



# **A computational model of binaural speech intelligibility level difference**

Kalle Palomäki<sup>1</sup> and Guy Brown<sup>2</sup>

1 Adaptive Informatics Research Centre

Helsinki University of Technology

2 Dept. of Computer Science

University of Sheffield



## **Acknowledgement**

- This work was supported in part by the IST Programme of the European Community, under the PASCAL2 Network of Excellence, IST-2007-216886. This publication only reflects the authors' views.



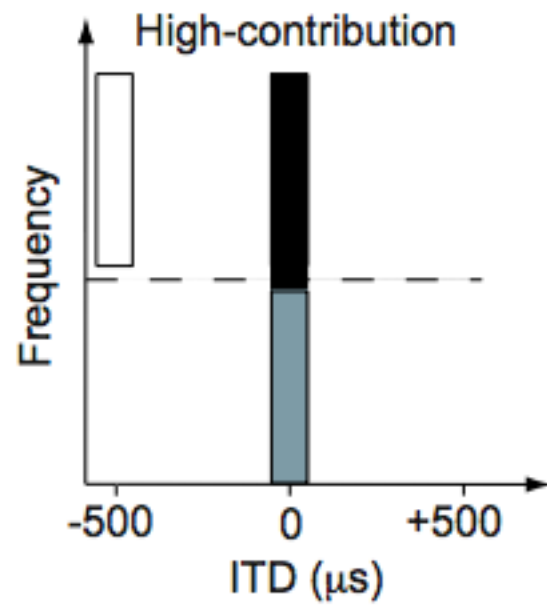
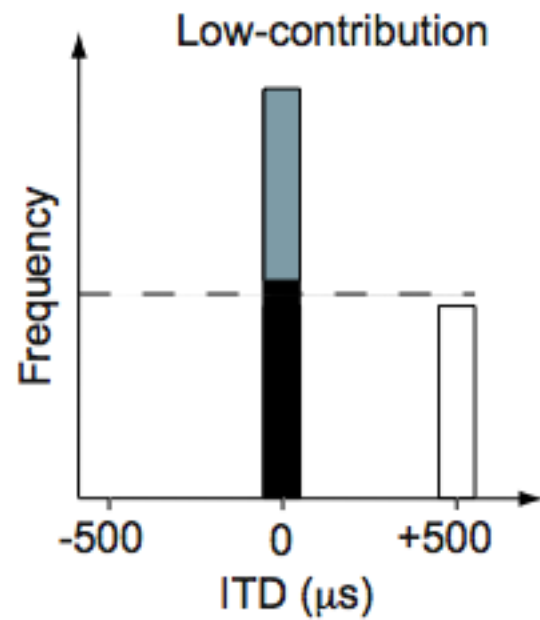
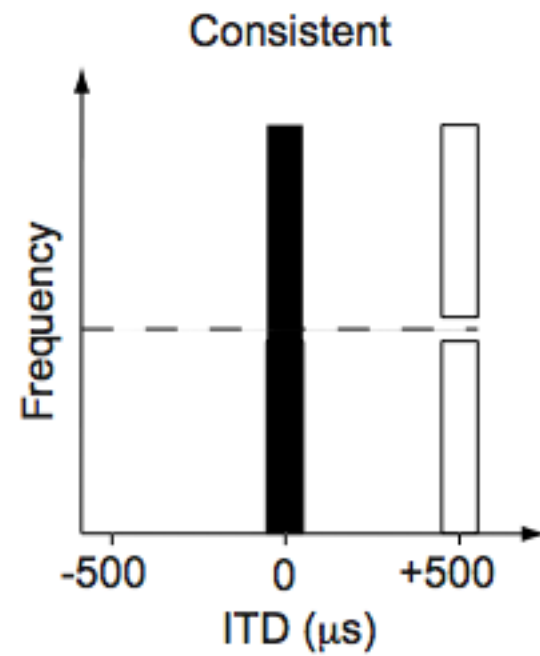
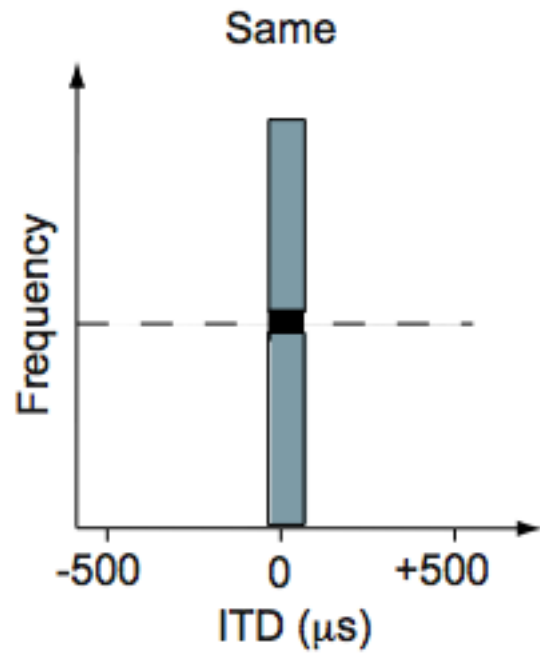
## Introduction

