William Demant

Binaural intelligibility prediction for hearing aid systems

Asger Heidemann Andersen

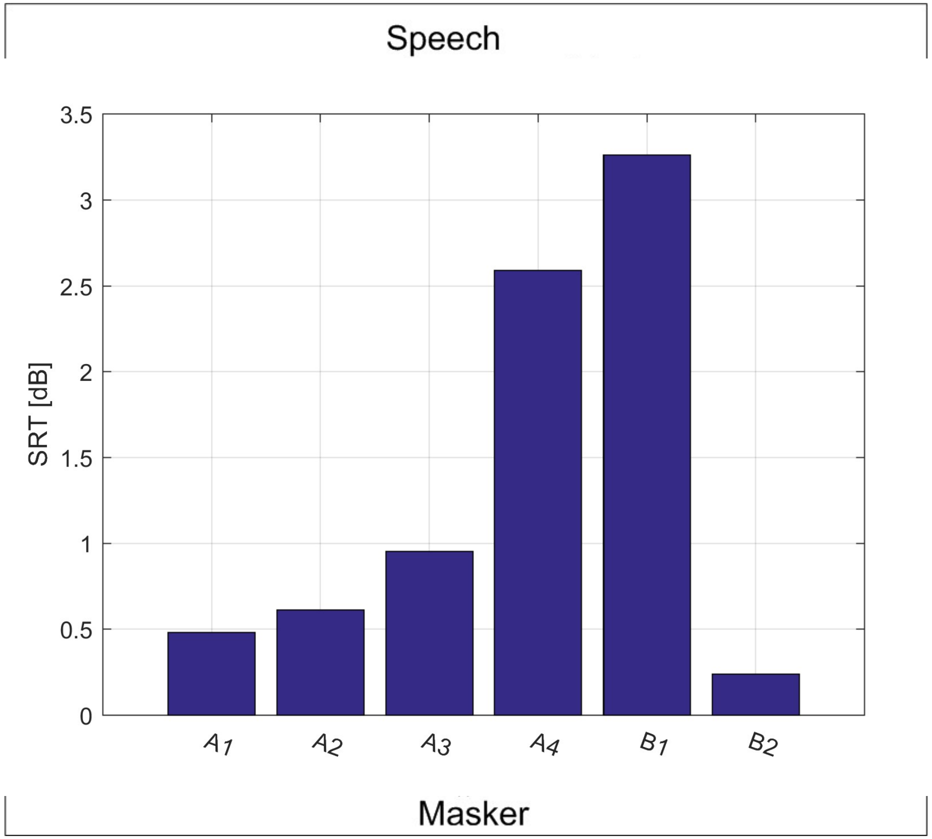
Oticon A/S




Intelligibility Prediction at Oticon

Internal Test of Hearing Aid Prototypes

- Setup:
  - Frontal target speaker and two ISTS maskers in a room with little reverberation.
- Hearing Aids:
  - Two prototypes: A in 4 settings, B in 2 settings.
  - Fitted to the subjects.
  - 14 hearing impaired subjects.
- Clean and noisy/processed signals were also recorded with a HATS.



William Demant

Key Takeaways

There exists a couple of binaural versions of STOI

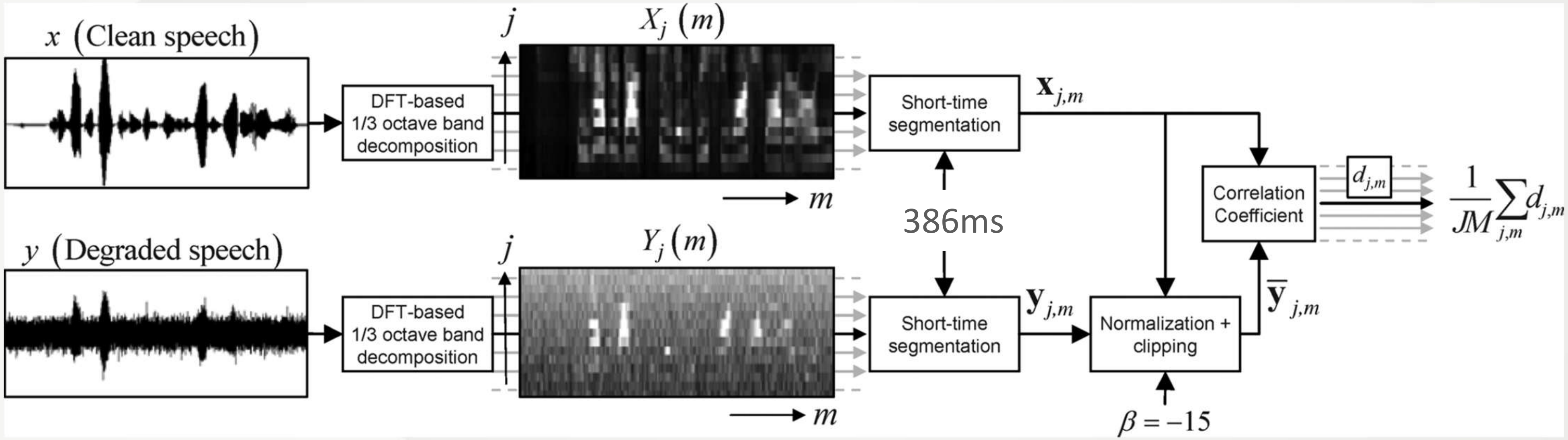
These work well for some applications

Most of the STOI-family is freely available and easy to use

William Demant

What is the STOI Measure?!?

The Short-Time Objective Intelligibility Measure



C. Taal et al., “An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech,” TASLP, 2011.

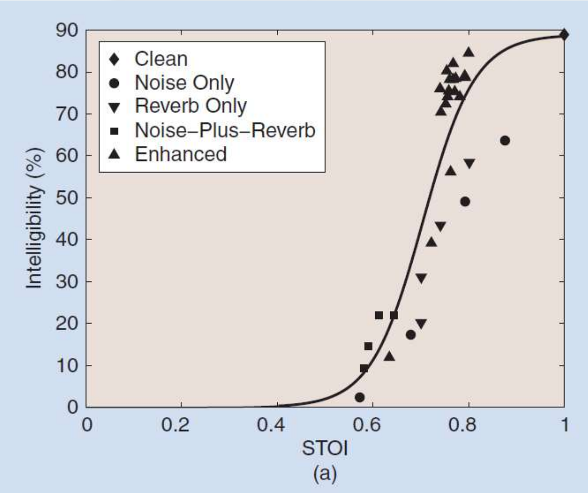
William Demant

Why STOI?!?

Tiago H. Falk, Vijay Parsa, João F. Santos, Kathryn Arehart, Oldooz Hazrati, Rainer Huber, James M. Kates, and Susan Scollie

Objective Quality and Intelligibility Prediction for Users of Assistive Listening Devices

IEEE SIGNAL PROCESSING MAGAZINE [114] MARCH 2015



[TABLE 1] PER-CONDITION PERFORMANCE CRITERIA FOR THE CI INTELLIGIBILITY DATABASE. THE NUMBERS IN BOLD REPRESENT THE BEST ATTAINED PERFORMANCES (STATISTICALLY INDIFFERENT) AMONG ALL TESTED INTRUSIVE AND NONINTRUSIVE ALGORITHMS.

