Validation of a spatial speech-in-speech test that takes Signal-to-Noise Ratio (SNR) confounds into account

Søren Laugesen, Filip Marchman Rønne, Niels Søgaard Jensen, Maria Grube Sorgenfrei
Eriksholm Research Centre, Rørtangvej 20, DK-3070 Snekkersten, www.eriksholm.com

Contact: Søren Laugesen, slu@eriksholm.com

Presented at the 4th International Symposium on Auditory and Audiological Research (ISAAR), Nyborg Strand, Denmark, August 28-30, 2013

The proposed Spatial Fixed-SNR (SFS) test was used to compare a linear hearing-aid setting to a setting with aggressive compression limiting. Two sub-groups of listeners were tested in a fixed-SNR paradigm at −5 and +5 dB SNR, respectively.

Introduction

Adaptive Speech-Reception Threshold (SRT) measures are popular for good reasons. However, the unbounded nature of the SNR at which the SRT is achieved often leads to a wide spread in SRT [1], certainly for aided hearing-impaired listeners. Thus, if hearing aids are under test they will be subjected to very different SNRs among listeners. This has the possibility of causing SNR confounds which may lead to faulty conclusions [2,3]. Furthermore, the SRT is often much lower than the SNR found in realistic listening conditions [4,5], particularly when testing normal-hearing listeners. This may compromise the ecological validity of the test results.

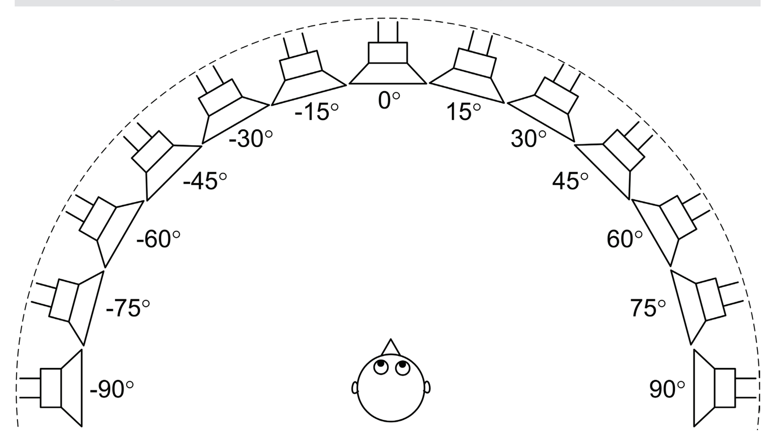
To address these problems, the SFS (Spatial Fixed-SNR) speech-in-speech intelligibility test is proposed, which uses a fixed-SNR paradigm. Percent-correct scores within the informative 20-90% range are obtained for the individual by selecting among four test conditions with different test difficulty [6]. Thus, the SFS test is aimed at within-subjects comparisons. As an option, Target Location Uncertainty (TGLU) can be added to the test [7].

Aim of the study

To validate the SFS test with emphasis on:

1. Do the four SFS conditions change test difficulty as expected?
2. What is the reliability of the SFS test?
3. Can the SFS test measure an expected contrast?
4. Does TGLU provide any added insight?

Figure 1. Sketch of the loudspeaker setup.



Method and material

Target speech was the Danish HINT corpus [8], played at 70 dB SPL (C). The masker speech signals were recordings of speakers reading from a fairy tale: two females and two males, used in pairs arranged symmetrically around the listener, see Figure 1. Target and maskers were all spectrally matched to a female reference spectrum and masker speech pauses were cut down to 65 ms. The four SFS test conditions are outlined in Table 1. In all non-TGLU conditions the target came from 0°. All reported SNR values are referenced to the centre of the loudspeaker ring with the listener absent. The test protocol is outlined in Table 2.

Table 1. SFS test conditions. *30mW* is the baseline. Use of TGLU is optional.

