "Audition and cognition: What audiologists need to know about listening". In: Palmer, C, Seewald, R (Eds), *Proceedings of the Adult Conference Phonak*, pp. 77-85.
- Pichora-Fuller, M.K. and Singh, G. (2006). "Effects of age on auditory and cognitive processing: Implications for hearing aid fitting and audiologic rehabilitation" *Trends. Amplif.*, **10**, pp. 29-59.
- Rönnerberg, J., Arlinger, S., Lyxell, B., and Kinnefors, C. (1989). "Visual evoked potentials: Relation to adult speechreading and cognitive function" *J. Speech Lang. Hear. Res.*, **32**, pp. 725-735.
- Rönnerberg, N., Stenfelt, S., and Rudner, M. (2011). "Testing listening effort for speech comprehension using the individuals' cognitive spare capacity" *Audiology Research*, 1(1), e22. doi:10.4081/audiores.2011.e22.
- Rudner, M., Foo, C., Rönnerberg, J., and Lunner, T. (2009). "Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids" *Scand. J. Psychol.*, **50**, pp. 405-418.
- Rudner, M., Rönnerberg, J., and Lunner, T. (2011). "Working memory supports listening in noise for persons with hearing impairment" *J. Am. Acad. Audiol.*, **22**, pp. 156-167.
- Sarampalis, A., Kalluri, S., Edwards, B., and Haftner, E. (2009). "Objective measures of listening effort: Effects of background noise and noise reduction" *J. Speech Lang. Hear. Res.*, **52**, pp. 1230-1240.
- Schneider, B.A. (2011). "How age affects auditory-cognitive interactions in speech comprehension" *Audiology Research*, 1(1), e10. doi:10.4081/audiores.2011.e10.
- Stenfelt, S. and Rönnerberg, J. (2009). "The Signal-Cognition interface: Interactions between degraded auditory signals and cognitive processes" *Scand. J. Psychol.*, **50**, pp. 385-393.
- Tun, P.A., McCoy, S., and Wingfield, A. (2009). "Aging, hearing acuity, and the attentional costs of effortful listening" *Psychol. Aging*, **24**, pp. 761-766.
- Zekveld, A.A., Kramer, S.E., and Festen, J.M. (2010). "Pupil response as an indication of effortful listening - the influence of sentence intelligibility" *Ear Hear.*, **31**, pp. 480-490.

Auditory and cognitive contributions to hearing-impaired listeners' spatial speech recognition performance

TOBIAS NEHER, SØREN LAUGESEN, NIELS S. JENSEN, AND LOUISE KRAGELUND

Eriksholm Research Centre, Oticon A/S, Kongevejen 243, DK-3070 Snekkersten, Denmark

This study investigated the auditory and cognitive processes affecting speech recognition in spatially complex, multi-talker situations. Twenty-three elderly hearing-impaired (HI) listeners were tested on a number of competing-speech tasks, a measure of monaural spectral ripple discrimination, a measure of binaural temporal fine structure (TFS) sensitivity, and two cognitive measures indexing working memory and attention. All auditory test stimuli were spectrally shaped to restore (partial) audibility for each listener on each listening task. Eight younger normal-hearing (NH) listeners served as a control group. Data analyses revealed that the chosen auditory and cognitive measures were unable to predict speech recognition when the target and maskers were separated along the front-back dimension. When the competing talkers were separated along the left-right dimension, however, speech recognition was correlated with the measures of attention and binaural TFS sensitivity as well as with low-frequency hearing thresholds. Altogether, these results support the notion that both bottom-up and top-down deficits are responsible for the impaired functioning of elderly HI listeners in cocktail party-like situations.

INTRODUCTION

Spatial hearing is an important capacity of the auditory system, which is mediated by different acoustic cues: interaural phase and level differences are crucial for left-right (L-R) spatial hearing, while monaural spectral cues introduced by pinna filtering are crucial for front-back (F-B) spatial hearing (e.g. Blauert, 1997). The benefits offered by spatial hearing are particularly large in noisy environments where considerable speech recognition improvements can occur, especially if the interferers are also speech signals. Compared to NH listeners, however, HI listeners generally obtain much less spatial hearing benefit in such situations, especially if they are also older (e.g. Marrone *et al.*, 2008).