	ALL				NONENHANCED (NOISE/REVERB)				ENHANCED			
METRIC	ρ	ρ_{spear}	ρ_{sig}	ε-RMSE	ρ	ρ_{spear}	ρ_{sig}	ε-RMSE	ρ	ρ_{spear}	ρ_{sig}	ε-RMSE
NCM	0.68	0.74	**0.87**	**9.03**	0.96	0.93	**0.93**	8.41	0.47	0.68	0.77	10.33
STOI	0.81	0.76	**0.89**	**7.05**	0.97	0.96	**0.97**	**0.6**	0.66	0.69	**0.92**	**3.82**
PESQ	−0.09	0.01	−0.02	26.85	−0.25	0.4	0.14	26.14	−0.09	0.21	−0.02	23.89
PEMO-Q	0.67	0.53	0.68	15.68	0.72	0.8	0.69	15.67	0.38	0.53	0.44	13.52
P.563	0.05	0.38	0.33	23.59	0.76	0.6	0.78	11.77	−0.79	0	−0.43	25.23
ModA	0.78	0.59	0.78	16.88	0.82	0.76	0.8	13.59	−0.13	−0.17	−0.07	18.42
SRMR	0.49	0.53	0.68	18.41	0.93	0.89	0.92	9.6	−0.35	−0.03	−0.37	23.16
SRMR-CI	0.86	0.77	**0.93**	**5.67**	0.98	0.98	**0.98**	**2.06**	0.65	0.5	**0.88**	**4.65**

[TABLE 2] PER-CONDITION PERFORMANCE CRITERIA FOR THE HA NONLINEAR FREQUENCY COMPRESSION QUALITY DATABASE. THE NUMBERS IN BOLD REPRESENT THE BEST ATTAINED PERFORMANCES (STATISTICALLY INDIFFERENT) AMONG ALL TESTED INTRUSIVE AND NONINTRUSIVE ALGORITHMS.

METRIC	ρ	ρ_{spear}	ρ_{sig}	ε-RMSE
NCM	0.67	0.67	**0.89**	7.46
STOI	0.77	0.67	**0.92**	**2.24**
PESQ	0.62	0.56	0.79	**5.73**
HASQI	0.71	0.71	**0.93**	7.67
HASPI	0.83	0.72	0.81	9.9
PEMO-Q	0.67	0.6	0.79	**5.06**
PEMO-Q-HI	0.89	0.71	**0.92**	**1.83**
P.563	−0.27	−0.38	−0.33	23.25
ModA	0.52	0.48	**0.54**	**8.86**
SRMR	0.49	0.59	0.4	17.06
SRMR-HA	0.51	0.58	0.46	14.39

William Demant

Why STOI?!?

Speech Intelligibility Evaluation for Mobile Phones

Søren Jørgensen[1,2], Jens Cubick[1], Torsten Dau[1]
[1] Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[2] Oticon A/S, Kongebakken 9, 2765 Smørum, Denmark. jr@oticon.com

Table I. Root mean squared error (RMSE) in dB between measured and predicted SRT_{70} from the three mobile phones in the three noises considered in the present study.

	SSN	Pub	Traffic
sEPSM	0.58	8.2	4.5
ESII	2.2	5.8	1.9
STOI	1.0	0.46	2.2

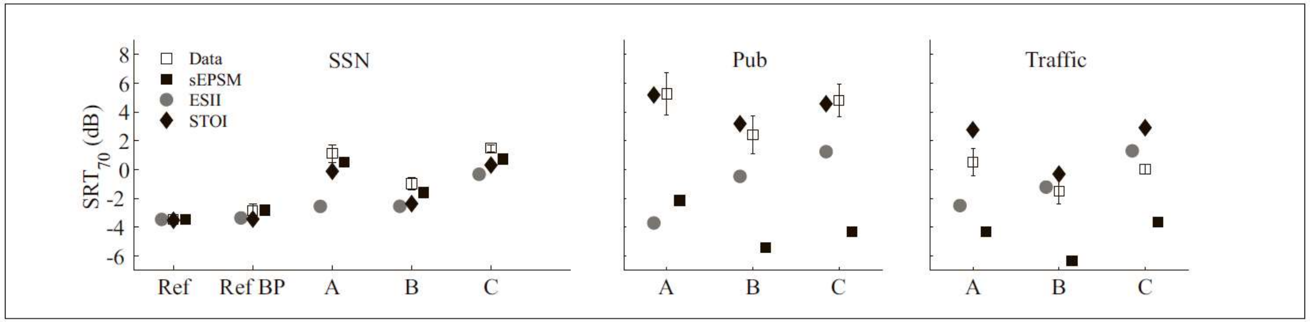


Figure 6. SRT_{70} obtained from the perceptual psychometric functions (open squares), for all the conditions with SSN (left panel), with Pub noise (middle panel), and with Traffic noise (right panel). The vertical bars denote one standard error. Predictions from the sEPSM are indicated by the filled black squares, predictions from STOI are shown as filled black diamonds, and predictions from the ESII are indicated by the filled gray circles.

William Demant

Why STOI?!?

Comparison of predictive measures of speech recognition after noise reduction processing

Karolina Smeds,[a] Arne Leijon,[b] Florian Wolters, Anders Hammarstedt, Sara Båsjö, and Sofia Hertzman
ORCA Europe, Widex A/S, Maria Bangata 4, SE-118 63 Stockholm, Sweden

Only one of the predictive measures, CSII (Kates and Arehart, 2005), correctly predicted the effect of the currently tested noise reduction algorithms on the speech recognition threshold within both groups of listeners. In general, measures using correlation between the clean speech and the processed noisy speech, as well as other measures that are based on short-time analysis of speech and noise, seemed most promising.

TABLE I. SRT prediction errors for two groups of listeners, 20 with impaired hearing (HI) and 10 with NH. The median prediction error was calculated across listeners, and the result with the largest magnitude among the three noise reduction algorithms is shown. Bold numbers indicate that the prediction errors were significantly ($p < 0.05$) different from zero for at least one noise reduction algorithm, as indicated by the Friedman test described in Sec. III B 2. Underlined numbers indicate that the predicted benefit was also in the wrong direction compared to the measured benefit and that this discrepancy was statistically significant ($p < 0.05$), as indicated by the correlation test described in Sec. III B 2.

		Prediction error (dB)	
Measure	Reference	HI	NH
SII	ANSI (1997)	**3.8**	**2.5**
ESII	Rhebergen *et al.* (2006)	**6.4**	**7.0**
STSII	This paper	1.2	**2.4**
Glimpses	Cooke (2006)	**1.4**	**–1.5**
fwSNRseg-a	Ma *et al.* (2009)	**–1.6**	**4.1**
fwSNRseg-b	This paper	**1.0**	**4.8**
STOI	Taal *et al.* (2011a)	–0.6	**–1.6**
CSII	Kates and Arehart (2005)	0.9	0.4
sEPSM	Jørgensen and Dau (2011)	**–1.2**	**2.0**
mr-sEPSM	Jørgensen *et al.* (2013)	–1.2	**–5.0**