Label	Masker gender	Scoring rule	Masker positions	Nominal SRT shift	TGLU locations
15mS	Male	Sentence	±15°	+5.0 dB	0°, ±30°
30mS	Male	Sentence	±30°	+2.5 dB	0°, ±60°
30mW	Male	Word	±30°	0 dB	0°, ±60°
45fW	Female	Word	±45°	-2.5 dB	0°, ±90°

Listeners

N = 26 hearing-impaired listeners with sensorineural and mixed hearing loss took part. PTA (Pure Tone Average hearing threshold levels at 0.5, 1, 2, and 4 kHz) ranged from 29 dB to 66 dB, with a mean value of 46 dB. Subjects were listening binaurally through Oticon Agil Pro miniRITE (Receiver In The Ear) hearing aids, fitted with closed 'power domes'. Directionality and noise management were disabled.

Table 2. Validation study protocol. The order of HAsettings, the order of paradigms in trial pairs (9,10) and (13,14), as well as use of HINT test lists were balanced across listeners.

Visit	Trial	SFS condition	Trial type	HAsetting	Paradigm	#HINTlists
1	1	*30mW*	Training	LIN	Adaptive SNR	1T
	2		Training-TGLU		Adaptive SNR	1T
	3		Baseline		Adaptive SNR	2
	4	Individual	Test-TGLU		Fixed SNR	3
	Break					
	5	*30mW*	Training	CLM	Adaptive SNR	1T
	6		Baseline		Adaptive SNR	2
	7	Individual	Test-TGLU		Fixed SNR	3
Between visits, about 1½ weeks						
2	8	Individual	Training	LIN	Adaptive SNR	2T
	9		Test		Adaptive SNR	2
	10		Test		Fixed SNR	2
	11		Retest		Adaptive SNR	1
	Break					
	12	Individual	Training	CLM	Adaptive SNR	1T
	13		Test		Fixed SNR	2
	14		Test		Adaptive SNR	2
	15		Retest		Fixed SNR	1

Experimental contrast

A comparison was made between an individually prescribed linear hearing-aid setting (LIN), and a setting with 8 dB additional gain and fast-acting compression limiting (CLM). Frequency-specific limiting thresholds were individually set to the expected output from the LIN setting with a representative SFS-test input signal.

Fast-acting compression changes the longterm SNR at the output, depending on input SNR and signal characteristics [3]. With the present SFS-test signals and the CLM setting, a SNR decrease of about 2 dB was found for an input SNR of +5 dB (that is, ΔSNR = −2 dB), while a ΔSNR of about +2 dB was found for an input SNR of −5 dB. This indicates a potential SNR confound in the LIN/CLM comparison. Thus, in an attempt to demonstrate a SNR confound, the listeners were assigned to one of two SNRgroups with target SNRs of −5 dB and +5 dB, respectively, depending on their baseline performance.

Results

THE EFFECT OF THE SFS CONDITIONS is evaluated by examining the differences between baseline and adaptive-SNR test conditions (trial pairs 3-9 and 6-14 in Table 2). The results are corrected for a 0.3 dB between-visit training effect [6], while the within-visit training effects are assumed to be balanced out. The results in Figure 2 show that the measured SRT shifts for the 45fW and 30mS conditions are very close to the nominal values in Table 1, while the 15mS condition is 1 dB below. The standard deviations are on par with the results from [6], except for the 45fW condition, which exhibits more variability. These deviations from [6] are yet unexplained.

