OPEN ACCESS Freely available online

Audio Engineering Society

Convention Paper 9842

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

The Adjustment / Satisfaction Test (A/ST) for the Subjective Evaluation of Dialogue Enhancement

Matteo Torcoli[1], Jürgen Herre[1,2], Jouni Paulus[1,2], Christian Uhle[1,2], Harald Fuchs[1], and Oliver Hellmuth[1]

[1] *Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany*
[2] *International Audio Laboratories Erlangen, a joint institution of Universität Erlangen-Nürnberg and Fraunhofer IIS*

Correspondence should be addressed to Matteo Torcoli (matteo.torcoli@iis.fraunhofer.de)

ABSTRACT

Media consumption is heading towards high degrees of content personalisation. It is thus crucial to assess the perceptual performance of personalised media delivery. This work proposes the Adjustment / Satisfaction Test (A/ST), a perceptual test where subjects interact with a user-adjustable system and their adjustment preferences and the resulting satisfaction levels are studied. We employ the A/ST to evaluate an object-based audio system which enables the personalisation of the level ratio between dialogue and background, i.e., a Dialogue Enhancement system. Both the case in which the original audio objects are readily available and the case in which they are estimated by blind source separation are compared. Personalisation is extensively used, resulting in clearly increased satisfaction, even in the case with blind source separation.

1 Introduction

The audio of broadcast material is traditionally produced by audio engineers that determine the level ratio between the foreground speech and all other sound sources. The goal is to deliver enjoyable and engaging mixes featuring full intelligibility and low listener fatigue. However, the achievement of this goal depends also on personal and contextual factors. The best understood ones are:

- listener's hearing acuity [1–3];
- listening environment, e.g., environmental noise [4] and reproduction system such as mobile device, TV set [5];
- listener's mother tongue and content language [6];
- personal taste, as suggested in [7] and more systematically analysed in this paper.

It follows that a unique *one-size-fits-all* mix can hardly satisfy the needs of the audience in all cases. This is also indicated by the increasing number of complaints about the difficulty in understanding what is said in the broadcast material, with too loud background sounds being the major cause of them [8].

Dialogue Enhancement (DE) methods provide a solution for this issue by giving the audience the possibility to control the relative level of dialogue and background.

In the context of DE, the term *dialogue* refers to all types of foreground speech, including monologues, narrations, and news reading. All the other sound sources are referred to as *background*, as they mostly consist of background music, sound effects, and ambient noise. DE can be implemented with object-based technologies such as MPEG-H Audio [9], Dolby AC-4 Audio [10], and MPEG-D SAOC-DE [11]. Today, object-based audio is reality: in May 2017, the major broadcasters in South Korea launched terrestrial UHD TV services based on ATSC 3.0 using MPEG-H Audio LC Profile as audio system [12, 13].

Object-based technologies require that the audio objects are separately available. Still, the audio is often only available as mono, stereo, or 5.1 mix, especially for archive content or low-budget productions. In these cases, methods for decomposing the mixture signals into separate signal components are needed to open the way for DE. Mixture decomposition strategies that can be adapted to DE are numerous in the literature on speech enhancement [14] and source separation [15].

These techniques are not able to perfectly reconstruct the original objects and artefacts and distortions may be introduced. For this reason, the assessment of the user experience is crucial. A set of objective measures for DE was proposed in [16]. However, the most reliable method for audio quality assessment remains subjective[1] evaluation. Moreover, no objective measure can answer the following research questions RQ1 and RQ2.

RQ1) Is DE functionality desired by the users in order to have an enjoyable mix where speech can be easily followed? If yes, to what extent?
RQ2) How satisfied are the users with a DE system with ideal quality and how satisfied are they with a particular system under development?

These research questions can be generalised to any user-adjustable system (instead of "DE") with a specific goal (instead of "an enjoyable mix where speech can be easily followed").

In order to investigate RQ1 and RQ2, we propose the Adjustment / Satisfaction Test (A/ST), which studies the satisfaction resulting from the direct interaction of subjects with a prototype of the final application. Related works are reviewed in Section 2. The A/ST is presented in Section 3 and it is used to evaluate a DE system: results are reported in Section 4. Conclusions are given in Section 5.

2 Literature review

In [1, 17–19], DE systems are subjectively assessed by providing pre-determined mixes with fixed dialogue-to-background ratios. The listeners are asked to rate the different mixes and to base their judgement on various criteria such as overall sound quality and speech clarity. In [2], speech intelligibility for hearing-impaired subjects is investigated on three different pre-determined mixes. The perceptual tests in these five works do not address RQ1 and the test participants cannot directly interact with the systems under test.

Direct interaction is the cornerstone of the tests employed in [3, 4, 7, 20]. The participants are free to choose their preferred dialogue-to-background ratios in [3, 4, 7], while the aesthetically preferred reverberation level is investigated in [20]. These works analyse the preferred levels, but they do not address RQ2 as they do not quantify the impact of the selections, e.g., on a satisfaction scale.

