

# CLEAN AUDIO FOR IMPROVED SPEECH INTELLIGIBILITY

## CONTENTS

1. DEFINITION	2
2. RATIONALE	2
3. APPROACH AND IMPLEMENTATION - BASIC MODE AND ADVANCED MODE	3
4. RECOMMENDATIONS	3
4.1. REQUIREMENTS TO INPUT SIGNAL	3
4.2. REALISATION ASPECTS	4



## 1. DEFINITION

Definition 1

We define Clean Audio (CA) as follows:

*Enhanced audio signal by means of signal processing, with improved intelligibility of the speech with respect to ambient noise, “atmo”, music etc.*

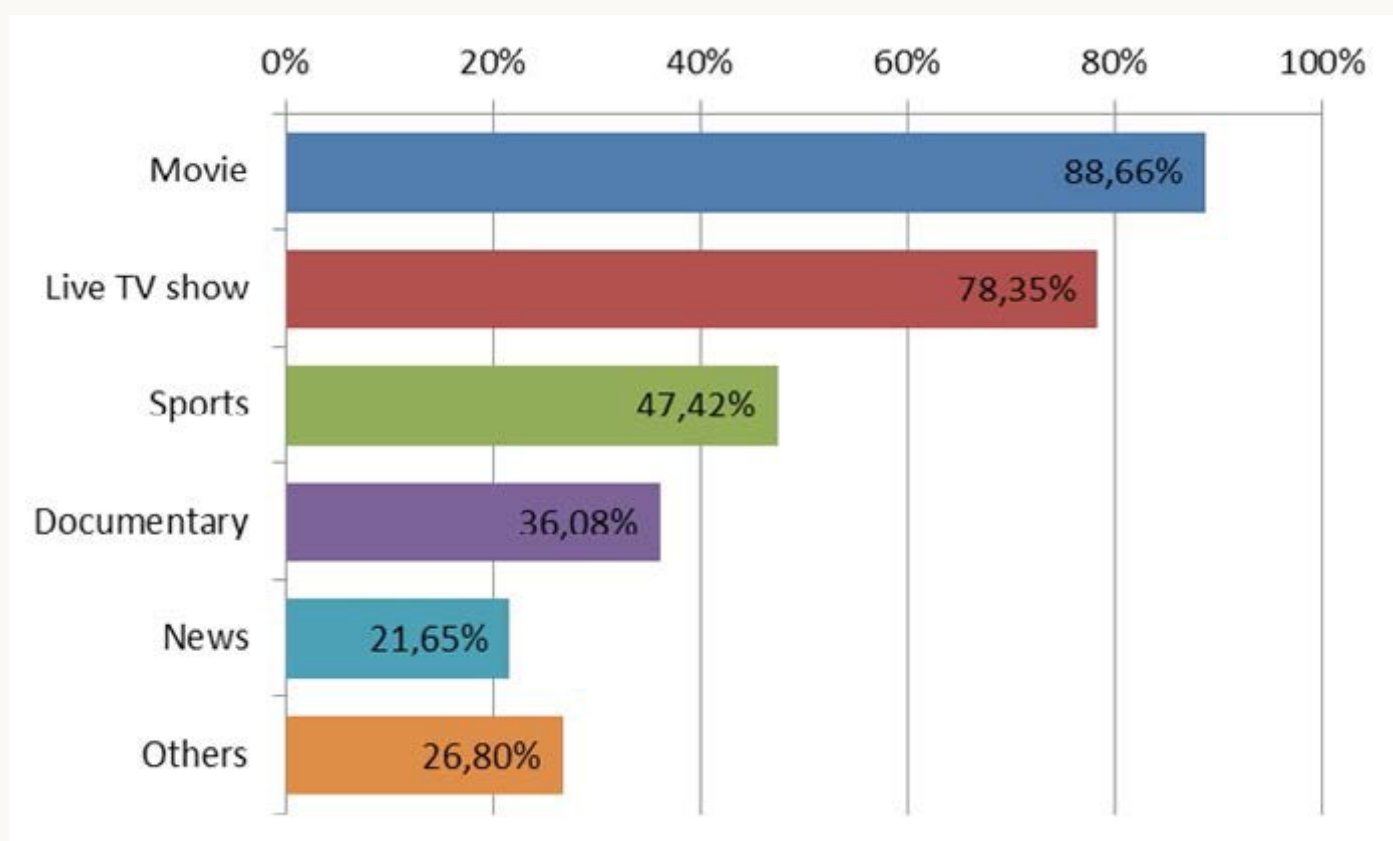
This is NOT a “speech only” audio track: ambient noise, atmo, music and sound effects are still present, but have been processed (reduced) in a way that the speech can be better understood.

