International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

FG AVA TR

Version 1.0
(10/2013)

Focus Group on Audiovisual
Media Accessibility

Technical Report

Part 12: Methods for improving the intelligibility of audio

FOREWORD

The procedures for establishment of focus groups are defined in Recommendation ITU-T A.7. The ITU-T Focus Group on Audiovisual Media Accessibility (FG AVA) was proposed by ITU-T Study Group 16 for creation in-between TSAG meetings and it was established on 22 May 2011. The Focus Group was successfully concluded in October 2013.

Even though focus groups have a parent organization, they are organized independently from the usual operating procedures of ITU, and are financially independent. Texts approved by focus groups (including Technical Reports) do not have the same status as ITU-T Recommendations.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Technical Report may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU-T Focus Group participants or others outside of the Technical Report development process.

CONTENTS

List of Tables

List of Figures

Summary

This Technical Report of FG AVA was prepared by Working Group C "Visual signing and sign language" and D "Emerging access services". It outlines the methods for improving the intelligibility of audio that FG AVA has foreseen as a future work item for ITU-T Study Group 16 (SG16) "Multimedia" and ITU-R SG9 "Broadcasting service".

1 Provision of appropriate audio to the elderly

When offering audio media or audiovisual media to the elderly, it should be taken into account that their auditory perceptual functions and audio-lingual processing functions may have been diminished. The following points should be taken into consideration.

1.1 Programme audio with an appropriate mixing balance should be provided for the elderly. That is, an appropriate balance should be achieved between foreground audio (a programme's main audio, consisting of the audio of the announcers' and actors' voices) and background audio (consisting of signals other than those of the main audio. It is mainly sounds for special effects and background music). This type of mixing balance tuning for the elderly is sometimes called "clean audio".

Hearing function declines with aging in general. Due to interference from background noise, the elderly often have difficulty in understanding the speech of a person they are talking with. When listening to television and radio programmes, they also experience a similar difficulty, due to background sound in the sound track of the programmes. Elderly persons sometimes say that performers on radio and TV programmes are difficult to understand because of interference from special effects sounds and background music. This Technical Report addresses this problem.

1.2 Programme audio at an appropriate speech rate for the elderly should also be provided. Generally, cognitive functions decline with aging. Since auditory perception function and audio lingual processing speed also decline with aging, the elderly have difficulty in understanding speech spoken at a fast rate. This Technical Report also addresses this problem.

1.3 It should be noted that the above declines in functions that occur with aging will vary considerably according to age and the individual.

2 Ways of providing a balanced mix

2.1 Methods of realization

There are four methods as ways of realizing 1.1.

2-1: Change the mixing balance on the transmitting side and transmit several versions of the audio stream with a different balance.

2-2: Generate an auxiliary signal to help separating foreground audio and background audio from the normal-balanced sound at the transmitting side, and balance the foreground audio and background audio at the receiving side using the auxiliary signal.

2-3: Separate the foreground audio and background audio at the receiving side without an auxiliary information and adjust the mixing balance.

2-4: Transmit foreground and background audio as independent audio streams and perform the mix at the receiver.

2.2 Merits and demerits of each method

Table 1 shows the criteria for evaluating the merits and demerits of the four methods and their evaluation.

Table 1-Evaluation of methods of providing mixing balance for the elderly

	User adaptability	F/B separation	Production load	Transmission compatibility	Transmission bitrate overhead	Receiver compatibility	Receiver complexity
2-1	-	++	--	+	--	++	++
2-2	++	+	-	++	+	+	-
2-3	+	-	++	++	++	+	-
2-4	++	++	--	-	--	--	-

2.3 Criteria definitions and evaluation reasons

Definitions of the evaluation criteria in Table 1 are given below. The reasons for the merits and demerits of each method are also given below.

User adaptability: Degree of freedom in adjusting the mixing balance of foreground audio and background audio on the receiving side.

2-1: The user can choose a normal broadcasting programme or a sub-audio channel programme (programme audio that reduces the background sound), but the choice is limited to a few number of stages, because a complete mix has to be transmitted for every stage.

2-2: The mixing balance of the foreground audio and the background audio can be changed continuously in a substantial range (typical setting is +/-12 dB). In principle, it is possible to enable complete foreground-only or background-only audio, but only with additional bitrate.