None of the mentioned works employs one of the numerous tests standardised or recommended by international organisations. Guidance through these standards is given in [21]. We were also not able to find a suitable method among them to address our research questions. These considerations motivated us to design a new test, which is described in the following section.

3 The Adjustment/Satisfaction Test

We now introduce the A/ST for the subjective evaluation of user-adjustable systems. The general rationale and design are presented in Section 3.1. Section 3.2 discusses its application to DE. Section 3.3 describes the independent variables of the test that we carry out.

3.1 Design

Let S be a system that can be personalised via the parameter p that is controlled by the user, e.g., via a rotating knob, a remote control, or a similar device. Let us evaluate S via the A/ST. After an introductory phase (Phase 0), adjustment (Phase 1) and satisfaction assessment (Phase 2) are repeated for each test item.

Phase 0) Explaining Envisioned Usage

First, the envisioned usage scenario and the goal of the personalisation are described to the participants. These

[1] In this work, the term *subjective* indicates that subjects (persons) are involved in the perceptual evaluation, in contrast to *objective* measures that estimate the outcome of subjective evaluation without the need of subjects. The data produced by subjective evaluation is anyway objective, as statistically similar data and subsequent conclusions are reproducible by another experimenter using a similar set-up, but different subjects at different place and time.

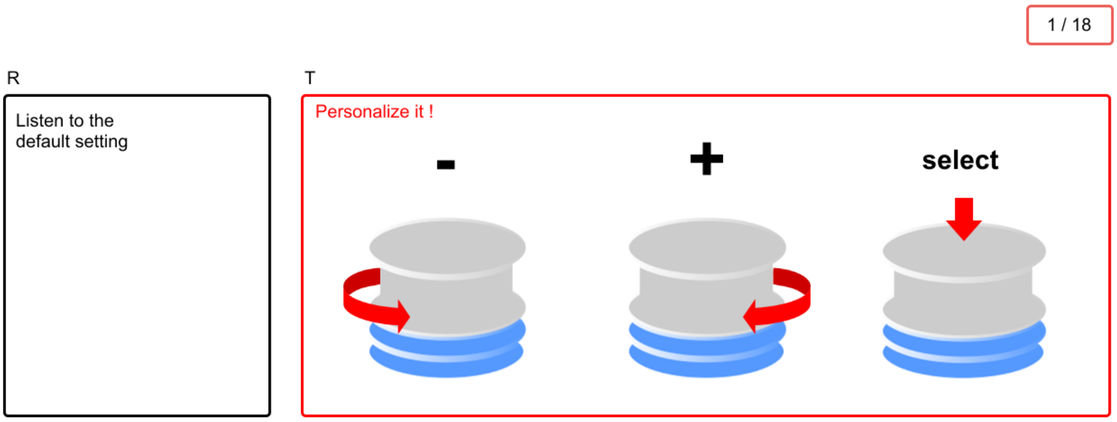
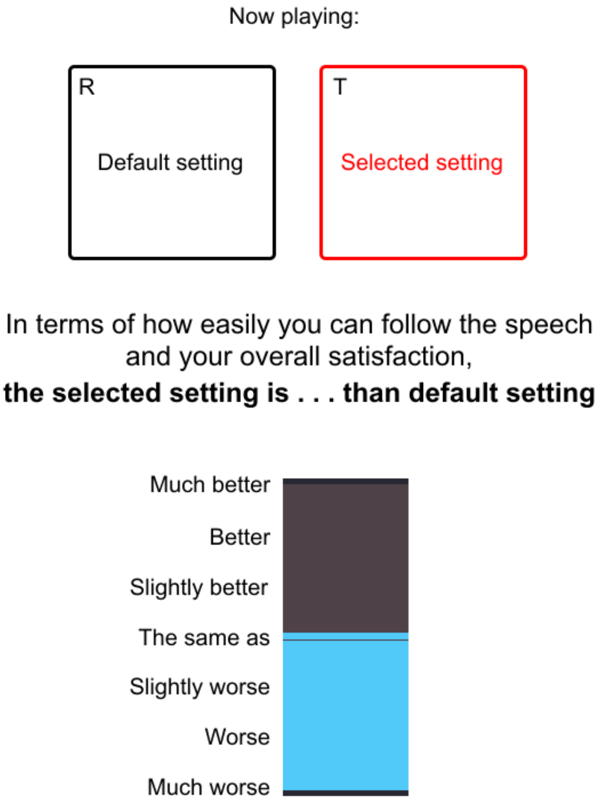


Fig. 1: User interface for the adjustment phase of the A/ST.


