Investigating the relationship between spectro-temporal modulation detection, aided speech perception, and directional noise reduction preference in hearing-impaired listeners

JOHANNES ZAAR[1,*], LISBETH BIRKELUND SIMONSEN[2], THOMAS BEHRENS[3], TORSTEN DAU[1] AND SØREN LAUGESEN[2]

[1] *Hearing Systems Section, Department of Health Technology, Technical University of Denmark, DK-2800 Lyngby, Denmark*

[2] *Interacoustics Research Unit, DK-2800 Lyngby, Denmark*

[3] *Oticon A/S, DK-2765 Smørum, Denmark*

In analogy to the restoration of reduced audibility via hearing-aid amplification, supra-threshold speech processing deficits may be partially compensated for by using state-of-the-art directional noise reduction (NR) techniques. However, while amplification is usually prescribed based on classical audiometry, a clinical test that represents supra-threshold speech processing and is thus useful for prescribing NR settings is yet to be established. The present study explored the potential of a suitably adapted spectro-temporal modulation detection (STMD) test for this purpose by means of laboratory-based tests and field tests with 30 hearing-impaired participants. In particular, it was investigated whether STMD performance (i) predicts aided speech intelligibility measured in a spatial multi-talker set up with different degrees of NR and (ii) predicts preference for moderate vs. aggressive NR. STMD thresholds were strongly correlated with (i) speech scores measured without NR, (ii) speech intelligibility benefit induced by aggressive NR, and (iii) the individual participants' NR preference. The latter relationship was mediated by performance in a reverse digit span task, which measures working memory capacity. Overall, the results suggest that a clinical test that assesses STMD sensitivity may be useful for prescribing NR settings in hearing-aid fitting.

INTRODUCTION

Hearing-aid amplification is typically tailored to the individual's hearing loss based on the pure-tone audiogram to restore audibility. However, some individuals experience severe supra-threshold difficulties with speech understanding in adverse conditions that cannot be resolved by audibility compensation. For those, additional help may be provided in the form of minimum variance distortionless response beamforming combined with single-channel noise reduction (MVDR-NR, in the following simply termed NR). Recent advances have yielded the possibility to substantially improve speech intelligibility (SI) and reduce listening effort – at least

*Corresponding author: jzaar@dtu.dk

Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR), Vol. 7: Auditory Learning in Biological and Artificial Systems, August 2019, Nyborg, Denmark. Edited by A. Kressner, J. Regev, J. C.-Dalsgaard, L. Tranebjærg, S. Santurette, and T. Dau. The Danavox Jubilee Foundation, 2019. ISSN: 2596-5522.

in laboratory-based scenarios – when using aggressively parametrized NR. However, these improvements may come at the cost of an impaired perceived naturalness of the sound scene and its acceptance may therefore be highly listener-specific. To use its full potential, NR thus needs to be carefully tailored to the individual, such that aggressive NR settings are only prescribed to those who truly need and therefore tolerate them.

To this end, a clinically viable measure that represents supra-threshold speech processing is required to identify listeners with severe supra-threshold deficits, who might benefit from aggressive NR. Bernstein *et al.* (2013) employed a spectro-temporal modulation detection (STMD) paradigm to assess such deficits in normal-hearing (NH) and HI listeners. They used broadband (354-5656 Hz) noise carrier signals modulated with various STM patterns (specified by spectral modulation rate in cycles/octave, c/o, and temporal modulation rate in Hz), generating spectral ripples that move upward or downward as a function of time. Bernstein *et al*. (2013) found a significant NH vs. HI difference for the combination of 2 c/o and 4 Hz, which has henceforth been widely used. Furthermore, Bernstein *et al*. (2013) and Mehraei *et al.* (2014) demonstrated that STMD performance was strongly correlated with speech-in-stationary-noise performance in HI listeners, measured at very high presentation levels but without individualized amplification. Bernstein *et al*. (2016) measured STMD performance in HI listeners using a bandlimited noise carrier (354-2000 Hz) with 2 c/o and 4 Hz. STMD thresholds were compared to speech reception thresholds (SRTs) measured in stationary noise and multi-talker babble with simulated hearing-aid processing (i.e., aided). While they found a significant correlation between STMD thresholds and SRTs, the relationship was not as strong as in previous studies, possibly because many listeners could not reach the required detection accuracy, such that thresholds could not be directly measured and instead had to be extrapolated. Introducing various changes to the measurement procedure, such as listener-specific frequency-dependent amplification, increased stimulus duration, and bilateral presentation mode, the authors of the current study proposed an STMD measurement procedure that all tested HI listeners were sensitive to (Zaar *et al*., 2018). Furthermore, the study demonstrated that STMD performance was strongly associated with speech-in-noise performance, measured in co-located stationary noise as well as in a spatial multi-talker set-up with audibility compensation.

