Sources of variability in consonant perception and their auditory correlates

Zaar, Johannes; Dau, Torsten

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Zaar, J., & Dau, T. (2015). Sources of variability in consonant perception and their auditory correlates. Poster session presented at 169th Meeting of the Acoustical Society of America, Pittsburgh, Pa, United States.

Sources of variability in consonant perception and their auditory correlates *(2pSC27)*

Johannes Zaar and Torsten Dau

Hearing Systems Group, Department of Electrical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

BACKGROUND AND OBJECTIVE

Responses obtained in consonant perception experiments typically show a large variability across stimuli of the same phonetic identity (Phatak at *al.*, 2008; Sing & Allen, 2012; Toscano & Allen, 2014).

The present study investigated the influence of different potential sources of this response variability. It was distinguished between *source-induced variability*, referring to perceptual differences caused by acoustical differences in the speech tokens and/or the masking noise tokens, and *receiver-related variability*, referring to perceptual differences caused by within- and across-listener uncertainty. It can be demonstrated that any physical change in the stimuli had a measurable effect. This holds even for slight time-shifts in the steady-state masking-noise waveform. Furthermore, responses obtained with identical stimuli differed substantially across different normal-hearing listeners, while individual listeners were able to reproduce their responses fairly reliably.

To determine how well the source-induced variability is reflected in different auditory-inspired internal representations (IRs), the corresponding perceptual distances were compared to the distances between the IRs of the stimuli. Several variants of an energy-based IR and a modulation-based IR were considered. The results suggest that a normalized modulation-based representation provides the best match to the perceptual data.

EXPERIMENTS

- 15 CVs: /bi, di, fi, gi, hi, ji, ki, li, mi, ni, pi, si, ʃi, ti, vi/
- Presented in white noise @ 12, 6, 0, -6, -12, and -15 dB SNR
- 8 young normal-hearing native Danish listeners

Experiment 1: Speech variability

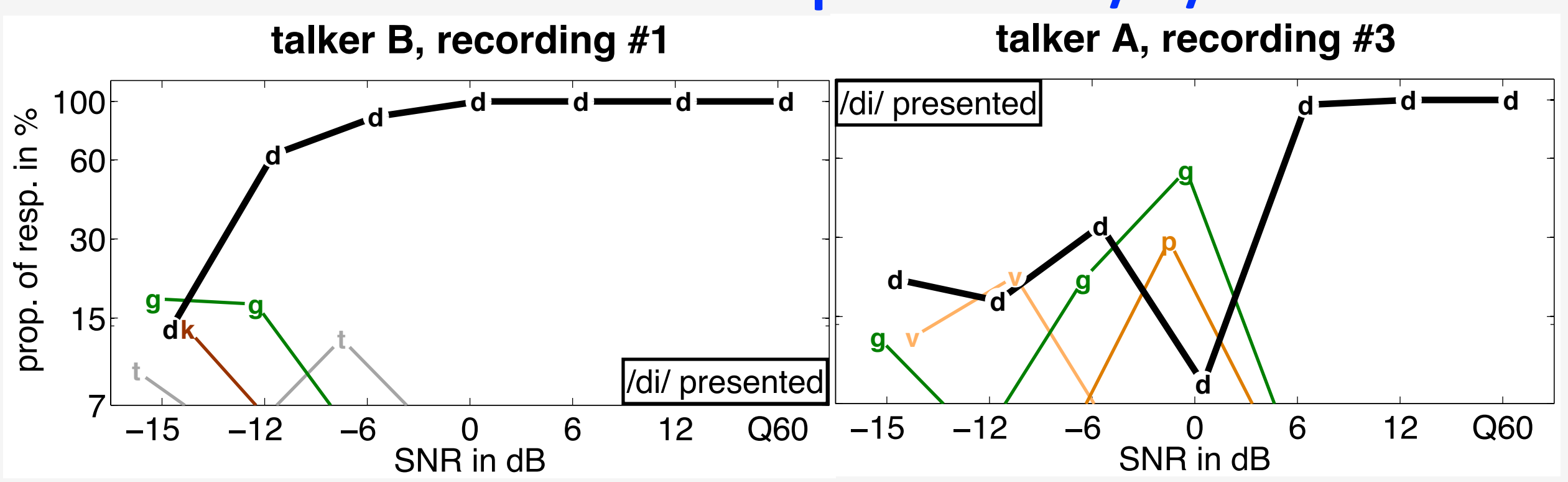
- 3 speech tokens of each CV spoken by a male talker (A)
- 3 speech tokens of each CV spoken by a female talker (B)
- Each token mixed with different frozen noise waveforms at 12, 6, 0, -6, -12, and -15 dB SNR
- Three observations per stimulus and listener