2-3: The mixing balance of the foreground audio and the background audio can be changed in several stages. The range of change is limited (e.g. to 4.5 dB as described in the example in Appendix 3) and depends on accuracy of separation of foreground and background audio.

2-4: The user has the full freedom to balance the mix and choose between foreground-only, background-only and all values in-between.

F/B separation: This is an evaluation of the remix sound quality and the precision of separation between foreground audio and background audio of the programme.

2-1: This method uses sound material that was originally recorded separately to change the mixing balance. Thus there are no factors that deteriorate due to separation, and there is no deterioration in the quality of the sound.

2-2: The sound quality is good because this method utilizes an auxiliary signal for separation. The signal is generated from normal broadcasting sound and input sound material at the broadcasting side.

2-3: This method separates the foreground audio and background audio on the receiving side without any additional signal from the broadcasting side. Thus, the degree of separation is limited and will depend on the precision of the signal processing used. The achievable separation is also signal dependent, e.g. if the background is too loud it is difficult to extract the foreground in sufficient quality.

2-4: Regarding the achievable foreground/background separation of this method is similar to method 2-1.

Production load: The load on the programme creator and on the transmitting side.

2-1: This method requires operations to mix and record with one or more alternative versions (programme audio with reduced background audio).

2-2: This method requires separate input of the normal audio mix and a separate foreground (or background) signal into the encoder. An extended encoder device is necessary for the generation of the side information.

2-3: No processing on the broadcasting side is necessary.

2-4: This method requires separate input of the foreground and background signals into the encoder. The mixing operation stage needs to be changed: both signals need to be balanced, but not mixed into a single signal. Additional information needs to be generated for transmission to potentially guide the receiving devices for the mix. Loudness normalization may also be more complex for this method.

Transmission compatibility: Compatibility with the current broadcasting transmission system.

2-1: Transmission is possible using current digital broadcasting with supplementary sound channels.

2-2: Transmission is possible using current digital broadcasting. No additional supplementary sound channel is needed. The additional side information is embedded as a part of the audio stream during audio encoding and thus it is transparent to the transmission system. No change to the transmission system is necessary.

2-3: Transmission is possible using current digital broadcasting. No change to the transmission system and equipment is required at all.

2-4: Transmission is possible using current digital broadcasting with two supplementary sound channels.

Transmission bitrate overhead: How much additional bitrate is necessary for additionally transmitted audio data?

2-1: One complete audio channel is necessary for every additional mix, so for example it doubles the required audio bitrate for one additionally delivered mix.

2-2: Additional bitrate is required, but substantially less than 2x of one audio stream.

2-3: No additional bitrate required.

2-4: Doubles the required audio bitrate. If a mixed signal needs to be provided to be backward compatible to legacy receivers: three times of the bitrate is required.

Receiver compatibility: Compatibility with current broadcast receivers.

2-1: Compatible. Reception is possible using current supplementary sound channel broadcasting.

2-2: Requires a receiver that can decode the side information and manipulate the mix. It is backward compatible for existing receivers that ignore the additional side information and play the default mix.

2-3: Requires use of a new audio processing system to separate foreground and background sound from the original one and change the mix. Backward compatible to existing receivers that play the default mix.

2-4: Requires a receiver that can decode multiple audio streams and mix those streams. Backward compatible only when mixed signal is provided in addition to separate foreground and background signals.

Receiver complexity: Complexity of the receiver.

2-1: No change necessary. Reception is possible with current receivers.

2-2: Requires a decoding chip with higher performance to decode the additional side information and change the mix of foreground and background.

2-3: Requires a processing function to separate foreground audio and background audio on the receiving side and therefore a decoding chip with higher performance to enable this function.

2-4: Requires two parallel audio decoders for the separate background and foreground audio and an additional mixing functionality.

3 Methods of adjusting the speech rate

3.1 Methods of realization

As a method of realizing 1.2, the following two methods have been proposed.

3-1: Adjusting the speech rate on the transmitting side.

3-2: Adjusting the speech rate on the receiving side.

3.2 Merits and demerits of each method

Table 2 shows the criteria for evaluating the merits and demerits of the two methods and their evaluations.

Table 2-Methods of providing speech rate conversion audio

	User adaptability	Production load	Transmission compatibility	Receiver compatibility	Receiver complexity
3-1	-	-	+	+	+
3-2	+	+	+	-	-

3.3 Evaluation criteria and reasons

User adaptability: Degree of freedom in adjusting the speech rate on the receiver side.

