Consonant confusions in frozen and random white noise

Zaar, Johannes; Jørgensen, Søren; Dau, Torsten

Publication date:
2014

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

Citation (APA):
Zaar, J., Jørgensen, S., & Dau, T. (2014). Consonant confusions in frozen and random white noise. Poster session presented at Concepts and computational models of robust bottom-up signal encoding, Copenhagen, Denmark.

Consonant confusions in frozen and random white noise

Johannes Zaar, Søren Jørgensen, Torsten Dau

Centre for Applied Hearing Research, Technical University of Denmark

Introduction

When speech intelligibility is degraded due to masking by background noise or distortion by transmission channels, a considerable part of this degradation is related to the consonants becoming unintelligible or ambiguous. Due to their short duration and low energy, consonants are more easily masked than vowels; at the same time, they carry a large amount of speech information and should hence be maintained (e.g. when passed through transmission channels) or restored (e.g. in signal enhancement algorithms or hearing aid signal processing).

Many studies have investigated consonant perception by means of consonant-vowel combinations (CVs) like /ti/, /bi/, etc. Typically, the CVs are **presented in random white noise** (WN) or speech-shaped noise maskers at different signal-to-noise ratios (SNRs). Listeners have to vote for the consonants they hear. The data are then analyzed in terms of (i) **detectability** and (ii) **confusability**. The specific confusions that listeners typically make are of special interest because they reveal the acoustic features used for consonant identification. It has been demonstrated that **different speech tokens of the same CV identity** lead to different confusions [Trevino & Allen, 2013]. Further, it has been shown that the **long-term spectrum of the masking noise** also has an effect [Phatak & Allen, 2007; Phatak et al., 2008].

Research Questions

In an attempt to reveal additional factors that might influence consonant perception, this study investigates whether there is an effect of the individual noise generations. This relates to the following two questions:

1) "Does the percept of a given CV speech token differ when presented in different masking noise realizations?"

2) "Do listeners respond more "systematically" for frozen noise than for random noise maskers?"

Experimental method

Speech tokens

- 15 Danish CVs /bi/, /di/, /fi/, /gi/, /hi/, /ji/, /ki/, /li/, /mi/, /ni/, /pi/, /si/, /Sji/, /ti/, /vi/
- Only one recording of each CV spoken by the same male talker

SNR conditions

- Speech tokens were presented **in quiet and in white noise** at SNRs of 12 dB, 6 dB, 0 dB, -6 dB, -12 dB, and -15 dB

Test subjects

- 8 young normal-hearing native Danish speakers

Masking noise conditions

1.CV & frozen WN "A"

➢ For each CV, a different frozen WN "A" was *generated*

2.CV & frozen WN "B"

➢ For each CV, frozen WN "B" was created by *temporally shifting frozen WN "A"* by 100 ms

