Speech Quality Improvement of Commercial Flat Screen TV-Sets

Speech Quality Improvement of Commercial Flat Screen TV-Sets

Arianna Astolfi, Fabrizio Riente, *Member, IEEE*, Louena Shtrepi, Alessio Carullo, Leonardo Scopece and Marco Masoero

***Abstract*—This work deals with the improvement of perceived speech quality of flat screen TV-sets on the various broadcasting platforms of the Italian radio and TV broadcasting company Rai. It is well known that the reduced thickness of flat TVs implied a degradation of the audio quality due to the adoption of miniaturised and cheap loudspeakers compared to the ones in the former Cathode Ray Tube televisions. The research activity gave rise to a Transfer Function (TF), which modifies in real-time the frequency spectrum of the audio signal from the TV station before the transmission to the broadcasting tower. In this way, the final users receive the processed audio signal without the need of additional hardware. A Digital Audio Optimizer dynamically equalizes the sound level, boosting the audio signal towards a flat frequency response without any increase in the loudness levels. Given that the majority of the broadcasted audio signals have speech or singing contents, the TF boosts the speech level in the frequency range that is more important for speech intelligibility, i.e. between 1 kHz and 4 kHz. Subjective evaluations of the proposed TF have been carried out in a laboratory in compliance with the standard ITU-R BS.1116. Three different commercial TV-set models, 30 audio excerpts from video tracks divided into the three genres Speech, Sing and Music, and Sport, and 72 subjects aged between 21 and 53 years, were part of the study. Overall, the perceived improvement in the audio quality compared to the non-processed signal was 25.3% on average among the three TV-set models and the three genres. In order to estimate the perceived improvement directly from the audio signal, regression analyses have been performed, which allow the estimation of subjective outcomes from objective indexes based on intensity features and frequency content of the signal, with standard errors between 10% and 14%.**

***Index Terms*—Speech quality, intelligibility, speech perception, enhanced audio, broadcasting, flat screen TV-sets, subjective tests.**

I. Introduction

THE national radio and TV broadcasting company of Italy, Rai (Radiotelevisione italiana), is committed to enhance the speech quality of their programs, as perceived by its clients. One of the critical aspects of perceived audio quality is due to the generalized diffusion in the market of flat screen TV-sets. While the video quality achieved with this solution has been constantly growing, marking a dramatic improvement with respect to the previous Cathode Ray Tube (CRT) televisions, the reduced thickness of flat TVs has implied a degradation of the audio quality. This is due to the limited space available in a flat TV-set that has brought to the use of miniaturized loudspeakers, which exhibit a poor acoustic response. Particularly, audio quality decreases at low frequencies due to distortions and reduced sound pressure levels [1].

The paper is organized as follows. Section II provides an overview of the related works in the literature. In section III the methodology adopted for the development of the transfer function is described. Moreover, the listening environment and the subjective tests method is presented. The results from the subjective tests and the evaluation of the objective indexes are described in section IV.

F. Riente, A. Carullo are with the Department of Electronics and Telecommunications Engineering, Politecnico di Torino, 10129 Torino, Italy.

A. Astolfi, M. Masoero, L. Shtrepi are with the Department of Energy, Politecnico di Torino, 10129 Torino, Italy.

L. Scopece is with the Direzione Rai Gold, Rai, Via Cavalli 6, 10138 Torino, Italy.

II. Background

The assessment of the perceived quality of a TV program involves both the audio and the video experience, even though the latter has been usually considered to a greater extent. On one side, subjective tests have been carried out in order to investigate relationships among video quality, screen resolution and bit rate [2], and on optimal encoding schemes that maximize viewer-perceived quality [3]. On the other side, the problem of low audio quality in flat TV-sets is a known fact and several methods have been developed in order to boost speech components and enhance speech intelligibility.

In Geiger *et al.* (2015) [4] the solution consists in executing a voice activity detection algorithm with the goal to isolate speech components and then processing speech by a speech enhancement filter.

Rumsey [5] underlined that there is a growing interest in speech intelligibility enhancement for people with hearing impairments when they listen to reproduced sound. This greatly affects the elderly more than the rest of the population, due to the higher incidence of this disability [6], although the modern techniques used for speech detection and enhancement of movie sound have shown to improve the sound experience also for normal hearing listeners.