3-1: The user can choose normal broadcasting or a sub-audio channel programme (audio with an altered speed rate), but the choice is limited to two stages.

3-2: The speech rate of the programme audio can be changed in several stages.

Production load: The load on the programme creator and on the broadcasting side.

3-1: This method requires a device that creates a sub-audio channel programme (programme audio with a changed speed rate).

3-2: Processing by the broadcasting side system is not necessary.

Transmission compatibility: Compatibility with the current broadcasting transmission system.

3-1: Transmission possible using current digital broadcasting with supplementary sound channel.

3-2: Transmission possible using current digital broadcasting.

Receiver compatibility: Compatibility with current broadcast receivers.

3-1: Compatible. Reception possible using current supplementary sound channel broadcasting.

3-2: Requires use of a new audio processing system

Receiver complexity: Complexity of the receiver.

3-1: No change necessary. Reception possible with current receivers.

3-2: This method requires a new speech rate conversion chip for the receiver.

4 Integrated broadcast and broadband (IBB) service

Recently, research on integrated broadcast and broadband (IBB) services has been making good progress. These services would use not only transmission channels for broadcasting but would also use complementary communication network. On a receiver, they would combine an audiovisual programme that is sent through broadcasting with supplementary information that is sent through a communication network. Audio services for the elderly can be candidates for such new services. In such a case, Tables 1 and 2 should be revised as Tables 3 and 4 below, respectively.

4.1 Use of an IBB service to offer a mixing balance for the elderly

Use of an IBB service is one means of offering a mixing audio balancefor the elderly. This method provides another transmission channel that uses a communication network instead of a supplementary sound channel on a digital television broadcast. An IBB service generally has wider bandwidth than a supplementary sound channel and therefore audio signals with more varieties of mixing balance can be provided. The row 2-1' in Table 3 shows the evaluations for this method.

Another advantage is that the additional audio only needs to be delivered to those viewers that request it on demand, contrary to a supplemental audio stream within the broadcast that is delivered to everyone. Therefore there is no burden on the transmission system with offering more options with different audio mixes or other additional audio versions.

However, the additional audio that is delivered on the broadband channel needs to be synchronized to the video that is delivered on the broadcast channel (lip sync). To enable this synchronization of transmission signals on both channels, broadcast and broadband need to be standardized. The more complex synchronization also results in an additional burden on the receiving device and may also increase the production load to provide the synchronization.

Table 3-Evaluation of an IBB service that provides mixing balance for the elderly

	User adaptability	F/B separation	Production load	Transmission compatibility	Transmission bitrate overhead	Receiver compatibility	Receiver complexity
2-1	-	++	--	+	--	++	++
2-1'	++	++	--	--	--	-	-

4.1.1 Reasons and changes in the evaluations

User adaptability: Degree of freedom in adjusting the balance of programme sound on the receiver side.

2-1': The mixing balance of the foreground audio and the background audio can be changed in several stages. Using a broadband connection on the communication network to deliver the alternative mixes allows offering a wider choice of mixes compared to using supplementary audio streams in the broadcast system.

F/B separation: This is an evaluation of the remix sound quality and the precision of separation between the programme's foreground audio and background audio.

2-1': This method provides audio signals with various mixing balance produced at the broadcasting side. Thus, there are no factors of deterioration in the quality of the sound.

Production load: The load on the programme creator and on the transmitting side.

2-1': It is necessary to make audio signals with various mixing balances, which may take additional load to production clause.

Transmission compatibility: Compatibility with the current broadcasting transmission system.

2-1': Necessary to use facilities and equipment of the communication network.

Receiver compatibility: Compatibility with current broadcast receivers.

2-1': Necessary to use a receiver that can receive and decode transmission signals from the communication network in addition to the broadcast network reception.

Receiver complexity: Complexity of the receiver.

2-1': Receiver complexity is increased through reception of signals over communication lines and through the addition of buffer memory for use in simultaneous broadcasting.

4.2 Use of an IBB service to provide speech rate conversion

Use of an IBB service is also another means of offering speech rate conversion audio. Speech rate conversion audio is assumed to be transmitted through communication lines instead of digital television supplementary audio channels. The row 3-1' in Table 4 gives the evaluation for this case.

