Estimation of signal-to-noise ratios in realistic sound scenarios Supplement: Additional results

Karolina Smeds, Florian Wolters, Martin Rung

Introduction

Based on recordings of everyday listening situations, recordings containing speech in noise were analyzed using a manual signal-to-noise (SNR) estimation method. Recordings were grouped into nine categories based on their background noise characteristics. An uncertainty measure was also developed (see online supplement "The uncertainty of SNR estimates"). The main results, with data from all noise categories, are presented in the main article. Below, additional results graphs for the worse ear and for the various noise categories are presented. For each noise category, the first two figures show frequency-specific SNRs for the better and worse ear. Next, RMS levels and SNRs are shown for each recording for A-weighted and un-weighted speech and noise estimates.

Contents

1 Additional results for worse ear

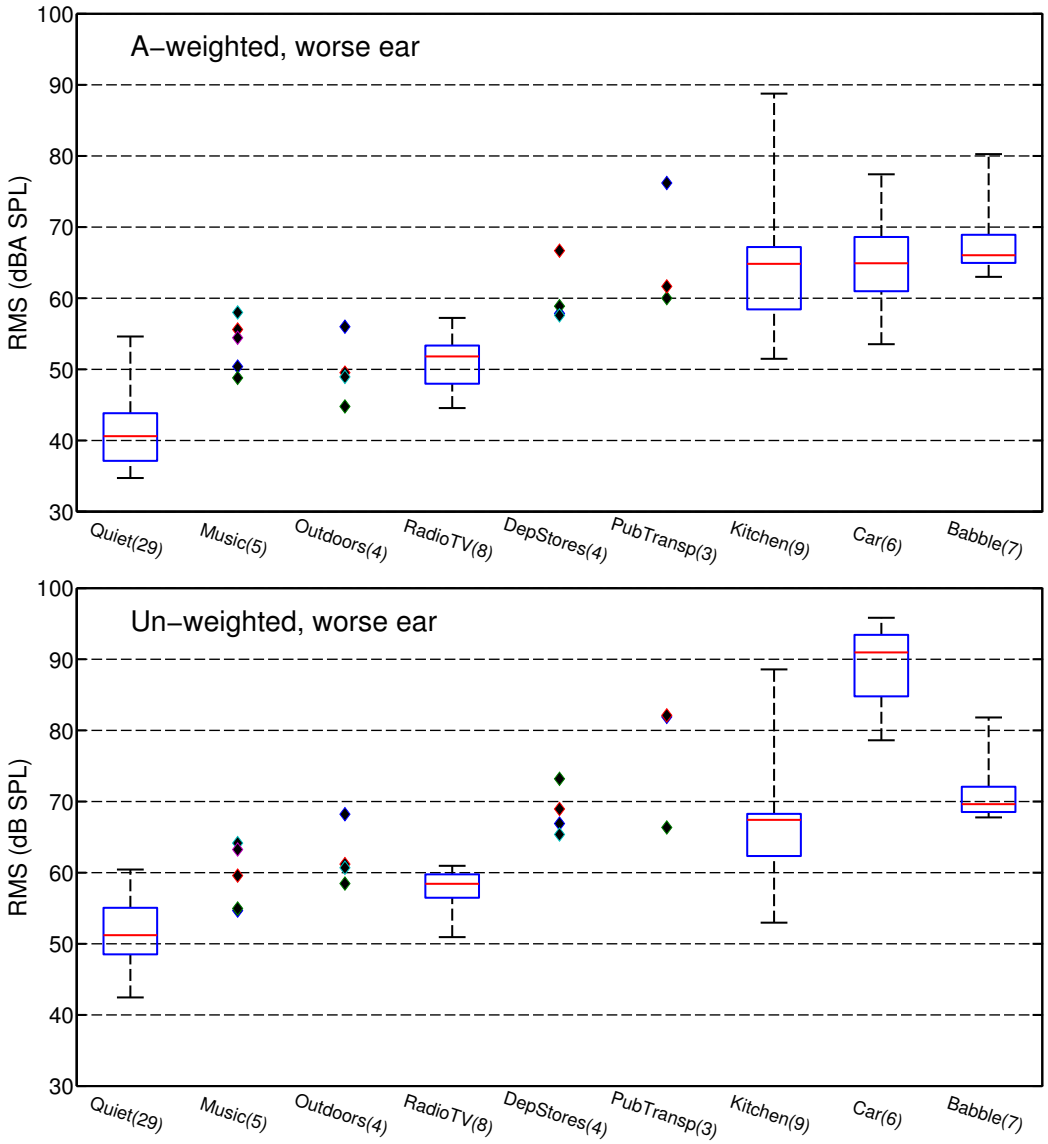

Figure 1: Long-term average RMS levels of the background noise at the ear with the worse SNR_A for each noise category. The upper panel shows the A-weighted RMS levels and the lower panel the un-weighted RMS levels. For noise categories with five recordings or less, the actual RMS estimates are presented (diamonds; the colors carry no particular meaning), whereas boxplots are used for the noise categories with more than five recordings. The central mark in each boxplot shows the median across the recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. The number in brackets after the noise category name gives the number of recordings in that category.

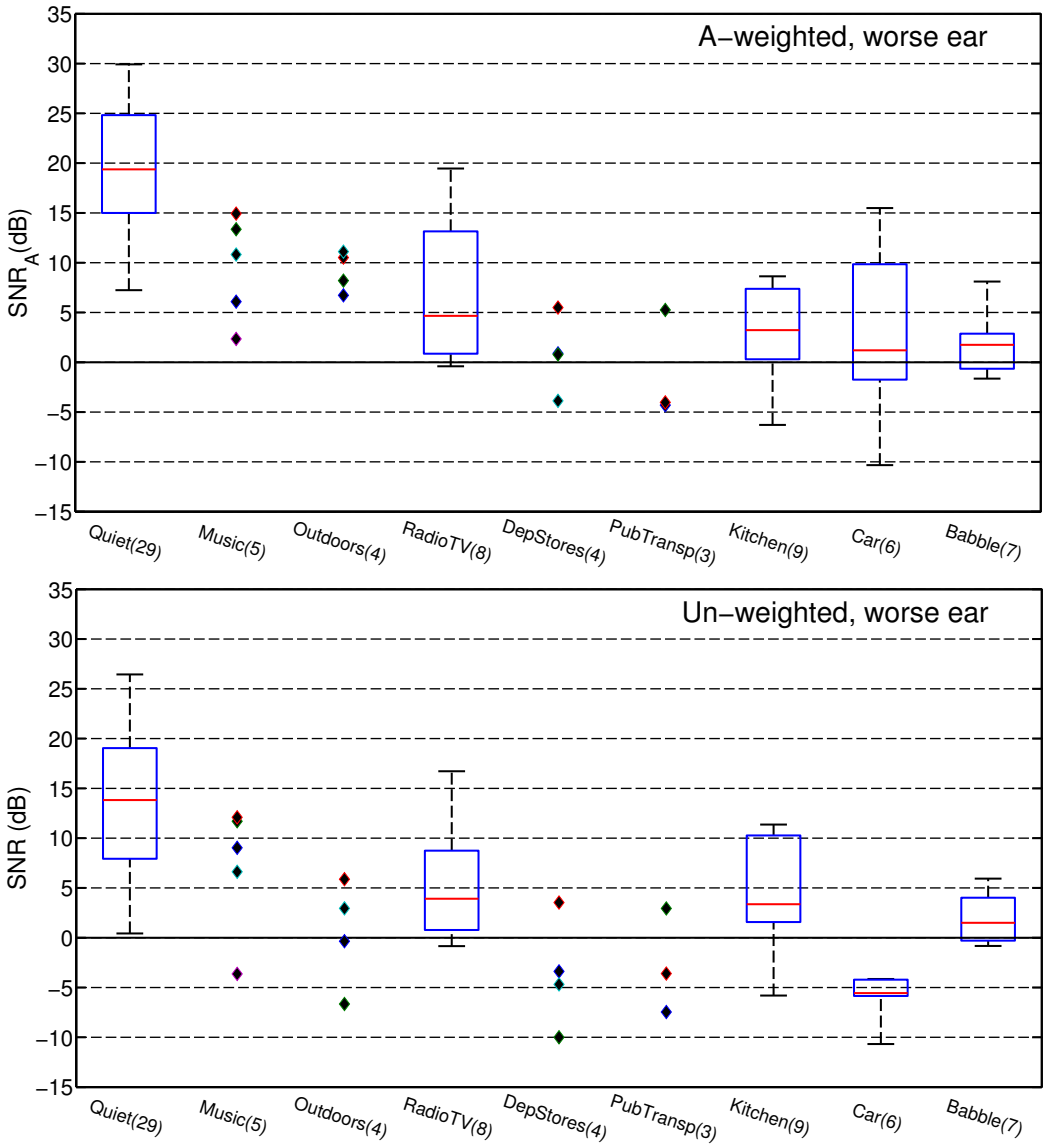

Figure 2: SNR_A (upper panel) and SNR estimates (lower panel) for the worse ear for all noise categories. For noise categories with 5 recordings or less, the actual SNR estimates are plotted (diamonds; the colors carry no particular meaning). For noise categories with more than 5 recordings, box plots are shown. The central mark in each boxplot shows the median across the recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. The number in brackets after each noise category name gives the number of recordings in that category.

2 Quiet

The Quiet category contained the largest number of recordings (29 recordings were analyzed) and displayed a broad variety of acoustical properties. Most of these recordings belonged to the original category "Conversation without background noise" (Wagener et al. 2008) and were judged to be free of background noise by the informants in the laboratory part of the original study. Thus, the description "without background noise" or "in quiet" was most often subjectively assessed and not related to any technical measurements. However, three of the recordings originally belonged to the category "Conversation in background noise", but were in the current study moved to the Quiet category because there was no prominent noise source which matched the selected categories and because the estimated SNR was greater than 20 dB.