In Uhle *et al.* (2008) [7] a method for improving sound quality of movie sound that showed to be more effective for hearing impaired than for listeners with normal hearing, is described. Particularly, speech signal is detected through a pattern recognition method and then it is further processed by a spectral weighting with the aim to suppress the background noise. The signal is also attenuated in the speech pauses.

Behrends *et al.* (2007) [1] propose different strategies such as Dynamic Virtual Bass and the adoption of resonant loudspeakers, for improving the overall audio quality at lower frequencies, where small and cheap loudspeakers included in flat TV-sets can determine level drops of 30 dB/octave below

200 Hz.
Fuchs *et al.* (2012) [8] propose a new dialog enhancement technology to personalize the balance between dialogue and ambient sound in TV programs, such the ones with sport and music contents where background noise from crowds or music can mask the dialogues.
Mapp (2016) [9] found that the frequency response of commertial flat TV-sets at typical listening distances shows variations between 10 dB and 20 dB and suggested that this could affect the perceived speech intelligibility. Furthermore, he showed that some settings of the TV-sets could improve the potential intelligibility by attenuating lower mid peaks and accentuating the speech consonant region that is in the range between 1 kHz to 5 kHz.
Niederjohn and Grotelueschen [10] [11] showed that a high-pass filtered speech enhances speech intelligibility in high noise levels over unprocessed speech at the same signal-to-noise ratio. The high-pass filter boosts the power of consonants which, although significantly weaker than the vowels (by as much as about 30-40 dB), convey information that is more significant for intelligibility.
Nowadays, in the case of Ultra High-Definition Television (UHD TV), which are largely diffused around the world, the tendency is to personalize the level ratio between dialog and background by the users. A user-adjustable system has proved to improve the overall quality of experience and increases the user's satisfaction, even if a very high variance between listeners occurs in terms of preferred relative levels of dialog and background [12].
In this work, we focused on improving the speech quality perceived from commercial flat TV-sets. It is well known that the frequency range from $0.5\,\mathrm{kHz}$ to $4\,\mathrm{kHz}$ is the most important for speech intelligibility [13] [14] [15], as it can be observed from Tab. I, which shows the weighting factors for males and females in octave bands that are used to calculate the Speech Transmission Index (*STI*). The *STI* is a metric ranging between

TABLE I: Factors that represent the octave-band contribution to the *STI* in accordance with the IEC 60268-11.

Gender	Octave-band center frequency [Hz]						
	125	250	500	1000	2000	4000	8000
Male	0.085	0.127	0.230	0.233	0.309	0.224	0.173
Female	-	0.117	0.223	0.216	0.328	0.250	0.194

0 and 1 representing the transmission quality of speech with respect to intelligibility by a speech transmission channel [14]. Therefore, we concentrated our effort on the most important frequencies for speech intelligibility.
The research activity described in this paper is aimed at developing a HW/SW system capable of improving the speech quality perceived by Rai clients on the various broadcasting platforms, such as Digital Terrestrial Television (DTT), Satellite Television (SAT) and Internet Protocol (IP). The system consists of two main parts:

- A HW apparatus, named Digital Audio Optimizer (DAO), equipped with a SW implementing the Transfer Function (TF) designed to improve the perceived speech quality of flat TVs without overcoming the loudness level specified by the norm EBU-R128 [16]. The idea is to insert the DAO in the Rai broadcasting chain for selected TV channels.
- A stationary system, based on a Workstation, which provides objective measures from the audio signal for the estimation of the perceived quality improvement through the Subjective Difference Grade (SDG), i.e. the evaluation index of the perceived speech quality variation when the TF is applied. It is expressed as a percentage of improvement or worsening.

The SDG is correlated to a subjective qualitative scale (e.g, very good, good, medium, poor, very poor), based on the comparison of the perceived quality with and without applying the TF to the incoming signal. The goal requested by Rai was to reach an overall improvement of the perceived audio quality of at least 20%, over different genres such as Speech, Music and Sport.