Table 4-Use of an IBB method to provide speech rate conversion audio

	User adaptability	Production load	Transmission compatibility	Receiver compatibility	Receiver complexity
3-1	-	-	+	+	+
3-1'	+	--	--	--	-
3-2	+	+	+	+	-

4.2.1 Reasons and changes in the evaluations

User adaptability: Degree of freedom in adjusting the speech rate on the receiver side.

3-1': If communication transmission lines have sufficient bandwidth, speech rate services for programme audio can be provided in several stages.

Production load: The load on the programme creator and on the broadcasting side.

3-1': This requires a device that can create programme audio with a speech rate conversion in several stages.

Transmission compatibility: Compatibility with the current broadcasting transmission system.

3-1': Transmission is possible using complementary communication network.

Receiver compatibility: Compatibility with current broadcast receivers.

3-1': Requires adoption of a receiver that can receive and decode transmission signals over communication lines.

Receiver complexity: Complexity of the receiver.

3-1': Receiver complexity is increased through reception of transmission signals over communication lines and through the addition of buffer memory for use in simultaneous broadcasting.

5 Conclusion

This Technical Report has provided information on two types of audio services for elderly people to understand and enjoy TV and radio programmes better: one is "clean audio" and the other is "speech rate conversion". Broadcasters and TV manufacturers are encouraged to provide better services and better receivers for elderly people by referencing to this Technical Report.

Additional information and examples of implementation of these services are partially provided in Appendices I to III.

Appendix I:

Using supplementary audio channel for improved audio quality for the elderly

Appendix I gives an example of providing audio with mixing balance for the elderly, using supplementary audio channel (see row 2-1 in Table 1).

Regarding the mixing balance of foreground and background audio, the experiments of NHK program production engineers has shown that lowering the sound level of background sound to 6 dB from the normal mixing level acquired better evaluation by elderly people. Lowering the sound to 6 dB made the audio easier to understand for X4 persons (among X3 number of evaluators) from ages X1 to X2 (see [I-1]).

I.1 References

[I-1] H. Nakamura, M. Sawaguchi, K. Masaoka, K. Watanabe, Y. Yamasaki, E. Miyasaka, M. Yasuoka, H. Seki (2003), *Better Audio Balance Broadcasting service for elderly people.-Back Ground Sound Levels of Television Programs for Easy Listening*, Proc. Spring Meet. Acoust. Soc. Jpn., 1-5-5, pp. 455-456, (in Japanese).

Appendix II:

Using auxiliary signal for balancing foreground and background audio

Appendix II gives an example of a system that generates an auxiliary signal at the broadcasting side, making the best balance of the foreground and background audio at the receiving side (see row 2-2 in Table 1).

II.1 Dialogue enhancement as an advanced clean audio solution

Dialogue enhancement is based on MPEG SAOC (ISO/IEC 23003-2), [see II-1]. MPEG SAOC (spatial audio object coding) is a powerful technology allowing the manipulation of any number of audio objects during rendering. The main principle is that all audio objects are mixed to one audio down-mix signal with one or more channels for transmission. Object-related side information is transmitted along with the down-mix. This side information enables the object manipulation in the receiver based on user interaction (see [II-2] for more details on SAOC and [II-3] for more information on dialogue enhancement).

Examples of objects are single instruments and the singer's voice in a music recording or the dialogue in a movie sound. SAOC allows gain changes and re-panning of objects in the stereo or surrounding scene. For dialogue enhancement (SAOC-DE), only a subset of the MPEG SAOC functionality is needed. To reduce decoder complexity, only two controllable objects are used (for dialogue and background). Furthermore, only loudness level attenuation or amplification is required for interactivity during rendering, and re-panning of objects is not used.

II.2 Basic principle

Figure II.1 shows the end-to-end signal flow from the encoder input to the receiver output. The SAOC-DE encoder analyses the objects from the source signals and produces a stream of parametric side information. The final mix is not touched and created externally beforehand. The encoding path from the mix input signal to the encoded advanced audio coding (AAC) audio bit stream is unchanged.