3.CV & random WN

➢ For each CV and each presentation, random WN was *newly generated*

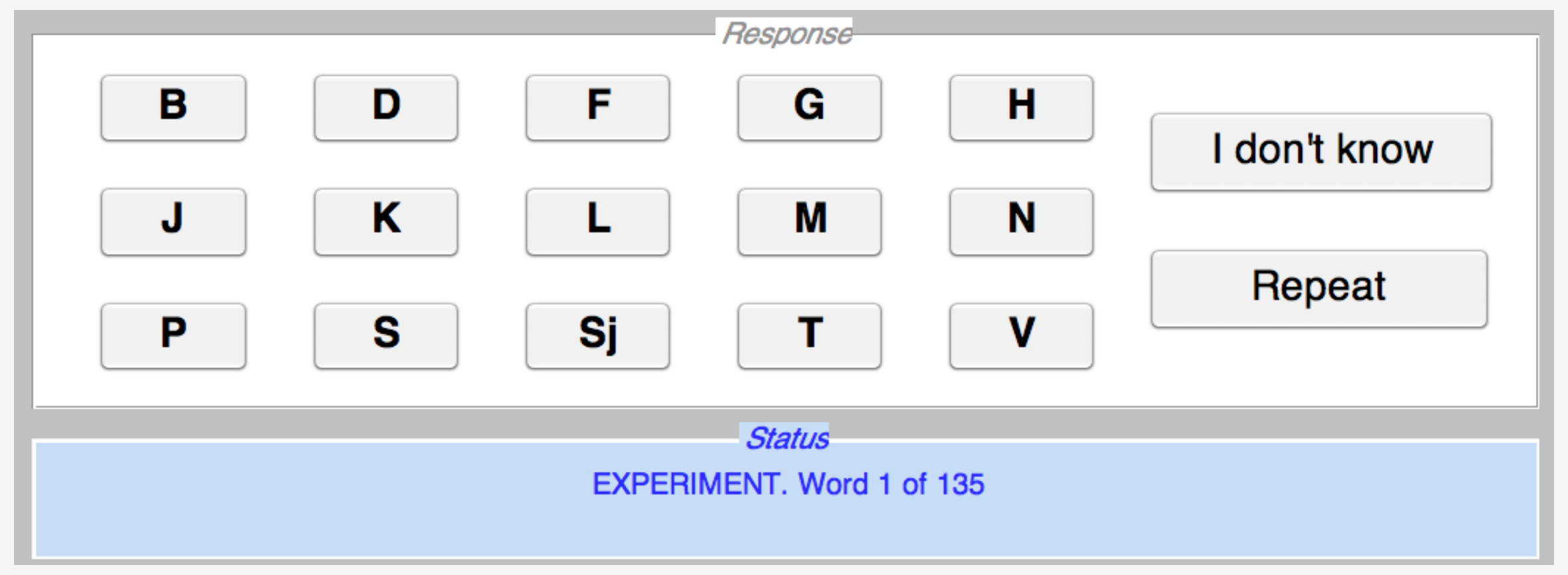

***Figure 1**: test interface.*

Test design

Each SNR condition was tested in one block, consisting of

➢ a **training run:**

each CV presented three times in random white noise (15 x 3 = 45 stimuli),

➢ the actual **experiment:**

each CV presented 5 times in each masking noise condition (15 x 3 x 5 = 225 stimuli).

➢ Listeners could **repeat each stimulus up to 2 times**

➢ Listeners were instructed to vote for the consonant they heard

➢ To minimize listener bias, listeners were instructed to vote for "I don't know" if they heard only the vowel