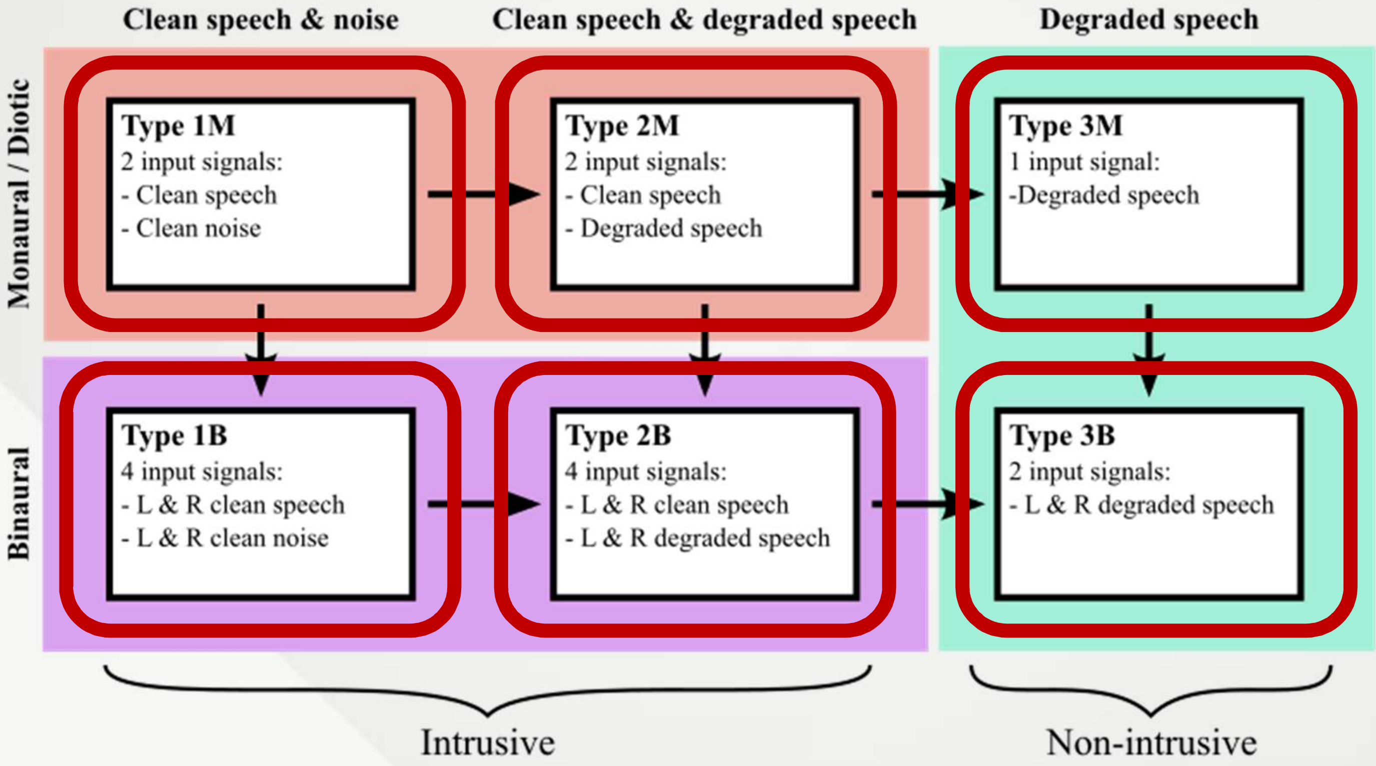
William Demant

Why STOI?!?

- Easy and free to use:
  - No patent, license, etc.
  - Available on Cees Taal's website.
  - 188 lines of MATLAB in one .m file.
  - Requires no toolboxes.

```
function d = stoi(x, y, fs_signal)
%   d = stoi(x, y, fs_signal) returns the output of the short-time
%   objective intelligibility (STOI) measure described in [1, 2], where x
%   and y denote the clean and processed speech, respectively, with sample
%   rate fs_signal in Hz. The output d is expected to have a monotonic
%   relation with the subjective speech-intelligibility, where a higher d
%   denotes better intelligible speech. See [1, 2] for more details.
%
```

A bit of Perspective



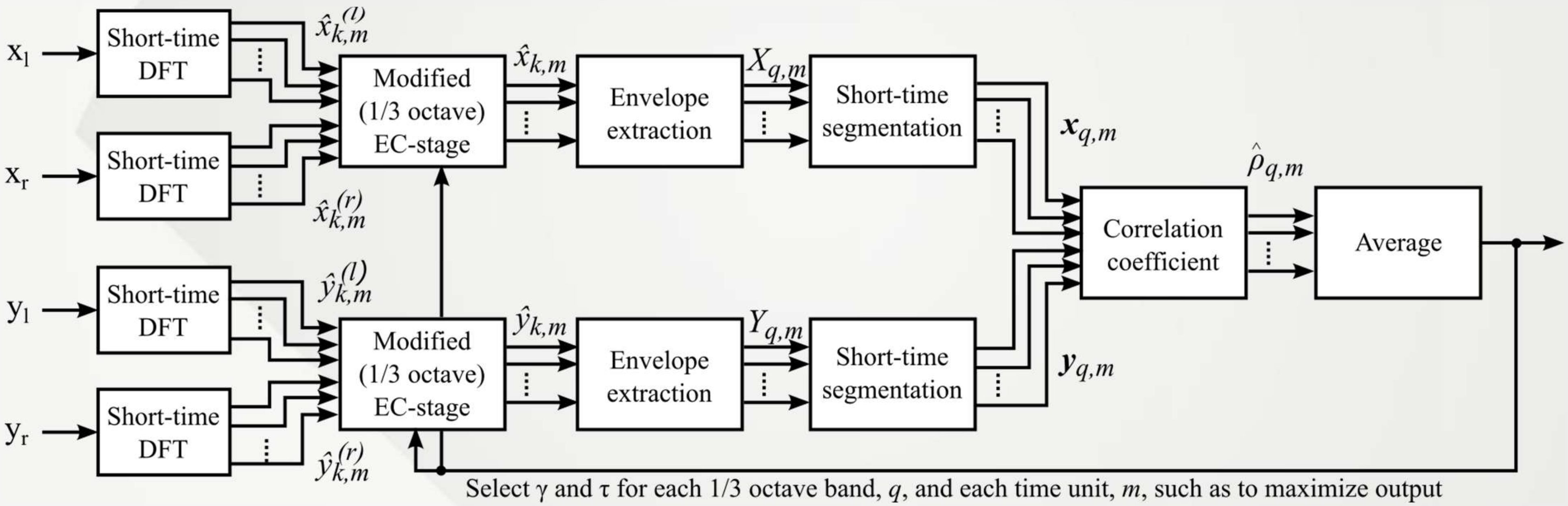
William Demant

The Binaural STOI Measure*

- Base idea: Combine the STOI measure with the equalization cancellation model.
  1. Combine left and right ear signal with EC stage.
  2. Compute STOI measure of output.
  - Find optimal EC-parameters by sweeping different values.
  - Handle internal noise by computing average STOI across many realizations.

* A. H. Andersen, et. al., "A binaural short time objective intelligibility measure for noisy and enhanced speech," Interspeech, Dresden, Germany, Sept. 2015.

The Binaural STOI Measure



The Deterministic Binaural STOI Measure

- The output of the measure is stochastic because the EC-stage adds internal noise to parameters.
- First version of BSTOI used Monte Carlo simulation to find expected value of the output.
- Second version computes the expectation analytically:
  - Fully deterministic predictions of intelligibility.
  - Lowered computational cost.

$$
\begin{aligned}
&E_\Delta\left[(\mathbf{x}_{q,m}-\mathbf{1}\mu_{\mathbf{x}_{q,m}})^\intercal(\mathbf{y}_{q,m}-\mathbf{1}\mu_{\mathbf{y}_{q,m}})\right]= \\
&\quad(e^{2\beta}\mathbf{l}^\intercal_{x_{q,m}}\mathbf{l}_{y_{q,m}}+e^{-2\beta}\mathbf{r}^\intercal_{x_{q,m}}\mathbf{r}_{y_{q,m}})e^{2\sigma^2_{\Delta\beta}} \\
&+\mathbf{r}^\intercal_{x_{q,m}}\mathbf{l}_{y_{q,m}}+\mathbf{l}^\intercal_{x_{q,m}}\mathbf{r}_{y_{q,m}}-2e^{\sigma^2_{\Delta\beta}/2}e^{-\omega^2\sigma^2_{\Delta\tau}/2}\times \\
&\quad\left\{\left(e^{\beta}\mathbf{l}^\intercal_{x_{q,m}}+e^{-\beta}\mathbf{r}^\intercal_{x_{q,m}}\right)Re\left[\mathbf{c}_{y_{q,m}}e^{-j\omega\tau}\right]\right. \\
&+\left.Re\left[e^{-j\omega\tau}\mathbf{c}^\intercal_{x_{q,m}}\right]\left(e^{\beta}\mathbf{l}_{y_{q,m}}+e^{-\beta}\mathbf{r}_{y_{q,m}}\right)\right\} \\
&+2\left(Re\left[\mathbf{c}^H_{x_{q,m}}\mathbf{c}_{y_{q,m}}\right]+e^{-2\omega^2\sigma^2_{\Delta\tau}}Re\left[\mathbf{c}^\intercal_{x_{q,m}}\mathbf{c}_{y_{q,m}}e^{-j2\omega\tau}\right]\right),
\end{aligned}
$$

where:

$$\mathbf{l}_{x_{q,m}}=[X^{(l)}_{q,m-N+1},\dots,X^{(l)}_{q,m}]^\intercal-\mathbf{1}\sum_{k=m-N+1}^{m}\frac{X^{(l)}_{q,k}}{N},$$

$$\mathbf{r}_{x_{q,m}}=[X^{(r)}_{q,m-N+1},\dots,X^{(r)}_{q,m}]^\intercal-\mathbf{1}\sum_{k=m-N+1}^{m}\frac{X^{(r)}_{q,k}}{N},$$