## 2. RATIONALE

Rationale 2

There is a clear demand for better speech intelligibility. The usual audio mixes in TV programs in various genres lead to comprehension problems, especially for those people who have hearing impairments (including elderly). An online survey amongst viewers with hearing impairments on speech intelligibility in TV programs, carried out within the HBB4ALL project by partners rbb and IRT, confirms this: almost every type of TV content is affected, but especially the categories movie and live TV show were rated as problematic (see the aside figure, which shows answers to the question “Are there any TV programs with dialogues that you find hard to understand?”).

There are many different reasons why the intelligibility of current TV audio mixes is assessed as insufficient. In the online survey, as most problematic factors of audio signals were identified: background music, background noise, sound effects, grubby articulation and speaking too fast. The former three factors potentially can be corrected by post-processing the produced audio mixes, which is exactly what our Clean Audio implementation tries to do.



[http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1\\_umfrage.pdf](http://www.schwerhoerigen-netz.de/RATGEBER/RUNDFUNK-FERNSEHEN/PDF/erg1_umfrage.pdf)  
(referenced on 28 September 2016, German only)

### 3. APPROACH AND IMPLEMENTATION - BASIC MODE AND ADVANCED MODE

The main requirement for a “Clean Audio generator” is that it should be able to process existing content from TV productions (i.e. it should take 5.1 as well as stereo content at its input); thus, it is suitable for potential integration in existing production workflows. To support as many standard TV receivers as possible, the output should be a stereo signal. Also, it should be able to run automatically. By means of setting certain parameters, the CA output by the generator should be controlled, to allow for adapting it to specific user requirements.

In the HBB4ALL project, IRT implemented the Clean Audio generator in software; it is available as a prototype. It takes a regular produced audio mix (in either stereo or 5.1 format) at its input, separates the speech, lowers the volume of the non-speech components and then creates a new audio mix of the speech and the processed non-speech; this is the basic mode. In an advanced mode of the tool, additional processing is added, to further emphasize the part of the signal in the speech frequency range.

Additional details on the functionality and implementation of the CA generator can be found in HBB4ALL deliverable D4.2, chapter 2.1<sup>1</sup>.

### 4. RECOMMENDATIONS

Recommendations 4

Various listening tests were carried out in the HBB4ALL project to obtain feedback from potential user groups on the effect of the CA generator on the speech intelligibility in TV programs, as well as on the “listening experience” of the modified audio version. The recommendations resulting from this work are presented here.

#### 4.1. REQUIREMENTS TO INPUT SIGNAL

Requirements to input signal 4.1

a) 5.1 content is the ideal input format for the CA generator

The conducted tests clearly show that an improvement in terms of speech intelligibility can be reached best with 5.1 content as input format. For stereo content the CA generator can achieve small improvements; in these cases, the level of enhancement (improvement of the speech intelligibility) highly depends on the original signal in terms of mix and capturing situation. Hence, if 5.1 content is available, it should be used as input for the CA generator.

b) The Center should be clean

If 5.1 content is available, the Center channel should be clean as this leads to the best results. This also applies to stereo content in a way: the speech should be recorded clean, meaning recorded without other noises. Both for 5.1 and stereo content is valid: in case the center part of the audio signal is not clean, the CA processing will enhance the noise as well as the speech parts and an improvement in speech intelligibility cannot be gained.