Experiment 2: Noise variability

- 1 speech token of each CV spoken by a male talker
- Each mixed with:
  - ➢ Frozen noise "A"
  - ➢ Frozen noise "B" (noise "A" shifted by 100 ms)
  - ➢ Random noise
- At 12, 6, 0, -6, -12, and -15 dB SNR
- Different frozen noises used for the different tokens
- Re-test with a subset of 4 listeners
- Five observations per stimulus and listener

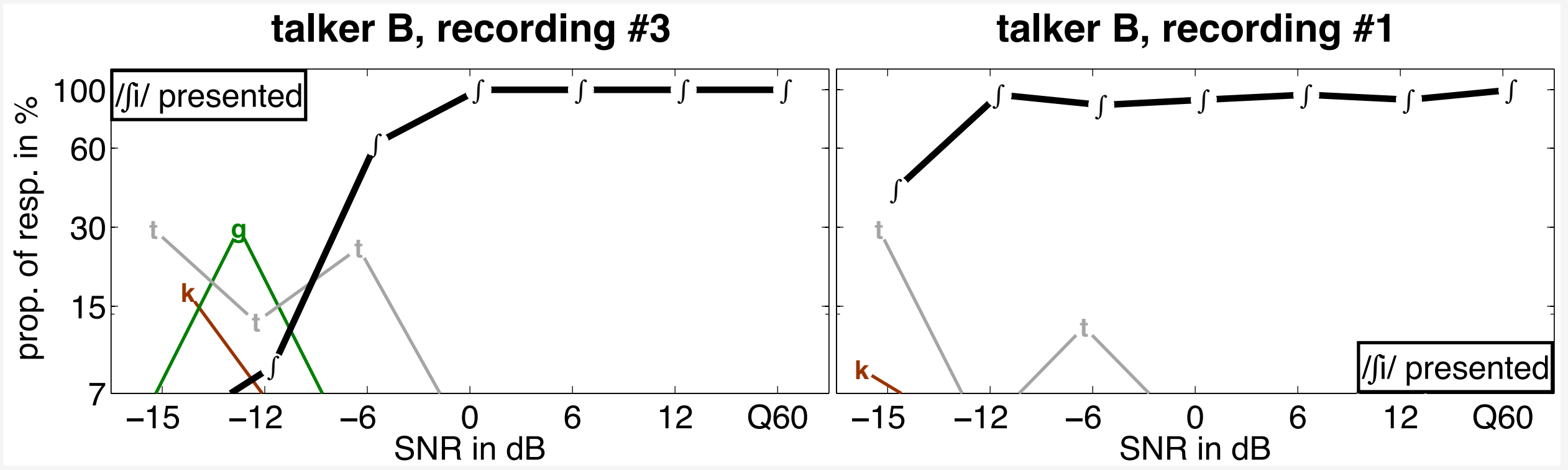
SELECTED RESULTS

Across-talker comparison for /di/



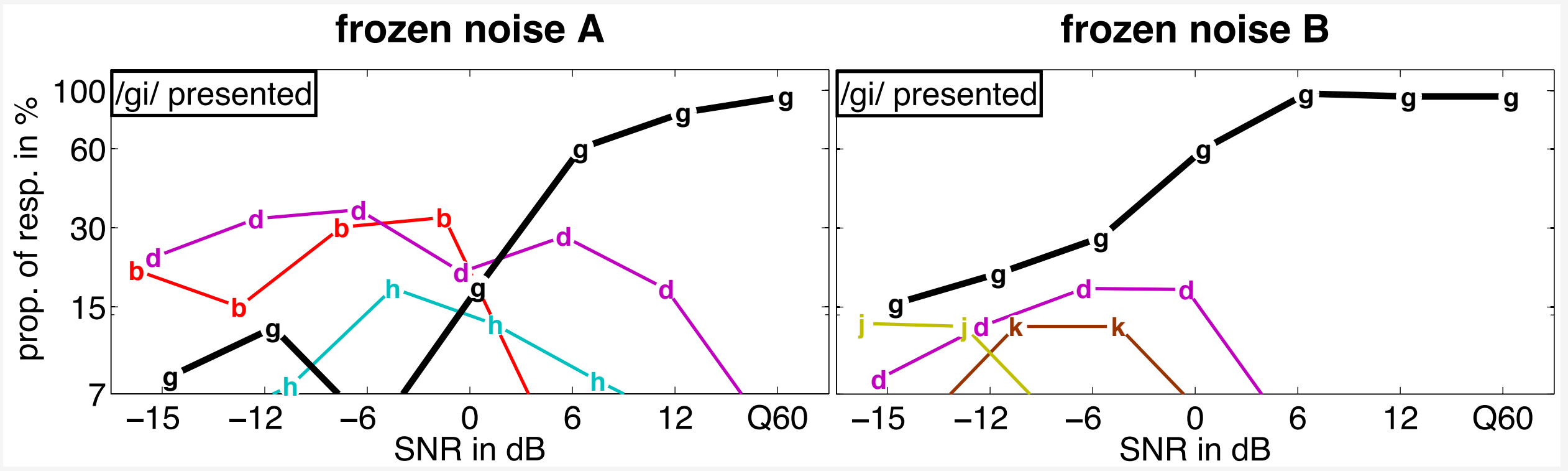
⇒ **Large influence of across-talker articulatory differences**

Within-talker comparison for /ʃi/



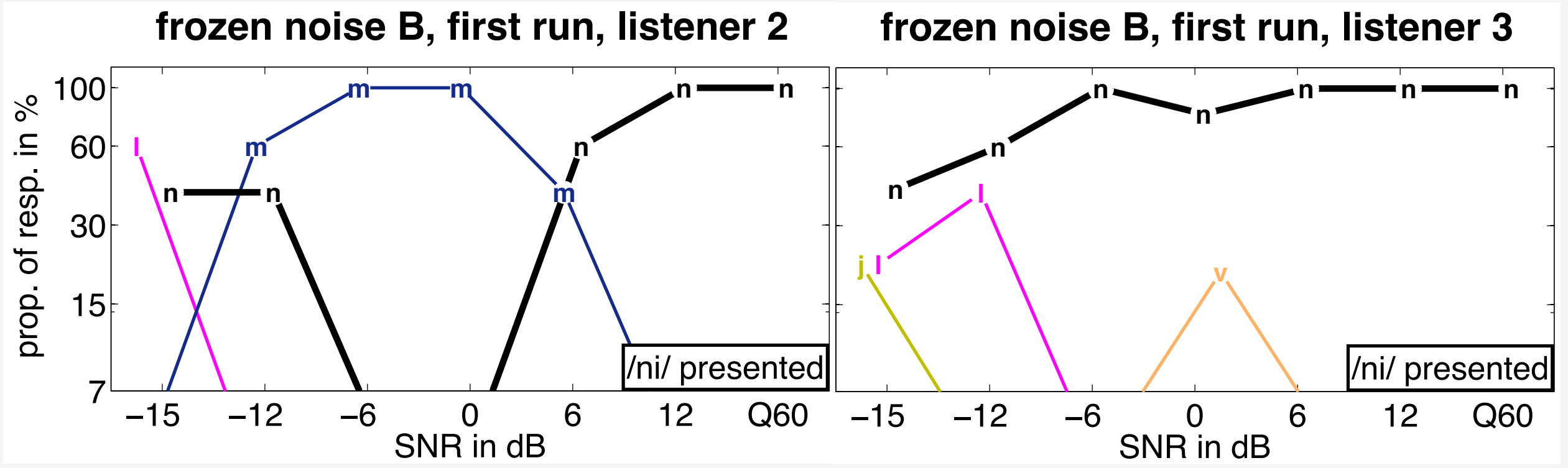
⇒ **Large influence of within-talker articulatory differences**