TEST RELIABILITY is evaluated by directly computing test-retest standard deviations (SD). First, the variances of the difference measure (trial pair 9-11 for the adaptive-SNR paradigm) is found as

$$V_{\Delta\text{adaptive-SNR}} = \frac{1}{N}\sum_{n=1}^{N}(\text{SRT}_{\text{test}} - \text{SRT}_{\text{retest}})^2$$

The test-retest SD of a single measurement is then

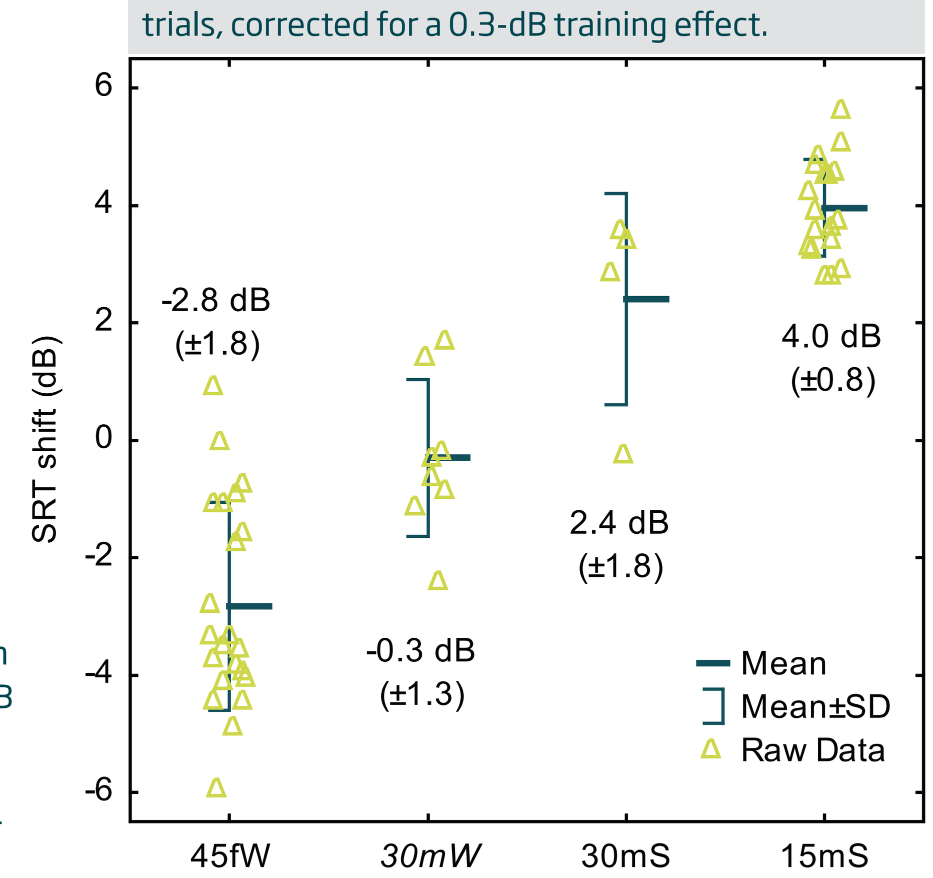
$$SD_{\text{adaptive-SNR}} = \sqrt{\tfrac{1}{2}V_{\Delta\text{adaptive-SNR}}} = 0.95\ \text{dB}$$

Similarly, the test-retest SD for the fixed-SNR paradigm is

$$SD_{\text{fixed-SNR}} = \sqrt{\tfrac{1}{2}V_{\Delta\text{fixed-SNR}}} = 8.6\%$$

For comparison, a value of 0.92 dB was found for the standard Danish HINT with hearing-impaired listeners [8]. In order to evaluate $SD_{\text{fixed-SNR}}$ the above value is converted to dB by means of the slope of the test's estimated psychometric function [6]. This yields 8.6%/13.7%/dB = 0.63 dB (best case), which indicates that reliability potentially is better for the fixed-SNR paradigm.

Figure 2. Effect on SRT between baseline and test trials, corrected for a 0.3-dB training effect.



Results, continued

MEASURES OF THE EXPERIMENTAL CONTRAST are shown in Figure 3 and 4 for the adaptive-SNR and fixed-SNR paradigms. The results from each paradigm and trial type were analysed with separate mixed-model ANOVAs. In all three models the random listener effect was highly significant (all $p < 0.0001$). In each plot, p-values for the main effect of HAsetting and the HAsetting*SNRgroup interaction are stated.