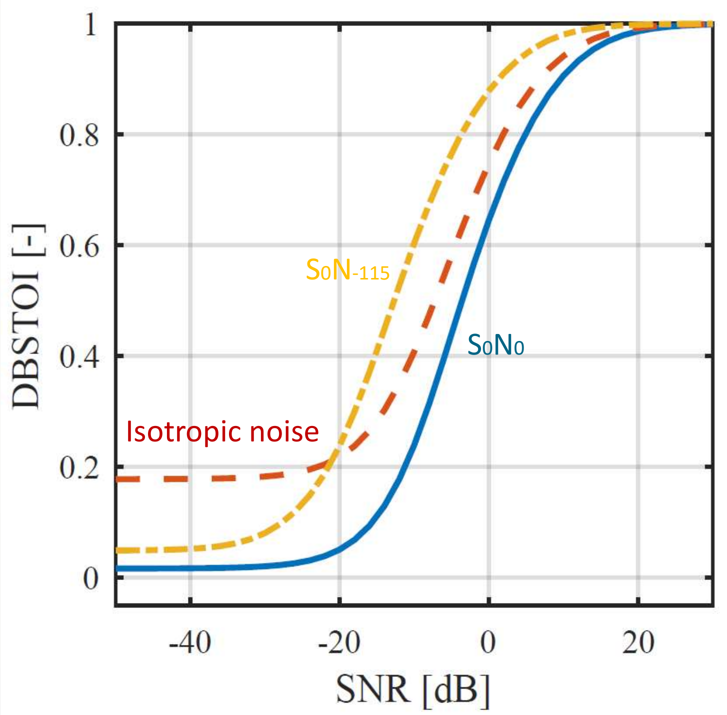
$$\mathbf{c}_{x_{q,m}}=[X^{(c)}_{q,m-N+1},\dots,X^{(c)}_{q,m}]^\intercal-\mathbf{1}\sum_{k=m-N+1}^{m}\frac{X^{(c)}_{q,k}}{N},$$

$$\beta=\frac{\ln(10)}{20}\gamma,\qquad \sigma^2_{\Delta\beta}=\left(\frac{\ln(10)}{20}\right)^2\sigma^2_{\Delta\gamma},$$

William Demant

The Deterministic Binaural STOI Measure

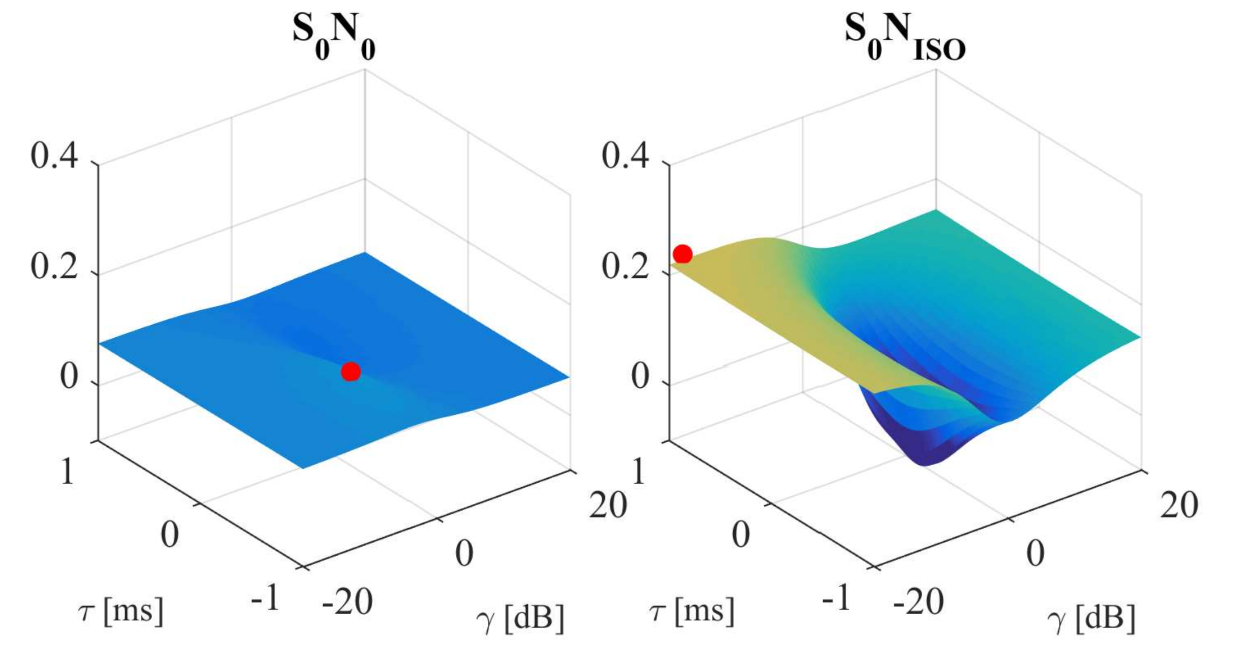
- The DBSTOI measure was shown to correlate well with speech intelligibility in many binaural listening conditions.
- Most significant inaccuracies were found at low SNRs for conditions with multiple or distributed interferers.