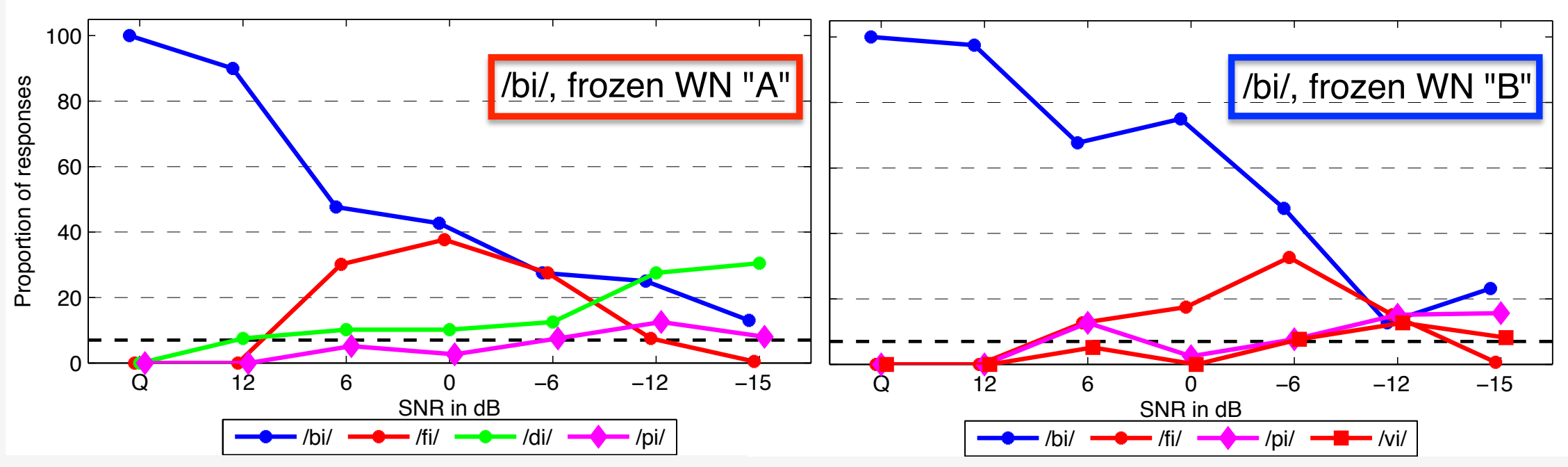

***Figure 2**: across-listener average confusion patterns for **/bi/** in frozen WN "A" and frozen WN "B".*

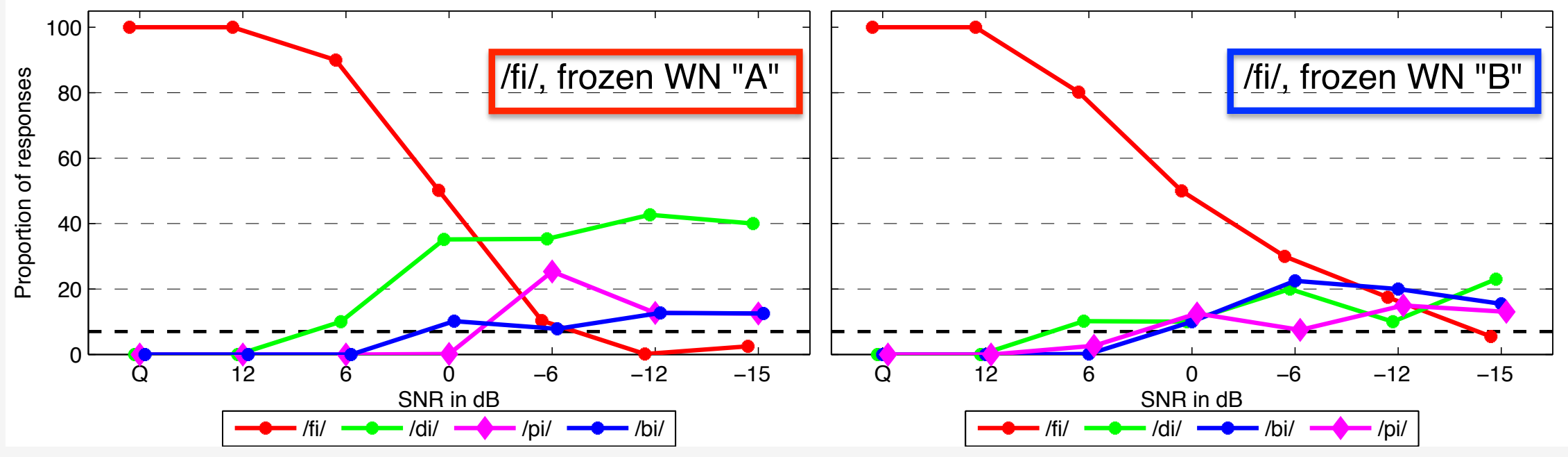

***Figure 3**: across-listener average confusion patterns for **/fi/** in frozen WN "A" and frozen WN "B".*

Results

I. Confusion pattern comparison across frozen noise conditions

- In **Figures 2-5**, the **average results across listeners** are plotted as *confusion patterns (CPs)* [Allen, 2005]
- CPs depict the percentage of responses to a given CV in a specific masking noise as a **function of the SNR**
- The example CPs in Figures 2-5 show only the respective **4 predominant responses** for the sake of clarity
- The example CPs show **huge perceptual differences**
- Note that the **only physical difference is a 100 ms shift in the noise!**

