

**Reverse correlation to derive  
listeners' internal templates in a word-in-noise test:  
Methodological aspects and relevance of a human readable goodness-of-fit metric**

**Alejandro Osses and Léo Varnet**

E-mail: [ale.a.osses@gmail.com](mailto:ale.a.osses@gmail.com)

Laboratoire des Systèmes Perceptifs, École Normale Supérieure (ENS),  
PSL University, Paris, France

21 April 2023

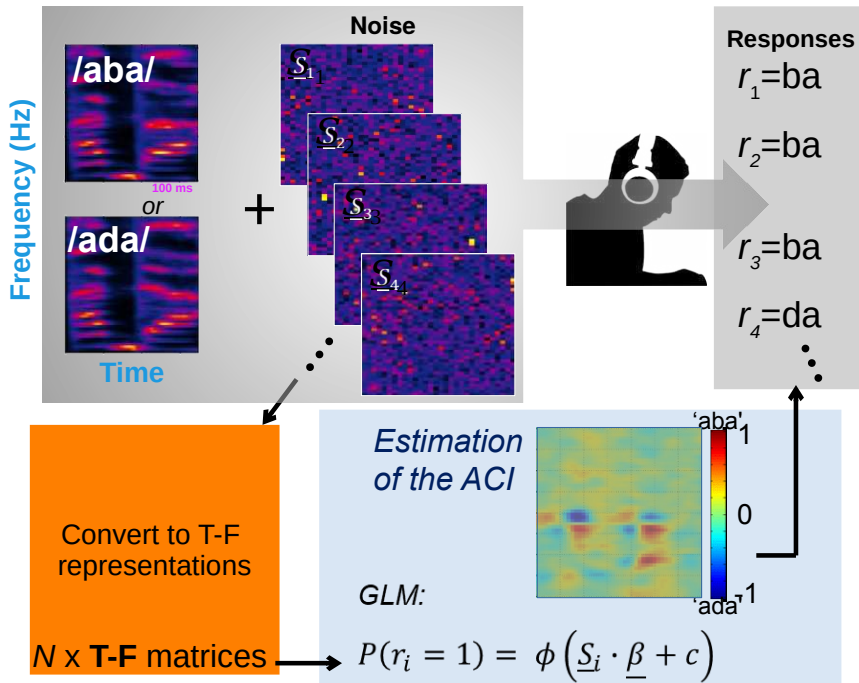
WASdag, Nijmegen, the Netherlands

## **Outline**

- Phoneme discrimination using the ACI method
- ACIs and the concept of template
- What to expect from the ACIs?
- Results for two participants
- ACI prediction performance and similarity between ACIs
- Outlook

# Phoneme discrimination using the ACI method

- The reverse correlation method allows to experimentally assess the cues that are relevant in a discrimination task: **Auditory Classification Images**



## Experimental design:

- Noise fixed, variable speech level
- 1-up, 1-down weighted procedure (Kaernbach, 1991) to track 70.7% correct responses: down step = 1 dB, up step = 2.4 dB
- Little level roving of  $\pm 2.5$  dB

Participant's response

	'ba'	'da'
Presented 'ba' ("target")	Hit	Miss
Presented 'da' ("no target")	False Alarm	Correct Rejection

ACI is built from  $(H+FA) - (M+CR)$

One weighting function per participant, <sup>3</sup> the Auditory Classification Image (ACI)

## ACIs and the concept of listeners' template

- From SDT: A template is an "expected signal  $s(t)$ " (Green & Swets, 1966), a reference for a detection using, e.g., template matching.
- The **template** is thus a weighting function:

$$y(T) = \int_{-\infty}^T x(\alpha) s(\alpha) d\alpha \quad (6.33)$$

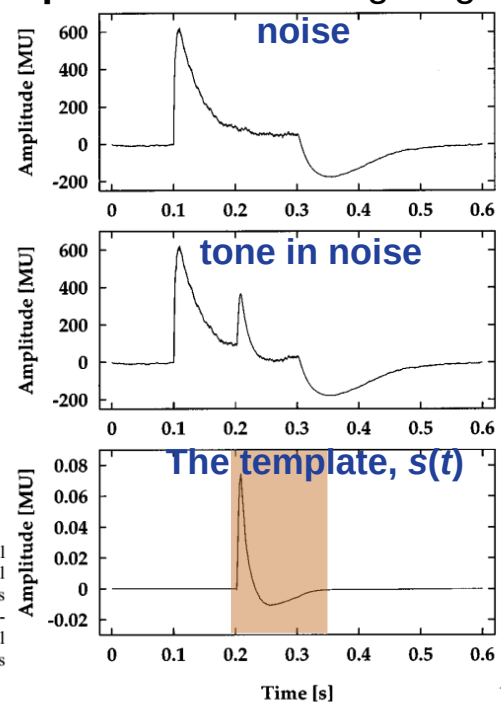


Fig. 7 from Dau *et al.* (1996)

FIG. 7. Top panel: Internal representation of the masker (without internal noise). Middle panel: Internal representation of the masker plus test signal (without internal noise). The test signal was a Hanning windowed 10-ms 3-kHz sinusoid that was added to the temporal center of the masker waveform of Fig. 4. Bottom panel: Normalized difference between the internal representations of the masker plus test signal and the masker alone. This difference is called the template.

## ACIs and the concept of listeners' template

- The previous concept of template:
  - Represents the average weighting function of an artificial listener
  - This concept has been used for a diversity of psychoacoustical tasks, including masking experiments, AM detection, and speech intelligibility

Dau et al. (1996), Jepsen et al. (2008), Biberger & Ewert, (2016), Relação-Iborra et al. (2019)

**A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements**  
 Torsten Dau<sup>a)</sup> and Dirk Püschel<sup>b)</sup>  
 Drittes Physikalisches Institut, Universität Göttingen, Bürgerstr. 42-44, 37073 Göttingen, Germany  
 Armin Kohrausch<sup>c)</sup>  
 Institute for Perception Research (IWR), 35115 Kassel, Germany  
 (Received 20 July 2007; revised 31 March 2008; accepted 17 April 2008)

**Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility**  
 Thomas Biberger<sup>a)</sup> and Stephan D. Ewert  
 Medizinische Physik und Cluster of Excellence Hearing@all, Universität Oldenburg, 26111 Oldenburg, Germany  
 (Received 23 August 2015; revised 15 July 2016; accepted 19 July 2016; published online 15 August 2016)

### A computational model of human auditory signal processing and perception

### A speech-based computational auditory signal processing and perception model

Morten L. Jepsen  
 Centre for Applied Hearing Research, Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Stephan D. Ewert  
 Medizinische Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

Torsten Dau<sup>a)</sup>  
 Centre for Applied Hearing Research, Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 2 August 2007; revised 31 March 2008; accepted 17 April 2008)

A model of computational auditory signal-processing and perception that accounts for various aspects of simultaneous and nonsimultaneous masking in human listeners is presented. The model is based on the modulation filterbank model described by Dau et al. [J. Acoust. Soc. Am. 102, 2892 (1997)] but includes major changes at the peripheral and more central stages of processing. The model contains outer- and middle-ear transformations, a nonlinear basilar-membrane processing

Helia Relação-Iborra,<sup>a)</sup> Johannes Zaar, and Torsten Dau  
 Hearing Systems Section, Department of Health Technology, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

(Received 9 March 2019; revised 29 August 2019; accepted 21 September 2019; published online 13 November 2019)