Previous research has been concerned with the supra-threshold deficits that might be responsible for HI listeners' poorer speech-in-noise performance. For example, reductions in TFS sensitivity (e.g. Strelcyk and Dau, 2009) and working memory capacity (e.g. Akeroyd, 2008) have been ascribed a role. However, in none of these studies was the speech target presented against a background of spatially separated speech maskers, and so it is unclear if these effects also apply to such situations.

In an earlier study, it was investigated if bilateral hearing-aid users are able to take advantage of spatial separation of competing talkers (Neher *et al.*, 2009). In that study, speech reception was measured with a frontal speech target and two speech maskers separated from the target along either the L-R or the F-B dimension. A number of auditory and cognitive measures were then tested in terms of their ability to predict speech recognition. For the L-R dimension, we observed correlations with average low-frequency hearing thresholds ($r = 0.48$, $p < 0.05$), working memory capacity ($r = -0.52$, $p < 0.05$), and attentional skills ($r = -0.64$, $p < 0.01$). For the F-B dimension, we observed correlations with working memory capacity ($r = -0.72$, $p < 0.01$) and attentional skills ($r = -0.52$, $p < 0.05$) as well. In addition, we found a correlation with average high-frequency hearing thresholds ($r = 0.60$, $p < 0.01$).

The purpose of the current study was to build on these findings. In particular, instead of trying to ensure audibility by fitting our listeners with hearing aids, we amplified all auditory stimuli in such a way that (partial) audibility was achieved for each listener on each listening task. We expected this approach to lead to “cleaner” supra-threshold effect estimates. Furthermore, we wanted to follow up on the correlations between L-R (or F-B) speech recognition and low (or high) frequency hearing thresholds. In particular, since low-frequency interaural phase differences (IPDs) mediate L-R spatial hearing and since pinna cues, which occur above about 4 kHz, mediate F-B spatial hearing (see above), we hypothesized that these correlations could reflect binaural TFS and pinna-cue deficits, respectively.

It should be noted that the data presented in this paper constitute a sub-set of a larger study into the spatial hearing abilities of HI listeners, and the interested reader is referred to (Neher *et al.*, 2011) for more information.

METHODS

Participants

Twenty-three HI listeners aged 60-78 yrs (mean: 67 yrs) participated in this study. They all had symmetrical, mild-to-moderate, sensorineural hearing losses. To facilitate the data analyses, these listeners were rank ordered according to their age followed by their average hearing loss across the audiometric frequencies of 0.5, 1, 2 and 4 kHz (4FAHL), and the notation “HI_{*n*}” was adopted to index the *n*-th listener.

In addition to 4FAHL, the average hearing loss across 0.125, 0.25, 0.5 and 0.75 kHz (4FAHL_{low}) and 4, 6 and 8 kHz (3FAHL_{high}) was calculated. The group means of the 4FAHL_{low}, 4FAHL, and 3FAHL_{high} measures were, respectively, 25, 41 and 60 dB HL (ranges: 6-49, 27-53, and 46-71 dB HL, respectively).

Eight NH listeners aged 26-44 yr (mean: 35 yr) also completed all auditory tests but, due to test protocol restrictions, not the cognitive ones.

Auditory predictors

To address the bottom-up processes involved in spatial speech recognition, two measures of auditory processing were used, one for each spatial dimension. The first

auditory measure was designed to gauge sensitivity to binaural TFS information and hence to predict L-R speech recognition. We adopted a measure devised by Ross *et al.* (2007) that allows determination of a listener's effective IPD frequency range (IPD_{FR}). More precisely, we used this measure to estimate the upper frequency limit for our listeners' ability to detect IPD changes of 180°.

The second auditory measure was designed to gauge sensitivity to pinna cues and hence to predict F-B speech recognition. We adopted a ripple phase-reversal paradigm devised by Supin *et al.* (1994) to quantify monaural spectral ripple discrimination (SRD). More precisely, we used this measure to determine the smallest spectral ripple spacing in a 4-8 kHz noise stimulus that our listeners were able to discriminate. Since pinna cues have ripple-like spectral patterns, we expected this measure to be a suitable, indirect index of pinna-cue sensitivity.

For both auditory measures, an adaptive 3-interval 3-alternative forced-choice procedure coupled with a 1-up 2-down staircase rule was applied.