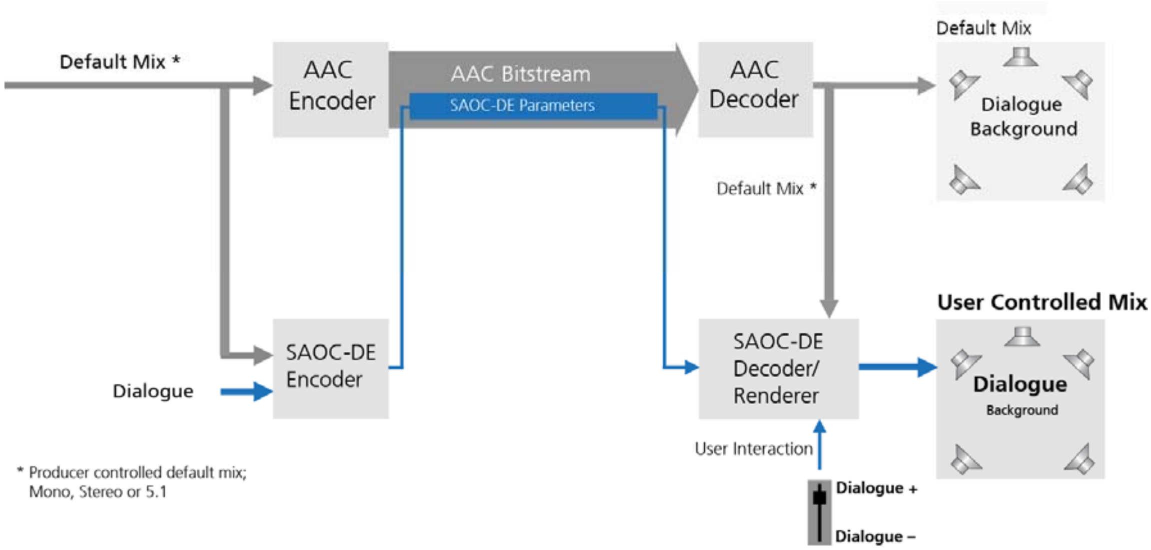

Figure II.1-Dialogue Enhancement end-to-end signal flow

The dialogue signal needs to be available as a second input signal. The background signal is re-created from mix and dialogue so that both source signals are available for the analysis in the SAOC-DE encoder.

The mixed signal is encoded with any audio codec. In this example, MPEG-4 advanced audio coding (AAC)/high efficiency (HE)-AAC is used. The stream of parametric side information can be embedded into the encoded audio bit stream, as in case of AAC/HE-AAC, or transported in a separate stream.

On the receiving side, the audio bit stream is decoded. The SAOC-DE decoder takes the decoded down-mix signal and uses the descriptive data from the parameter bit stream to enable access to the audio sources. The user is then able to adjust the volume of the dialogue and background sources individually, e.g. to improve the intelligibility of the dialogue or sports commentary.

II.3 Dialogue enhancement modes

SAOC-DE supports two different encoding modes for the side information data stream:

- The basic parametric mode as described above.
- The enhanced (residual) mode that additionally embeds residual waveform information for the dialogue object into the side information.

The parametric mode is typically useful for dialogue level modification up to 6 dB.

The residual mode allows for more attenuation or enhancement at a high quality (beyond 6 dB), but this entails that additional bitrate is needed for the residual information. From discussions with broadcasters, an upper limit of 12 dB for the alterations was found a good compromise between bitrate and flexibility. The value of 12 dB is also used in the experiments described below.

SAOC-DE can be used for any number of transmission channels (e.g. mono, stereo or multichannel). The dialogue can have one or more channels, e.g. mono or stereo. In case of multichannel audio, two typical modes became apparent from discussions with broadcasters: the dialogue is either present in the centre channel only or it is part of the front (left, centre and right) channels.

II.4 Advantages

The SAOC-DE technology is completely compatible with existing transmission and playback equipment. Legacy devices not capable of decoding the parametric side information will ignore it and play back the default mix signal.

Another advantage of the technology is that only one audio track with additional side information needs to be transmitted. This saves bandwidth compared to transmitting several pre-mixed versions or separate dialogue and background audio tracks. Furthermore, the broadcaster is able to control/limit the amount and the balance of the mix can be changed by their audience.

II.5 Speech intelligibility listening test

The goal of this test was to verify the applicability of SAOC-DE for a hearing-impaired audience. Therefore, a speech intelligibility test was conducted.