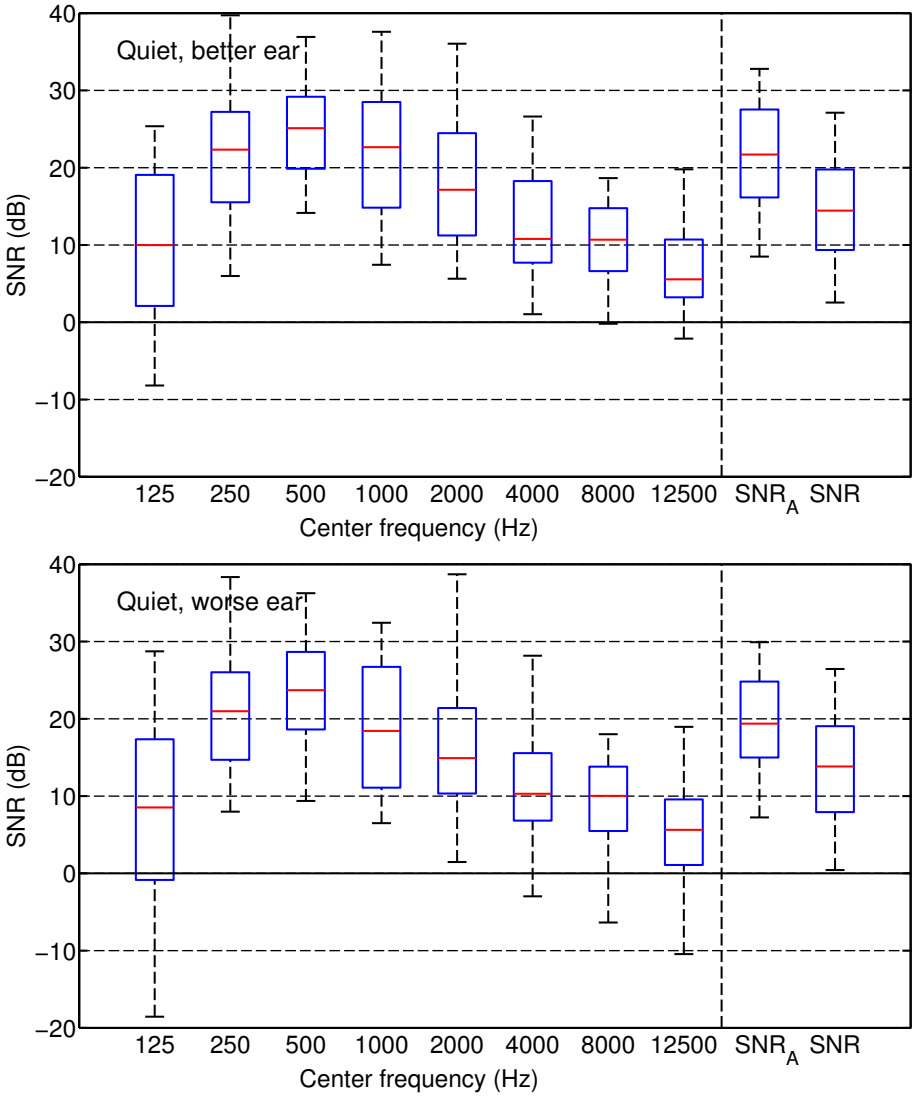

Figure 3: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Quiet category. The central mark in each boxplot shows the median across the 29 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

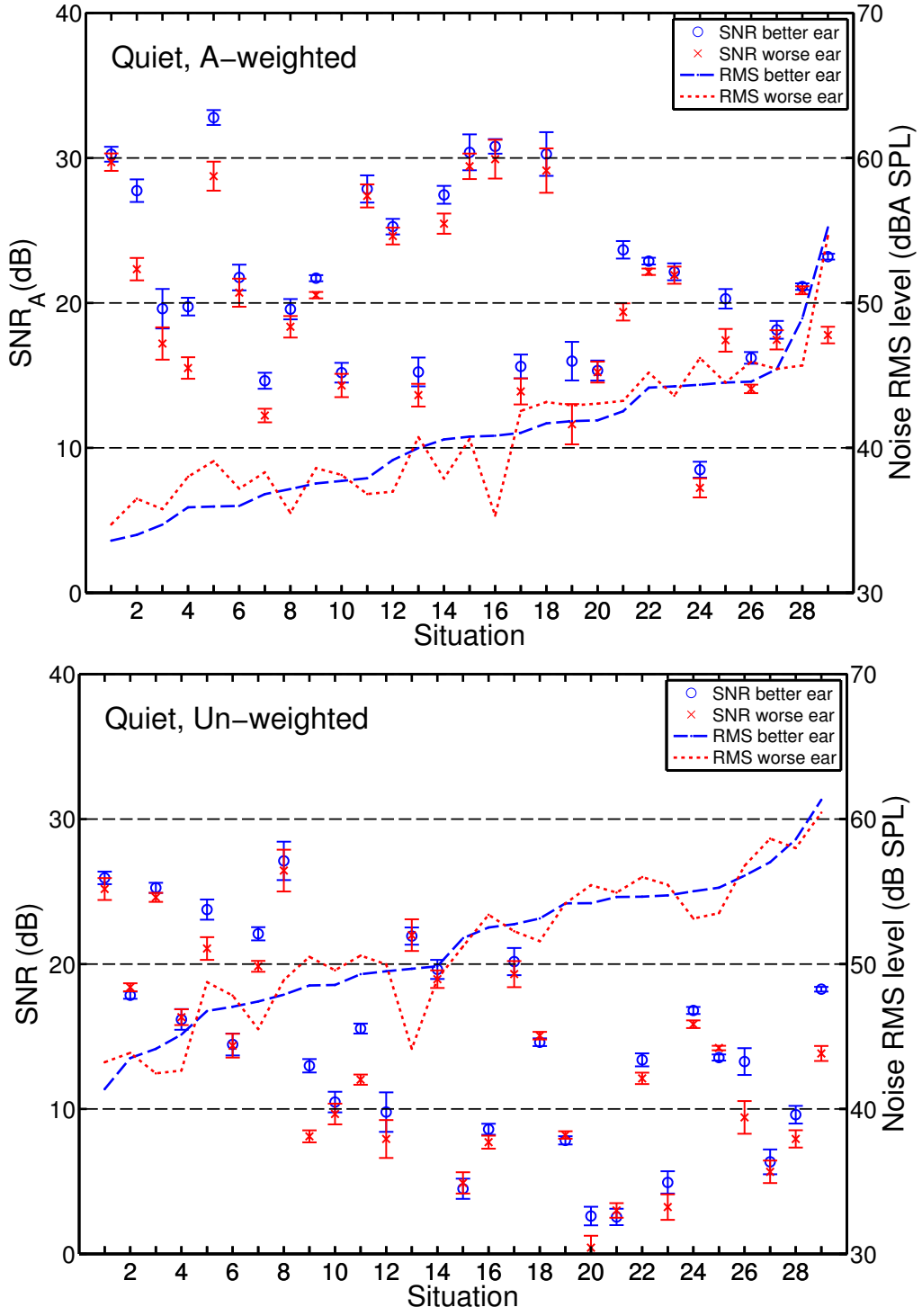

Figure 4: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 29 analyzed situations in the Quiet category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

3 Music

In the Music category, five recordings were analyzed. The target signal was, like in all other analyzed recordings, live speech and music constituted the background noise. Most of the recordings were made in the informants homes. Recordings with both female and male target talkers and were included.

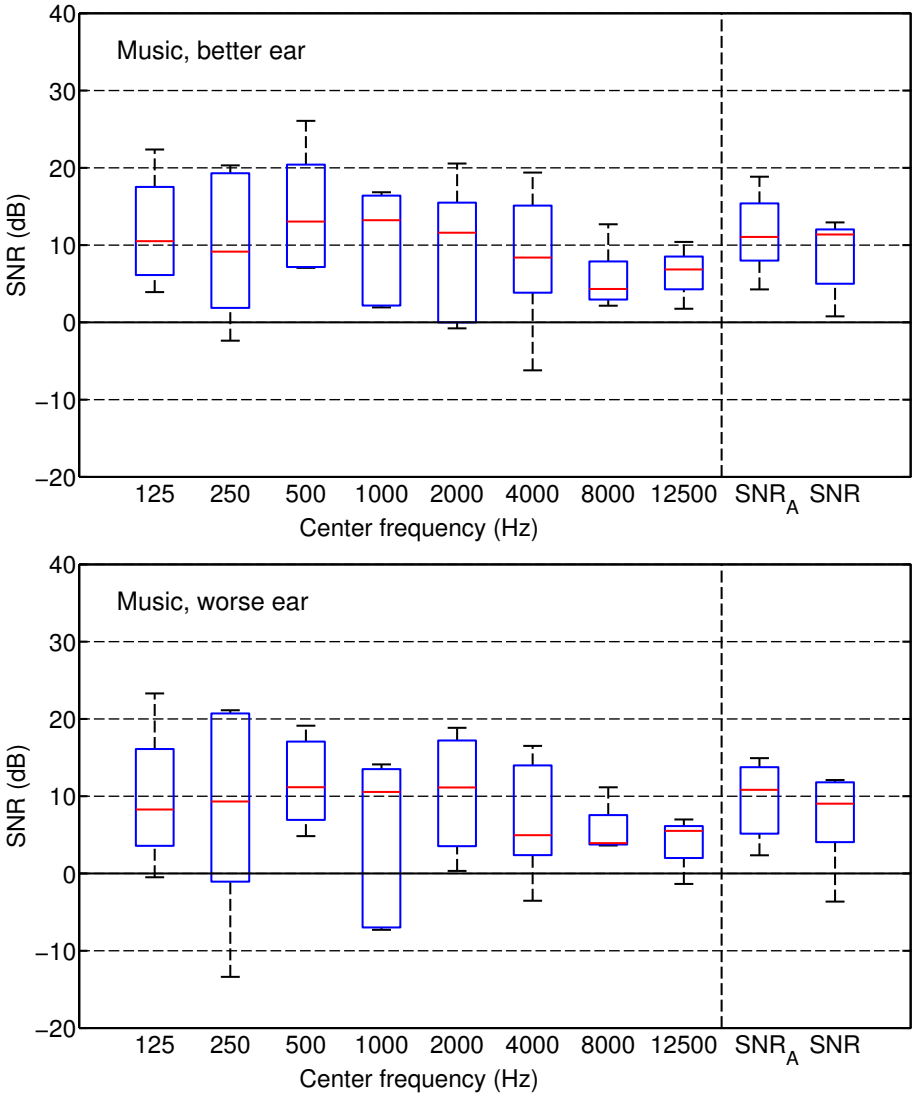

Figure 5: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Music category. The central mark in each boxplot shows the median across the 5 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