The results in Figure 3 shows that by using the four SFS conditions, the measured SRTs in the test trials (9,14, right panel) are forced further apart and closer to the target SNRs than in the baseline (trials 3,6, left panel). Further, the adaptive-SNR paradigm was able to detect the overall effect of HAsetting with LIN giving about 1 dB lower (better) SRTs than CLM. There is, however, no sign of the projected SNR confound, since the HAsetting*SNRgroup interaction is not significant.

Figure 3. Mean SRTs for the HAsetting*SNRgroup interaction, adaptive SNR, baseline and test trials.

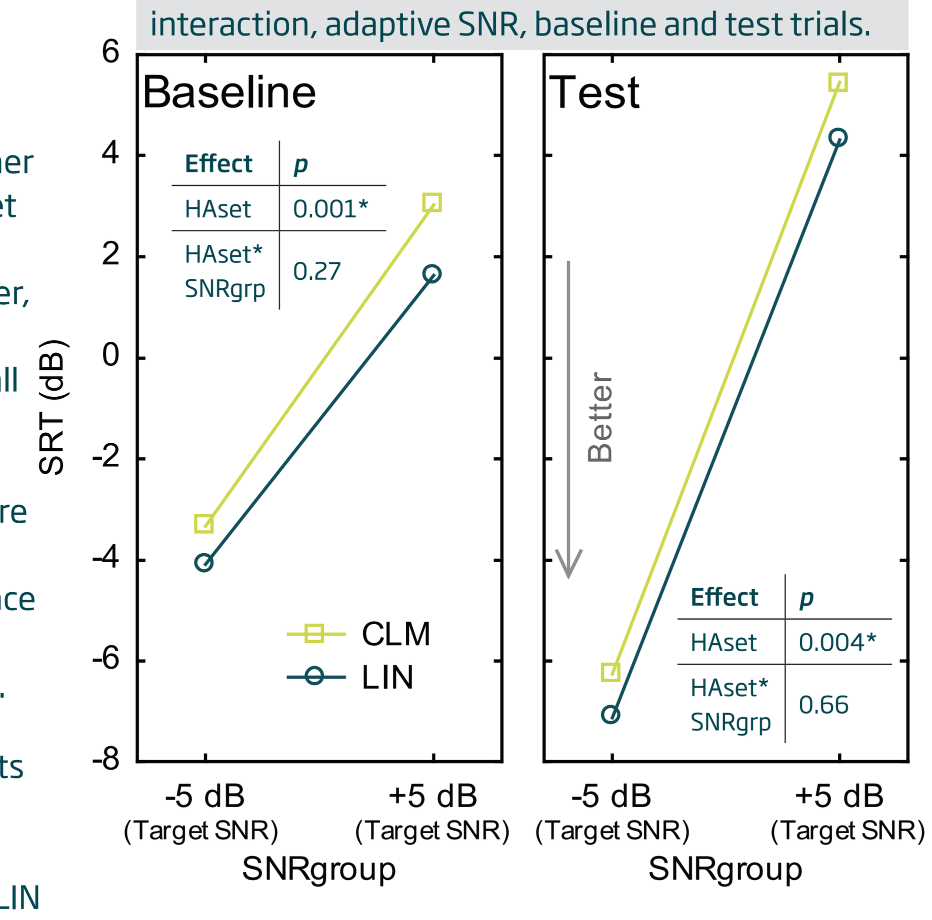


Figure 4. As Figure 3, fixed-SNR paradigm.