Based on the observations and findings described above, the goal of the present study was to explore whether STMD performance, as measured in Zaar *et al*. (2018), is indicative of supra-threshold speech processing deficits in HI individuals and thus predictive of the individuals' preference in terms of NR settings. The following three research questions (RQs) were addressed:

[RQ1] Does STMD performance predict aided SI without NR?

[RQ2] Does STMD performance predict SI benefit offered by aggressive NR?

[RQ3] Does STMD performance predict NR preference?

METHODS

Participants, hearing-aid fitting and NR settings

30 HI participants (mean age: 70.2 years, standard deviation: 9.1 years) were recruited, all of whom were native speakers of Danish and regular hearing-aid users. All participants underwent audiometric screening. Hearing aids were fitted based on the individuals' audiograms using the standard prescription offered by the fitting software. Three NR settings were defined: "*Off*" (NR algorithm inactive, hearing-aid directivity pattern in omni-directional mode), "*Default*" (mode-rate parametrisation of the NR algorithm); "*FullThrottle*" (custo-mized aggressive NR setting).

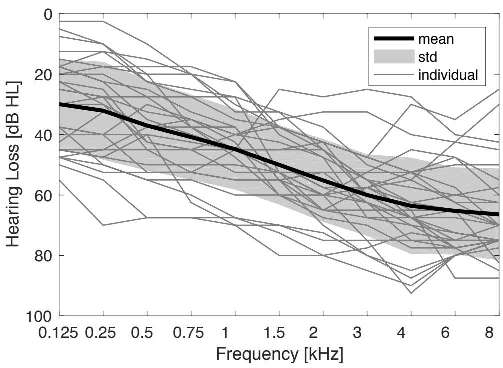

Fig. 1: Pure-tone thresholds (averaged across ears) for all participants (thin grey lines) and on average (thick black line).

Reverse digit span (RDS) test

A measure of working memory capacity was obtained using the reverse digit span (RDS) test, where randomly selected Danish digits from "1" to "9" were presented to the participants at their self-adjusted most comfortable level over Sennheiser HDA200 headphones in a sound attenuated booth. Initially, two digits were presented in each trial. The number of digits per trial was then increased by one after each second trial. The procedure ended after two incorrect responses or after 14 trials (with maximally 8 digits per trial). The participants were required to type the digits they had heard in reverse order on a computer keyboard. In each trial, two points could be obtained, one for the correct number of repeated digits and one for the correct placement of the digits. The maximum possible cumulative score was thus 28.