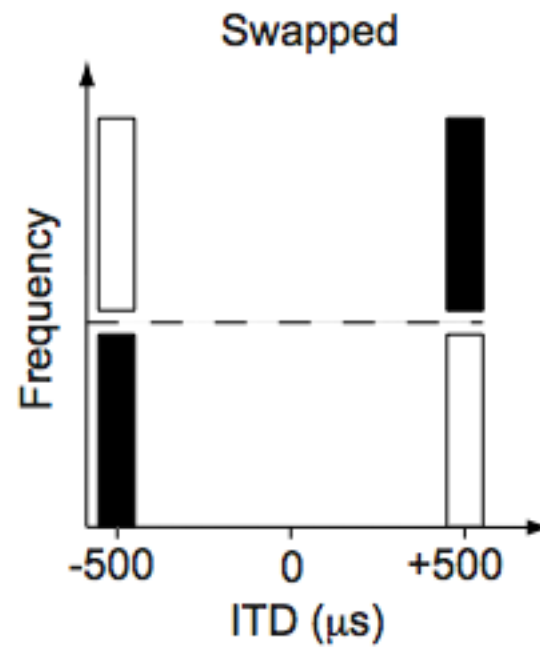
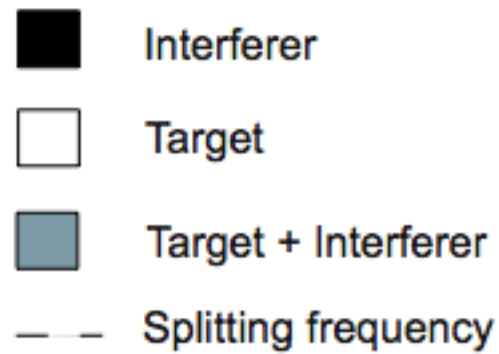
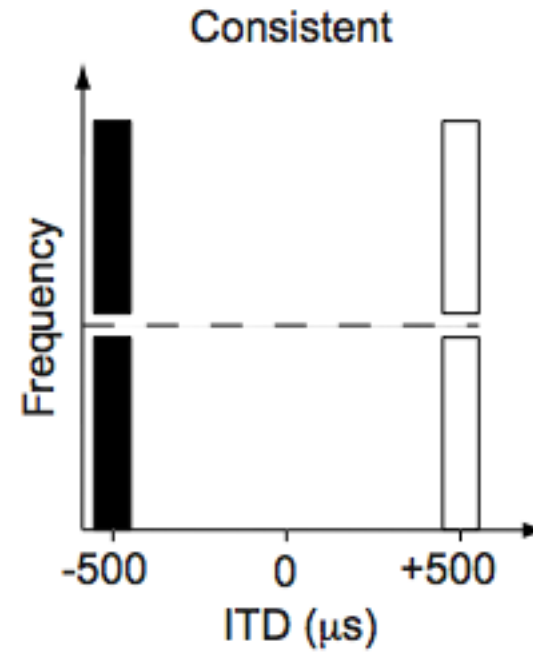
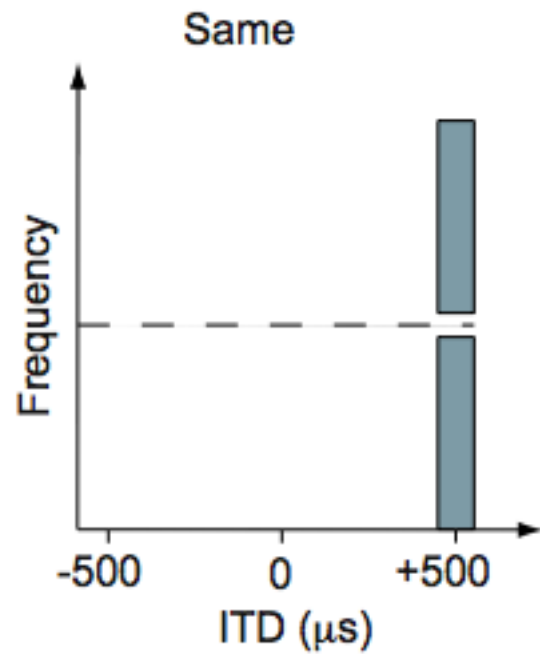
- Human speech intelligibility benefits from spatial separation between target and interference
- Binaural masking level differences (BMLD)
- Unmasking of the target occurs if target and interference are spatially separated
- Durlach (1963) explained BMLD in terms of an equalization cancellation (EC) model
  - Noise interference assumed louder than the signal
  - Equalize left and right signals by their energy
  - Cancel by left – right subtractions over range of delays
  - Best cancellation occurs at interaural time delay (ITD) of noise