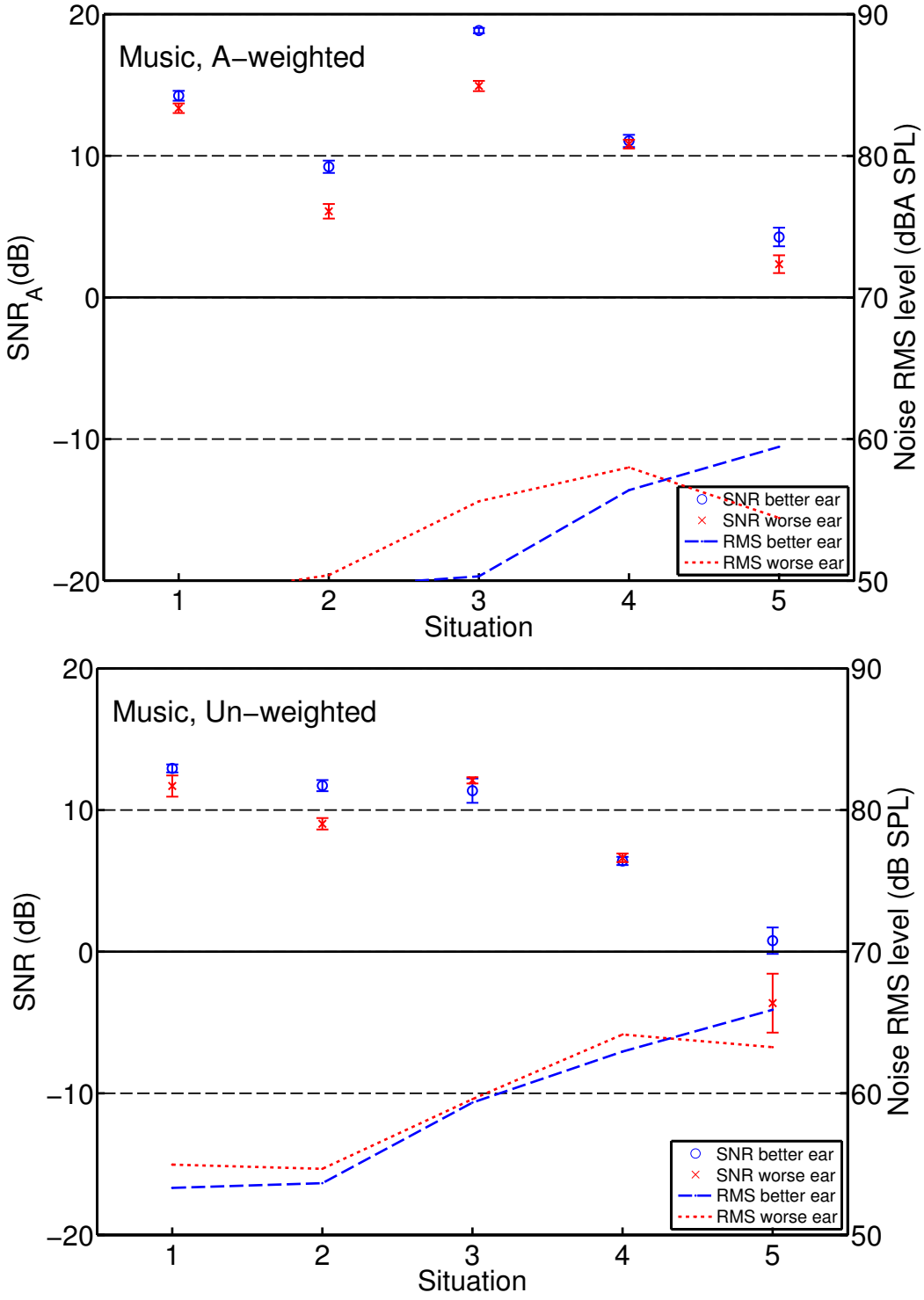

Figure 6: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 5 analyzed situations in the Music category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

4 Outdoors

In the Outdoors category, four recordings were analyzed. They were recorded in the informants gardens and generally contained common nature sounds, for instance birdsong, and some kind of traffic noise. Two recordings contained the sound of airplanes passing by. Recordings with target speech from children, male and female speakers from different source locations were included.

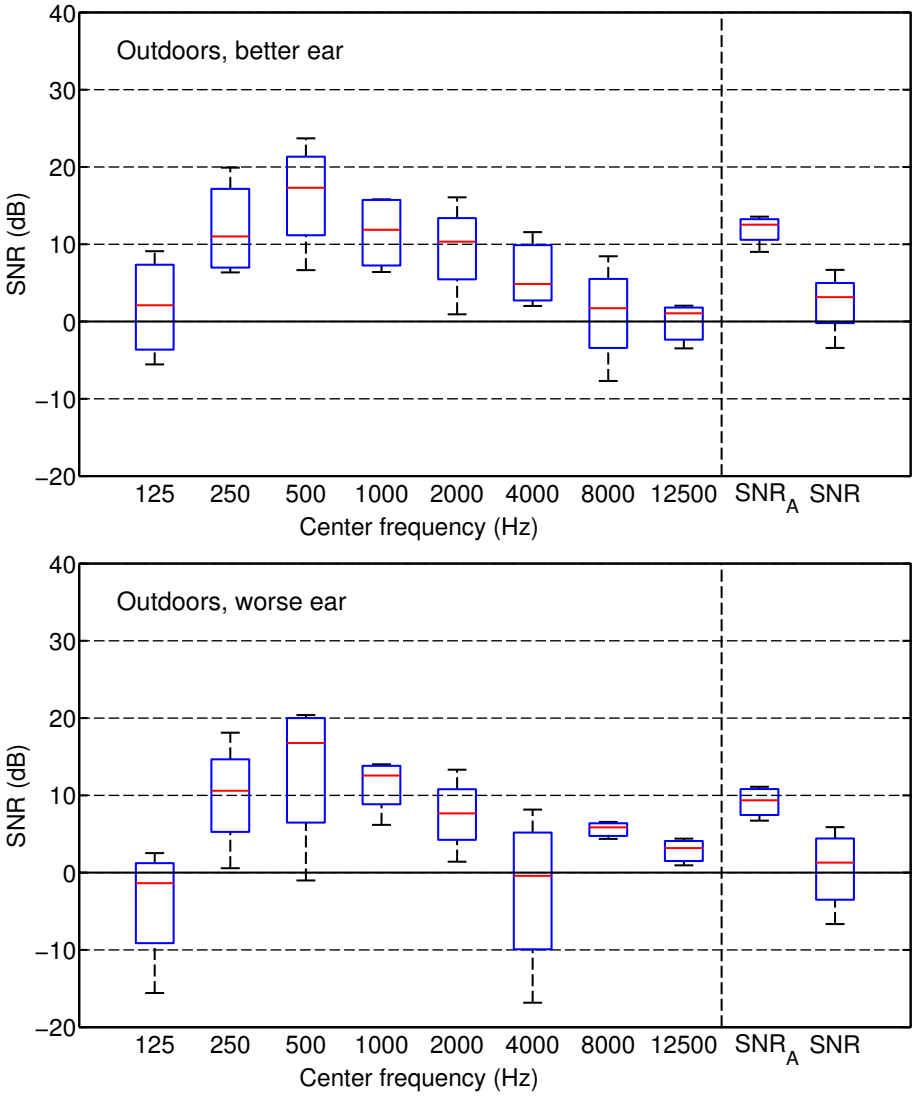

Figure 7: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Outdoors category. The central mark in each boxplot shows the median across the 4 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