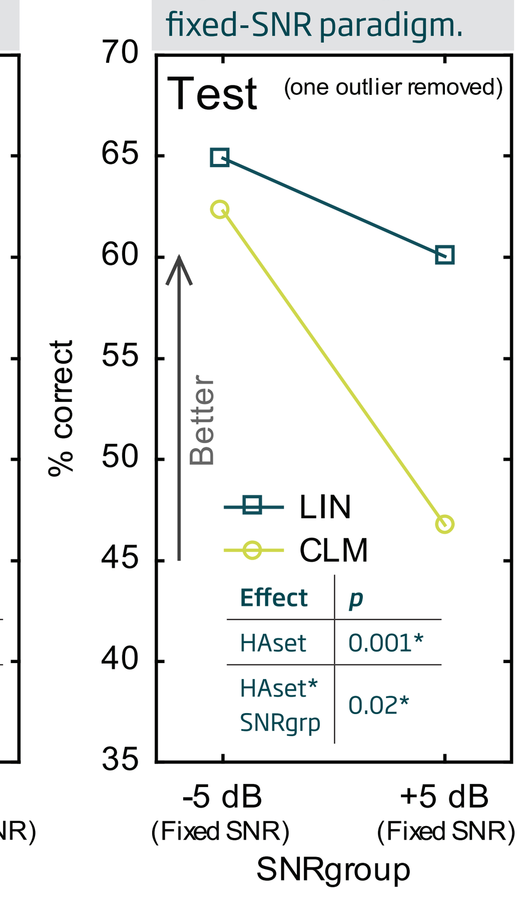


Figure 4 shows similar results for the fixed-SNR paradigm. Again a main effect of HAsetting is observed with LIN giving best performance. In addition, the fixed-SNR results show a significant HAsetting*SNRgroup interaction. Note that the mean %-correct values are between 45 and 65 (individual data range 10-84%), which is considered very satisfactory.

Target location uncertainty (TGLU)

The mean %-correct scores for the three TGLU target locations are shown in Figure 5, together with the corresponding mean value from the non-TGLU trials. A planned comparison between the two centre-location means indicates that the addition of TGLU makes the test more difficult, by about 4% (p = 0.03), in agreement with [7]. The results also indicate better speech recognition from the left and right targets compared to TGLU centre. When the target comes from the side locations, the SNR at the hearing aid becomes about 3 dB better at the ipsilateral side, due to the acoustic baffle effect of the head. If better-ear SNR alone governed speech recognition, this should translate to 30-40% better scores. The fact that the left-right improvement is only about 5% indicates that the target locations to the side are in fact associated with increased difficulty. This could be due to an attention bias towards the centre loudspeaker or loss of truly binaural spatial unmasking.

Figure 5. Mean %-correct values per target location, non-TGLU (10,13) and TGLU trials (4,7).