Across noise-token comparison for /gi/



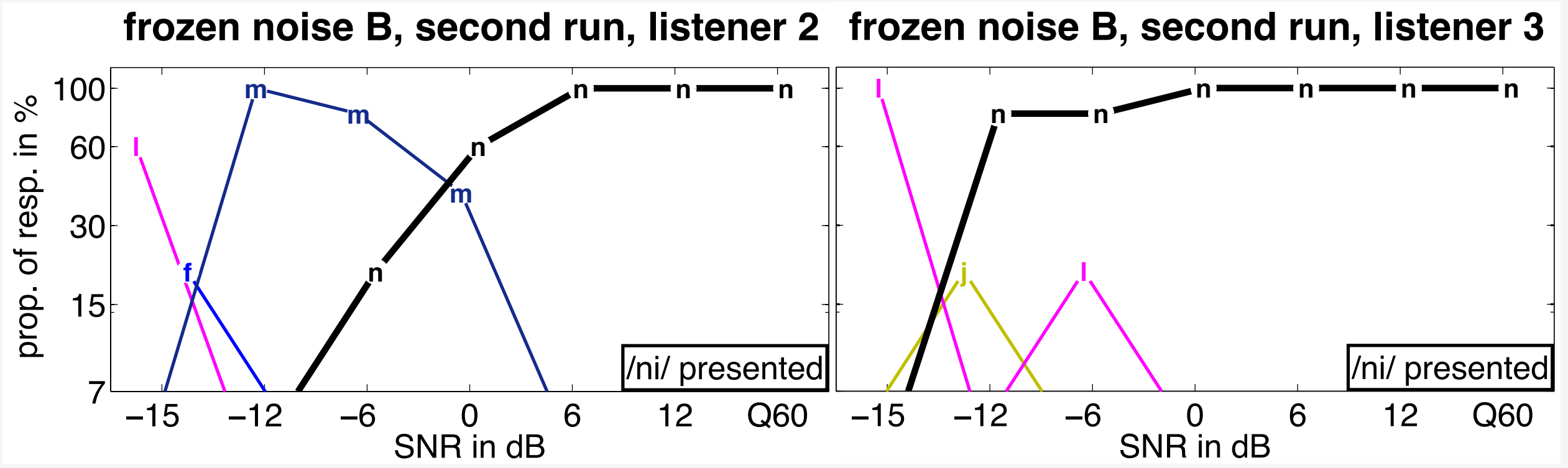
⇒ **Considerable influence even of a 100-ms time shift in the masking noise waveform**

Across-listener comparison for /ni/



⇒ **Large influence of across-listener differences for identical stimuli**

Re-test data for /ni/, same listeners



⇒ **Good reproducibility for individual listeners in test and re-test (for identical stimuli)**

ANALYSIS

Perceptual distance definition

To quantify the perceptual effect of the considered factors, a measure of the perceptual distance between responses was defined. The responses of a given listener, obtained with a given stimulus, were treated as vectors $r = [p_b, p_d, \ldots, p_v]$, where p_x denotes the proportion of response "x". The perceptual distance between two such response vectors r_1 and r_2 was defined as the normalized angular distance between them:

$$D(r_1, r_2) = \arccos\left(\frac{\langle r_1, r_2 \rangle}{\| r_1 \| \cdot \| r_2 \|}\right) \cdot \frac{100\%}{\pi/2}$$