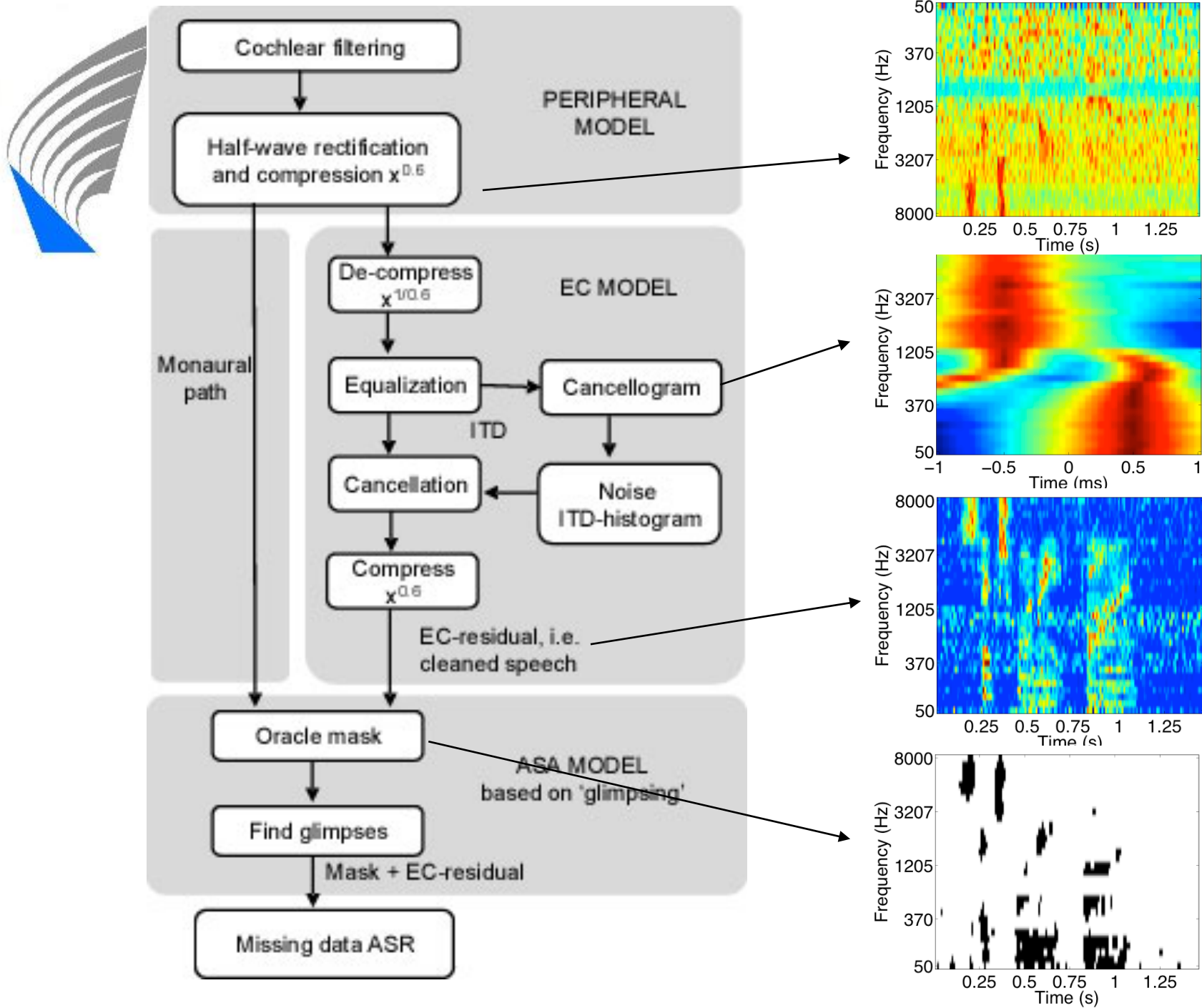


## Introduction

- Does the auditory system group sounds across frequency based on common azimuth or common ITD?
- Psychoacoustic evidence (Culling & Summerfield, 1995; Edmonds & Culling, 2005) suggests that grouping across common ITD does not occur (or is a very weak cue)
- Many computational binaural processors use grouping across common ITD
- Hypothesis: The use of ITD should be based on frequency independent processing