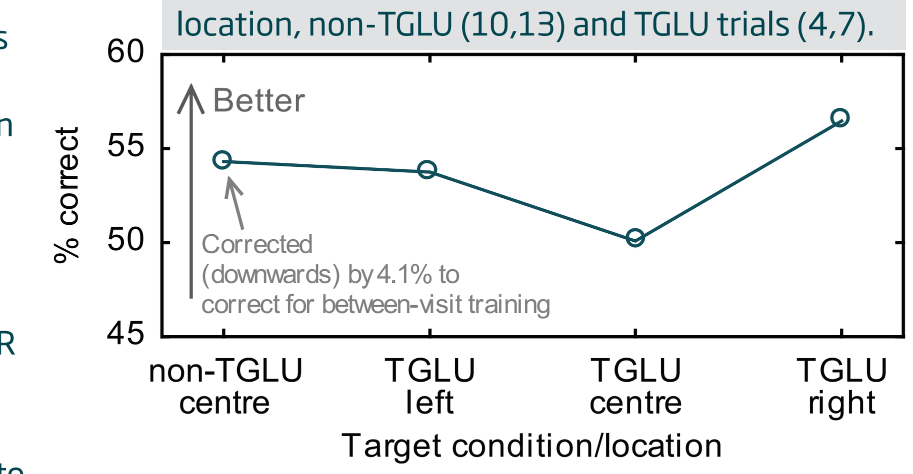


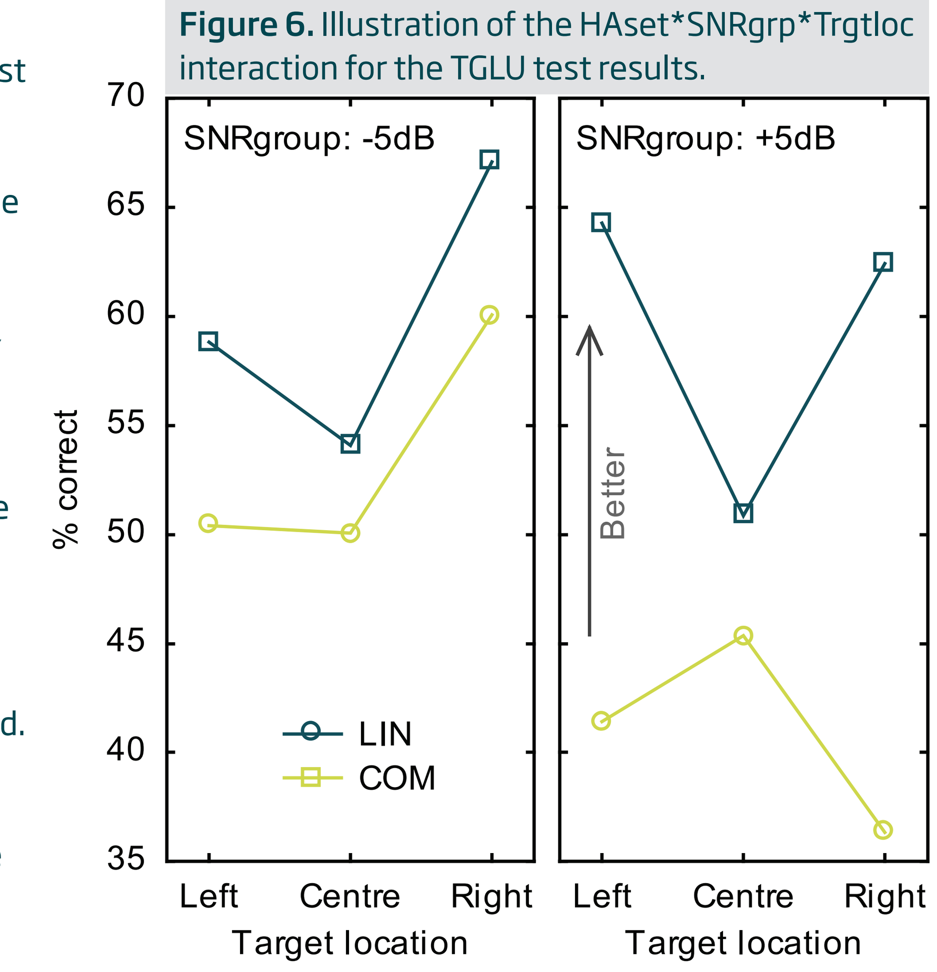
Table 3 shows the results of a repeated-measures ANOVA with SNRgroup as a between-subjects factor. The absence of a main effect of SNRgroup confirms the success of using the SFS conditions to align the listeners' %-correct scores. The TGLU results regarding the HAsetting main effect and the HAsetting*SNRgroup interaction confirms the observations from the fixed-SNR non-TGLU results in Figure 4, with similar statistical figures. This indicates that adding TGLU to the SFS test can be done without sacrificing validity.

Table 3. Repeated-measures ANOVA results regarding the TGLU test trials (4,7).