Spectro-temporal modulation detection (STMD) test

STMD thresholds were adaptively measured using a three-alternative forced choice procedure with a one-up/two-down tracking rule (approaching 70.7 percent correct). The STM stimulus was generated by modulating a bandlimited noise carrier (354-2000 Hz, 1000 log-spaced random-phase sinusoidal components per octave) with an upward moving ripple pattern defined by spectral and temporal modulation rates of 2 c/o and 4 Hz, respectively. In each trial, the two noise-only intervals contained an identical realization of the carrier noise and the signal interval contained the same carrier noise with the modulation imposed on it. The STM starting phase in the modulated stimulus was randomized across trials. The stimulus duration was 1 s with 500 ms inter-stimulus intervals. The modulation depth was considered as the tracking variable, which started at 0 dB (full modulation). The initial step size was 4 dB, which was halved after the first and again after the second upper reversal. After 8 reversals at a step size of 1 dB the procedure terminated. The threshold was calculated as the mean across the tracking variable at the 8 last reversals. The participants were seated in a sound attenuating booth in front of a computer screen and bilaterally presented with the stimuli using Sennheiser HDA200 headphones. The nominal presentation level was set at a sound pressure level (SPL) of 65 dB and ear-specific linear amplification was applied where necessary to ensure at least 15 dB sensation level in each 3^{rd}-octave band within the stimulus frequency range. The participants provided their responses using a touch screen, a computer keyboard, or a computer mouse, according to their preference. They received visual feedback after each response (correct/incorrect). A short training run was provided by means of a simple amplitude modulation detection task (using broadband noise with a 4-Hz modulation) in order to familiarize the participants with the procedure. Three adaptive measurements were conducted and the median of the resulting three thresholds was considered as the final threshold.

Speech-in-noise test

Speech intelligibility was measured using the Danish hearing in noise test (HINT, Nielsen and Dau, 2011) using a spatial loudspeaker set up in a quiet but slightly reverberant room. Target sentences spoken by a male talker were presented from a frontal location (0º azimuth angle) at 65 dB SPL(C). Running speech interferers spoken by two different male talkers, mixed with low-level speech-shaped noise (-6 dB relative to the running speech level), were played from two loudspeakers positioned at ±100º azimuth. The participants were seated in the middle of the loudspeaker arrangement wearing hearing aids and instructed to use a headrest to maintain a static head position. They were asked to verbally repeat the target-sentence words they had understood, which were then manually scored by an audiologist. SRTs were tracked by adjusting the level of the interferers (i.e., the signal-to-noise ratio, SNR) according to sentence correct scoring (see Nielsen and Dau, 2011). The resulting data were analysed using the method suggested by Rønne et al. (2017) to obtain SRTs relating to 50% sentences correct. Two trainings runs were conducted with NR *Off* using one HINT list (20 sentences) for the first and two lists (40 sentences) for the second run. SRTs were then measured for each NR setting (*Off*, *Default*, *FullThrottle*) using two HINT lists (40 sentences). The presentation order was balanced across participants.

Field testing and questionnaires

The participants were provided with the test hearing aids for two successive field trial periods (3-5 weeks each), with *Default* NR in one period and *FullThrottle* NR in the other one. The order was balanced across participants who were unaware of the difference between the two settings. After the first trial period, participants were asked to fill out the SSQ12 questionnaire (Noble *et al*., 2013); after the second trial period, they were asked to fill out the comparative version of the questionnaire (SSQ12-C; Jensen *et al*., 2009). The participants were asked to rate their preference for the first or second setting on a 5-point scale (-2, -1, 0, 1, 2), where the extremes indicated strong preference for either of the two settings and the midpoint indicated no preference. In addition, the participants were asked to indicate their level of certainty regarding their preference on a scale from 0 (very uncertain) to 10 (very certain). The preference ratings were multiplied with the certainty ratings (normalized by 10) to obtain the final preference score and then processed such that positive values reflect preference for *FullThrottle* NR and negative values preference for *Default* NR.

RESULTS & ANALYSIS

Effect of NR settings on speech intelligibility

Fig. 2 shows the average SRTs and across-participant standard deviations measured for the three NR settings. As can be seen, *Default* NR yielded about 2 dB and *FullThrottle* NR about 4 dB SRT benefit as compared to NR *Off*. The large standard deviations indicate substantial performance differences across participants. A two-way ANOVA with NR setting as a fixed factor and participant as a random factor showed highly significant ($p<<0.001$) main effects of NR and participant. A post-hoc analysis revealed that the different NR settings were all significantly ($p<0.001$) different from each other.