Now playing:

R Default setting

T Selected setting

In terms of how easily you can follow the speech and your overall satisfaction,

the selected setting is . . . than default setting

Fig. 2: User interface for the satisfaction assessment phase of the A/ST.

concepts define the expectations and thus the satisfaction of the subjects and so they have to be very clear to the users. The test environment reproduces the main characteristics of the envisioned usage environment and the test material is representative for the application. In this introductory phase, it is also explained how to operate the interface and how the following phases alternate for each item under test. In order to minimise the risk of a poor comprehension of the task, we give written instructions to the participants, we let them operate the test with a training item, and we verbally clarify any doubt that may rise.

Phase 1) Adjustment

In the adjustment phase, the test participants interact in real time with S by adjusting p. The user interface that we use is shown in Fig. 1. In order to ensure that the selection of the preferred p is not biased, as little additional information as possible is given to the user. For example, p is not shown as a numeric value and the adjustment steps are not perceivable while operating the knob by which the user controls p.

During the personalisation, it is possible to compare the adjusted setting with a default setting p_0 by instantaneously switching between the two versions. This p_0 is also included in the range of p. The possibility of comparing against p_0 is important for two reasons. First, it prevents the frustration that a user may experience for small adjustment steps. Second, a default value for p can help undecided users: if the user likes p_0, she/he is encouraged to find p_0 or a similar value in the available range of p; if the user does not like it, she/he is stimulated to find a different p.

Changing the value of p produces physically different outputs. If the physical differences are perceptually relevant, they may or may not result in differences in terms of user satisfaction. Phase 1 studies if and how p is adjusted. This has to be complemented by assessing if and how the modification of p impacts on the user satisfaction, as done by Phase 2.

Phase 2) Satisfaction Assessment

Phase 2 aims to assess the user satisfaction resulting from the adjustment of p. The participants are asked to quantify the difference in satisfaction between p_0 and the chosen p by means of a provided labelled scale. The Comparison Category Rating scale is used for this purpose [22]. The points and labels of this scale are displayed by the user interface that we use (Fig. 2).

This test design provides a post-screening criterion: the satisfaction experienced with the chosen p cannot be worse than the one with p_0. If p_0 is preferable, this should be selected in Phase 1. Hence, satisfaction levels lower than “The same as” reveal low reliability of the participant or the task being misunderstood. Subjects that select “Worse” or “Much worse” are excluded from the analysis of the results. We decide to accept “Slightly worse” because even if the selected p violates the test assumptions, it is likely to be close to what the participant actually prefers.

3.2 A/ST configuration for DE

The goal of the adjustment of a DE system is to find an enjoyable mix, where the dialogue can be easily followed. To this end, the control parameter p adjusts the dialogue-to-background ratio. All the outputs have equal integrated loudness [23].

We apply the A/ST on two DE systems: S_{OO} and S_{DS}. S_{OO} has access to the original dialogue and background objects and the adjustment of p does not introduce any distortion. S_{DS} estimates the audio objects from their stereo mixture by a blind source separation algorithm and may introduce distortions such as artefacts or changes in timbre. The default mixes (p_0) are used as inputs to the blind source separation.

The test participants are not explicitly informed that two systems are tested in the same session, but they can directly compare the adjusted mix with the default mix. This is free of artefacts and it can work as reference for the original timbre. The distortions potentially introduced by S_{DS} should be clear to the user and they can be taken into consideration for the choice of p.

The subjects are first asked to imagine being home and watching television for a long time. While listening to the training item, they are asked to adjust the overall volume. All subsequent items are loudness normalised to equal integrated loudness.

The core text of the instructions is as follows: *“Some test items may contain speech that is difficult or tiring to understand. If this is the case, you want to change the audio so that you can easily follow the speech, yet keeping the rest of the content (e.g., background music) enjoyable. To this end, you can adjust the relative level of the speech by means of the provided knob. Please note that the speech adjustment process may cause a degradation in quality. The graphic interface (Fig. 1) shows visual feedback (not shown in Fig. 1) while you are operating the knob: a blue frame around one of the turn knob icons indicates that the audio is changed according to the direction of rotation; a red frame around the icon indicates that the audio cannot be modified further in that direction. You can switch between the personalised setting and the default setting by pressing R and T on the keyboard. When you find the parameter setting that allows you to follow the speech easily, yet keeping the rest of the content enjoyable, please select it by pressing the knob from the top. If you like it, please feel free to select an extreme setting like minimal or maximal dialogue level (red frame). A new window (Fig. 2) will display a question about your satisfaction with the selected setting. Also here, you can compare the selected setting with the default setting. Please answer and press Enter to confirm. Now, you are done with the first test item. Please complete all the items in the same way. If you need a break at any time, you can pause the audio by pressing the spacebar.”*

The participants are asked to consider at the same time the enjoyment and the ease of listening to the dialogue. There are cases where these two goals diverge [1]. In these cases, the preferred trade-off has to be found.

3.3 Independent variables of the carried out test

Room set-up. The experiment is carried out in a listening room that resembles a quiet low-reverberant living room. Other listening environments could be simulated in future works. Stereo signals are reproduced over two Genelec 8250A studio monitors, which are positioned approximately at the height of the listener's head, 2 meters away from her/him, and 30° from her/his looking direction. The user interface is displayed on a TV positioned between the loudspeakers. The participants sit on a chair with fixed position, and the knob and the keyboard controlling the interface are on a little table nearby.