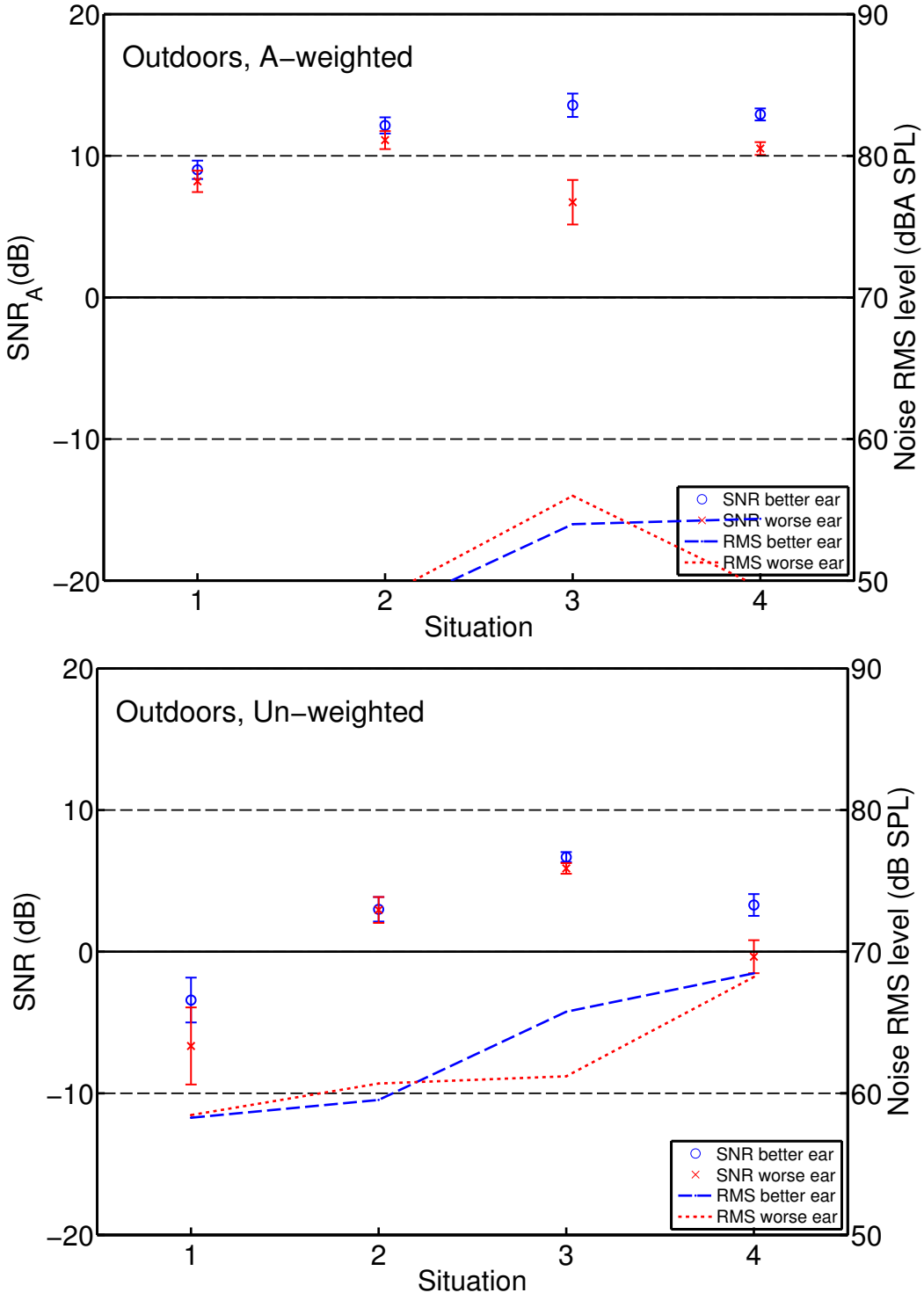

Figure 8: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 4 analyzed situations in the Outdoors category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

5 Radio or TV

In the Radio or TV category (RadioTV), eight recordings were analyzed. The target signal was live speech and one single talker on the Radio or the TV constituted the background noise. Most of the recordings were made in the informants' homes. In five of the recordings, the source of the background noise was a television. In the remaining three recordings, the noise originated from a radio. The radio recordings differed from the television recordings in that the radio noise was at a larger distance from the talker than the TV noise. Accordingly, the Radio recordings showed lower noise RMS levels and higher SNRs than the TV recordings. Recordings with both female and male target talkers and competitive talkers were included. The recorded target speech was often more prominent in one of the recording channels, most likely since the visual target determines the head orientation in the TV situations.

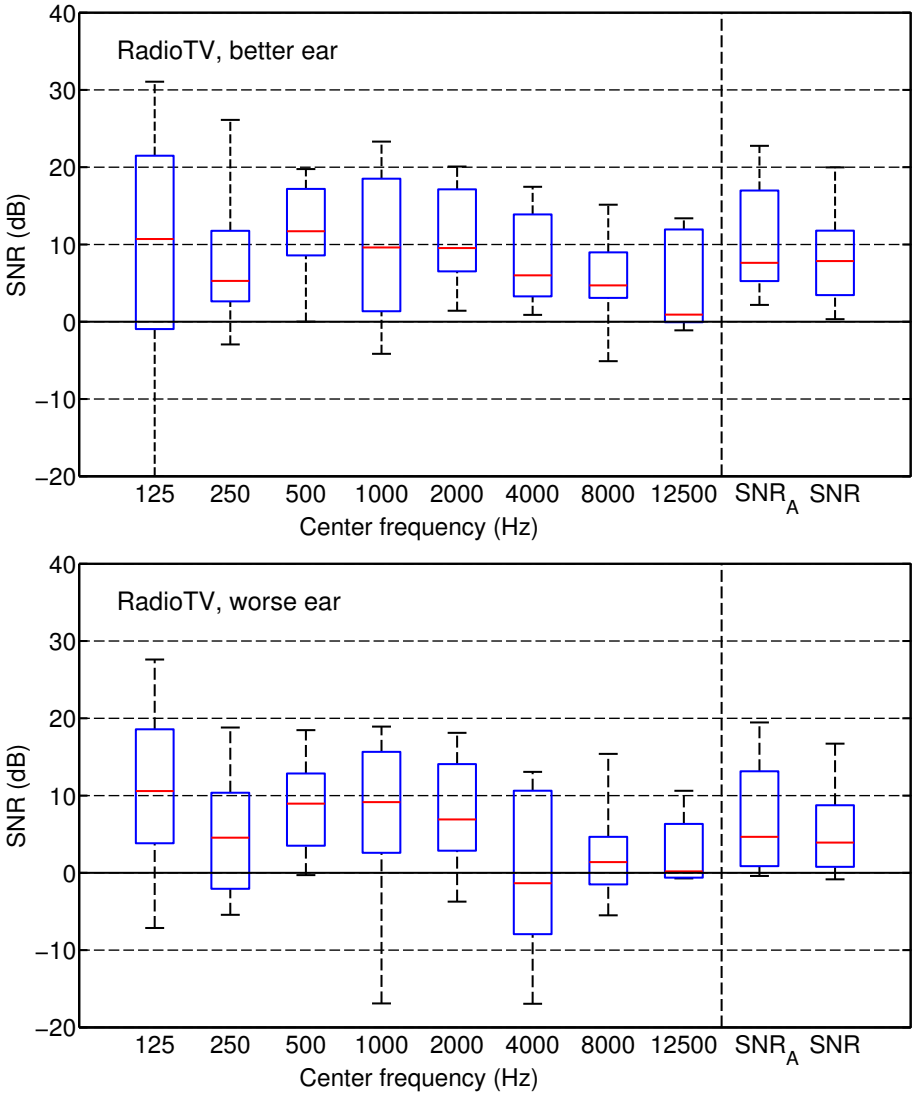

Figure 9: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the RadioTV category. The central mark in each boxplot shows the median across the 8 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

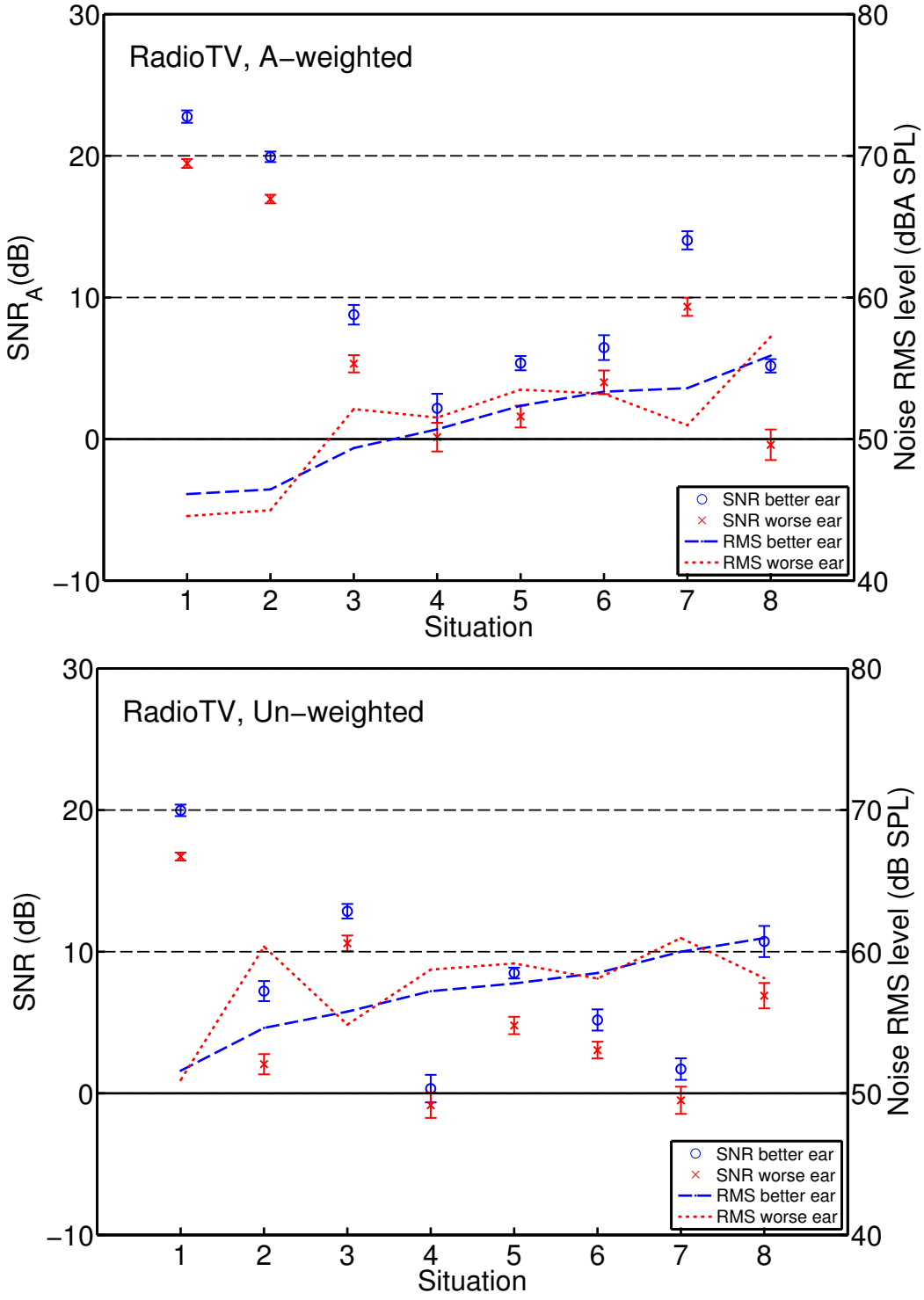

Figure 10: SNR_As and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 8 analyzed situations in the RadioTV category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

6 Department stores

In the Department Stores category (DepStores), four recordings were analyzed. The recordings were characterized by a rather stationary noise (multi-talker babble and noise from ventilation and other electrical devices). One recording was made in a beverage store and this recording also included a large number of transients from bottles hitting another bottle. Male and female target speech was recorded from different source locations.