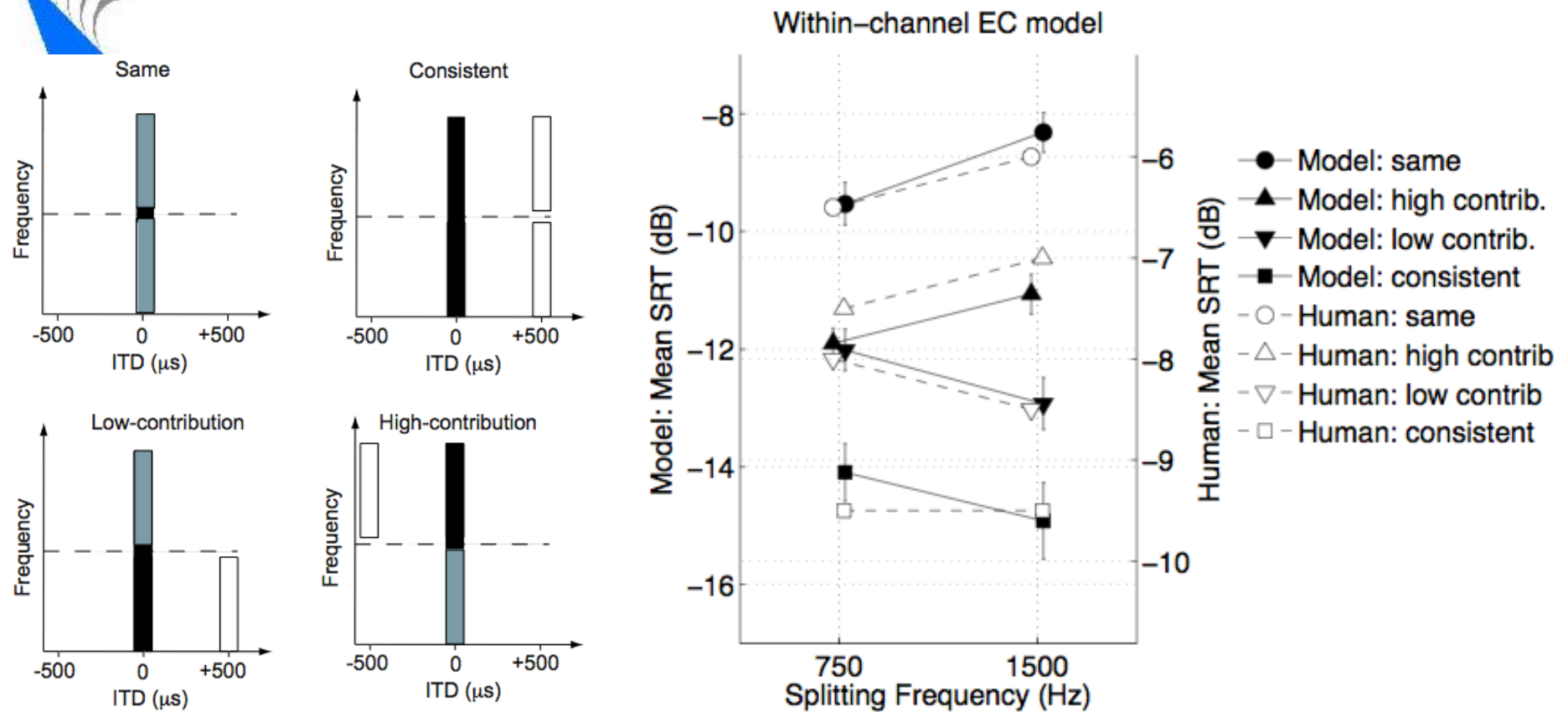
## Speech reception threshold test

- The model is tested against human performance in a speech reception threshold (SRT) test (Edmonds & Culling, 2005)
- Model is used as a subject in the SRT test and compared against 12 subjects
- SRT measures signal-to-noise-ratio (SNR) between speech and noise when half of speech is recognised correctly (we had a slightly relaxed criterion)
  - if accuracy above 75 % increase noise by 2 dB
  - if accuracy below 75 % decrease noise by 2 dB
- Speech: digit strings from TIDigits corpus, one syllable digits (seven and zero excluded), four digits per utterance
- Interference: Speech shaped noise