Cognitive predictors

To address the top-down processes involved in spatial speech recognition, the two cognitive measures we previously had found to be correlated with spatial speech recognition (see above) were included. Specifically, the measure indexing working memory capacity was the reading span test (Daneman and Carpenter, 1980) and the measure indexing attention was the visual elevator test from the Test of Everyday Attention (Robertson *et al.*, 1996). Note that by using visually administered cognitive measures, we were able to decouple our estimates of the listeners' top-down skills from their auditory abilities.

Spatial speech recognition measurements

Spatial speech recognition was assessed with the help of a Danish multi-talker speech corpus (Behrens *et al.*, 2008). This corpus consists of a large set of Dantale II sentences (Wagener *et al.*, 2003) spoken by several trained female talkers that all follow the form “name verb numeral adjective object”. Using this corpus, speech recognition was measured in three (anechoic) test conditions: Co-located (CO), F-B, and L-R. In each condition, three speech signals were presented concurrently, one of them serving as target and the other two as maskers. In the CO condition, all speech signals came from the same, frontal loudspeaker. In the F-B condition, the target was presented from in front, while the two maskers were presented from a loudspeaker directly behind. In the L-R condition, the target was also presented from in front, while the two maskers were presented from loudspeakers located at ±45°. The target sentence was cued using its first word (a name), which was displayed on a computer screen located above the frontal loudspeaker. The words repeated by the listener were scored individually, and the obtained scores were used to estimate the target-to-masker ratios (TMRs) corresponding to 50%-correct speech recognition. The resulting estimates will be referred to as SRT_{CO}, SRT_{F-B}, and SRT_{L-R} for the CO, F-B, and L-R conditions, respectively.

Audibility criteria

On each listening task, the listeners were matched closely in terms of audibility. For the IPD_{FR} measure, a sensation level of 30 dB was used. For the SRD measure, a nominal presentation level of 75 dB SPL was chosen. The stimuli were then spectrally shaped such that their 1/3-octave band RMS spectra were at least 15 dB above hearing threshold. For the spatial speech recognition measurements, a nominal presentation level of 65 dB SPL for three equal-level speech signals presented simultaneously from one loudspeaker was chosen. The speech stimuli were then spectrally shaped such that their 1/3-octave band RMS spectra were at least 15 dB above hearing threshold for all frequencies up to 3 kHz. Due to the large dynamic range of the speech signals, the audibility criterion had to be lowered to at least 12 dB at 4 kHz, to at least 8 dB at 6 kHz, and to at least 4 dB at 8 kHz.

RESULTS AND DISCUSSION

Auditory and cognitive predictors

Fig. 1 displays the data from the auditory and cognitive predictors. Concerning the two auditory predictors, the data from the HI group exhibited more spread than those from the NH group. Furthermore, the HI group's performance on both the SRD [$t(24) = -5.5$, $p < 0.0001$] and the IPD_{FR} [$t(29) = 4.2$, $p < 0.001$] measure was significantly worse than that of the NH group (note that the SRD data from those five HI listeners with thresholds higher than 1 kHz were excluded from all statistical analyses, as they might have relied on overall level rather than spectral-ripple cues to complete the task; cf. Supin *et al.*, 1994).

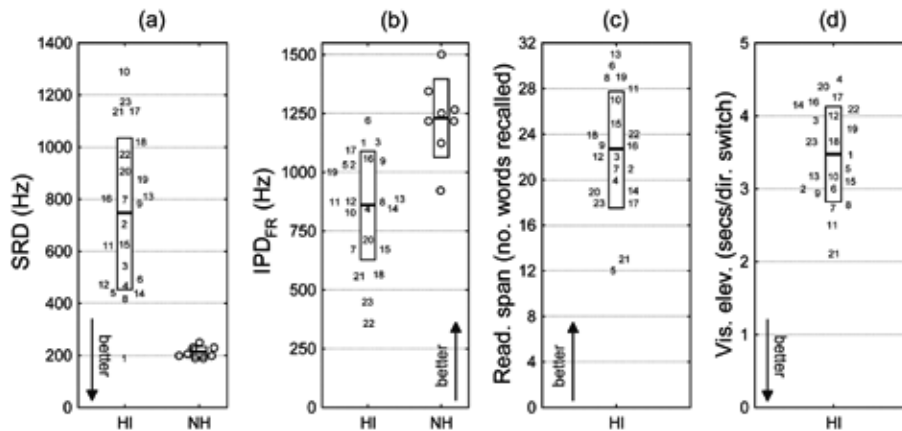


Fig. 1: (a) SRD, (b) IPD_{FR} , (c) reading span, and (d) visual elevator data. HI listeners are denoted by numbers and NH controls by circles. Black bars correspond to group means. Boxes represent ± 1 standard deviation (SD). Arrows indicate direction of better performance.