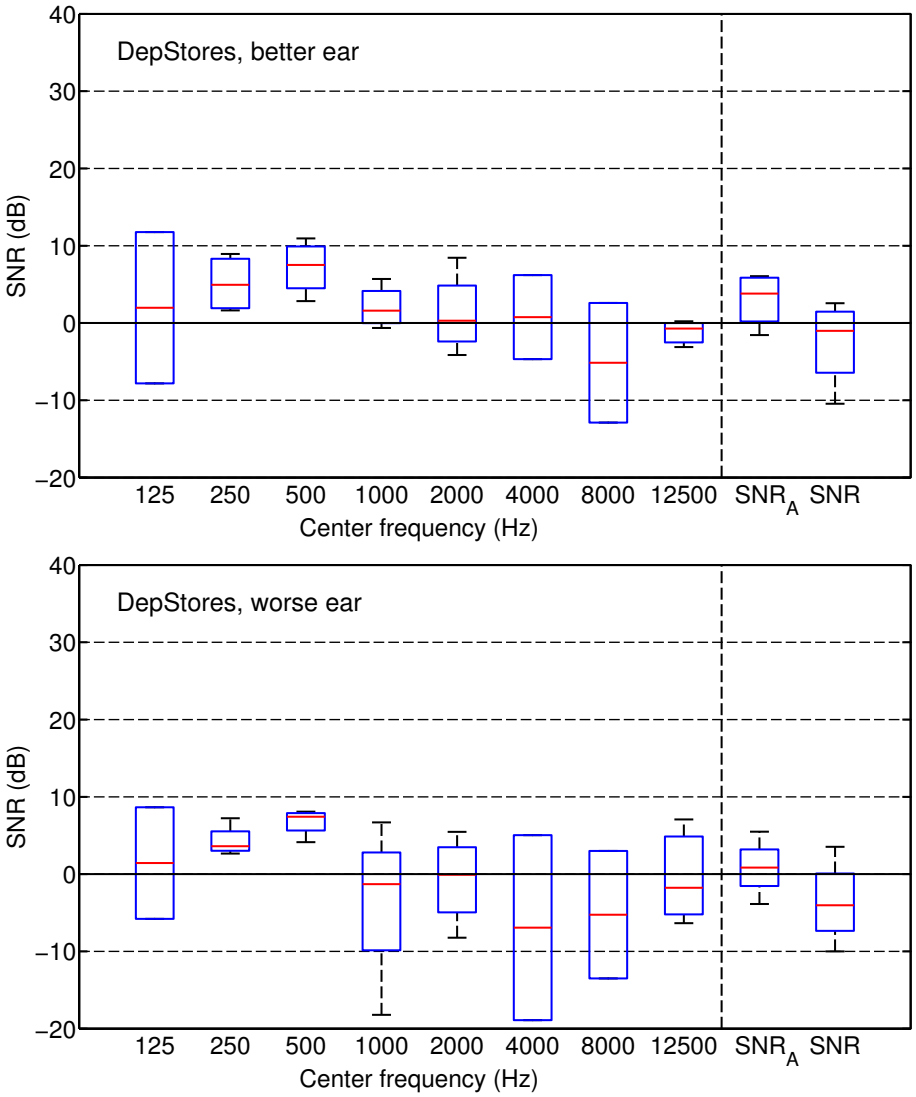

Figure 11: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the DepStores category. The central mark in each boxplot shows the median across the 4 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

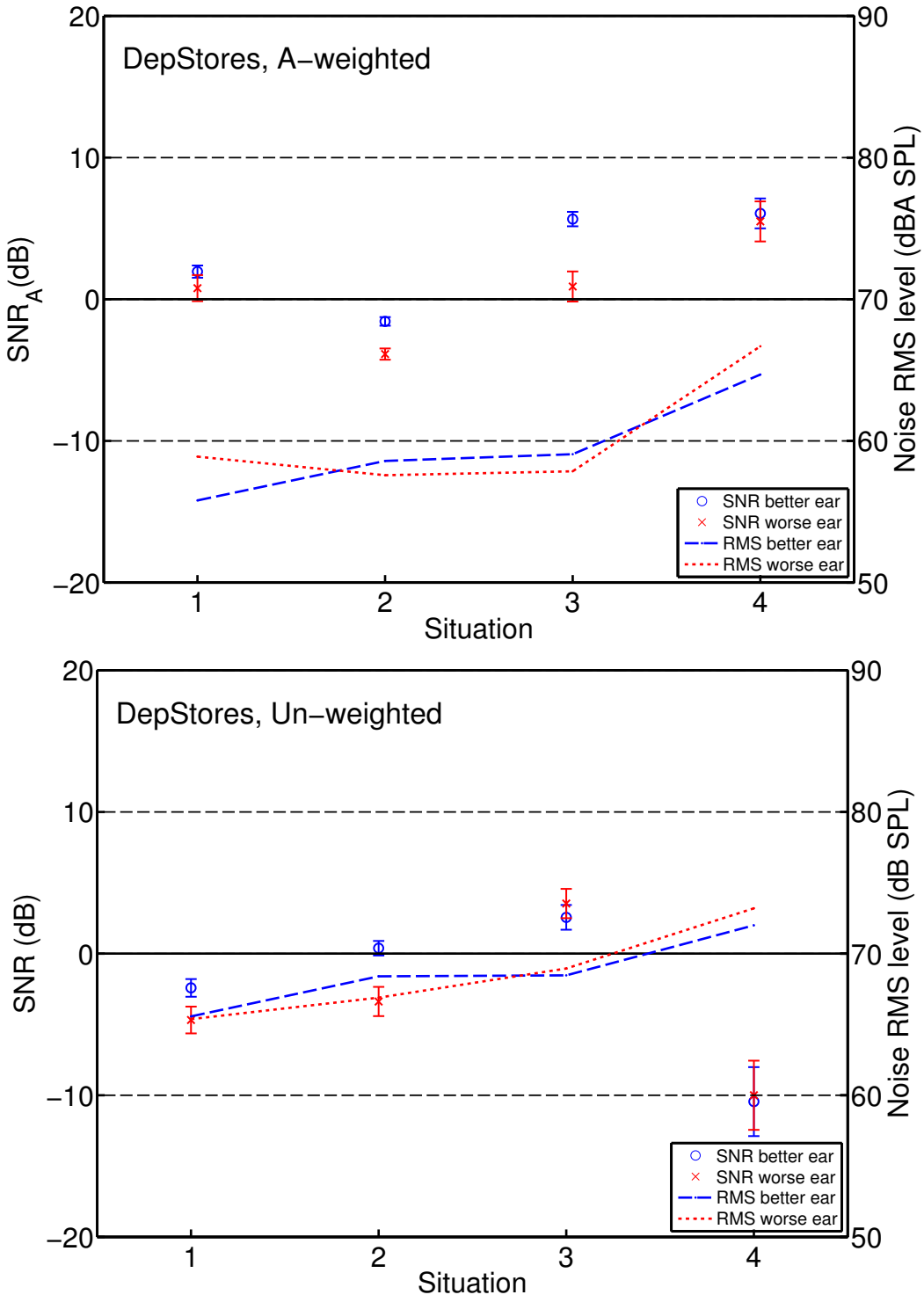

Figure 12: SNR_As and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 4 analyzed situations in the DepStores category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

7 Public transport

In the Public Transport category (PubTransp), three recordings were analyzed. The recordings were from different acoustical environments: one in a station building, one on a platform near a train, and one inside a commuter train. This makes the category the most heterogeneous in terms of acoustical properties and noise sources. The informants were facing the talker, as the recorded target speech is equally prominent in both recording channels for all items in this category. The material contains target speech from female and male talkers.