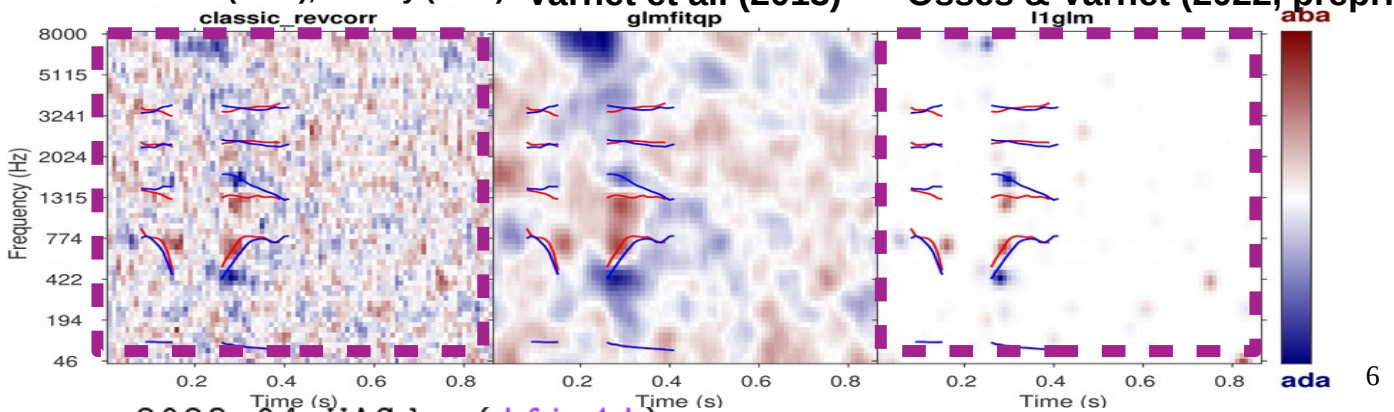
A new speech intelligibility prediction model is presented which is based on the Computational Auditory Signal Processing and Perception model (CASP) of Jepsen, Ewert, and Dau [(2008), J. Acoust. Soc. Am. 124(1), 422–438]. The model combines a non-linear auditory-inspired prepro-

cessing stage. Models extract the long-term (envelope) power to account for SI in fluctuating noise on various time scales (multi-resolution speech-PSM) to account for SI in fluctuating noise on various time scales (multi-resolution speech-PSM) and the mr-sPSM to jointly account for psych. Am. 108, 1181–1196] has been successfully applied to account for SI in fluctuating noise on various time scales (multi-resolution speech-PSM) and the mr-sPSM to jointly account for psych. Am. 134, 436–446]. Here, a generalized model was extended to consider the local AM depth in the relative role of long-term and short-term SNR. The model is shown to account for a large range of speech intelligibility in various types of background noise. <http://dx.doi.org/10.1121/1.4960574>

## ACIs and the concept of listeners' template

- The ACIs are “templates” that are behaviourally derived from the participant’s data.
  - How interpret an ACI?
  - Accuracy of the fitting between ACI predictions and the actual participant responses?
- I will show results on two participants using two ACIs obtained with two methods.

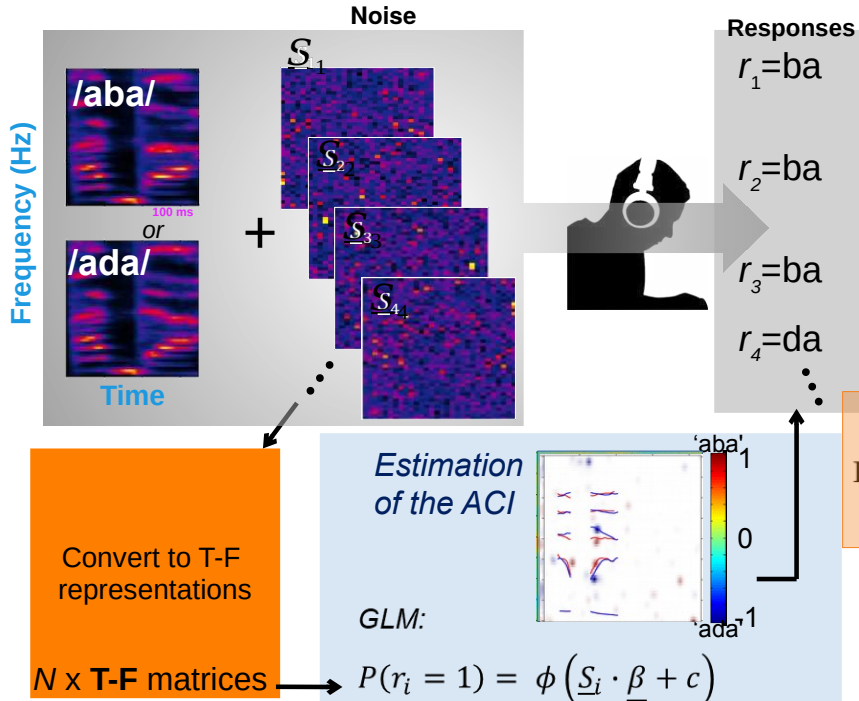
Similar to Ahumada et al. (1971), Murray (2011) Varnet et al. (2013) Osses & Varnet (2022, preprint)