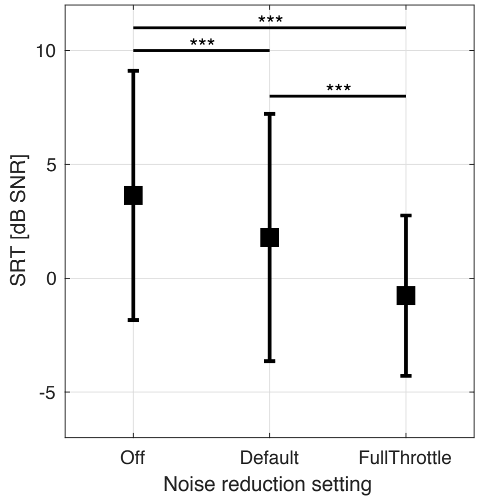

Fig. 2: Mean and standard deviations of SRTs across participants. ***: p<0.001.

RQ1: Does STMD performance predict aided SI without NR?

The left panel of Fig. 3 shows the SRT_{Off} (SRTs measured with NR *Off*) as a function of the STMD thresholds. A highly significant positive correlation between the two measures can be observed, with an R-squared of 0.63 and $p<0.001$. The answer to RQ1 is thus “yes”. However, other measures may also yield good predictions of SRTs. The middle panel of Fig. 3 shows the percentage of variance explained (i.e., the R-squared in percent) for various predictors. While STMD thresholds accounted for the largest amount of variance in SRT_{Off} (63%), the average pure-tone thresholds between 0.125 and 8 kHz (PTA) also accounted for a substantial 58% ($p<0.001$). Performance on the RDS test accounted for a much smaller yet significant ($p<0.05$) 26% of the variance in SRT_{Off}, whereas age showed no effect. Finally, the right panel of Fig. 3 addresses the question of how much additional predictive power can be provided by the individual predictors beyond that provided by the PTA. Linear regression models were employed with PTA as the first predictor and STMD thresholds, RDS scores, or age as the second predictor. It can be seen that the STMD thresholds accounted for an additional 15% ($p<0.01$) of SRT_{Off} variance explained, amounting to 73% overall. The RDS scores also added a significant amount ($p<0.05$) of 10% SRT_{Off} variance explained. The contribution of age was again not significant. All reported significance levels were Bonferroni-corrected for multiple comparisons.

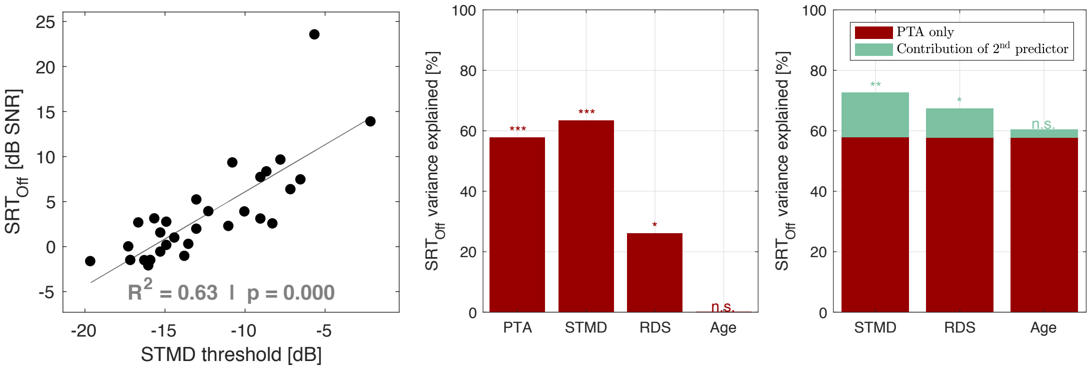

Fig. 3: Left: SRT_{Off} as a function of STMD thresholds along with regression line fit. Middle: percent of SRT_{Off} variance explained by various predictors. Right: percent of SRT_{Off} variance explained by various predictors in addition to PTA in two-predictor linear regression model.

RQ2: Does STMD performance predict SI benefit offered by aggressive NR?