Ten hearing-impaired listeners[1] showing a typical high frequency age-related hearing loss participated in this experiment. The test was conducted without hearing aids. Ten normal-hearing

[1] Mean age: 73.3 (± 3.7 years), PTA between 25 and 45 dB (PTA = Pure-tone average: average of hearing loss at 500, 1000, 2000 and 4000 Hz).

listeners[2] served as a reference group. All listeners performed the Oldenburg sentence test (OLSA). The OLSA test is also used to determine the improvement of speech intelligibility with hearing aids [II-5]. During the OLSA test, listeners have to understand sentences of five words (e.g. "Peter has five green cars."), and the number of the correctly identified words is counted. Ten sentences were presented for each test condition.

The signal of the spoken sentences was mixed with a noise signal. Two different noise signals were used:

– Speech shaped noise (SSN) with a frequency characteristic similar to speech.

– "Applause" of clapping audience.

The default mix of speech and noise was set to a level of approximately 50% intelligibility, i.e. the hearing-impaired listeners could correctly identify 50% of the words during the test. The assumed default mix for SSN was at -4 dB signal-to-noise ratio (SNR), and for the applause noise at -14 dB SNR. As a reference point, the normal-hearing listeners performed the test also with this default mix. The dialogue enhancement technology was used to process the (speech and noise) mix signal in order to change the speech-to-noise balance of the mix in the test. Two different loudness levels for remixing were tested: Attenuation of the noise by 6 dB and by 12 dB.

II.6 Test results

The results for the different conditions are shown in Figure II.2 for the SSN noise and in Figure II.3 for the "applause" noise. Note that in the experiment slightly lower levels than the intended 50% intelligibility for the default mix were achieved for hearing-impaired listeners (46% for SSN and 34% for "applause"). Therefore, for each noise signal a different reference point for comparison of the results has to be taken into account.

The main findings are:

– SAOC-DE at 12 dB increases the speech intelligibility from 46% to 91% for SSN and from 34% to 81% for the "applause" noise condition. The achieved intelligibility level corresponds to the intelligibility of normal-hearing listeners with the default mix. Note that normal-hearing listeners showed a considerable difference between both noise signals: 95% for SSN and only 72% for "applause".

– SAOC-DE at 6 dB showed a substantial improvement in intelligibility to 86% for SSN and to 62% for applause.

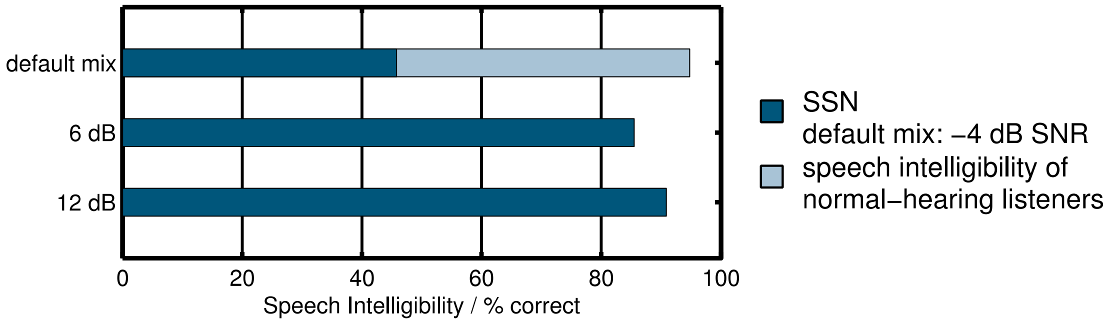

Figure II.2-Results of the OLSA test for hearing-impaired listeners for different SAOC-DE conditions with stationary speech shaped noise (SSN)

[2] Mean age: 23.3 (± 2.2 years) with audiometric thresholds of 20 dB Hearing Level or better at the audiometric test frequencies [II-4].