pres\_osses2023\_04\_WASdag('fig1');

# ACIs: What were we expecting? (Ossees & Varnet, 2022, preprint)

- H1: The ACIs can be used to predict “aba” and “ada” responses for each participant
- H2: The ACIs from different participants will be somewhat similar but not completely different



## Goodness-of-fit metrics:

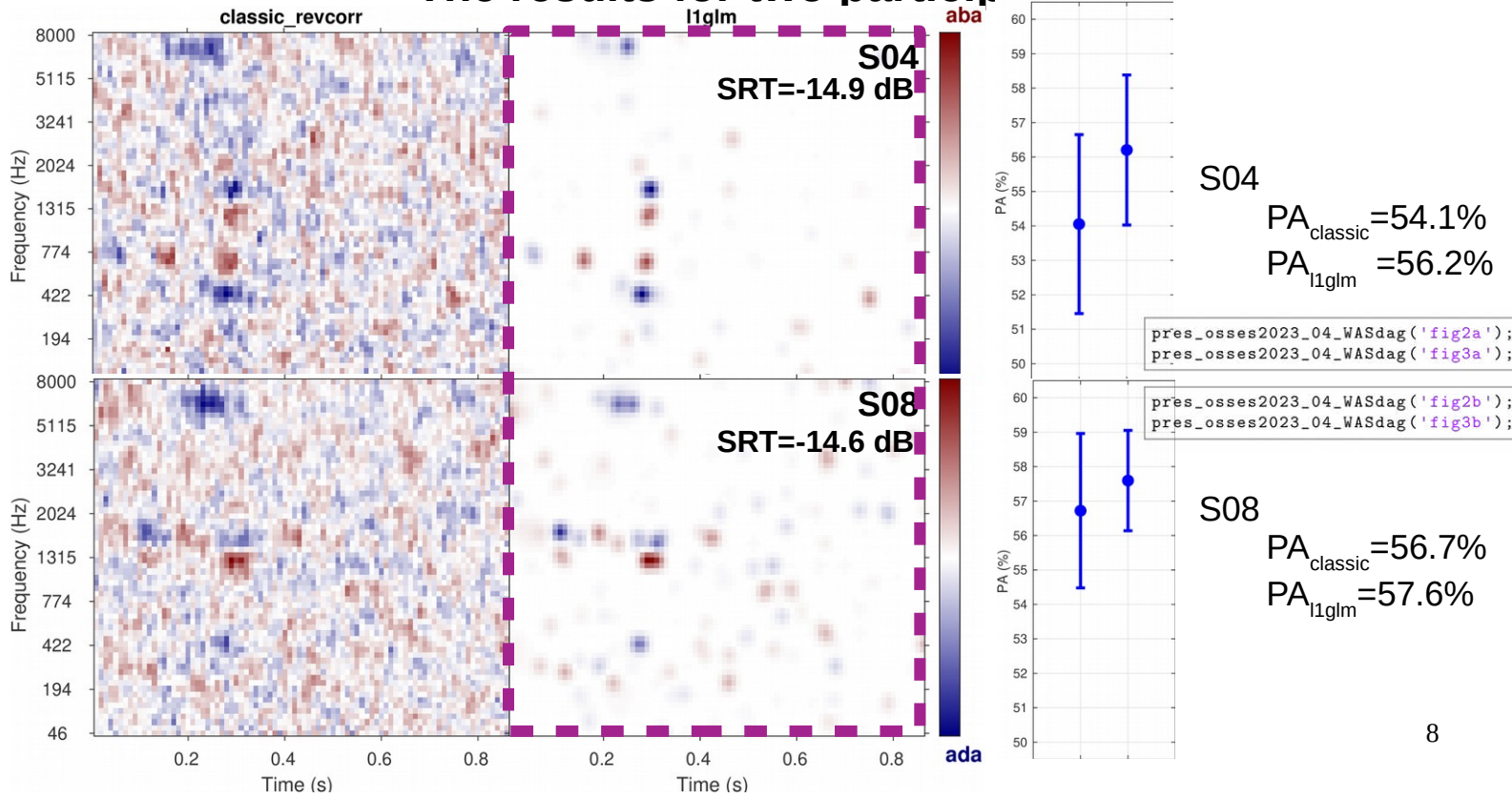
- $R^2$
- Deviance

Used during the fitting

But, the GLM is used for a binary classification

$$PA_i = \begin{cases} 1 & \text{if } P(r_{k,i} = \text{“aba”} | /aba/ \text{ presented}) \geq 0.5 \\ 1 & \text{if } P(r_{k,i} = \text{“aba”} | /ada/ \text{ presented}) < 0.5 \\ 0 & \text{otherwise} \end{cases}$$

## The results for two participants



## Similarity between listening strategies

Possibility to predict “aba” / “ada” judgements from other participants:

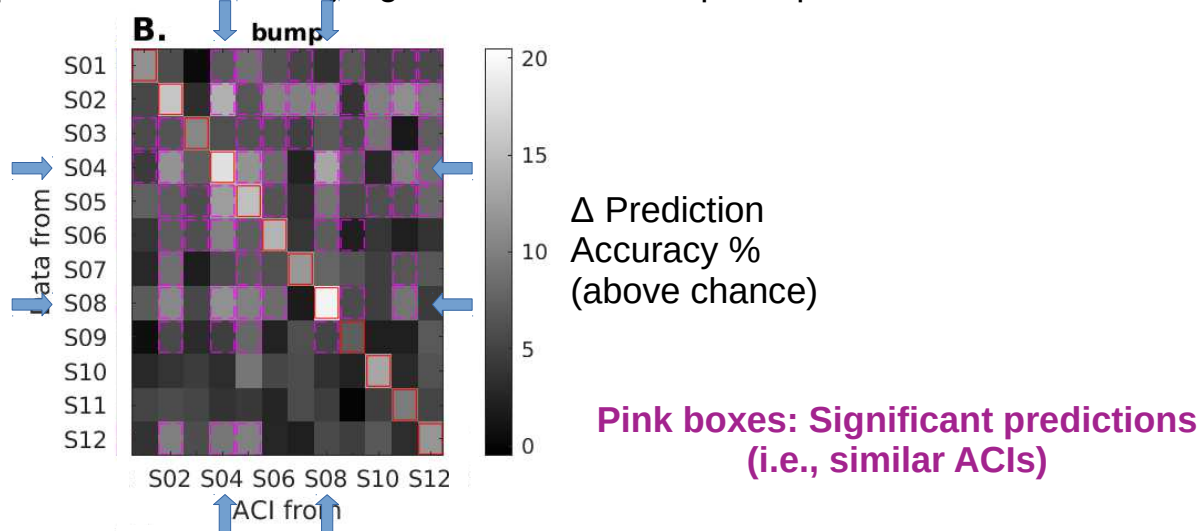


Fig. 8B from Osses & Varnet (2022, preprint)

From our four study hypotheses (2 hypotheses mentioned in the presentation):  
The ACIs across participants were too heterogeneous (H2 was not supported).