Perceptual Distance calculation across six factors

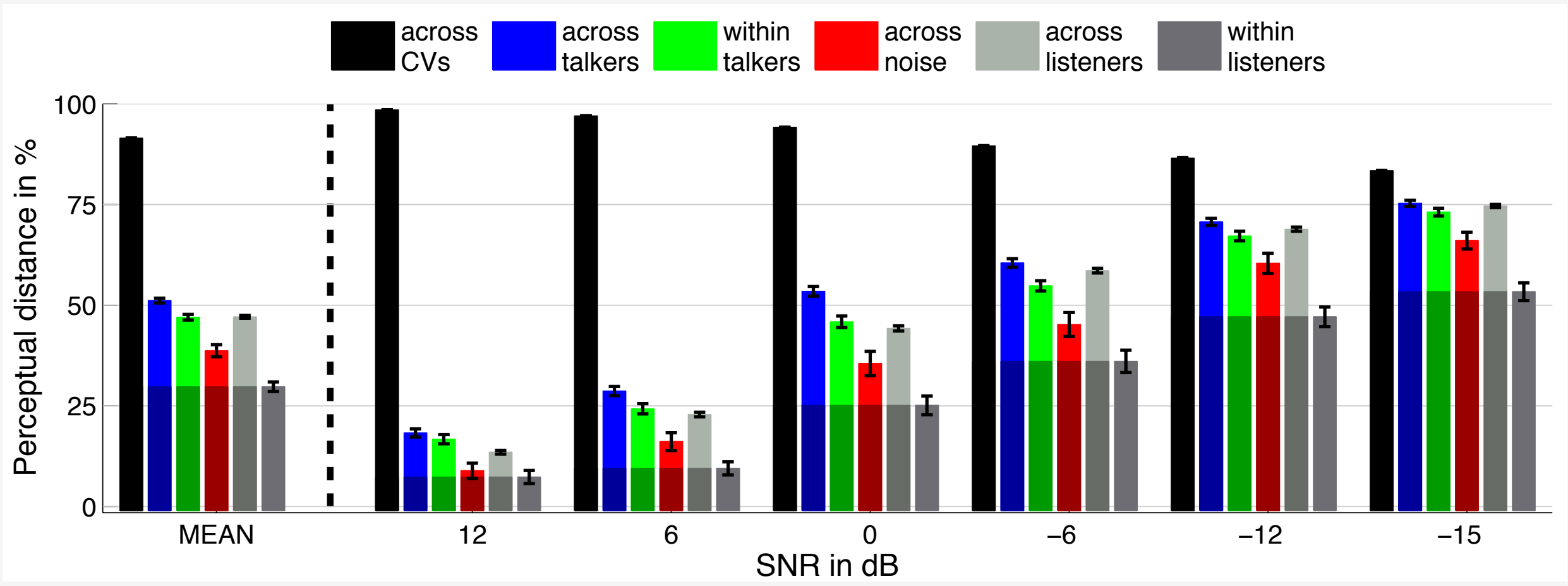
Reference: **across CVs**

Source-induced: **across talkers, within talkers, across noise tokens**

Receiver-related: **across listeners, within listeners**

Apart from the across-CV factor, only responses obtained with stimuli of the same phonetic identity were compared. For each considered factor, the perceptual distance was calculated across all pairwise comparisons of response vectors representative of that factor. The calculation was performed for each SNR condition separately and the individual distance values were averaged across the considered response pairs and across listeners.

QUANTIFICATION OF FACTORS



⇒ **CV-in-noise perception critically depends on**

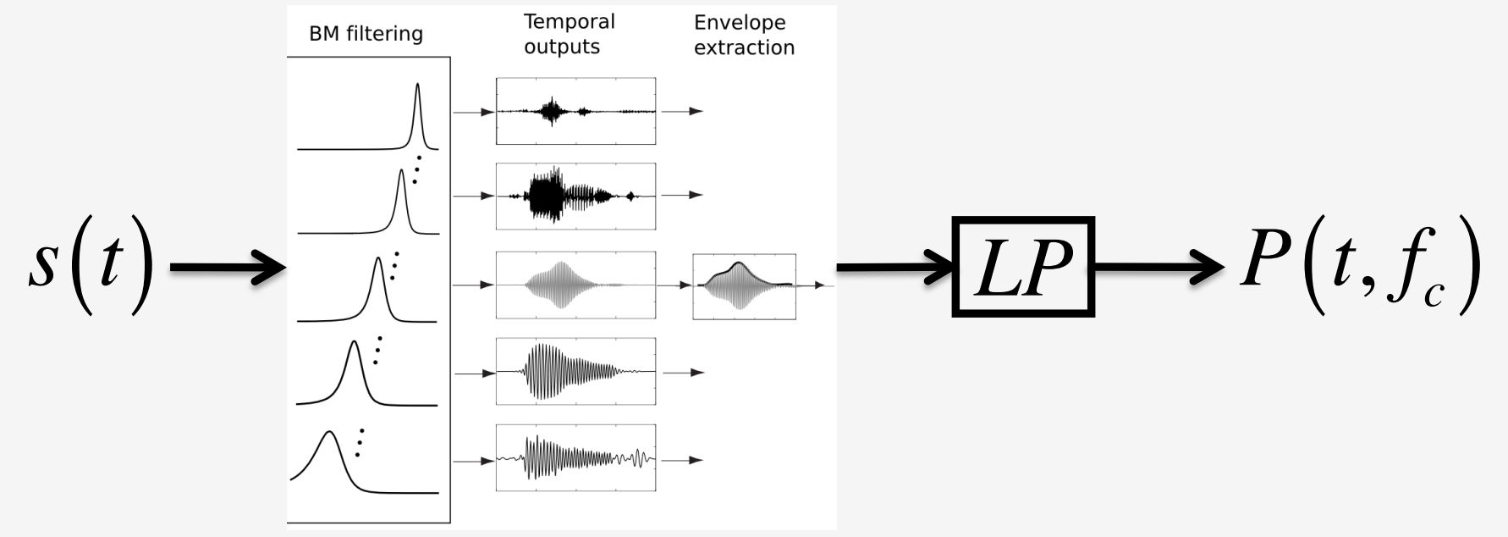
- ➢ **Speech-token specific effects**
- ➢ **Masking-noise-token specific effects**