Subjects. The test involves 11 participants with normal hearing. They are voluntary, remunerated, non-expert, initiated[2], between 19 and 32 years old (median age is 25), and mostly German university students. Six of them claim to be passionate about Hi-Fi, music, or audio/video production: these subjects are referred to as *Hi-Fi lovers*. The other five claim no particular interest in audio besides using regularly the main platforms for music or film streaming: these subjects are referred to as *naive listeners*.

Test items. As test items, we use material that was broadcast in Germany or in the UK as well as artificially created mixes. In total, 13 signals are employed: one is used as the training item and 5 items are presented twice, once with S_{OO} and once with S_{DS}. The repetitions of one item are not presented one after the other, but interleaved with other items. Sampling frequency is 48 kHz. The length of the items varies between 8 and 17 seconds and the playback loops over the entire duration until the subject decides to proceed to the next item. The stereo backgrounds of the items comprise music (classical, ambient, jazz, and pop) and environmental recordings (rain, sea waves, cheering crowd, train station hall, and construction site). The dialogue is panned to the centre and feature German and English language, male and female talkers. The accompanying video for this material is not shown, as its quality can influence the perception of audio quality [24].

The item names are composed as follows. The name starts with "TV" for the recorded broadcast material, while it starts with "AR" for the artificially created mixes. A numerical ID and an underscore follow. Finally, the language of the content is indicated by "en" for English or by "de" for German.

Default mixes (p_0). The original broadcast signals are used as default mixes. They were selected so to have an original Source-to-Interference Ratio (SIR) of about 0 dB. Similar SIR values are used as p_0 for the artificial mixes. In this work, the SIR is the ratio of the power of the dialogue and the power of the background (considered as interference). The SIR can be computed only if the original dialogue and background objects are available. The SIR is known for the signals processed by S_{OO}, while it is estimated via the BSS Eval toolbox for the signals processed by S_{DS} [25].

Item	SIR_0 [dB]
AR2_de	0.97
AR4_en	1.17
AR1_de	2.43
TV3_en	−4.18
AR3_de	2.47
Mean	0.57

Table 1: SIR_0 values corresponding to p_0 for the mixes for which original audio objects are available.

The original SIR of the default mix is referred to as SIR_0. The difference between the adjusted SIR and SIR_0 is referred to as ΔSIR. The user-adjustable parameter p controls ΔSIR, while p_0 refers to SIR_0. Table 1 lists the SIR_0 values of the signals that are presented with both S_{OO} and S_{DS}. One signal with $\mathrm{SIR}_0 \gg 0$ dB is also presented in the test. This is AR5_de and consists of the same dialogue and background signals of AR2_de, but they are mixed with $\mathrm{SIR}_0 = 18$ dB.

Available range for *p*. Values of ΔSIR ranging from 0 to +15 dB are used for S_{OO}, with steps of 0.5 dB. The same range is used for the so-called *nominal* ΔSIR of S_{DS}. However, the performance of S_{DS} is item-dependent and the actual amplification of the dialogue varies. The (actual) ΔSIR values corresponding to the maximum nominal ΔSIR (+15 dB) are shown later on (Fig. 6). When S_{DS} is used with nominal $\Delta\mathrm{SIR} = +15$ dB, the Source-to-Artefacts Ratio (SAR) [25] ranges between 9 and 13 dB.

Implementation. The test software is implemented in Max/MSP and it is made available for general non-commercial use at `https://www.audiolabs-erlangen.de/resources/2017-AES-AST`.

4 Results

For the adjustment phase and the satisfaction assessment phase, the collected data points lie in a space with three dimensions: items, listeners, and selections. In the visualisation of the results, a symbol and a colour

[2] A person who has already taken part in a sensory test is referred to as *initiated*.

are fixed for each item. The entire collected data is shown in Fig. 3 (adjustment phase) and Fig. 4 (satisfaction assessment phase). The left-hand plots of Figs. 3 and 4 show the projection of the 3D data space onto the item plane, while the right-hand plots show the same data projected onto the listener plane. Figs. 3 and 4 also show a comparison between the case with S_{DS} and the one with S_{OO}, please refer to the plot titles. Bigger markers are used in case data points overlap. The size of the markers is proportional to the number of overlapping points, which is printed inside the marker. In the plots on the right, dark grey circles are used when many different markers overlap.

Following the post-screening criterion introduced in Section 3.1, one naive listener is excluded from the analysis of the results, as she/he selected "Worse" once in the satisfaction assessment phase. The remaining ten subjects are represented by numerical labels: from 1 to 4 for naive listeners and from 5 to 10 for Hi-Fi lovers. The ordering of the items from left to right reflects the order in which they were presented in the test.