Concerning the two cognitive predictors, in terms of average performance the data obtained agreed well with those collected as part of our previous study (Neher *et al.*, 2009). Nevertheless, the current study's participants performed in a more homogeneous manner, probably because they covered a smaller age span than the group tested previously (60-78 yrs vs. 28-84 yrs). It is also interesting to note that one listener (HI₂₁) obtained the second lowest (i.e. poorest) score on the reading span test, while at the same time obtaining the lowest (i.e. best) score on the visual elevator test. This finding hints at some unexpected effects in the cognitive data, which will be addressed further below.

Spatial speech recognition measurements

Fig. 2 displays the data from the competing-speech tasks. The poorest performance was observed in the CO condition, while the F-B and L-R conditions were characterized by better performance, especially for the NH group. Furthermore, the F-B and L-R data exhibited more spread than the CO data, especially for the HI group. A repeated-measures analysis of variance with test condition as within-subject factor [$F(2, 58) = 236.5$, $p < 0.00001$] and listener group as between-subject factor [$F(1, 29) = 150.7$, $p < 0.00001$] showed both main effects to be significant. The same was true for their interaction [$F(2, 58) = 46.3$, $p < 0.00001$]. According to a Scheffé *post hoc* analysis, all within-listener group means were significantly ($p < 0.00001$) different from each other, as were the across-group means for the F-B and L-R (but not the CO) test condition.

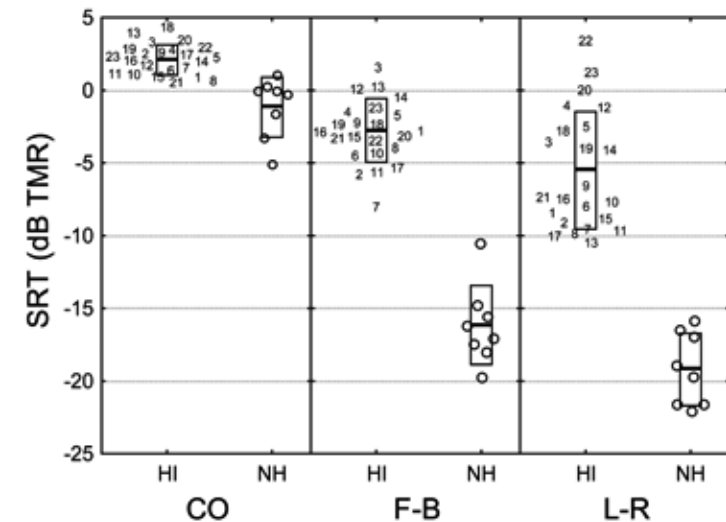


Fig. 2: Speech recognition results for the CO, F-B, and L-R test condition. HI listeners are denoted by numbers and NH controls by circles. Horizontal black bars correspond to group means. Boxes represent ± 1 SD.

Inter-correlation analysis of predictor data

Before trying to model F-B and L-R speech recognition performance, the predictors were examined by means of a product-moment correlation analysis. Only two significant correlations were observed: between age and SRD ($r = 0.70$, $p = 0.0012$), and age and IPD_{FR} ($r = -0.67$, $p = 0.0005$). In contrast, neither SRD and $3FAHL_{high}$ nor IPD_{FR} and $4FAHL_{low}$ were significantly correlated. Thus, these results do not lend support to our hypotheses that elevated high- and low-frequency hearing thresholds lead to poorer SRD and binaural TFS sensitivity (see above). Instead, they imply that higher age impairs a listener's ability to resolve high-frequency monaural spectral details and to detect IPD changes over a wide frequency range.

Importantly, none of the auditory measures was correlated with any of the cognitive measures. This finding implies that our estimates of the listeners' auditory skills were not confounded with their cognitive abilities. Thus, we considered them to be suitable predictors of bottom-up contributions to spatial speech recognition.

It should also be noted that the two cognitive measures were not correlated with age. This finding, which is at odds with previous research that has shown cognitive function to decline with age (e.g. Salthouse, 1982), suggests that our HI group was somewhat atypical as far as cognitive aging effects are concerned.