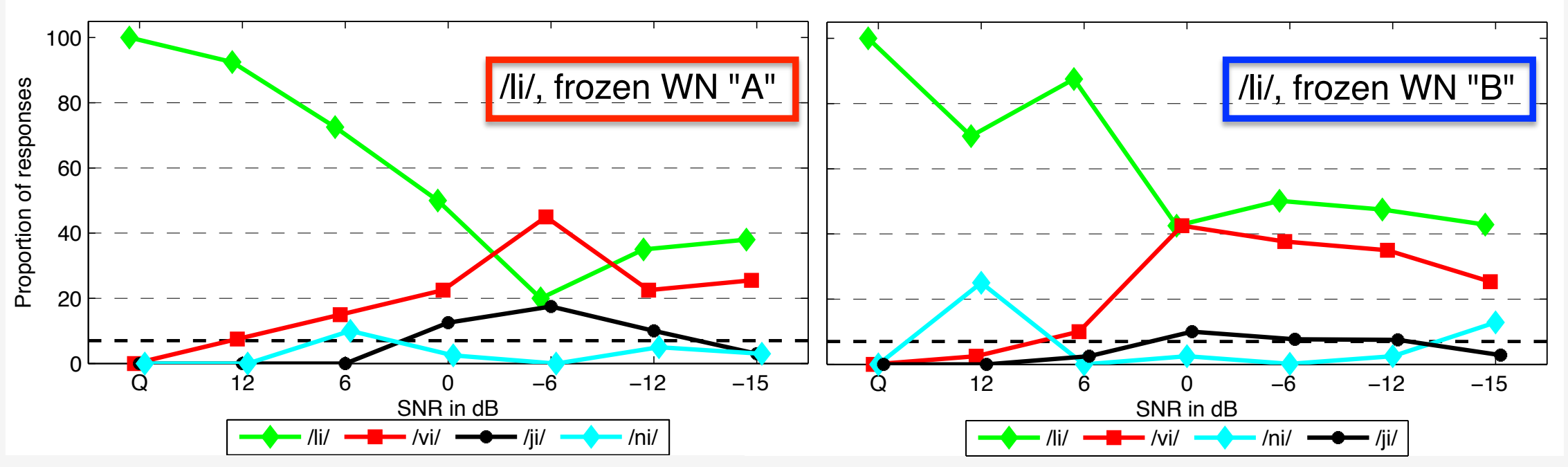

***Figure 4**: across-listener average confusion patterns for **/li/** in frozen WN "A" and frozen WN "B".*

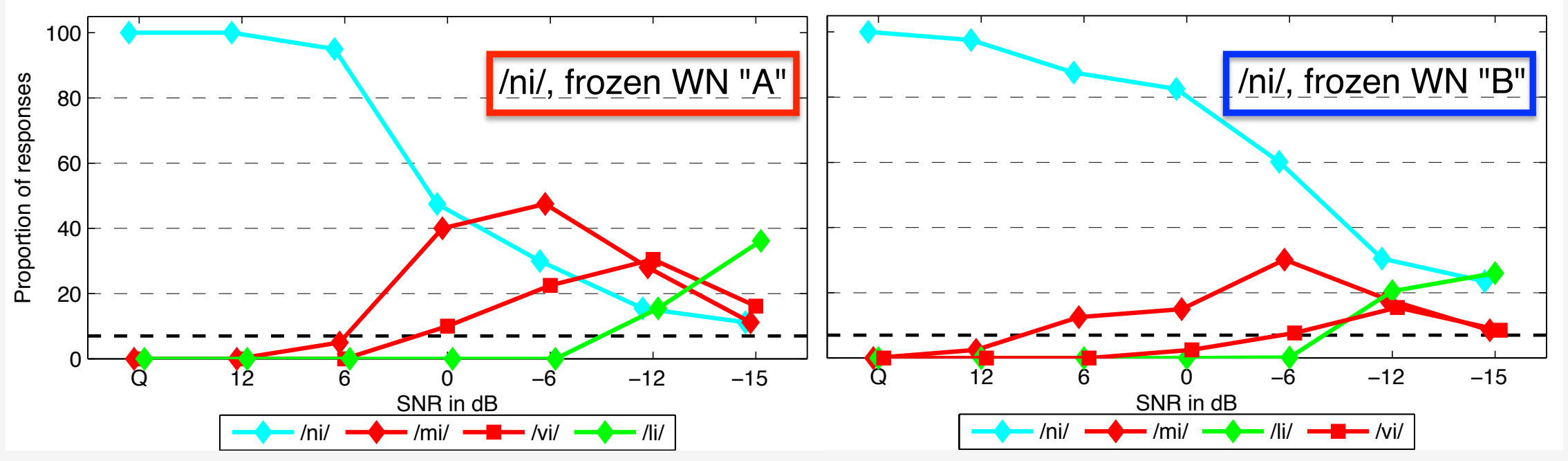

***Figure 5**: across-listener average confusion patterns for **/ni/** in frozen WN "A" and frozen WN "B".*

II. Analysis of within-listener consistency

- In each SNR condition, **each CV was presented 5 times** per masking noise condition
- A listener is considered to be **certain** about his/her response to a given stimulus if he/she makes the **same choice at least 3 out of 5** times
- The analysis does not take into account whether the response is **correct or not**
- The **certainty** is calculated as the **percentage of certain responses:**

$$P_{certain} = \frac{N_{certain}}{N_{stimuli}} \cdot 100\%$$

- The results in Figures 6-8 suggest that **listeners on average respond more systematically when presented with frozen white noise** as compared to random white noise