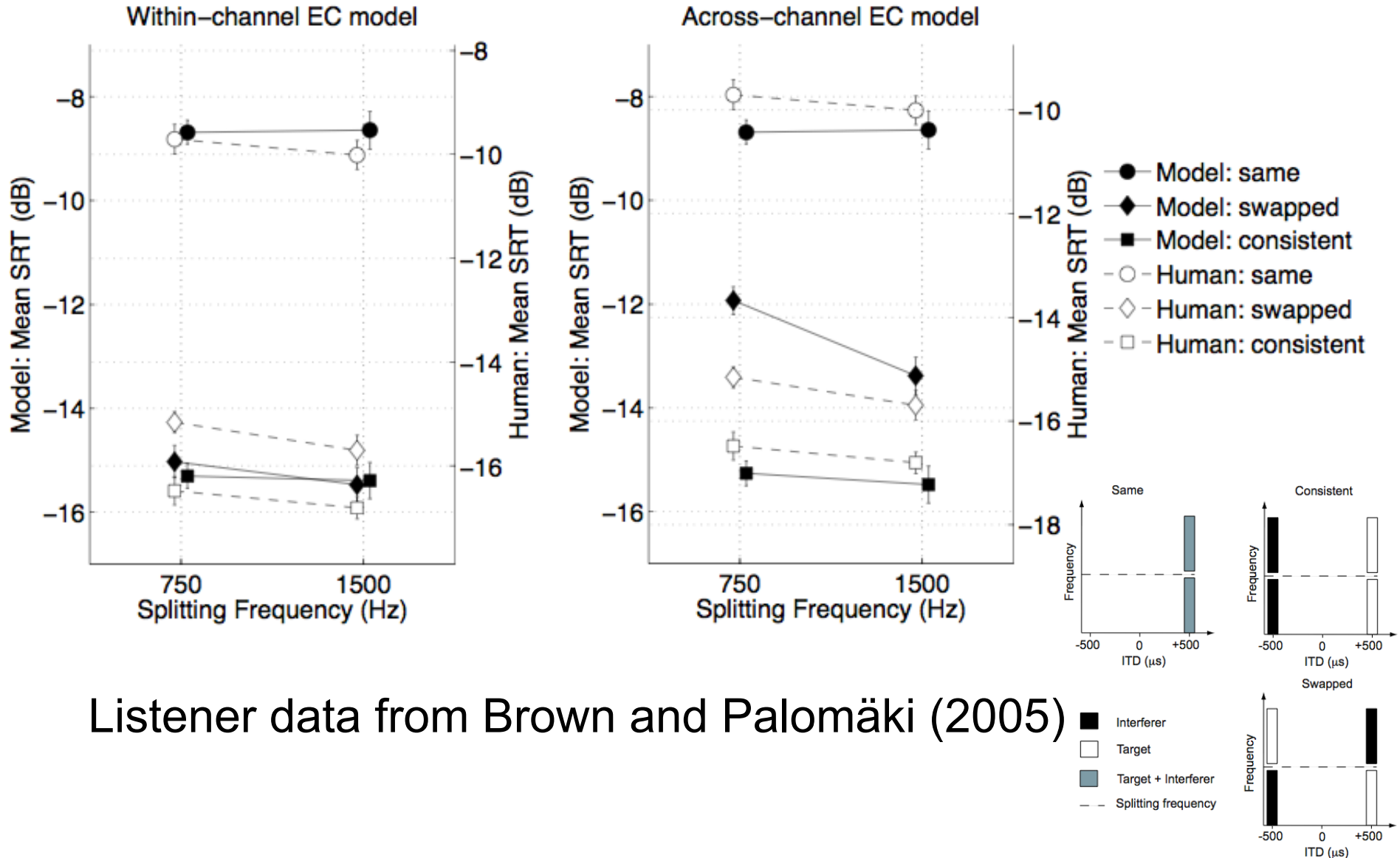


## Edmonds and Culling, Experiment 1: low-, high- contribution



Listener data from Edmonds and Culling (2005)

# Edmonds and Culling, Experiment 3: same, consistent and swapped





## Conclusions

- Edmonds and Culling (2005) suggested entirely frequency independent processing in the use of ITD in sound separation
  - They did not find significant difference between consistent and swapped cases
- Brown and Palomäki (2005) found small (but significant) difference between consistent and swapped conditions
- Unlike Edmonds & Culling we suggest a process that is in between purely frequency dependent and independent approaches
  - Frequency independent model predicts too little difference between consistent and swapped
  - Frequency dependent model predicts too much difference



Edmonds B. A. and Culling J. F. (2005). “The spatial unmasking of speech: evidence for within-channel processing of interaural time delay,” *J. Acoust. Soc. Am.* 117(5), 3069-3078.

Brown G. J. and Palomäki K. J. (2005). “A computational model of the speech reception threshold for laterally separated speech and noise,” *Proc. Interspeech, Lissabon 4th-8th Sep, 2005*, 1753-1756.

Palomäki K. J. and Brown G. J. “A computational model of binaural speech recognition: role of across-frequency vs. within-frequency processing and internal noise” Submitted.