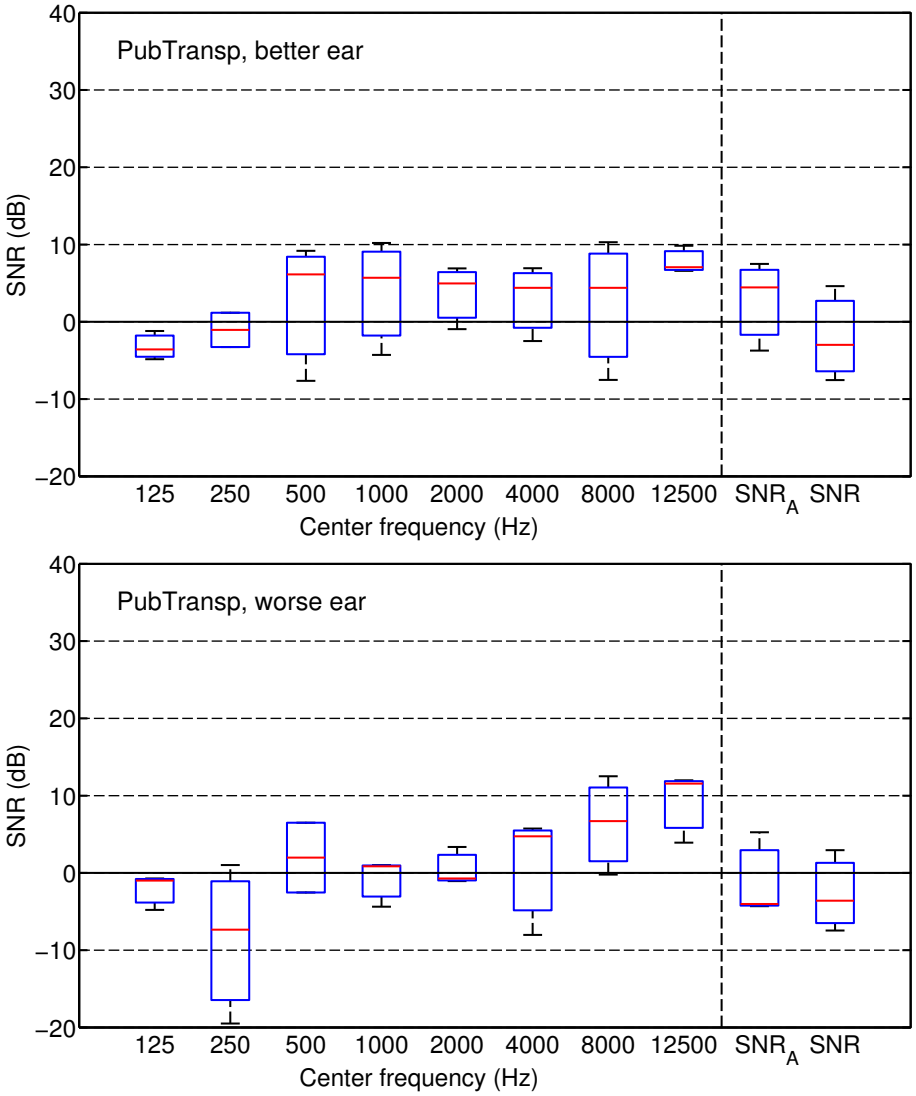

Figure 13: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the PubTransp category. The central mark in each boxplot shows the median across the 3 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

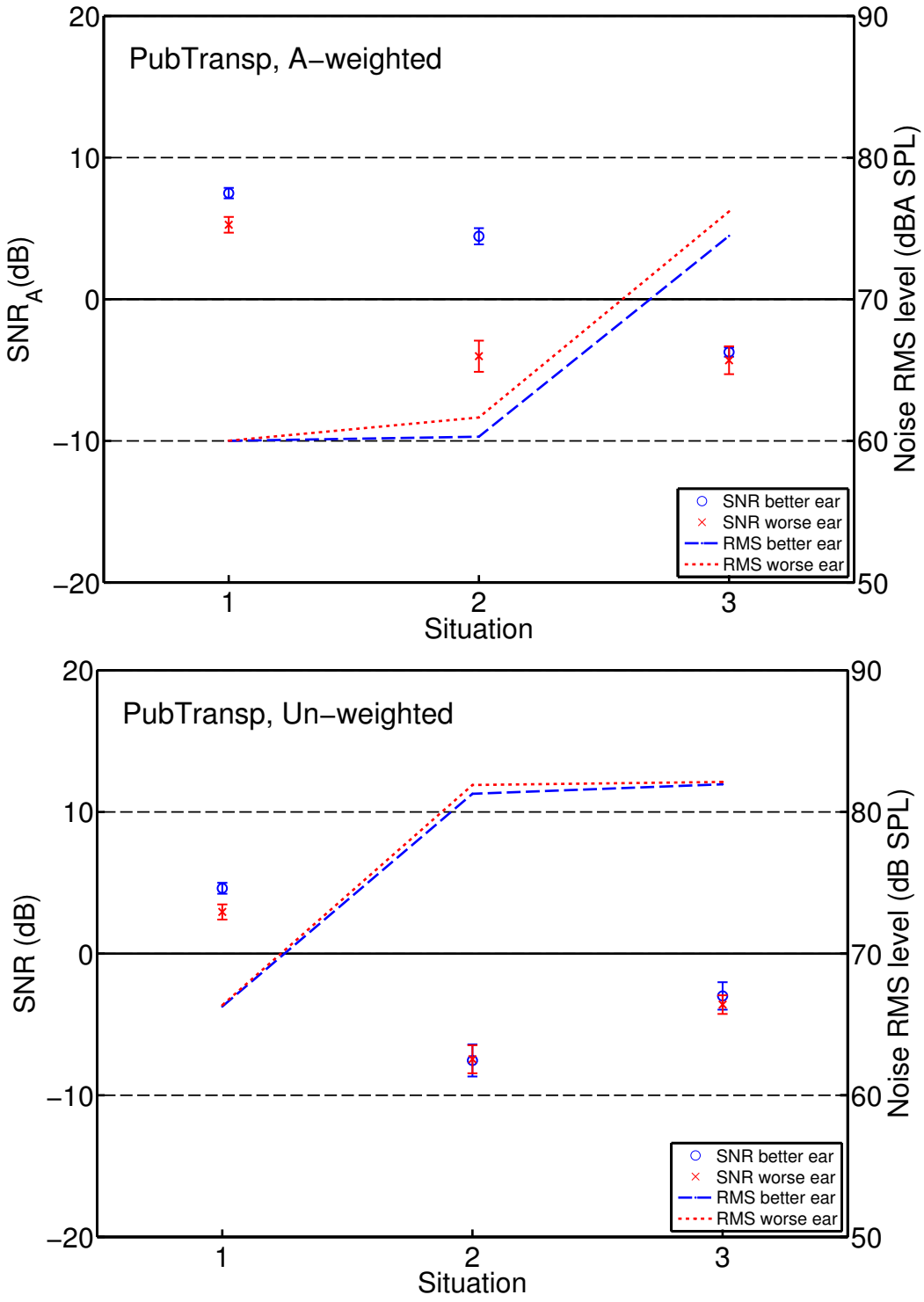

Figure 14: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 3 analyzed situations in the PubTransp category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

8 Kitchen

In the Kitchen category, nine recorded situations were analyzed. These recordings generally contained stationary background noise from exhaust hoods, running water, electrical kitchen tools or similar. The recordings differed from most other recordings in that the Kitchen recordings generally contained more high-frequency noise. The RMS level of the background noises varied considerably. The target speakers are children and adult female and male talkers. The target speech location varied both during a recorded situation and across recordings. This was mostly due to the target speaker moving around in the kitchen.

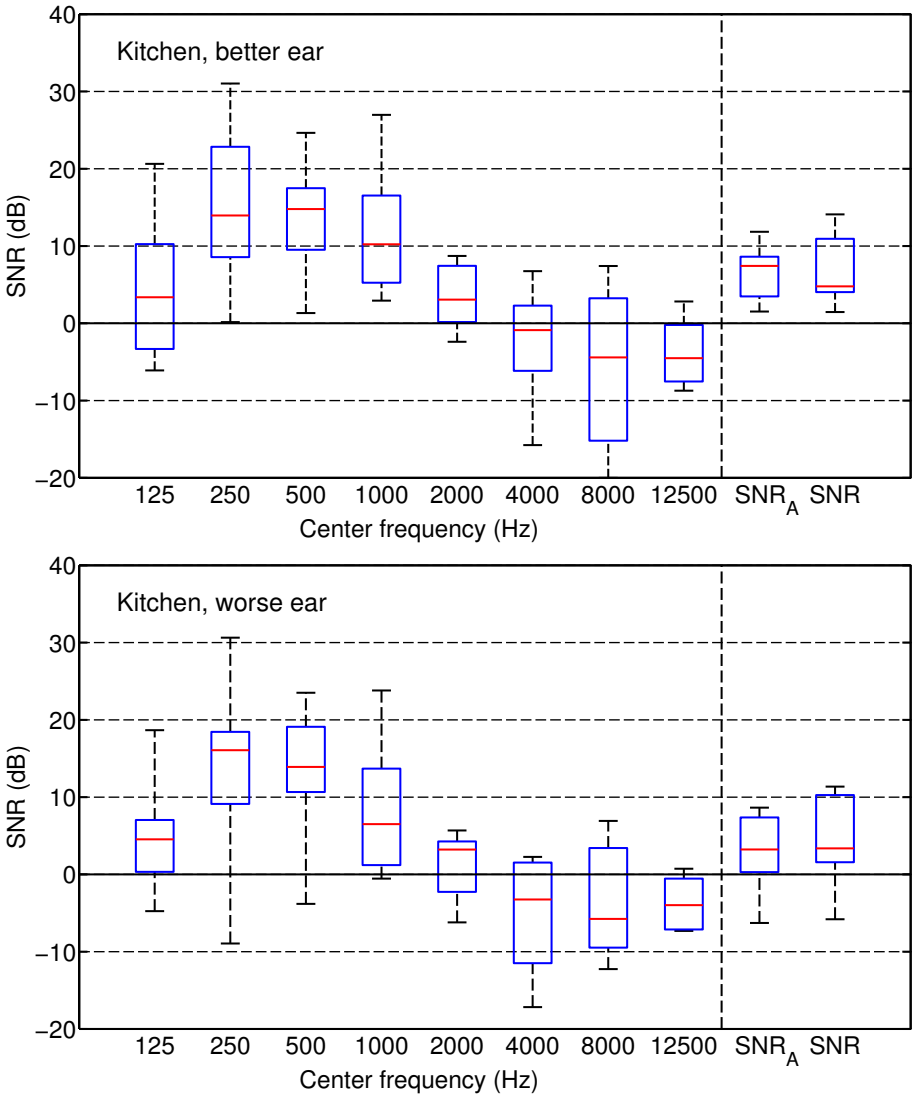

Figure 15: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Kitchen category. The central mark in each boxplot shows the median across the 9 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