III. Method

In this section, the methodology adopted to improve the speech intelligibility in flat TVs is presented. First, the three TVs were characterized both in the anechoic chamber and in the listening environment named Audio Space Lab (ASL). Afterwards, the experimental material, which consisted of video tracks provided by Rai, were selected and divided in three different genres: Speech (movies, news, TV fictions, documentaries), Music (musical contests) and Sport (sport events and running commentaries). The spectrum of these tracks was analyzed to define a transfer function for the Digital Audio Optimizer. Subsequently, subjective tests were carried out according to the standard ITU-R BS.1116 [17].

A. System Overview

The block scheme of the measurement setup for the characterization of the TV-sets is reported in Fig. 1. It is composed of a workstation (video and audio source), a Digital Audio Optimizer (DAO), the TV and one microphone. The microphone is used to characterize the spectral response of the TV-sets. Two converters were introduced in the processing chain to connect the workstation with the DAO, and the DAO with the TV. The same setup was used in both the anechoic chamber and the ASL measurements. During the subjective tests, the microphone was replaced by the participant. The workstation

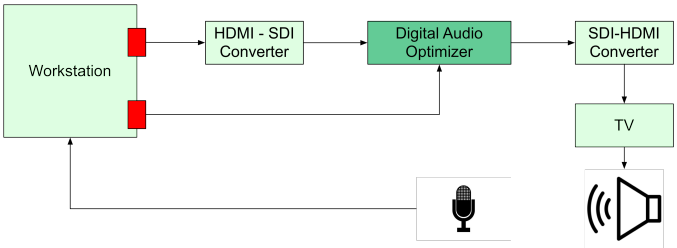


Fig. 1: Measurement setup for characterizing the TVs and running the subjective tests.

reproduces the video over the HDMI connection. It can be considered as the broadcasting signal.
The DAO receives the video on the SDI input connection. It divides the audio from the video as depicted in Fig. 2.

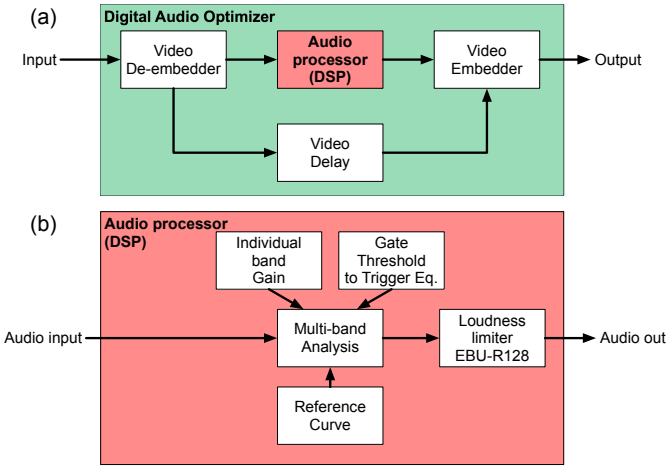


Fig. 2: a) Schematic representation of the Digital Audio Optimizer; b) Schematic representation of the audio processor integrated within the Digital Audio Optimizer.

The audio signal is processed according to the developed transfer function, while the video signal is delayed by a fixed time interval of about $6\,\mathrm{ms}$, which takes into account the audio processing time. The input signal is compared against a reference curve, which represents the transfer function. The signal level in every frequency band determines the activation of that band only if the signal level is above a certain threshold (Fig. 2.b). The processed audio then passes through the limiter block, it is merged with the video signal and routed to the SDI output connector. The output of the DAO is directly sent to the broadcasting station without additional processing. The limiter block is compliant with the EBU-R128 loudness standard [16]. The last converter makes it possible to reproduce the processed signal on the TV.
The DAO can be controlled over the network through an Ethernet connection. It is compatible with the Ember+ control protocol, which makes it possible to send commands and configurations to the DAO. Ember+ is an initiative of the Lawo Group [18], which makes openly available this communication protocol. The described system was adopted for the TV characterizations and the subjective tests.
In summary, the final production environment of the DAO is depicted in Fig. 3. The idea is that the TV broadcasting company processes the audio signal before it is transmitted to the broadcasting tower. In this way, the signal reaches the final user already processed, without any additional hardware.

B. TV Choice and Characterization

The goal is to enhance the speech quality perceived by the users when listening to TV programs. Three commercial TVs produced by three different manufacturers were chosen:

- TV model A: 55” display size, ultra-HD 4K, 3840x2160 pixels. Dolby Digital audio decoder. 2.0 ch loudspeakers with 20W power, facing downward.