The reader should not be overwhelmed if Figs. 3 and 4 appear complex at a first sight: a large amount of information is indeed displayed. The most evident fact is, however, that there is a very high variance among listeners and items in terms of preferred ΔSIR (Fig. 3). In fact, preference clusters cannot be observed. The implication is that an expert would not be able to determine a dialogue-to-background ratio that would be also the chosen one for most of the subjects. The adjustment usually translates into positive satisfaction levels (Fig. 4), meaning that it has a positive noticeable effect.

This means that the personalisation offered by DE is desired, even by subjects with normal hearing in quiet and controlled listening conditions. Personal taste is likely to be the main reason behind this discovery. Personalisation is extensively used not only for the item processed by S_{OO}, but also for the items processed by S_{DS}. This suggests that S_{DS} offers a useful service, despite the potential artefacts or change in timbre.

Clusters of data points can be observed only for the extreme setting ΔSIR= 15 dB. We suspect that this is due to the limited provided range: the results are likely to be clipped and they could be even more spread if a larger range were provided, especially for the naive listeners 2 and 3. The low number of participants in each listener category (naive, Hi-Fi lovers) makes a detailed analysis of the categories impossible. Still, the observed trend would suggest that naive listeners prefer higher levels of dialogue than the Hi-Fi lovers, even if with high personal variations. This should be investigated in future.

An indication of the coherent behaviour of the participants throughout the test is given by the ΔSIR selected for AR2_de and AR5_de, i.e., the items created by mixing with different SIR_0 the same dialogue and background objects. Almost all the listeners select a ΔSIR for AR2_de (SIR_0=0.97 dB) that is significantly bigger than the one that they select for AR5_de (SIR_0=18 dB).

Fig. 5 depicts the mean of the listeners' adjustments and satisfaction levels for S_{DS}, together with 95% confidence intervals[3] as well as minimum and maximum selections (filled circles). A clear correlation between the levels of ΔSIR and satisfaction can be observed (Pearson's $r = 0.81$). Furthermore, Fig. 6 compares the selections for the items presented with both S_{OO} and S_{DS}. The upper plot of Fig. 6 depicts also the actual maximum ΔSIR achievable by S_{DS}.

There is a clear correlation also in the adjustment of S_{OO} and S_{DS} ($r = 0.93$) and in the satisfaction they provide ($r = 0.99$). Still, lower levels of ΔSIR are preferred for S_{DS}: on average 3 dB lower than S_{OO}, resulting in lower satisfaction. As confirmed by interviewing the participants, this is due to the fact that the subjects have to trade-off between the desired ΔSIR (selected while operating S_{OO}) and the change in timbre and the artefacts, which S_{DS} introduces for high values of ΔSIR.

5 Conclusions

We proposed the A/ST, a novel perceptual test for the evaluation of user-adjustable systems. The aim is to understand if and to what extent personalisation is desired for such systems. The impact of the personalisation is studied on a satisfaction scale. In addition, the performance of the system under test can be compared with the one of an ideal system.

The A/ST was employed to evaluate a DE system where dialogue and background are estimated from their stereo mixture via blind source separation. This is compared with a DE system with ideal quality.

[3] The 95% confidence interval is defined as $\left[\overline{X} \pm \frac{s}{\sqrt{N}} t_{(.975,N-1)}\right]$, where $\overline{X}$ is the sample mean, s is the corrected sample standard deviation, t is read from the Student's t distribution table, and N is the number of considered data points. We do not adopt the common approximation of using $t_{(.975,\infty)} = 1.96$ independently of N.

Personalisation was extensively used and resulted in clearly increased user satisfaction, despite the distortions that the source separation may introduce. Even if the test was carried out with a homogeneous group of normal hearing subjects in quiet and controlled listening conditions, we observed very high variance among listeners and items in terms of the preferred dialogue-to-background ratio. This can be explained as a consequence of personal taste. The full available personalisation range was found to be useful: an even larger range could be desired and can be considered in the future.

References

[1] B. Shirley and P. Kendrick, "ITC Clean Audio Project," in *AES 116th Conv., Berlin*, 2004.

[2] H. Fuchs and D. Oetting, "Advanced clean audio solution: Dialogue enhancement," *SMPTE Motion Imaging J.*, vol. 123, no. 5, 2014.

[3] B. Shirley, M. Meadows, *et al.*, "Personalized Object-Based Audio for Hearing Impaired TV Viewers," *J. of the AES*, vol. 65, no. 4, 2017.

[4] T. Walton, M. Evans, *et al.*, "Does Environmental Noise Influence Preference of Background-Foreground Audio Balance?," in *AES 141st Conv., Los Angeles*, 2016.

[5] P. Mapp, "Intelligibility of Cinema & TV Sound Dialogue," in *AES 141st Conv., Los Angeles*, 2016.

[6] M. Florentine, "Speech Perception in Noise by Fluent, Non-native Listeners," *J. Acoust. Soc. Am.*, vol. 77, no. S1, 1985.