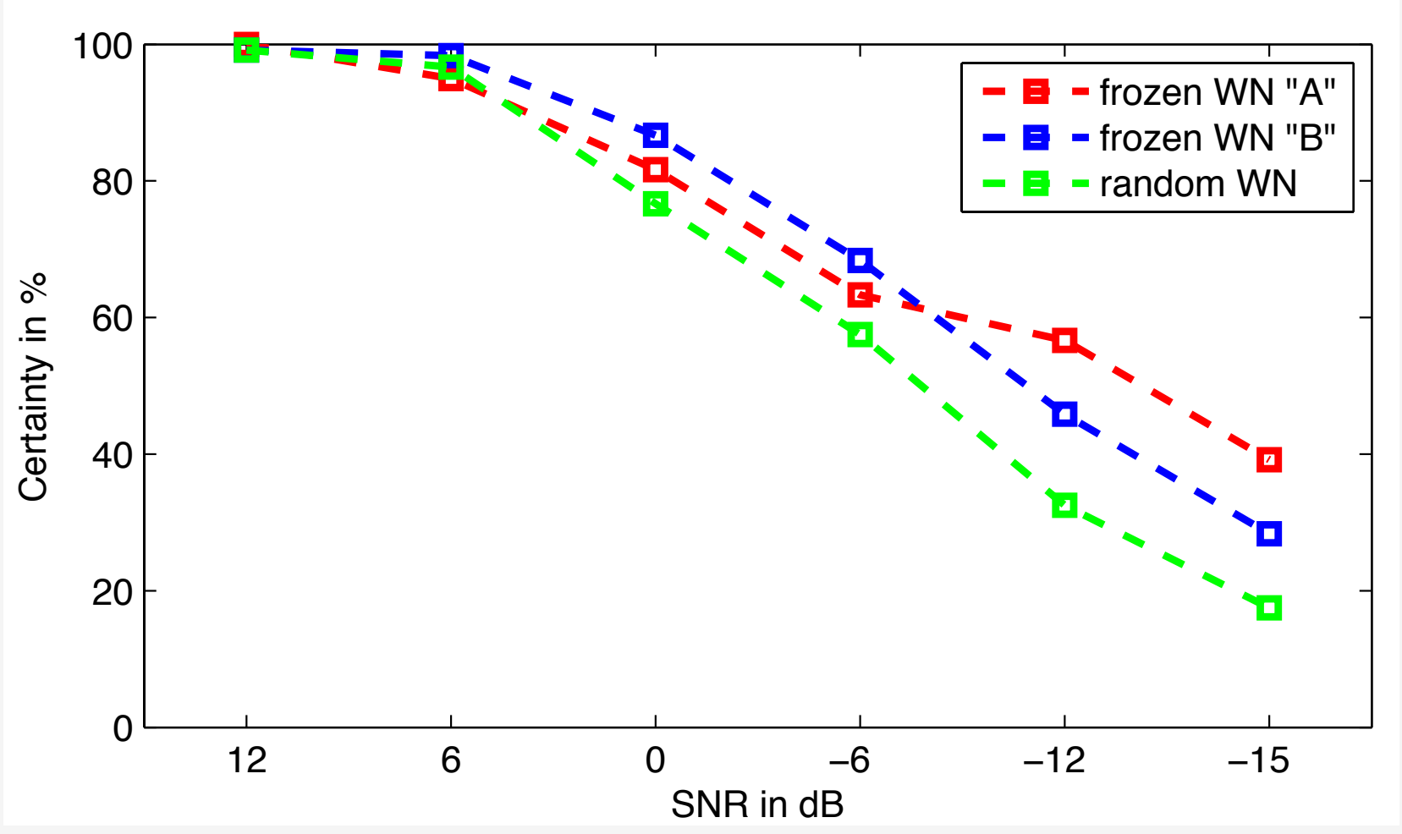

***Figure 6**: average listener certainty $P_{certain}$ in percent as a function of SNR for the 3 masking noise conditions.*