Spatial speech recognition models

To test the ability of our auditory and cognitive measures to predict the HI listeners' spatial speech recognition performance, product-moment correlation coefficients were calculated between (i) SRT_{F-B} and $3FAHL_{high}$, SRD, and the two cognitive measures, and (ii) SRT_{L-R} and $4FAHL_{low}$, IPD_{FR} , and the two cognitive measures. Unexpectedly, the F-B predictors were basically unable to predict SRT_{F-B} . It therefore seems that the SRD measure failed to tap the skills required for discriminating pinna cues. Furthermore, matching the listeners closely in terms of audibility might have led to the removal of any inter-listener differences that previously had given rise to a correlation between SRT_{F-B} and high-frequency hearing thresholds (see above).

Concerning L-R spatial speech recognition, however, a correlation pattern was obtained that agreed well with previous findings (Lunner *et al.*, 2010; Neher *et al.*, 2009). In particular, SRT_{L-R} was found to be significantly correlated with $4FAHL_{low}$, IPD_{FR} , and the visual elevator test (see Table 1).

	$4FAHL_{low}$	IPD_{FR}	Read. span.	Vis. elev.
SRT_{L-R}	0.52**	-0.47*	-0.36	0.60**

Table 1: Product-moment correlation coefficients for the HI listeners' SRT_{L-R} , $4FAHL_{low}$, IPD_{FR} , reading span, and visual elevator data (* $p < 0.05$, ** $p < 0.01$).

To obtain an indication concerning the statistical independence of these effects, multiple-regression analyses were carried out. For reasons of statistical rigor, we restricted our analyses to simple models based on no more than two predictors. In order of variance accounted for, the three models tested were (i) the visual elevator test and IPD_{FR} ($R^2 = 0.60$, adjusted $R^2 = 0.56$, all $p < 0.003$), (ii) the visual elevator test and $4FAHL_{low}$ ($R^2 = 0.47$, adjusted $R^2 = 0.42$, all $p < 0.045$), and (iii) $4FAHL_{low}$ and IPD_{FR} ($R^2 = 0.39$, adjusted $R^2 = 0.33$, all $p < 0.048$). Since in each case both predictors were found to be significant, one could interpret these results as suggesting three bottom-up or top-down effects in SRT_{L-R} performance: (i) the relation to $4FAHL_{low}$ might indicate a more peripheral (cochlear) effect; (ii) the relation to IPD_{FR} might indicate a more central auditory effect; and (iii) the relation to the visual elevator measure might indicate a cognitive effect. Further research would have to be conducted to completely establish the independence of these effects as well as to determine their causality.

SUMMARY

The purpose of the present study was to further our understanding of the supra-threshold processes involved in spatial speech recognition. The HI listeners tested as part of this study differed widely in terms of their performance on a number of auditory and cognitive predictors as well as on two competing-speech tasks featuring spatial complexity. Concerning performance on the predictor measures, the analyses indicated a negative effect of age on the HI listeners' ability to detect spectral ripple and IPD changes. The same was not true for the listeners' hearing thresholds, however. Concerning performance on the competing-speech tasks, the chosen auditory and cognitive measures generally failed to predict recognition of a frontal speech target presented against two speech maskers located directly behind the listener; when the two speech maskers were located at $\pm 45^\circ$, however, the analyses suggested effects of attention, the frequency range over which the listeners were sensitive to IPD changes, and average low-frequency hearing thresholds. Altogether, these findings lend credence to the involvement of both auditory and cognitive factors in spatially complex, multi-talker speech recognition tasks.