[7] H. Fuchs, S. Tuff, and C. Bustad, "Dialogue Enhancement - technology and experiments," *EBU Technical Review*, vol. Q2, 2012.

[8] M. Armstrong, "Audio Processing and Speech Intelligibility: a literature review," in *BBC Research & Development White Paper, WHP190*, 2011.

[9] R. L. Bleidt, D. Sen, *et al.*, "Development of the MPEG-H TV Audio System for ATSC 3.0," *IEEE Trans. Broadcasting*, vol. 63, no. 1, 2017.

[10] J. Riedmiller, K. Kjörling, *et al.*, "Delivering Scalable Audio Experiences using AC-4," *IEEE Trans. Broadcasting*, vol. 63, no. 1, 2017.

[11] J. Paulus, J. Herre, *et al.*, "MPEG-D Spatial Audio Object Coding for Dialogue Enhancement (SAOC-DE)," in *AES 138th Conv., Warsaw*, 2015.

[12] L. Claudy, "What's going on in Korea with UHD Broadcasting?." PILOT, 2017. `https://nabpilot.org/whats-going-on-in-korea-with-uhd-broadcasting/`.

[13] Advanced Television Systems Committee, "ATSC 3.0 Standards: A/342 Part 3, MPEG-H System," 2017.

[14] P. Loizou, *Speech Enhancement: Theory and Practice*. Boca Raton, FL: CRC, 2007.

[15] S. Makino, T.-W. Lee, and H. Sawada, *Blind Speech Separation*. Springer, 2007.

[16] M. Torcoli and C. Uhle, "On the Effect of Artificial Distortions on Objective Performance Measures for Dialog Enhancement," in *AES 141st Conv., Los Angeles*, 2016.

[17] C. Uhle, O. Hellmuth, and J. Weigel, "Speech enhancement of movie sound," in *AES 125th Conv., San Francisco*, 2008.

[18] J. T. Geiger, P. Grosche, and Y. L. Parodi, "Dialogue Enhancement of Stereo Sound," in *IEEE 23rd EUSIPCO, Nice*, 2015.

[19] A. Craciun, C. Uhle, and T. Bäckström, "An evaluation of stereo speech enhancement methods for different audio-visual scenarios," in *IEEE 23rd EUSIPCO, Nice*, 2015.

[20] J. Paulus, C. Uhle, *et al.*, "A Study on the Preferred Level of Late Reverberation in Speech and Music," in *AES 60th Conf. (DREAMS), Leuven*, 2016.

[21] S. Bech and N. Zacharov, *Perceptual Audio Evaluation - Theory, Method and Application*. John Wiley & Sons, 2006.

[22] ITU-T Rec. P.800, "Methods for subjective determination of transmission quality," 1996.

[23] ITU-R Rec. BS.1770-2, "Algorithms to measure audio programme loudness and true-peak audio level," 2011.

[24] J. Beerends and F. De Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *J. of the AES*, vol. 47, no. 5, 1999.

[25] E. Vincent, R. Gribonval, and C. Févotte, "Performance Measurement in Blind Audio Source Separation," *IEEE Trans. Audio, Speech and Language Process.*, vol. 14, no. 4, 2006.