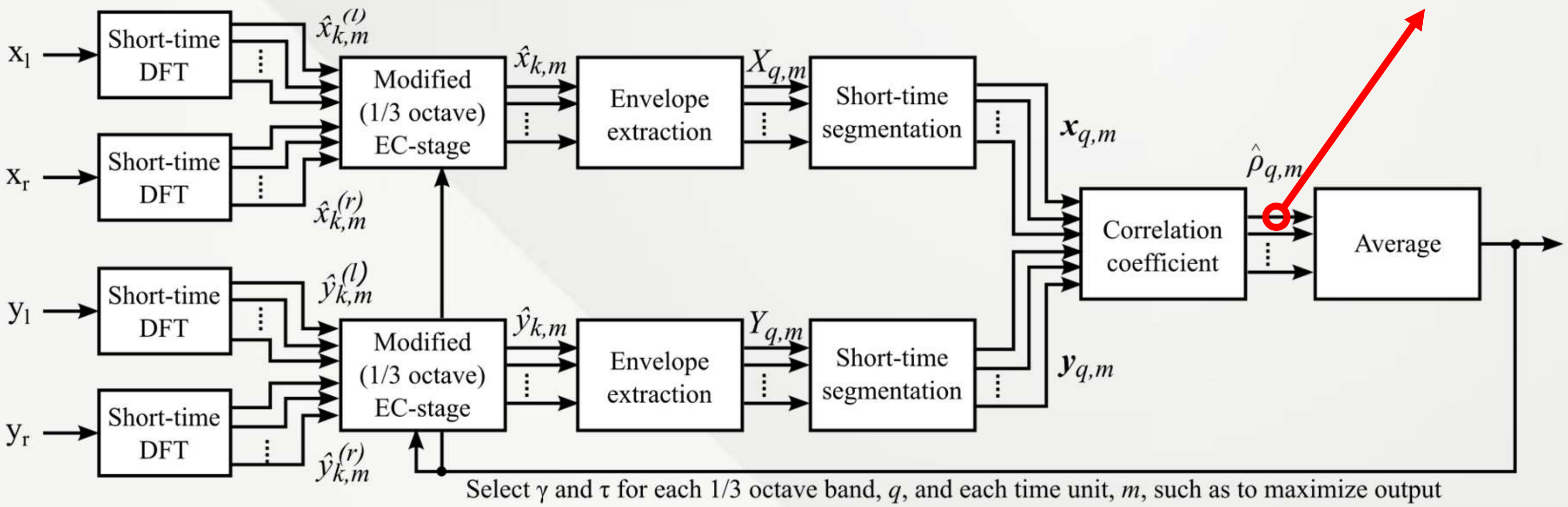


William Demant

The Deterministic Bina

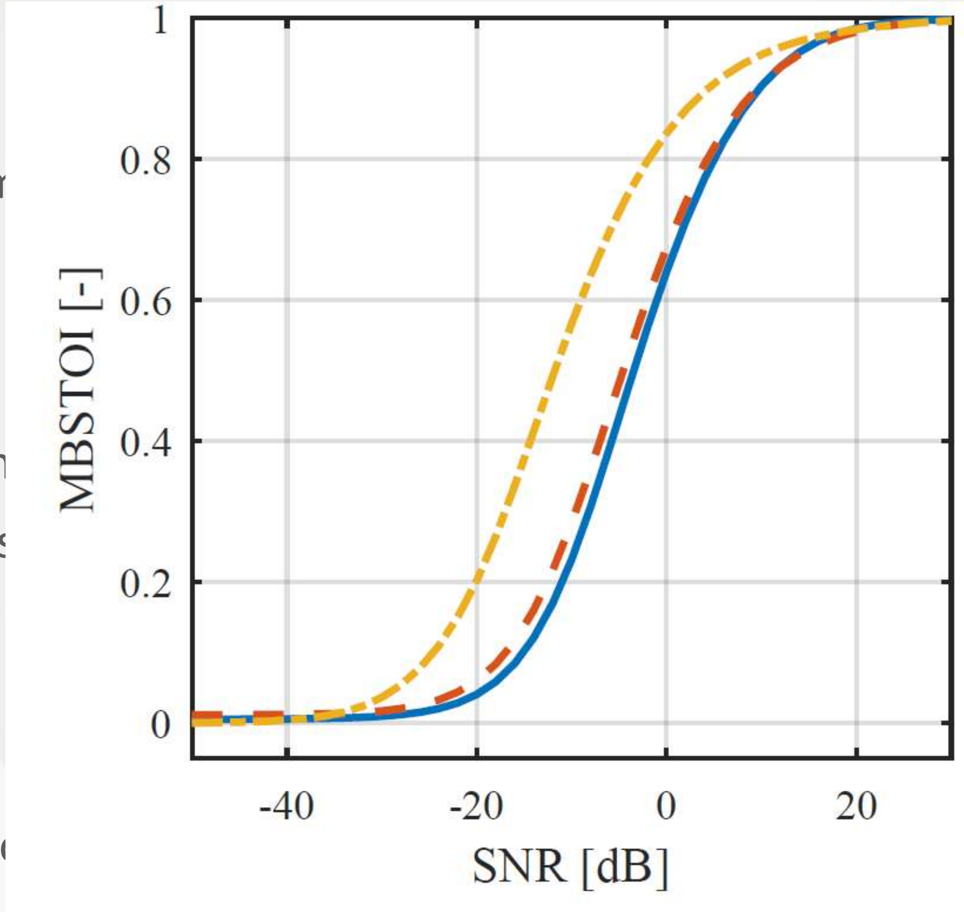
A look under the hood





The Modified Binaural STOI Measure

- DBSTOI finds EC param
- Bias term can be math
- To remove bias, we ins
- Loosely corresponds t



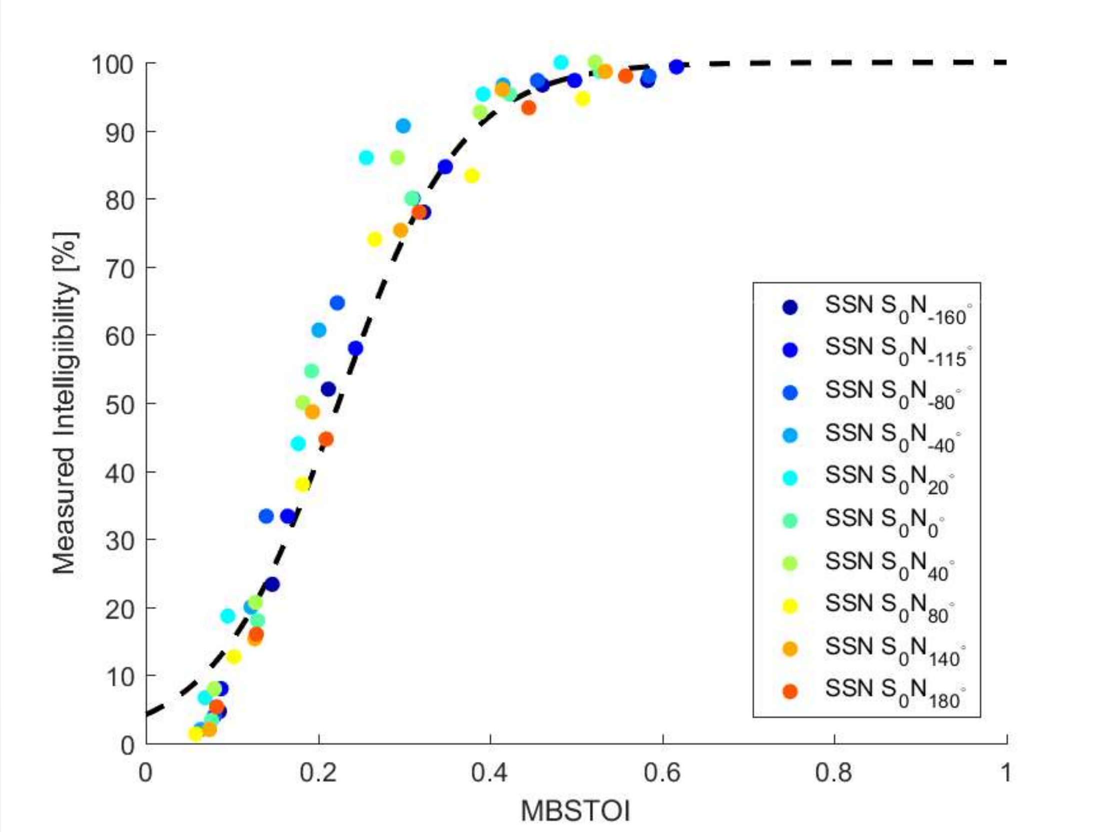
tions.

William Demant

The Modified Binaural STOI Measure

Dataset 1 (Andersen et al.)

- Danish matrix sentences from front.
- SSN from one of ten directions in horizontal plane.
- 10 NH listeners.
- Simulated anechoic with HRTFs.
- (10 listeners) x (10 conditions) x (6 SNRs) x (3 reps) = 1800 sentences.