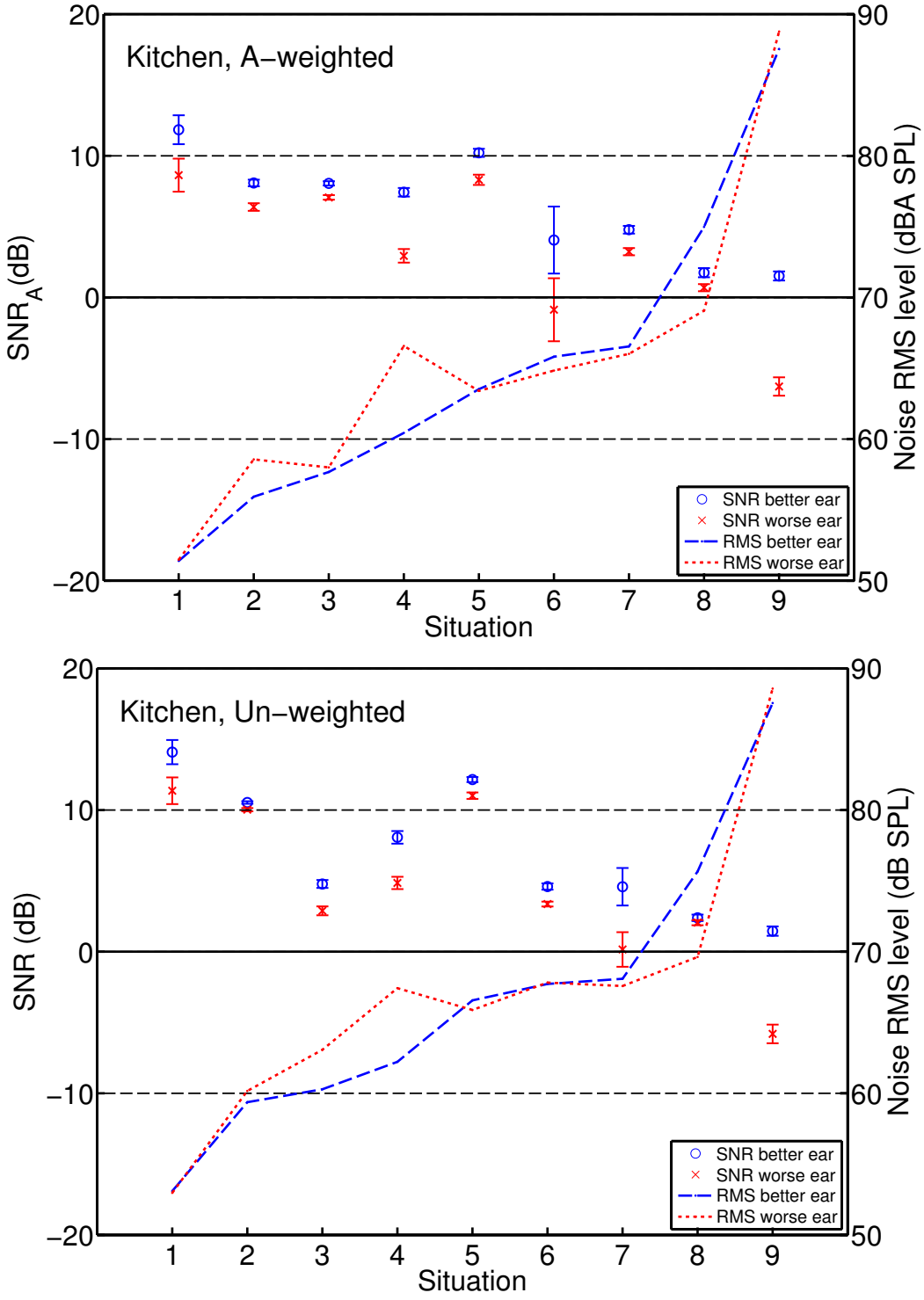

Figure 16: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 9 analyzed situations in the Kitchen category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

9 Car

In the Car category, six recorded situation were analyzed. The noise was generally stationary, but could change its properties both within a particular recording (e.g., during acceleration) and across recordings (due to different car brands). In some recordings the informant is the driver and in others the passenger and in most cases apparently focused on the traffic. Consequently, the target speech is more prominent in one of the recording channels. However, in two of the recordings, the informant is obviously facing the target speech source. Recordings with both female and male target talkers were included. In one of the recordings the target speech was a child speaking with raised vocal effort.

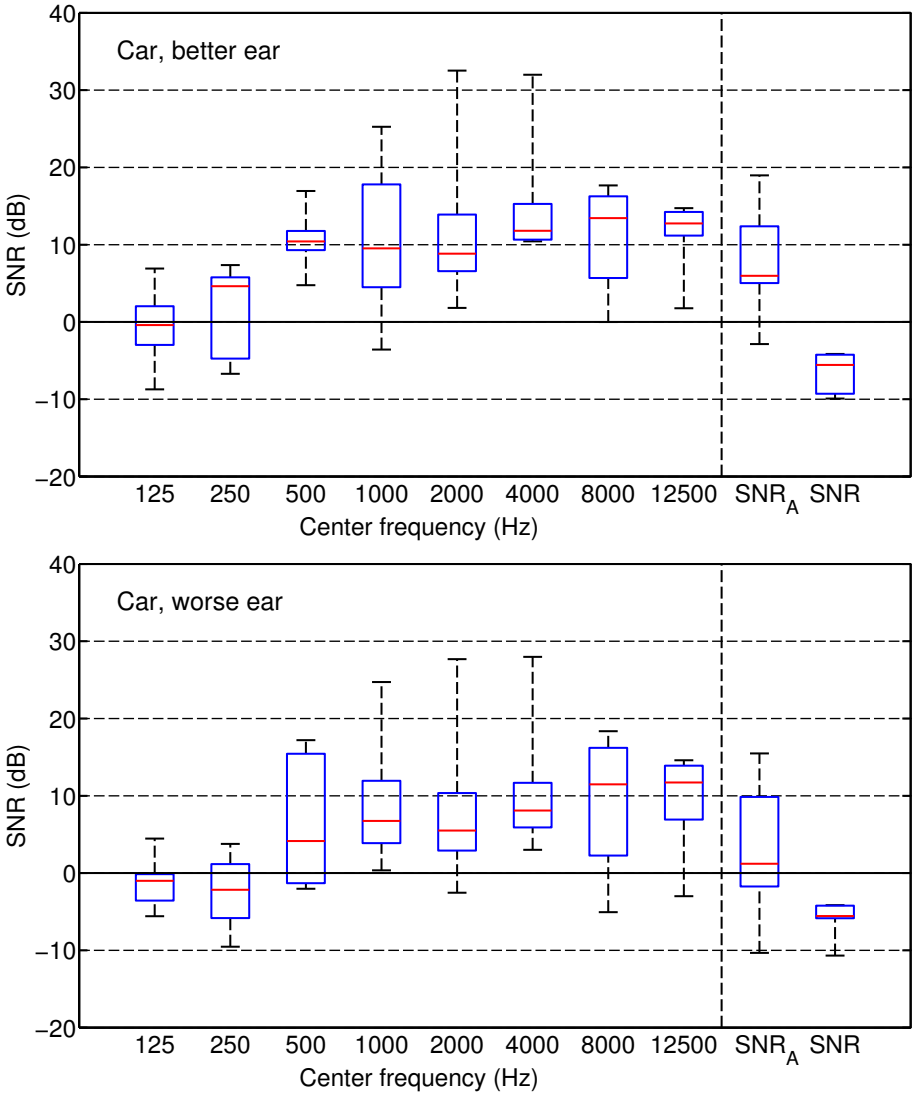

Figure 17: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Car category. The central mark in each boxplot shows the median across the 6 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

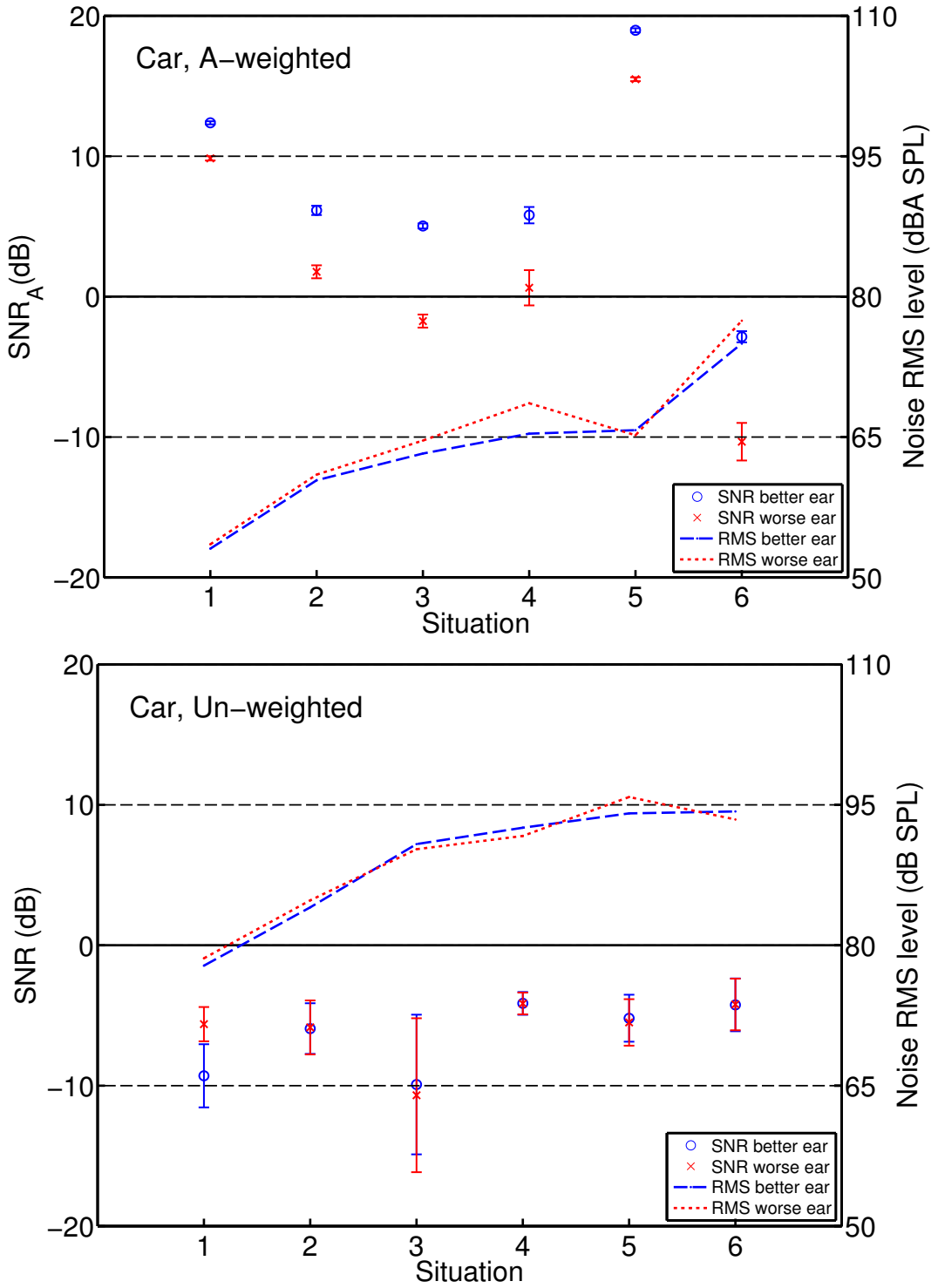

Figure 18: SNR$_A$s and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 6 analyzed situations in the Car category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