9

## Things I did not discuss in detail (1 of 2)

The reverse correlation method requires “random fluctuations” in the trials: Noises

The target sounds (here aba / ada) need to be contrasted: signal preparation

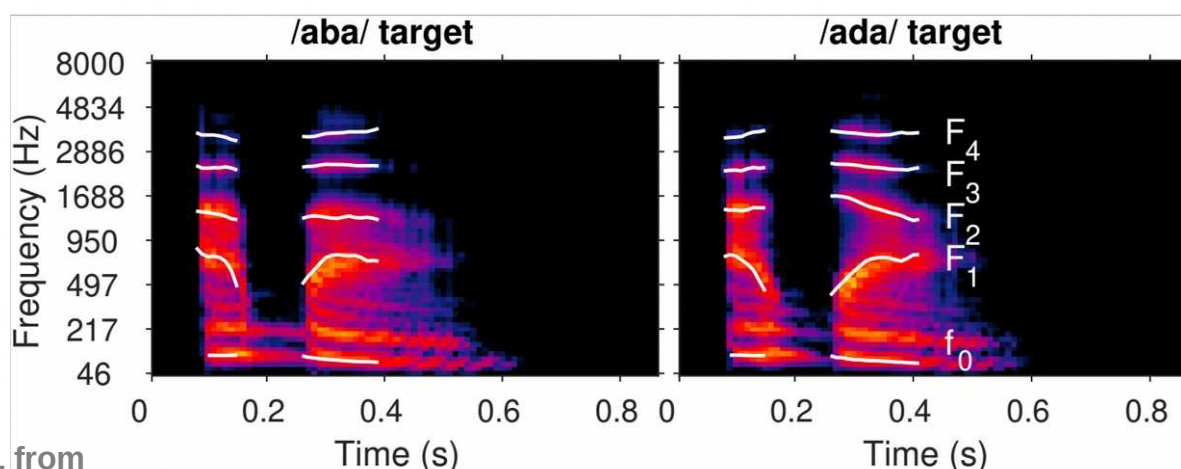


Fig. 1 from Osses & Varnet (2022, preprint)

10

## Things I did not discuss in detail (2 of 2)

Many trials are required (words-in-noise ~3000, prosody ~800)

The noise types that we have evaluated: white noise, SSN, MPS noise, **bump noise**

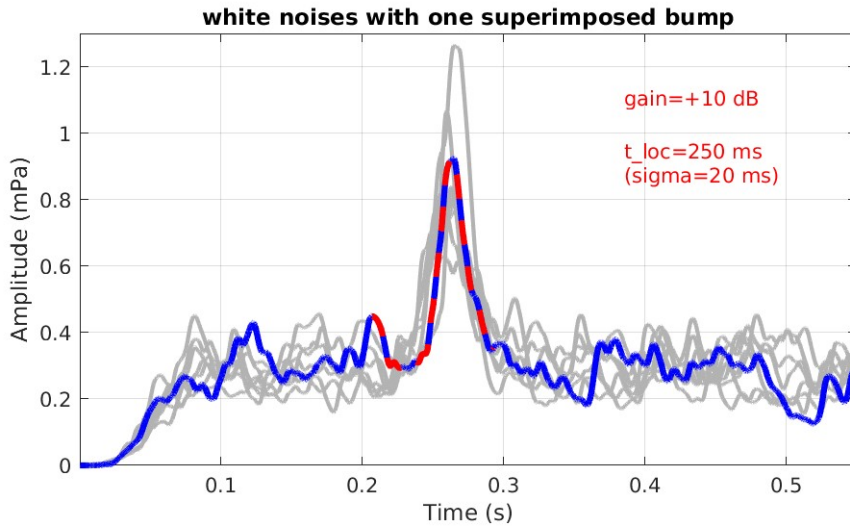


Fig. 2 from  
Osses & Varnet (2022, ARCHES)

11



**Bedankt!**

**Hebben jullie nog vragen of opmerkingen?**

Reproduce some of my figures:

```
1 pres_osses2023_04_WASdag('fig1');
```

```
2 pres_osses2023_04_WASdag('fig2a');
```

```
3 pres_osses2023_04_WASdag('fig3a');
```

```
4 pres_osses2023_04_WASdag('fig2b');
```

```
5 pres_osses2023_04_WASdag('fig3b');
```



Osses & Varnet (2022).