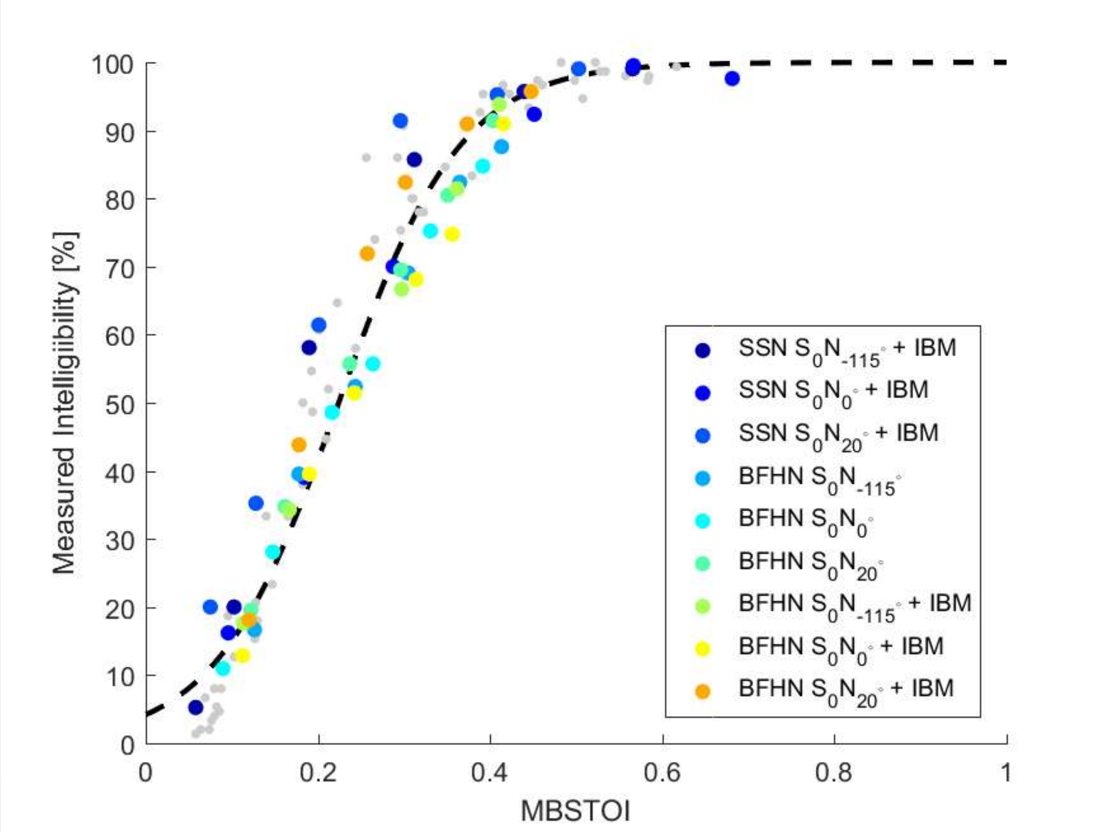


William Demant

The Modified Binaural STOI Measure

Dataset 2 (Andersen et al.)

- Danish matrix sentences from front.
- SSN or bottling factory hall noise (BFHN) from one of 3 directions.
- 14 NH listeners.
- Simulated anechoic with HRTFs.
- (14 listeners) x (9 conditions) x (6 SNRs) x (3 reps) = 2268 sentences.

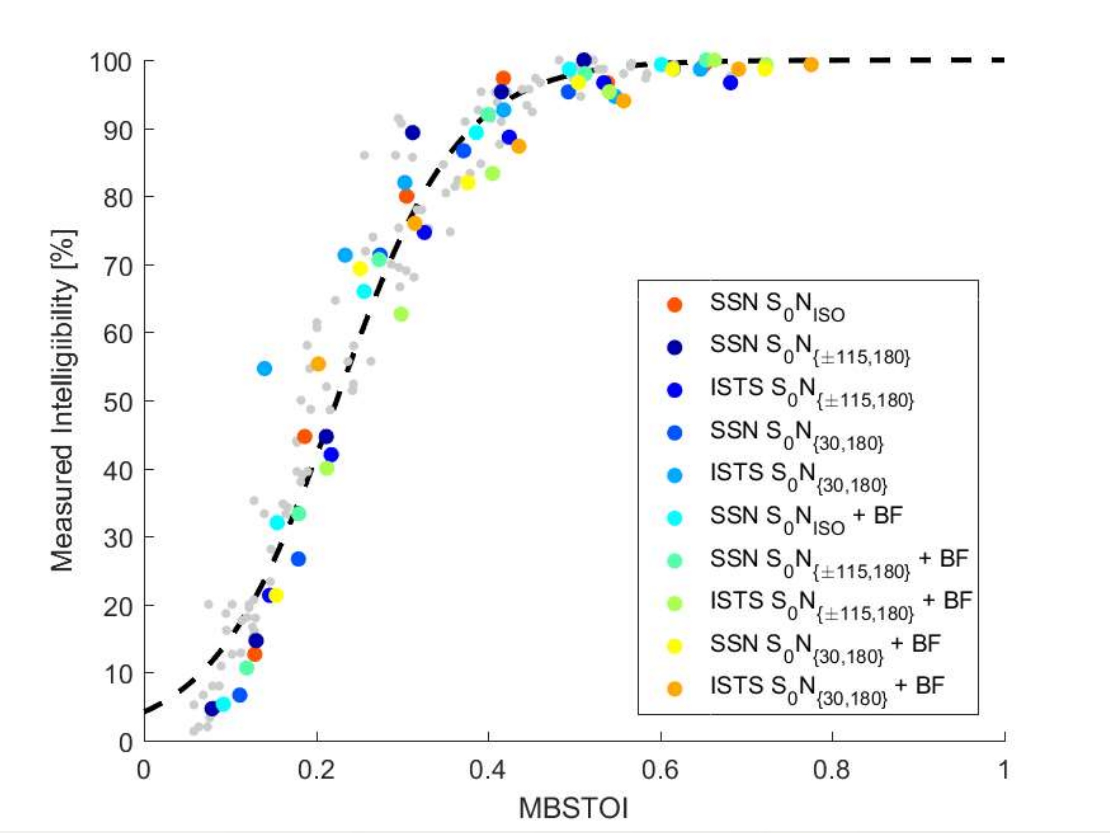


William Demant

The Modified Binaural STOI Measure

Dataset 3 (Andersen et al.)

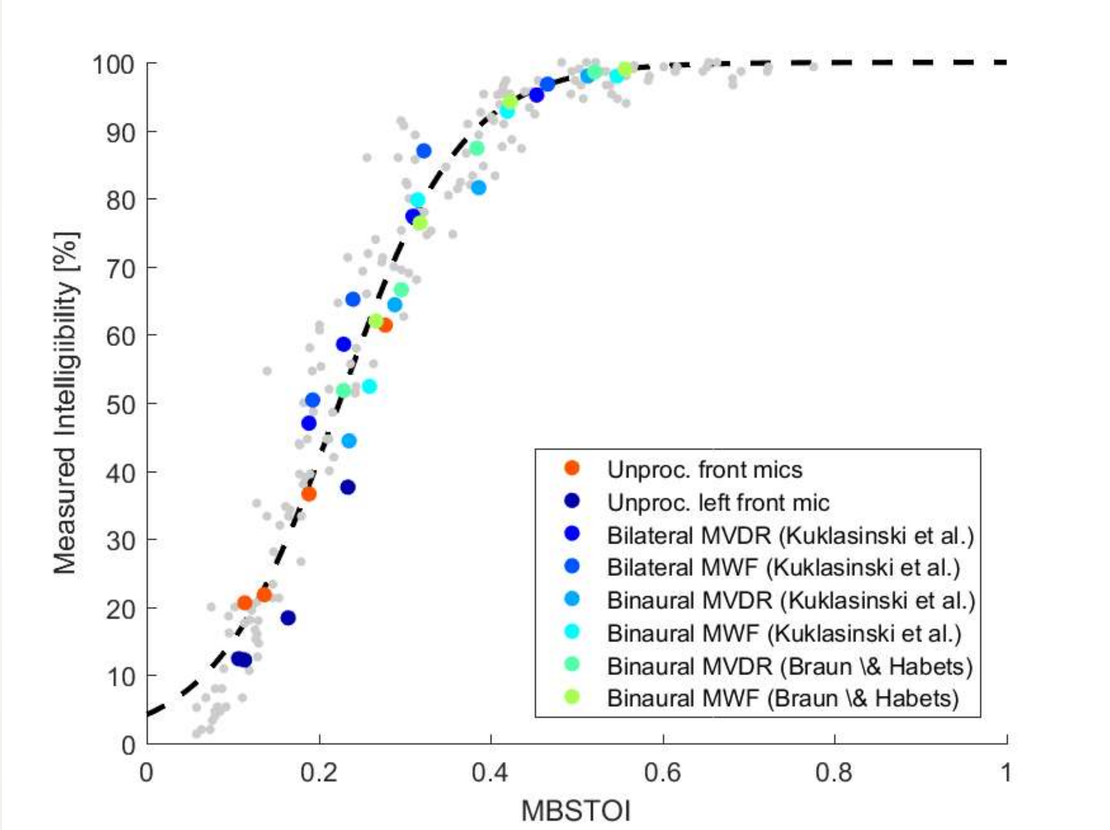
- Danish matrix sentences from front.
- SSN or ISTS noise. Isotropic or multiple sources.
- Beamforming as used in hearing aids.
- 10 NH listeners.
- Simulated anechoic with HRTFs.
- (10 listeners) x (10 conditions) x (6 SNRs) x (3 reps) = 1800 sentences.