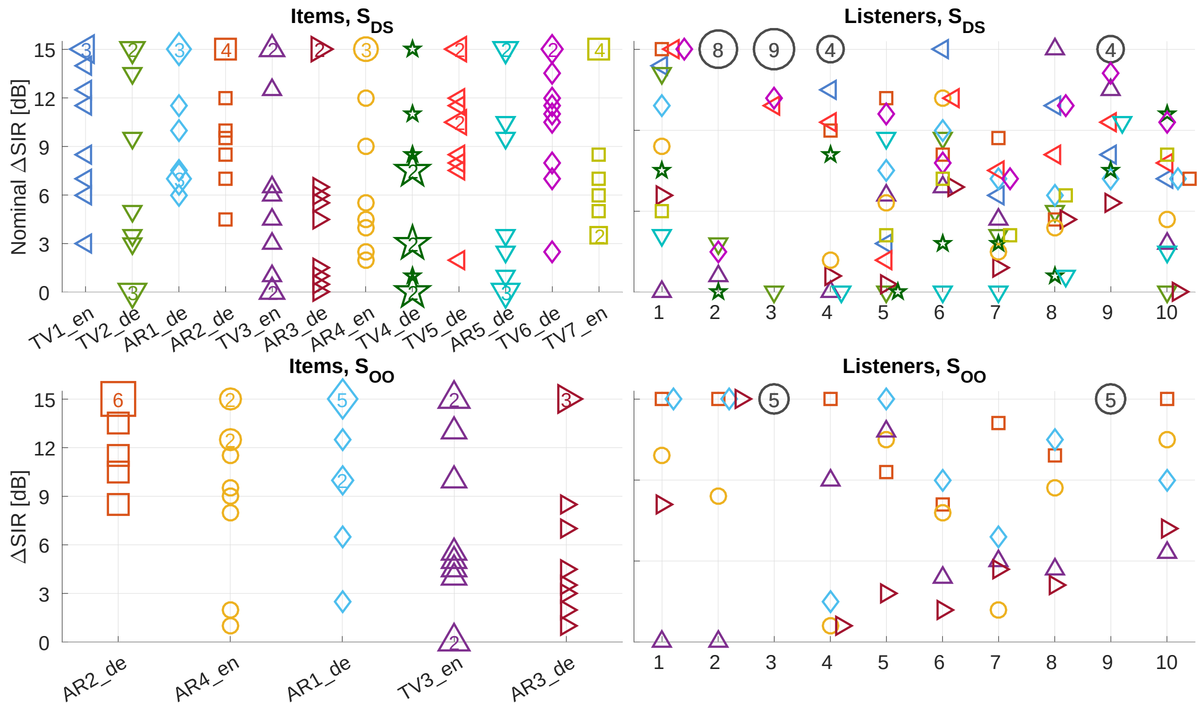


Fig. 3: Adjustment phase: the selections for the dialogue-to-background ratio (ΔSIR) while operating S_{DS} (upper plots) and S_{OO} (lower plots) are projected onto the item plane (plots on the left) and on the listener plane (plots on the right). A symbol and a colour are fixed for each item.

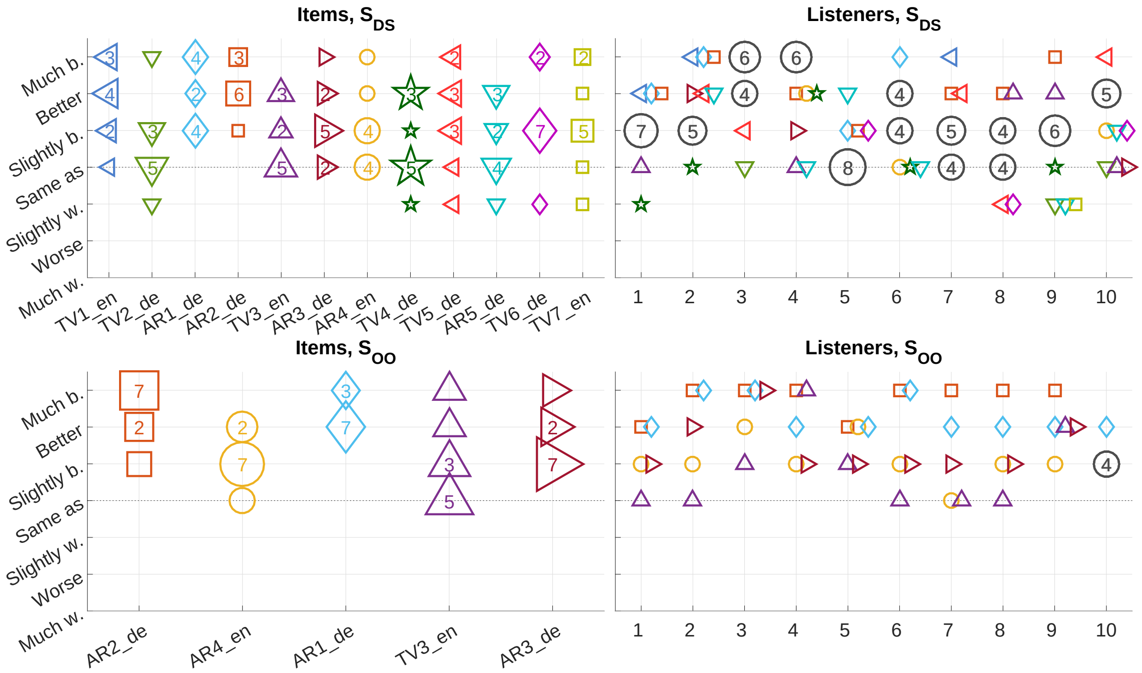


Fig. 4: Satisfaction assessment: satisfaction levels related to the adjustment of S_{DS} (upper plots) and S_{OO} (lower plots) projected onto the item plane (plots on the left) and onto the listener plane (plots on the right).