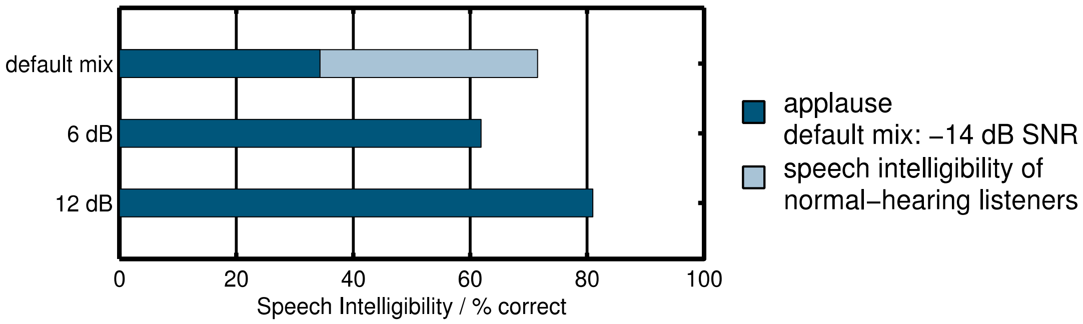

Figure II.3-Results of the OLSA test for hearing-impaired listeners for different SAOC-DE conditions with applause noise

II.7 References

[II-1] ISO/IEC 23003-2:2010, *Information technology − MPEG audio technologies − Part 2: Spatial Audio Object Coding (SAOC*).

[II-2] Hellmuth, O. *et al.* (2010), *MPEG Spatial Audio Object Coding-The ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes*, 129th AES Convention, Nov.

[II-3] Fuchs, H., Tuff, S. and Bustad C. (2012), *Dialogue Enhancement-technology and experiments*, EBU Technology Review, June.. http://tech.ebu.ch/webdav/site/tech/shared/techreview/trev_2012-Q2_Dialogue-Enhancement_Fuchs.pdf.

[II-4] ANSI S3.22-2003, *Specification of Hearing Aid Characteristics*.

[II-5] Wagener, K., Brand, T. and Kollmeier, B. (1999), Entwicklung und Evaluation eines Satztests für die deutsche Sprache Teil III: Evaluation des Oldenburger Satztests. Zeitschrift für Audiologie, 38(3), pp. 86-95.

Appendix III:

Separating the foreground audio and background audio at the receiver side

Appendix III gives an example of separating the foreground audio and background audio at the receiving side without any additional information from the transmission side (see row 2-3 in Table 1).

III.1 Sound level adjustment system in TV programmes at receiver side

As people get older, they tend to become distracted by background sound in TV programmes and find that the voices of announcers and actors become harder to make out. In fact, it was reported that programmes can be made easier for elderly people to listen to by suppressing the background sound components within the programmes by 3 to 6 dB [III-1]. Investigative research is being conducted in order to make the voices in broadcast programmes easier for elderly people to listen to. In this appendix, a novel method for controlling the voice level against background sound in broadcast programmes is described, and the results of an evaluation with a prototype device are shown.

A schematic diagram of the proposed method is shown in Figure III.1. The novel method controls the voice level against background sound at the receiver side by assuming that voice is located at the centre position in the stereo sound of broadcast programmes. The method automatically identifies speech segments in which voices and background sound are mixed and non-speech segments with background sound only and then controls the suppression of the sound and the emphasis of the voice in each segment [III-2]. The reason we applied two different processes to speech and non-speech segments is that any deterioration in audio quality such as that of music is readily noticed in non-speech segments.

In speech segments, we extract correlated components between the left and right signals of the stereo signal with an adaptive filter and suppress uncorrelated components. As a result, we obtain voice signals that follow our assumption mentioned above [III-3]. In background sound segments, however, the sound signal is processed by gain control alone, which causes the least deterioration in the sound quality of the programme overall. To further improve the background sound suppression effect, we added a method that applies a technique that emphasizes the acoustic characteristics of phonemes by filter processing [III-4].

Evaluation experiments showed that the background sound suppression effect of the novel method corresponded to real suppression of background sound at about 4.5 dB. Additional improvement to the suppression effect is expected at 1 or 2 dB by using voice emphasis performed with suitable parameter settings [III-5].

Even when the objective of the experiments was not identified (blind test), 35% of 20 elderly evaluators, whose ages were over 60 to 80 years old except for one female evaluator whose age was 50s, replied in a questionnaire that the processed programmes "Have become easier to listen to" or "The volume of the background sound is different," which showed the effectiveness of the method. In addition, after experiencing the effects of playing with the prototype device, 90% of the evaluators replied that the programme audio had become easier to listen to. These results indicate that the elderly people highly appreciated the prototype device.

An external view of the prototype device used for the evaluation is shown in Figure III.2. The magnitude of the background sound and the clearness of the voice are adjusted separately with the control inputs of the controller.