A measure of the SRT benefit induced by *FullThrottle* as compared to *Default* NR is $\Delta SRT = SRT_{Default} - SRT_{FullThrottle}$. Positive ΔSRT values thus indicate an increase in SI induced by *FullThrottle* NR. Fig. 4 shows ΔSRT as a function of the STMD thresholds. A substantial and highly significant positive correlation ($p<0.001$) can be observed, with 51% of the variance in ΔSRT explained by STMD performance. The answer to RQ2 is therefore "yes".

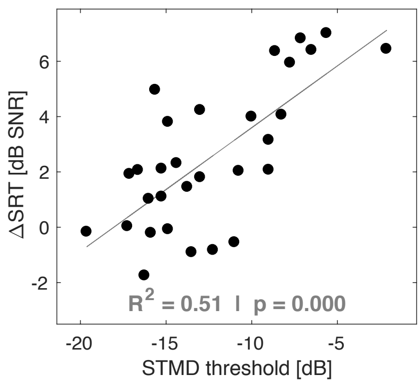

Fig. 4: ΔSRT as a function of STMD thresholds along with regression line.

RQ3: Does STMD performance predict NR preference?

Real-world benefit was evaluated using both the SSQ12-C questionnaire and the participants' preference ratings. Tab. 1 shows the correlations between the preference ratings and the average SSQ12-C score as well as the speech, spatial, and quality subsets SSQ12-C. The preference ratings were almost perfectly correlated with the SSQ12-C scores, except for the spatial subset of SSQ12-C, indicating that the preference ratings were mainly driven by advantages/disadvantages in terms of speech understanding and sound quality. The left panel of Fig. 5 shows the NR preference ratings as a function of STMD thresholds, indicating no obvious relationship between the two measures. However, six participants for whom the hearing-aid logging data indicated little exposure to non-quiet acoustical scenarios were discarded, as well as another two participants with technical problems related to the hearing-aid fitting. For the remaining 22 participants, a significant correlation was found ($p<0.05$, R-squared of 0.23; middle panel of Fig. 5). Additionally, two parallel "correlated patterns" can be observed in the middle panel of Fig. 5, which were connected to the RDS scores (i.e., working memory capacity), as indicated by diamonds ("good" RDS>11.8) and squares ("poor" RDS<11.8). The working memory capacity thus appeared to influence the STMD performance but not the NR preference ratings. The right panel of Fig. 5 shows the preference ratings as a function of RDS-corrected STMD thresholds (obtained by subtracting the STMD thresholds predicted by RDS scores using linear regression from the actual STMD thresholds), indicating a highly significant ($p<0.001$, R-Squared of 0.5) correlation with the preference ratings.

	SSQ12-C_{All}	SSQ12-C_{Speech}	SSQ12-$C_{Spatial}$	SSQ12-$C_{Quality}$
NR Pref.	0.88***	0.88***	0.48**	0.83***

Table 1: Pearson's correlation between NR preference ratings and SSQ12-C scores averaged across all 12 questions and across the respective subsets related to speech, spatial, and quality aspects. ***: $p<0.001$; **: $p<0.01$.

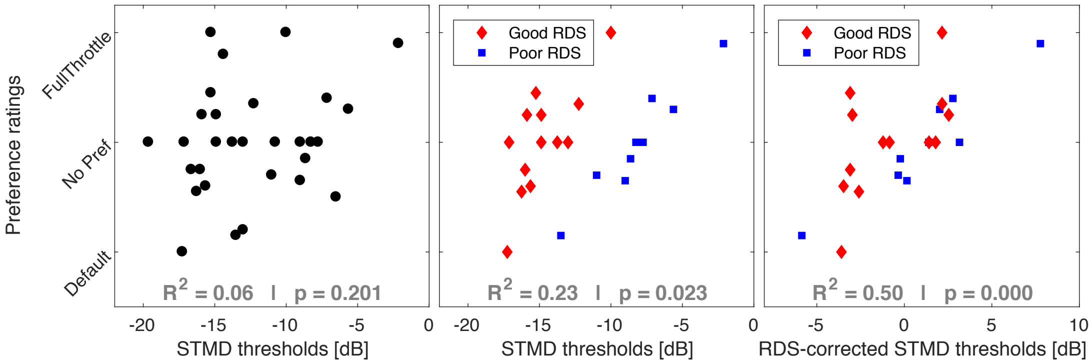

Fig. 5: Left: preference ratings as a function of STMD thresholds. Middle: same as left with 8 participants excluded, symbols according to RDS scores. Right: same as middle with RDS scores factored out of STMD thresholds.