The Modified Binaural STOI Measure

Dataset 4 (Kuklasiński et al.)

- Danish matrix sentences from front.
- ISTS sources at ±90°, ±135°, 180°.
- Reverberant cellar BRIRs.
- Direct-to-reverberant ratio modified to make reverberation more or less detrimental.
- Variations of hearing aid noise reduction.
- 20 NH listeners.
- (20 listeners) x (8 conditions) x (4 DRRs) x (5 reps) = 3200 sentences.

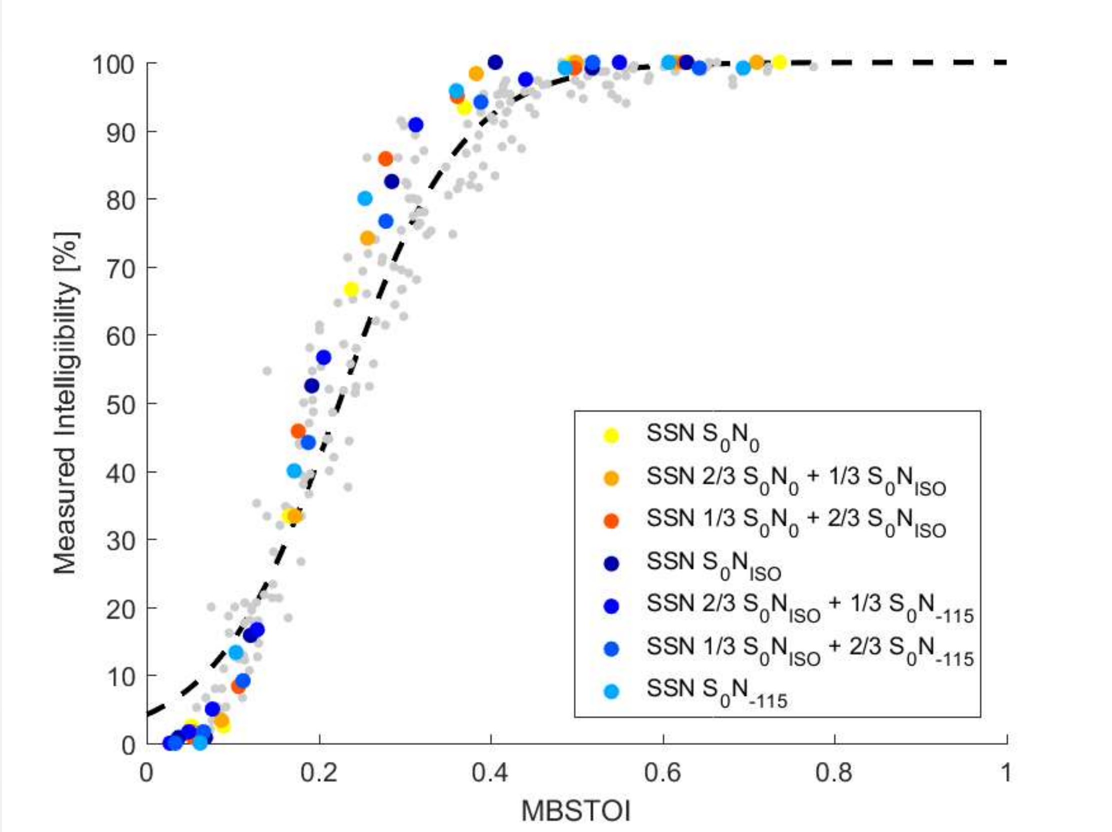


A. Kuklasiński et al., “Maximum likelihood PSD estimation for speech enhancement in reverberation and noise”, IEEE Tran. on Audio, Speech and Language Processing 24 (9), 2016.

The Modified Binaural STOI Measure

Dataset 5 (Andersen et al.)

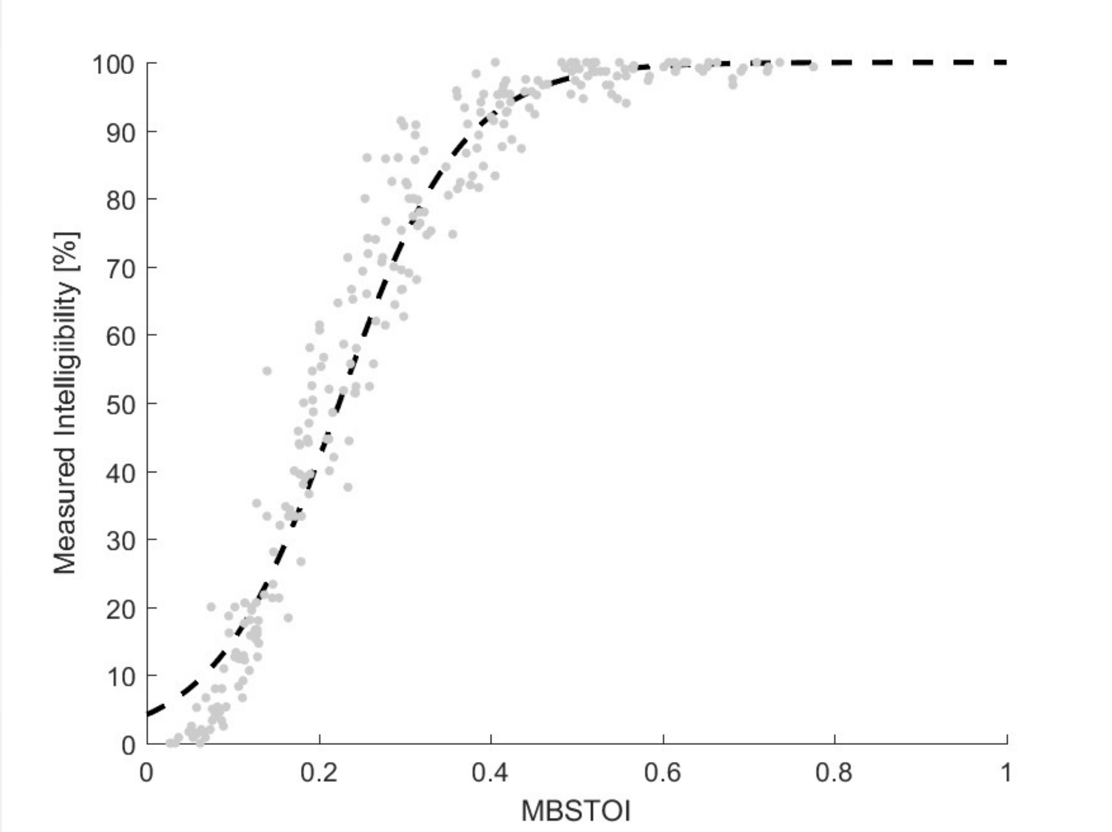
- Danish matrix sentences from front.
- SSN. Mixture of isotropic and point source at either 0 ° or 115°.
- 8 NH listeners.
- Simulated anechoic with HRTFs.
- (8 listeners) x (7 conditions) x (8 SNRs) x (3 reps) = 1344 sentences.