⇒ **Perceptual distances across listeners much more pronounced than within listeners**

⇒ **Within-listener perceptual distance (internal noise) inversely related to SNR**

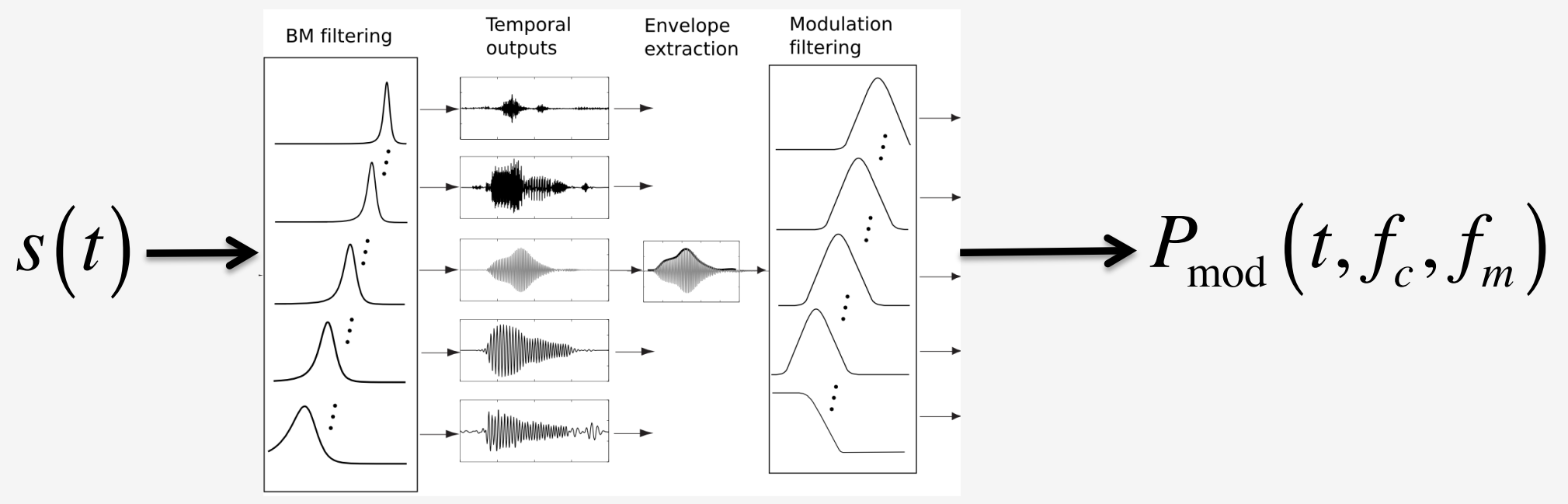
MODELING

Energy-based representation



Configurations: *LP* @ 2, 4, 8, 16 ,32, 64, 128, 256 Hz

Modulation-based representation



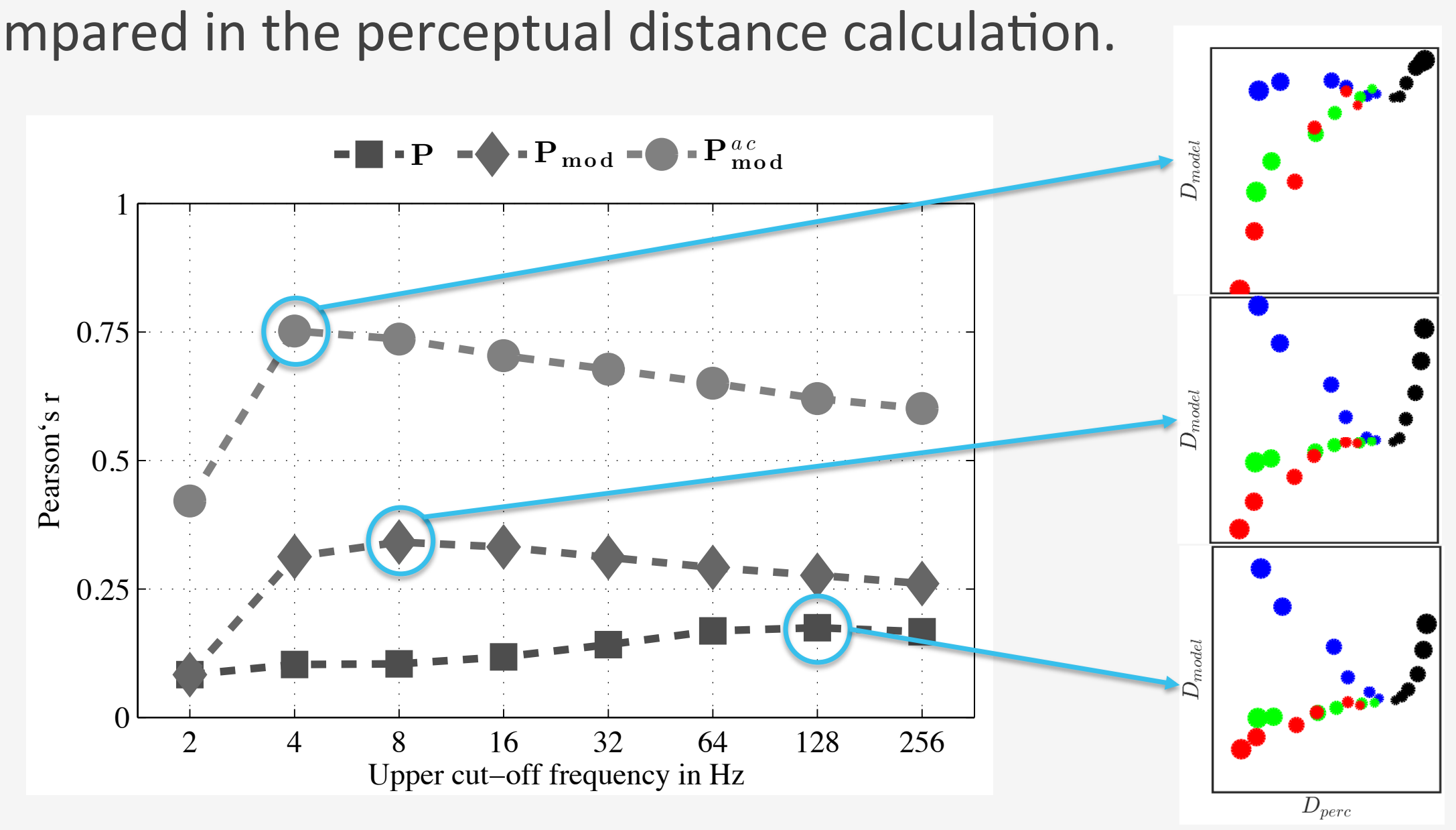
Configurations: f_m = 2 Hz, f_m = [2,4] Hz, …, f_m = [2,4,8,16,32,64,128,256] Hz

AC-coupled modulation-based representation

$$P_{\mathrm{mod}}^{ac} = \frac{P_{\mathrm{mod}}}{DC_{subband}}$$

Modeled distance versus perceptual distance

The modeled distance was calculated between the model representations of the stimuli using a dynamic time warping algorithm. Only the source-induced factors were considered (across CVs, across talkers, within talkers, across noise tokens), using the same pairwise comparisons of stimuli that had been compared in the perceptual distance calculation.



⇒ **AC-coupled modulation representation closest to the perceptual data (least overestimation of across-talker distances)**

ACKNOWLEDGEMENTS

This research was supported by the European Community FP7 Marie Curie Initial Training Network INSPIRE.

REFERENCES

Phatak, S.A., Lovitt, A., Allen, J.B. (**2008**): "Consonant confusions in white noise," J. Acoust. Soc. Am. 124 (2): 1220–1233.

Singh, R. and Allen, J.B. (**2012**): "The Influence of Stop Consonants' Perceptual Features on the Articulation Index Model," J. Acoust. Soc. Am. 131 (4): 3051–3068.

Toscano, J.C. and Allen, J.B. (**2014**): "Across- and Within-Consonant Errors for Isolated Syllables in Noise," Journal of Speech, Language, and Hearing Research 57: 2293–2307.