<sup>1</sup> This document is publicly available from the HBB4ALL website: <http://www.hbb4all.eu/wp-content/uploads/2015/03/D4.2-Pilot-B-Solution-Integration-and-Trials-2015.pdf>

### c) Stereo should be stereo

If stereo content is used as input format, the stereo signal should be really stereo and not e.g. dual-mono. It is also important that the speech signal is mixed in the stereo middle of the stereo mix, otherwise an extraction of the speech parts for the clean audio processing would not be possible. Another essential factor for an effective processing is that the speech signal is recorded without any background noises or other interfering elements (see requirement b).

## 4.2. REALISATION ASPECTS

Realisation aspects 4.2

### a) Differentiate between users wearing a hearing aid and those who do not

The HBB4ALL test results show that it is sensible to distinguish between people wearing a hearing aid and those who do not. Due to the audio processing in hearing aids, the optimal CA parameter settings are different between these user groups. The following three recommendations are based on this distinction.

### b) 5.1 audio material as input for users wearing a hearing aid: the basic mode is advisable

Our results and interviews with test subjects show that when using 5.1 material, people wearing a hearing aid mostly benefit from the basic mode processing (where the volume of the non-speech components is lowered with respect to the speech without additional processing). The listening experience also was rated positively by this user group for this processing mode. For further details, please refer to HBB4ALL deliverable D4.1<sup>2</sup>, chapter 7.1.1.

### c) 5.1 audio material as input for users not wearing a hearing aid: the advanced mode is advisable

Our results show, that for users not wearing a hearing aid, when using 5.1 material, the advanced mode, including additional frequency and dynamic processing got the best ratings. The basic mode as described in case b) also led to good results. For further details, please refer to HBB4ALL deliverable D4.1, chapter 7.1.1.

### d) Stereo audio material as input for users not wearing a hearing aid: the advanced mode is advisable

Our results show, that for users not wearing a hearing aid, when using Stereo material, the advanced mode, including additional amplification for the speech frequency band, was slightly favored<sup>3</sup>. For further details, please refer to HBB4ALL deliverable D4.4, chapter 7.1.2 .

<sup>2</sup> This document is publicly available from the HBB4ALL website: <http://www.hbb4all.eu/wp-content/uploads/2015/03/D4.1-Pilot-B-Progress-Report.pdf>

<sup>3</sup> Note: for Stereo input material, and users wearing a hearing aid, our tests did not yet allow to conclusively derive generic and stable knowledge regarding the effect of the CA processing.

e) Compromise between processing maximum and deformation of the audio signal

The CA settings should not be “pushed to the max”. Rather a compromise in the parameter settings is needed, so that the overall audio is not deformed to an extent which makes it unpleasant to listen to. This is specifically important in case viewers with and without hearing impairments simultaneously listen to a TV program with the Clean Audio track (e.g. in the living room). Additional tests with targeted user groups would be useful to further optimize the parameter settings.

f) Offer several pre-produced CA versions to the end-user

Ideally, to allow adjustment of the speech intelligibility on a personal basis, audio re-mixing would be done in the receiver under end-user control. Current off-the-shelf receiver technology and audio production workflows do not support such a solution. Therefore, as a compromise between full personalization and technical feasibility, it is recommended to offer several pre-produced CA versions with different mixing levels and processing varieties to the user.

In case only a single CA version is provided to the end-user, a compromise is required: just one set of parameters can be used for the CA generator, either in basic or advanced mode, simultaneously trying to support both user groups, with and without hearing aids. As an example we refer to the settings used in the HBB4ALL Clean Audio field trial in Berlin-Brandenburg, see D4.4, section 3.3.2.