Effect	F-test	p-value
HAsetting	$F(1,24)=21$	0.0001*
SNRgroup	$F(1,24)=1.1$	0.3
Target location	$F(2,48)=3.7$	0.03*
HAsetting*SNRgroup	$F(1,24)=4.7$	0.04*
HAsetting*Trgtloc	$F(2,48)=7.1$	0.002*
SNRgroup*Trgtloc	$F(2,48)=3.9$	0.03*
HAset*SNRgrp*Trgtloc	$F(2,48)=3.4$	0.04*

The three remaining significant interactions in Table 3 suggest additional insights due to adding TGLU. As an example, the HAset*SNRgrp*Trgtloc three-way interaction is illustrated in Figure 6. This result indicates that in the +5-dB SNRgroup, the LIN/CLM difference is more pronounced for the left and right targets compared to the centre target. This could be due to a distortion effect of the very aggressive compression limiting, which causes a disbenefit of CLM. As mentioned, the left and right targets are 3 dB louder at the ipsilateral ear. In the +5-dB SNRgroup the target dominates and increased target level means increased overall level, which in turn means more distortion and additional disbenefit with CLM.

Figure 6. Illustration of the HAset*SNRgrp*Trgtloc interaction for the TGLU test results.



Discussion

THE EXPERIMENTAL CONTRAST did not behave as expected. The objectively measured 4-dB swing in ΔSNR between the LIN and the CLM settings had a much smaller impact on the perceptual data, which furthermore was visible only with the fixed-SNR paradigm (compare Figures 3 and 4). The reason for this difference between the paradigms is currently under investigation.

Conclusion

The Spatial Fixed-SNR (SFS) test was validated. With regard to the study's aims it was shown that:

1. The four proposed SFS conditions were able to change test difficulty, such that informative %-correct data could be measured for all listeners at the target SNRs.
2. The reliability of the SFS test was found to be on par with the standard HINT test, and possibly even better when used with the fixed-SNR paradigm.
3. The SFS test was able to detect relevant differences between the tested LIN and CLM hearing-aid settings with high statistical significance.
4. Target location uncertainty (TGLU) appears to offer additional insights (at the cost of extra testing time).

REFERENCES [1] Naylor G (2010). Limitations of Speech Reception Threshold (SRT) as an outcome measure in hearing-aid research. International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA, oral presentation. [2] Bernstein JGW (2011). Controlling signal-to-noise ratio effects in the measurement of speech intelligibility in fluctuating maskers, In: Dau T, Jepsen ML, Poulsen T, Dalsgaard JC. (Eds.) Speech Perception and Auditory Disorders. Proceedings of the Third International Symposium on Audiological and Auditory Research, ISAAR. Centertryk A/S: Denmark. [3] Naylor G, Burmand Johannesson R (2009). Long-term signal-to-noise ratio at the input and output of amplitude-compression systems. J. Am. Acad. Audiol. 20, p. 161-171. [4] Pearsons KS, Bernett RL, Fidell S (1977). Speech levels in various noise environments. Bolt, Beranek and Newman. Inc. Canoge Park, California. [5] Smeds K, Wolters F, Rung M (2012). Realistic signal-to-noise ratios. International Hearing Aid Research Conference (IHCON), Lake Tahoe, CA, USA, poster presentation. [6] Rønne FM, Laugesen S, Jensen NS, Hietkamp RK, Pedersen JH (2013). Magnitude of speech-reception-threshold manipulators for a spatial speech-in-speech test that takes signal-to-noise ratio confounds and ecological validity into account. ICA 2013 Montreal, Canada, vol. 19, 1-8. [7] Jensen NS, Burmand Johannesson R, Laugesen S, Hietkamp RK (2011) Measuring speech-in-speech intelligibility with target location uncertainty, In: Dau T, Jepsen ML, Poulsen T, Dalsgaard JC. (Eds.) Speech Perception and Auditory Disorders. Proceedings of the Third International Symposium on Audiological and Auditory Research, ISAAR. Centertryk A/S: Denmark. [8] Nielsen JB, Dau T (2011). The Danish hearing in noise test. Int. Journ. Aud. 50, p. 202-208.

Eriksholm Research Centre
PART OF OTICON