REFERENCES

- Akeroyd, M.A. (2008). “Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults” *Int. J. Audiol. Suppl.* **2**, 47, S53-S73.
- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (The MIT Press, Cambridge, MA).
- Behrens, T., Neher, T., and Johannesson, R.B. (2008). “Evaluation of speech corpus for assessment of spatial release from masking” in *Proceedings of ISAAR 2007: Auditory Signal Processing in Hearing-Impaired Listeners*. 1st International Symposium on Auditory and Audiological Research. Elsinore, Denmark. Edited by T. Dau, J.M. Buchholz, J. Harte, and T.U. Christiansen. ISBN: 87-990013-1-4. (The Danavox Jubilee Foundation, Copenhagen), pp. 449-457.
- Daneman, M., and Carpenter, P.A. (1980). “Individual differences in integrating information between and within sentences” *J. Exp. Psychol. Learn. Mem. Cogn.* **9**, 561-584.
- Lunner, T., Neher, T., Hopkins, K., and Moore, B.C.J. (2010). “Sensitivity to low-frequency temporal fine structure is correlated with aided spatial release from masking” presented at the 2010 International Hearing Aid Research Conference (IHCON), Lake Tahoe, California, Aug. 11-15.
- Marrone, N.L., Mason, C.R., and Kidd, G.Jr. (2008). “The effects of hearing loss and age on the benefit of spatial separation between multiple talkers in reverberant rooms” *J. Acoust. Soc. Am.* **124**, 3064-3075.
- Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A.S., and van Schaik, A. (2009). “Benefit from spatial separation of bilateral hearing-aid users: Effects of hearing loss, age, and cognition” *Int. J. Audiol.* **48**, 758-774.
- Neher, T., Laugesen, S., Jensen, N.S., and Kragelund, L. (2011). “Can basic auditory and cognitive measures predict hearing-impaired listeners’ localization and spatial speech recognition abilities?” *J. Acoust. Soc. Am.* **130**, 1542-1558.
- Robertson, I.H., Ward, T., Ridgeway, V., and Nimmo-Smith, I. (1996). “The structure of normal human attention: The Test of Everyday Attention” *J. Int. Neuropsychol. Soc.* **2**, 525-534.
- Ross, B., Tremblay, K.L., and Picton, T.W. (2007). “Physiological detection of interaural phase differences” *J. Acoust. Soc. Am.* **121**, 1017-1027.
- Salthouse, T. (1982). *Adult Cognition* (Springer Verlag, New York).
- Strelcyk, O., and Dau, T. (2009). “Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing” *J. Acoust. Soc. Am.* **125**, 3328-3345.
- Supin, A.Y., Popov, V.V., Milekhina, O.N., and Tarakanov, M.B. (1994). “Frequency resolving power measured by rippled noise” *Hear Res.* **78**, 31-40.
- Wagener, K., Josvassen, J.L., and Ardenkjær, R. (2003). “Design, evaluation and optimization of a Danish sentence test in noise” *Int. J. Audiol.* **42**, 10-17.

Measures of the ecological loudness of speech

MARY FLORENTINE^{1,3,4} AND MICHAEL EPSTEIN^{2,3,4}¹ *Communication Research Laboratory*² *Auditory Modeling and Processing Laboratory*³ *Dept. of Speech-Language Pathology and Audiology and Institute for Hearing, Speech, and Language*⁴ *Communications and Digital Processing Center, ECE Department Northeastern University, Boston, Massachusetts, 02115 USA*

Most laboratory studies of binaural loudness summation show ample amounts of summation (e.g., a tone presented binaurally is clearly louder than the same tone presented monaurally), but classroom demonstrations of this phenomenon in typical daily environments yield negligible loudness summation for most listeners. To gain insight into this difference, experiments were performed with different degrees of ecological validity. Statistical analysis indicates that the most ecologically valid condition results in less binaural loudness summation than traditional laboratory procedures. Implications for normal-hearing listeners and impaired listeners with hearing aids are discussed.

INTRODUCTION

Motivation for the present work came from the observation that classroom demonstrations of binaural loudness summation never yielded the magnitude of the effect that was reported in the literature. According to the literature, most earphone studies suggest binaural-to-monaural loudness ratios ranging from about 1.3 to 1.7, or almost 2 (Reynolds and Stevens 1960; Scharf and Fishken 1970; Marks 1978; Hellman 1991; Zwicker and Zwicker 1991; Schneider and Cohen 1997; Marozeau *et al.* 2006; Whilby *et al.* 2006; Epstein and Florentine 2009).

CLASSROOM DEMONSTRATIONS

In real-world classroom demonstrations of binaural loudness summation—first performed in 1975 at the Acoustics Laboratory of the Technical University of Denmark—a lecturer asked her students to estimate the loudness of sounds while they sat in their usual seats in a typical classroom. She recited memorized passages while attempting to keep her voice at a constant level that was typical for her lectures. The students’ task was to look at her and estimate the loudness of her voice while listening with both ears compared to the loudness of her voice while they were blocking one ear by pressing on a tragus with an index finger. They were encouraged to make several observations for each of the two conditions before making a judgment. The students’ subjective reports indicated that the loudness of speech changed a negligible amount, if at all. This phenomenon has been dubbed