**fastACI toolbox**: the MATLAB toolbox for investigating auditory perception using reverse correlation (v1.2). doi:[10.5281/zenodo.7314014](https://doi.org/10.5281/zenodo.7314014). GitHub: [github.com/aosses-tue/fastACI](https://github.com/aosses-tue/fastACI)

## References

- Ahumada, A., & Lovell, J. (1971). Stimulus features in signal detection. *J. Acoust. Soc. Am.*, 49, 1751–1756. [10.1121/1.1912577](https://doi.org/10.1121/1.1912577)
- Biberger, T., & Ewert, S. (2016). Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility. *J. Acoust. Soc. Am.*, 140, 1023–1038. [10.1121/1.4960574](https://doi.org/10.1121/1.4960574)
- Dau, T., Püschel, D., & Kohlrausch, A. (1996a). A quantitative model of the effective signal processing in the auditory system. I. Model structure. *J. Acoust. Soc. Am.*, 99, 3615–3622. [10.1121/1.414959](https://doi.org/10.1121/1.414959)
- Dau, T., Püschel, D., & Kohlrausch, A. (1996b). A quantitative model of the effective signal processing in the auditory system. II. Simulations and measurements. *J. Acoust. Soc. Am.*, 99(6), 3623–3631. [10.1121/1.414960](https://doi.org/10.1121/1.414960)
- Green, D., & Swets, J. (1966). *Signal Detection Theory and Psychophysics*. John Wiley & Sons Inc.
- Jepsen, M., Ewert, S., & Dau, T. (2008). A computational model of human auditory signal processing and perception. *J. Acoust. Soc. Am.*, 124, 422–438. [10.1121/1.2924135](https://doi.org/10.1121/1.2924135)
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept. Psychophys.*, 49(3), 227–229. [10.3758/BF03214307](https://doi.org/10.3758/BF03214307)
- Murray, R. (2011). Classification images: A review. *J. Vis.*, 11, 1–25. [10.1167/11.5.2](https://doi.org/10.1167/11.5.2)
- Osses, Alejandro, & Varnet, L. (2022b). Using reverse correlation to study individual perception: Including an auditory model in the experimental design loop. *ARCHES Meeting*. [hal-03866646](https://hal.archives-ouvertes.fr/hal-03866646)
- Osses, A., & Varnet, L. (2022a). A microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception. *BioRxiv*, 1–24. [10.1101/2022.12.27.522040](https://doi.org/10.1101/2022.12.27.522040)
- Osses, A., & Varnet, L. (2022). *Raw and post-processed data for the microscopic investigation of the effect of random envelope fluctuations on phoneme-in-noise perception*. [10.5281/zenodo.7476407](https://doi.org/10.5281/zenodo.7476407)
- Osses, Alejandro, & Varnet, L. (2022b). *fastACI toolbox: the MATLAB toolbox for investigating auditory perception using reverse correlation (v1.2)*. [10.5281/zenodo.7314014](https://doi.org/10.5281/zenodo.7314014)
- Relaño-Iborra, H., Zaar, J., & Dau, T. (2019). A speech-based computational auditory signal processing and perception model. *J. Acoust. Soc. Am.*, 146, 3306–3317. [10.1121/1.5129114](https://doi.org/10.1121/1.5129114)
- Varnet, L., Knoblauch, K., Meunier, F., & Hoen, M. (2013). Using auditory classification images for the identification of fine acoustic cues used in speech perception. *Front. Hum. Neurosci.*, 7, 1–12. [10.3389/fnhum.2013.00865](https://doi.org/10.3389/fnhum.2013.00865)