10 Babble

In the Babble category, seven recorded situations were analyzed. They were mostly recorded in restaurants and cafeterias. The number of interfering voices was large, which gave the babble noise a fairly stationary character. In one recording there was also music, but the babble was the dominating noise source. Recordings with both female and male target talkers were included. In situations with many talkers, only one "prominent" speaker (based on speech duration) was selected as the target speech. The location of the target speech in relation to the informant varied among the recordings and the informant was not necessarily facing the talker, which made the target speech more prominent in one of the recording channels.

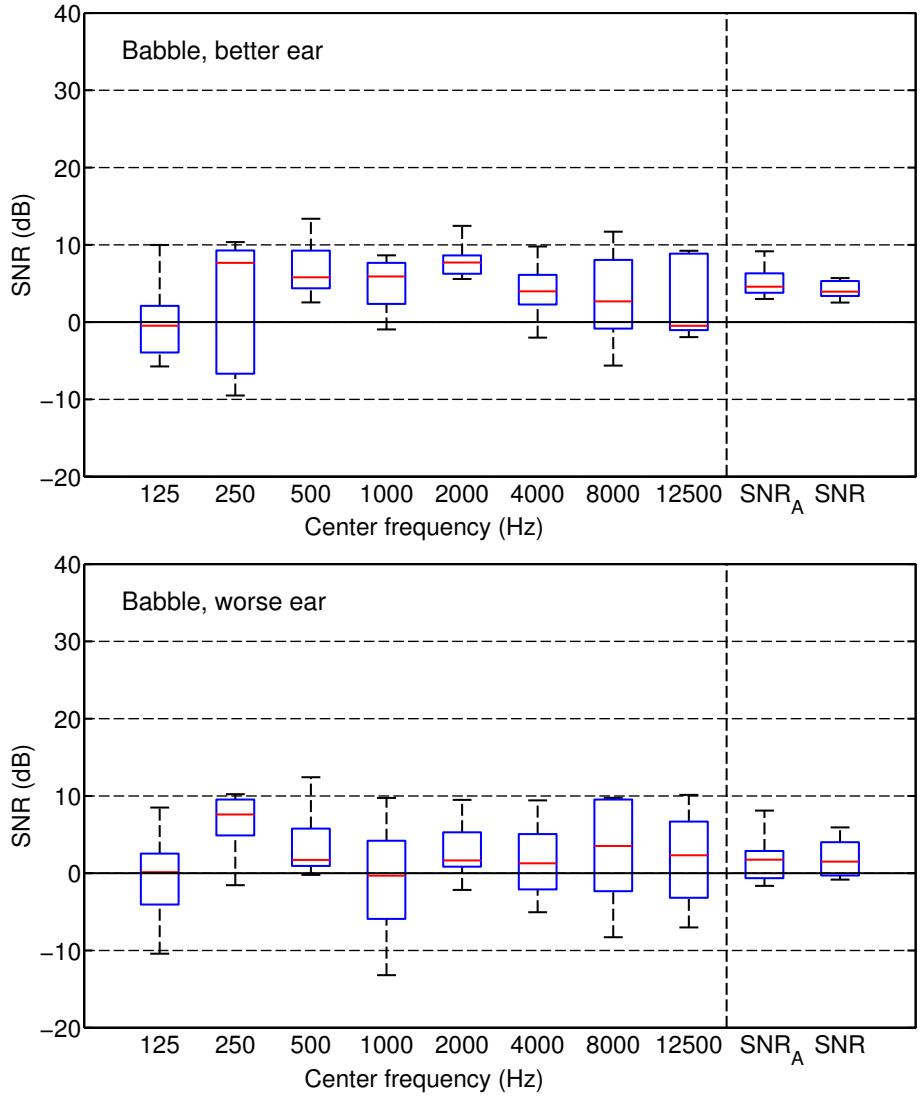

Figure 19: Estimated SNRs for the better ear (upper panel) and worse ear (lower panel) calculated in one-third octave bands around the frequencies on the abscissa for the Babble category. The central mark in each boxplot shows the median across the 7 recordings, the box the inter-quartile range, and the whiskers extend to the most extreme data points. To the right, the overall SNR_As and SNRs are shown.

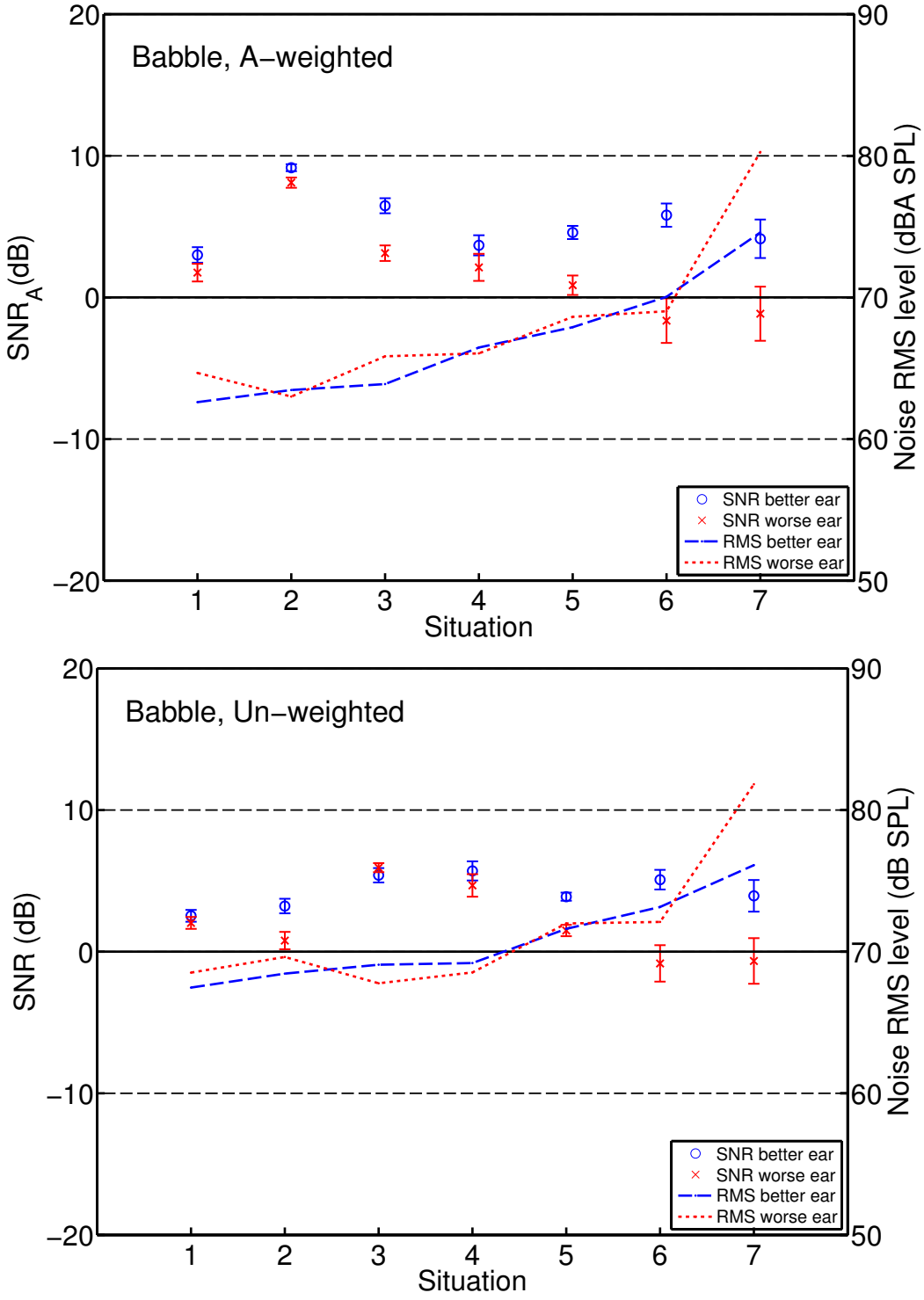

Figure 20: SNR_As and A-weighted noise RMS levels (upper panel) and SNRs and un-weighted noise RMS levels (lower panel) for the 7 analyzed situations in the Babble category. The SNRs are read on the left ordinate and the RMS levels on the right. The SNRs at the better ear are marked by (blue) circles and the SNRs at the worse ear are marked by (red) crosses. The RMS levels at the better ear are marked by a dashed (blue) line, and the RMS levels at the worse ear are marked by a dotted (red) line. The situations are sorted based on the A-weighted RMS levels in the better ear. The bars illustrate the SNR estimation uncertainty.

References

Wagener KC, Hansen M, Ludvigsen C. (2008) Recording and classification of the acoustic environment of hearing aid users. *J Am Acad Audiol* 19: 348-70.