The RDS scores were significantly negatively correlated both with STMD thresholds ($r = -0.53$, $p<0.01$) and SRT_{Off} ($r = -0.51$, $p<0.01$) to the same extent, indicating that working memory capacity positively affected performance both in STMD and speech-in-noise. However, the preference ratings showed no correlation with the RDS scores ($r = 0.01$, $p>0.05$), which merely acted as a mediator variable between the STMD thresholds and the preference ratings. The answer to RQ3 is thus "yes, for most listeners and best in combination with RDS".

CONCLUSIONS

The present study demonstrated that STMD performance as measured with the proposed paradigm (i) can serve as a highly reliable proxy for aided speech-in-noise perception in a spatial multi-talker set up (RQ1), (ii) is strongly correlated with SRT improvement offered by the aggressive NR considered here (RQ2), and (iii) appears to be associated with NR preference reported by participants of a field study (mediated by working memory capacity, RQ3). These findings suggest that a clinical measure of STMD sensitivity may yield a powerful predictor of supra-threshold speech processing, which could be translated to prescriptions of NR settings in hearing aids.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support from the William Demant Foundation, as well as contributions to this study from Bue Kristensen, James Harte, Valentina Campagnaro, Thomas Lunner, Fares El-Azm, Jacob Aderholt, Erengül Sabedin, and Sébastien Santurette.

REFERENCES

Bernstein, J.G.W., Mehraei, G., Shamma, S., Gallun, F.J., Theodoroff, S.M., and Leek, M.R. (**2013**). "Spectro-temporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners," J. Am. Acad. Audiol., **24**(4), 293-306. doi:10.3766/jaaa.24.4.5

Bernstein, J.G.W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., and Lunner, T. (**2016**). "Spectro-temporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids," Trends Hear., **20**, 1-17. doi: 10.1177/2331216516670387

Jensen, N.S., Akeroyd, M.A., Noble, W., and Naylor, G. (**2009**). "The Speech Spatial and Qualities of Hearing scale (SSQ) as a benefit measure," Poster presented at 4th NCRAR international conference, Portland, US.

Nielsen, J.B. and Dau, T. (**2011**). "The Danish hearing in noise test," Int. J. Audiol., **50**, 202-208. doi: 10.3109/14992027.2010.524254

Mehraei, G., Gallun, F.J., Leek, M.R., and Bernstein, J.G.W. (**2014**). "Spectro-temporal modulation sensitivity for hearing-impaired listeners: dependence on carrier center frequency and the relationship to speech intelligibility," J. Acoust. Soc. Am., **136**(1), 301-316. doi: 10.1121/1.4881918

Noble, W., Jensen, N.S., Naylor, G., Bhullar, N., and Akeroyd, M.A. (**2013**). "A short form of the Speech, Spatial and Qualities of Hearing scale suitable for clinical use: The SSQ12," Int. J. Audiol., **52**, 409-412. doi: 10.3109/14992027.2013.781278

Rønne, F.M., Laugesen, S., and Jensen, N.S. (**2017**). "Selection of test-setup parameters to target specific signal-to-noise regions in speech-on-speech intelligibility testing," Int. J. Audiol., **56**, 559-567. doi: 10.1080/14992027.2017.1300349

Zaar, J., Simonsen, L.B., Behrens, T., Dau, T., and Laugesen, S. (**2018**). "Towards a clinically viable spectro-temporal modulation test," Poster presented at the international hearing aid research conference (IHCON), Lake Tahoe, US.