William Demant

The Modified Binaural STOI Measure

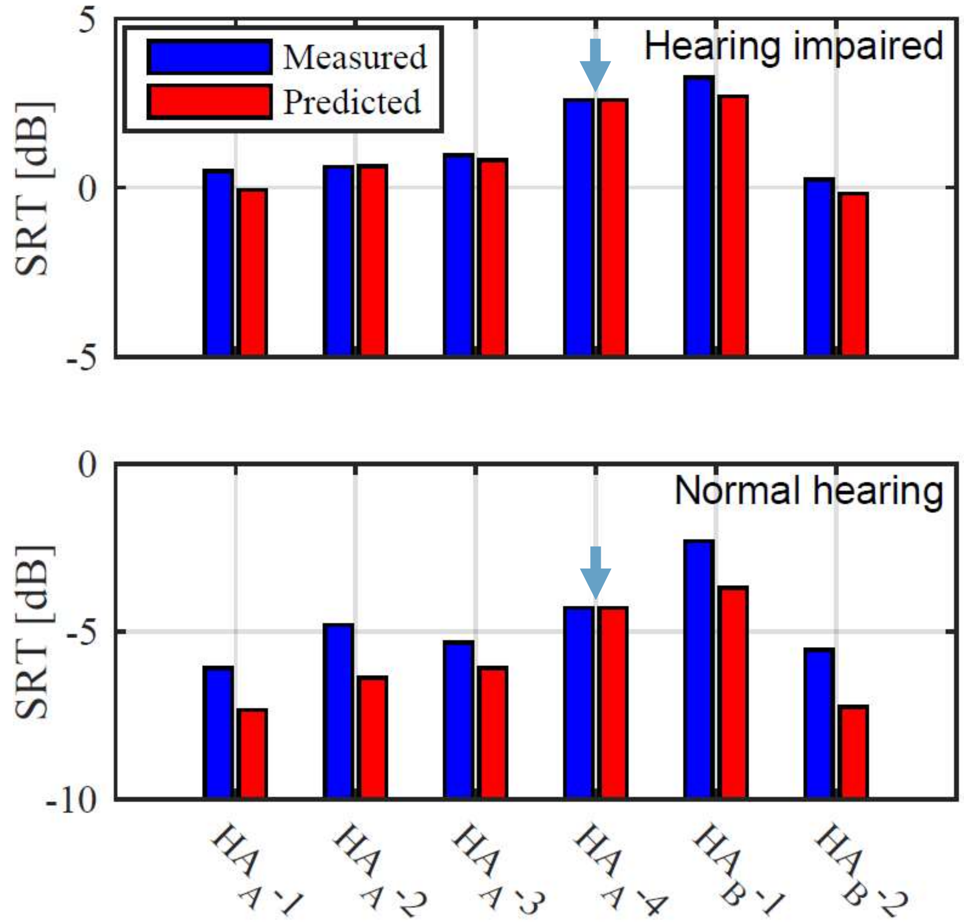
	D_1	D_2	D_3	D_4	D_5
RMSE	**7.34%**	**6.92%**	**7.07%**	**6.30%**	**7.69%**
Kendall's Tau	**0.887**	0.858	0.838	**0.906**	**0.893**
Pearson Correlation	**0.981***	0.974*	**0.973***	**0.975***	**0.987***



Intelligibility Prediction at Oticon

Internal Test of Hearing Aid Prototypes

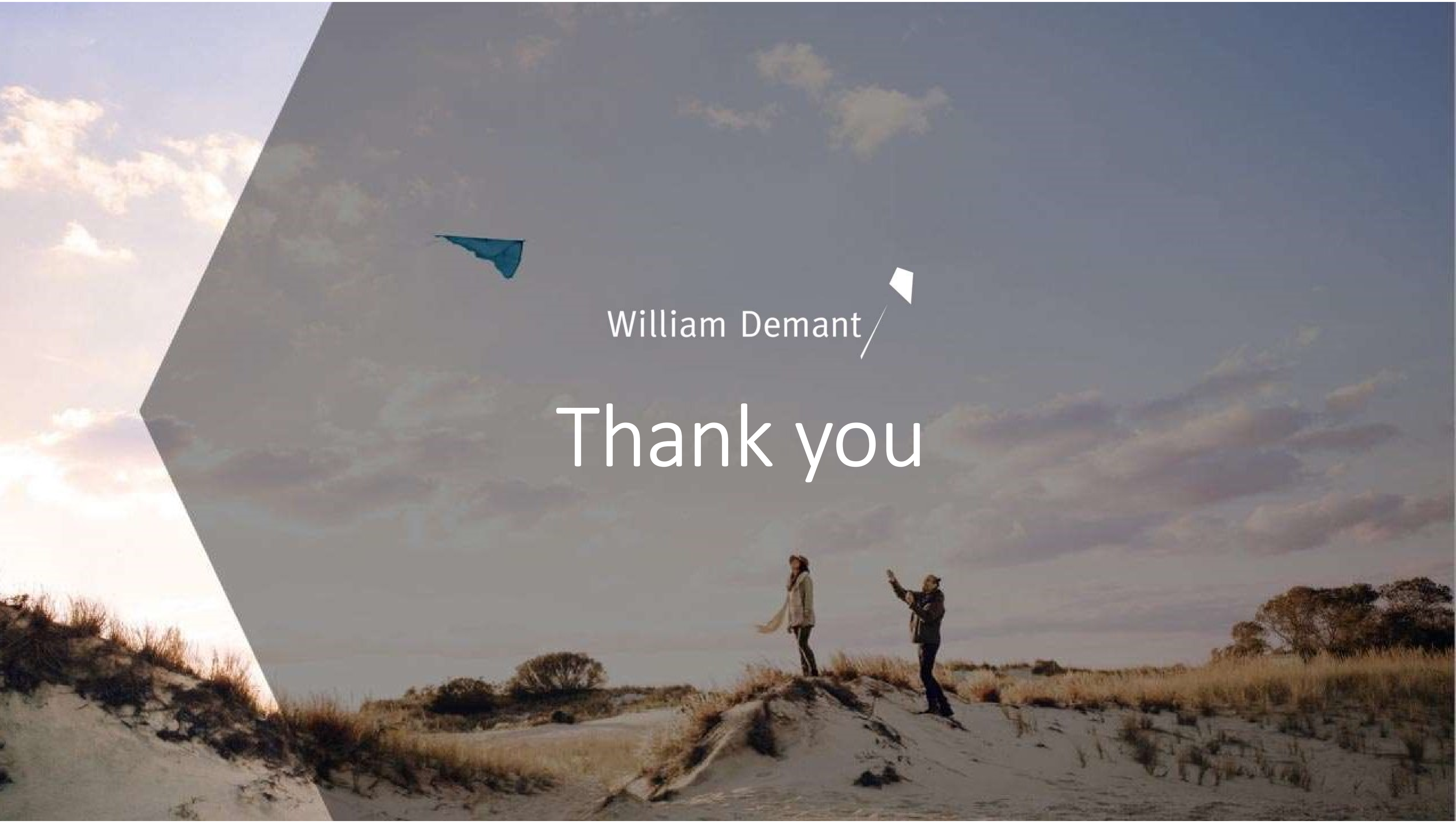
- Setup:
  - Frontal target speaker and two ISTS maskers in a room with little reverberation.
- Hearing Aids:
  - Two prototypes: A in 4 settings, B in 2 settings.
  - Fitted to the subjects. Minor amplification provided for normal hearing subjects.
- Subjects:
  - 14 hearing impaired subjects.
  - 14 normal hearing subjects.
- Clean and noisy/processed signals were also recorded with a HATS.



William Demant

References

Algorithm	Paper	Download
STOI	C. Taal et al., “An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech,” TASLP, 2011.	http://www.ceestaal.nl/code/
DBSTOI	A. H. Andersen et al., “Nonintrusive Speech Intelligibility Prediction Using Convolutional Neural Networks,” TASLP, 2018.	http://ah-andersen.net/code/
MBSTOI	A. H. Andersen et al., “Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions”, Sp. Comm., 2018.	http://ah-andersen.net/code/
NI-STOI	A. H. Andersen, et al., “A Non-Intrusive Short-Time Objective Intelligibility Measure,” ICASSP, 2017.	http://ah-andersen.net/code/
ESTOI	J. Jensen et al., “An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers,” TASLP, 2016.	http://kom.aau.dk/~jje/code/estoi.m
DNN-SIP	A. H. Andersen et al., “Nonintrusive Speech Intelligibility Prediction Using Convolutional Neural Networks,” TASLP, 2018.	


William Demant
Thank you