More detailed information is shown in [III-5].

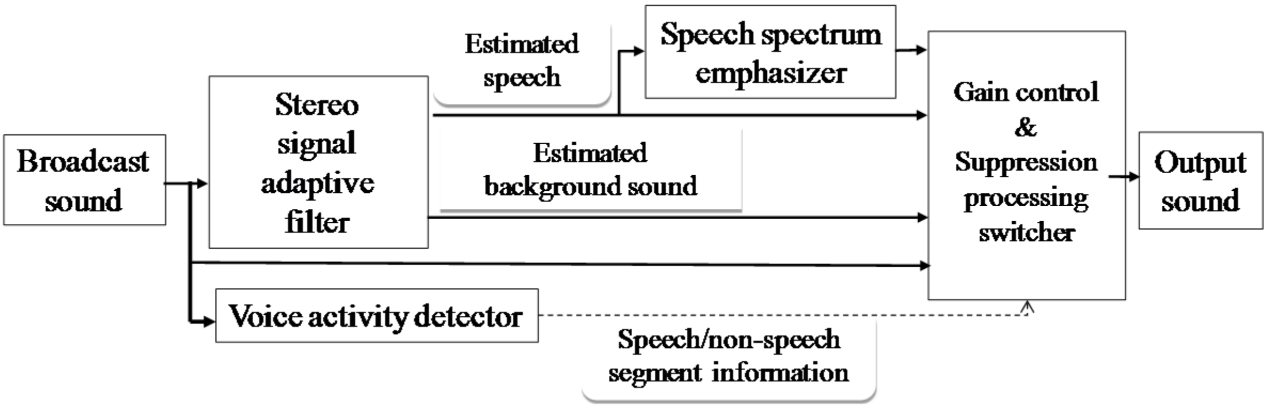
Broadcast sound
Stereo signal adaptive filter
Estimated speech
Speech spectrum emphasizer
Estimated background sound
Gain control & Suppression processing switcher
Output sound
Voice activity detector
Speech/non-speech segment information

Figure III.1-Block diagram of prototype adjustment system

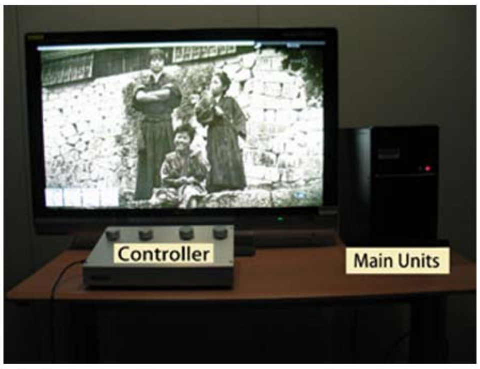

Figure III.2-Prototype device

III.2 References

[III-1] H. Nakamura, M. Sawaguchi, K. Masaoka, K. Watanabe, Y. Yamasaki, E. Miyasaka, M. Yasuoka, H. Seki (2003), *Better Audio Balance Broadcasting service for elderly people.-Back Ground Sound Levels of Television Programs for Easy Listening*, Proc. Spring Meet. Acoust. Soc. Jpn., 1-5-5, pp. 455-456, (in Japanese).

[III-2] T. Komori, A. Imai, N. Seiyama, R. Takou, T. Takagi, Y. Oikawa (2012), *Development of a Broadcast Sound Receiver for Elderly Persons*, Proc. ICCHP 13th, pp. 681-688, 2012 (Springer-Verlag).

[III-3] Y. Murayama, H. Hamada, S. Komiyama, Y. Kawabata (2007), *Adaptive control for advanced re-production of narration voice*, 13th AES Regional Convention, Tokyo, August.

[III-4] R. Takou, N. Seiyama, A. Imai, T. Komori, T. Takagi (2012), *A study on spectrum contrast enhancement for sentence speech intelligibility in noise for elderly*, 2012 Proc. Autumn Meet. Acoust. Soc. Jpn., 2-Q-a8, pp. 531-532, (in Japanese).

[III-5] T. Komori, A. Imai, N. Seiyama, R. Takou, T. Takagi, and Y. Oikawa (2013), *Development of volume balance adjustment device for voices and background sounds within programs, for Elderly people*, AES 135th Convention paper 9010.