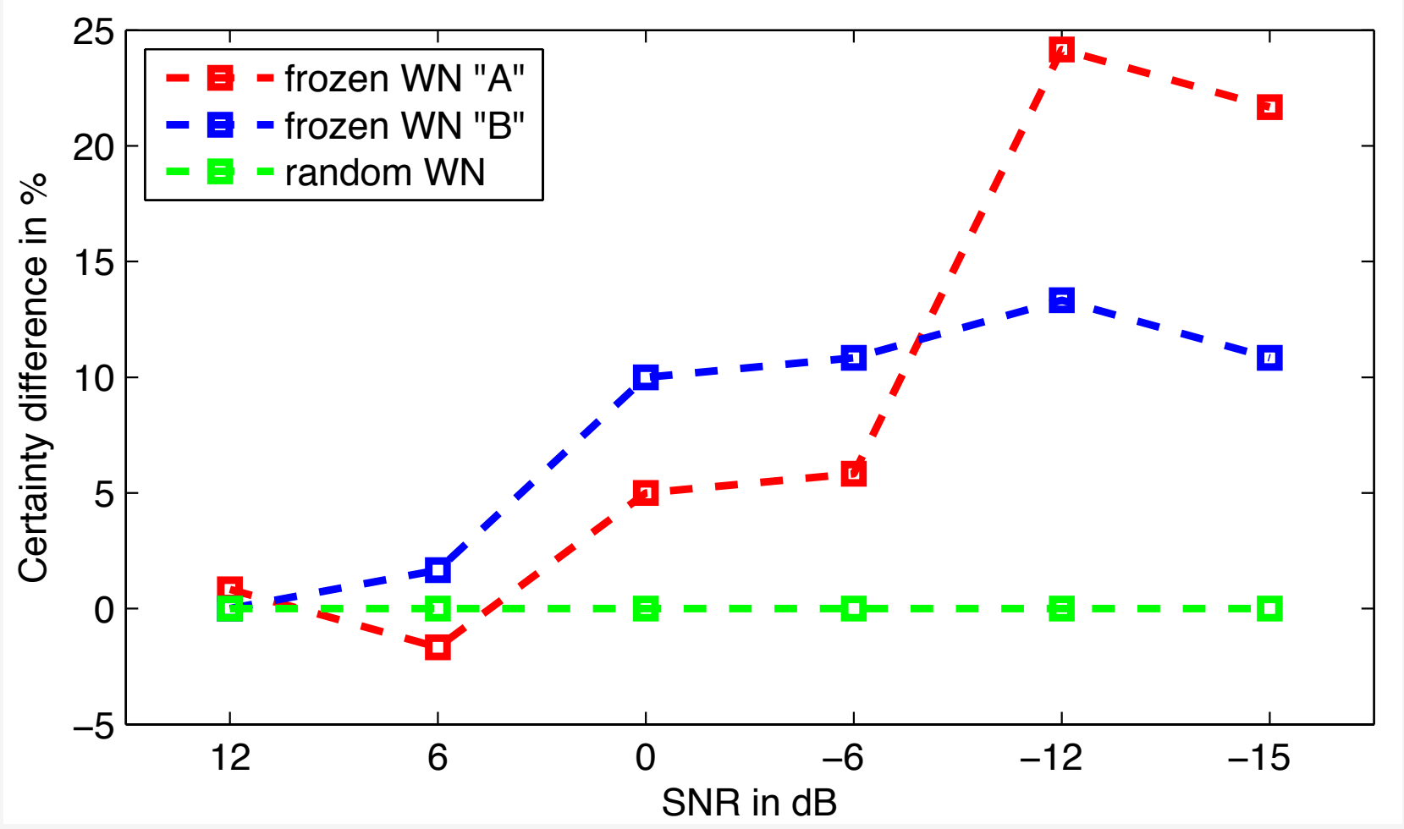

***Figure 7**: Deviation from $P_{certain}$ of the random WN condition (green curve in Figure 6). Positive values indicate more certainty, negative values less certainty.*

Conclusions

- Consonant perception in noise seems to depend strongly on the individual masking noise realization:
  - ➢ Even with the same noise file shifted by 100 ms, huge differences can be observed
- Listeners appear to respond more systematically when presented with frozen noise as compared to random noise
- The effect of the masking noise on a token-by-token basis was not taken into account in prior studies
- The findings presented here are relevant for microscopic speech perception modeling approaches since
  - ➢ the data and stimuli of this study allow for an in-depth acoustic analysis that takes the individual noise tokens into account
- No effect of noise learning is assumed given the presentation of 150 different noise realizations in each experimental block (consisting of 270 stimuli all in all)

References

[Trevino & Allen, 2013] A. Trevino, J. Allen: *Within-consonant perceptual differences in the hearing impaired ear*. J. Ac. Soc. Am. **134** (2013) 607-617.

[Phatak & Allen, 2007] S. Phatak, J. Allen: *Consonant and vowel confusions in speech-weighted noise*. J. Ac. Soc. Am. **121** (2007) 2312-2336.

[Phatak et al., 2008] S. Phatak, A. Lovitt, J. Allen: *Consonant confusions in white noise*. J. Ac. Soc. Am. **124** (2008) 1220-1233.

[Allen, 2005] J. Allen: *Consonant recognition and the articulation index*. J. Ac. Soc. Am. **117** (2005) 2212-2223.