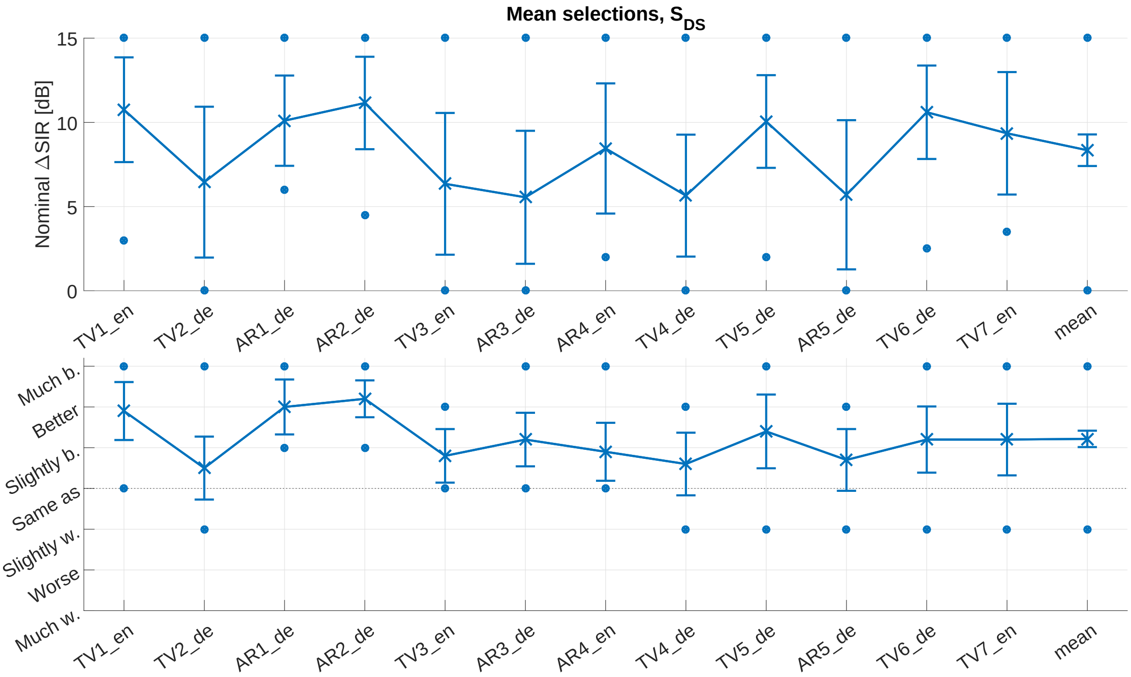


Fig. 5: Mean selected nominal ΔSIR and satisfaction level together with 95% confidence intervals and minimum and maximum selections (filled circles) while operating S_{DS}, i.e., DE employs source separation.

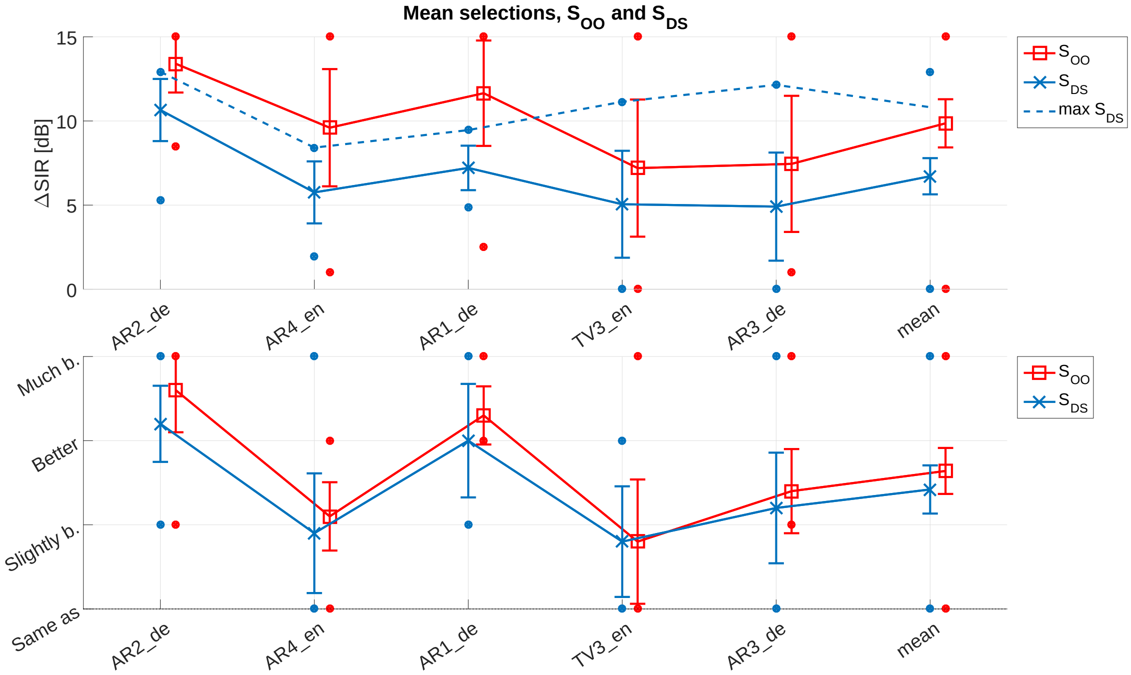


Fig. 6: Mean selected ΔSIR and satisfaction level together with 95% confidence intervals and minimum and maximum selections (filled circles). S_{OO} is compared with S_{DS}. The maximum ΔSIR achievable by S_{DS} is also depicted (dashed line).