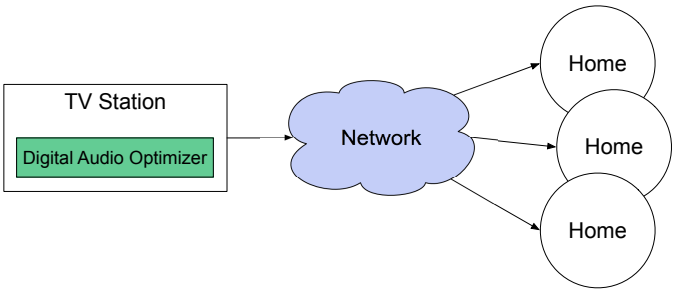


Fig. 3: System applied to TV broadcasting.

- TV model B: 43” display size, ultra-HD 4K, 3840x2160 pixels. Dolby Digital Plus decoder. 2.0 ch loudspeakers with 20W power, facing downward.
- TV model C: 50” display size, full-HD, 1920x1080 pixels. Dolby Digital Plus decoder. 2.0 ch loudspeakers with 10W each, facing downward.

To detect possible failures in the spectrum that can worsen the overall speech intelligibility, the impulse response of each TV-set was computed. The procedure was initially carried out in an anechoic chamber in order to minimize the effects of reflections, and then repeated inside the Audio Space Lab, which is a non-anechoic room, to verify how the TV-sets would perform in a typical indoor environment, similar to the one of a final user.
The impulse response was obtained by means of a convolution technique and using a sweep signal. The sweep had the following characteristics: start frequency $50\,\mathrm{Hz}$, stop frequency $20\,\mathrm{kHz}$, exponential type, and duration equal to $5\,\mathrm{s}$. Three repetitions of such a signal were sent to each TV and its output was recorded using the Schoeps CMC 5U MK6 microphone. The recording was then convolved with the inverse of the original sweep signal, obtaining the impulse response from which the spectrum was computed.
As shown in Fig. 4, TV model A shows a flat frequency

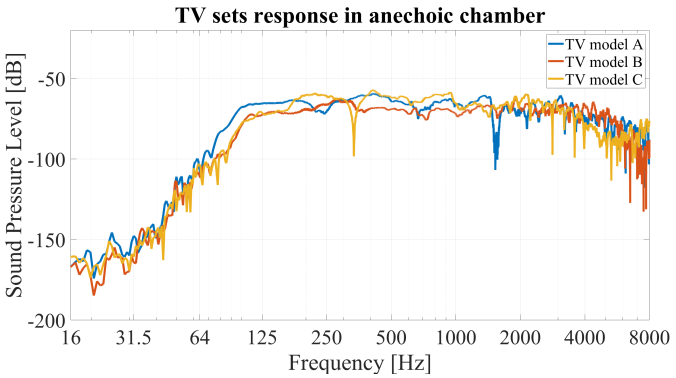


Fig. 4: Impulse response of the three TV-sets in anechoic chamber.

response from $0.1\,\mathrm{kHz}$ up to $3\,\mathrm{kHz}$, with a $40\,\mathrm{dB}$ drop at about $1.5\,\mathrm{kHz}$. Above $3\,\mathrm{kHz}$ the spectrum decreases with a slope of about $-14\,\mathrm{dB}/oct$ rightwards. TV model B shows a flat frequency response over the frequency range $0.1\,\mathrm{kHz}$-$4\,\mathrm{kHz}$, above which the sound pressure decreases with a slope

of about $-35\,\mathrm{dB}/oct$. TV model C presents a 30 dB drop at about 300 Hz and the spectrum was less flat, especially above 2 kHz. From 2 kHz to 5 kHz the sound pressure level decreases with a slope of about $-20\,\mathrm{dB}/oct$, and then increases with a slope of about $16\,\mathrm{dB}/oct$. Very similar, but less flat frequency responses have been obtained in the ASL, as shown in Fig. 5. In the same frequency ranges reported for the anechoic chamber, the sound pressure level decreases with a slope of $-7\,\mathrm{dB}/oct$, $-13\,\mathrm{dB}/oct$ and $-15\,\mathrm{dB}/oct$ and $+5\,\mathrm{dB}/oct$ for the TV-sets A, B and C respectively.

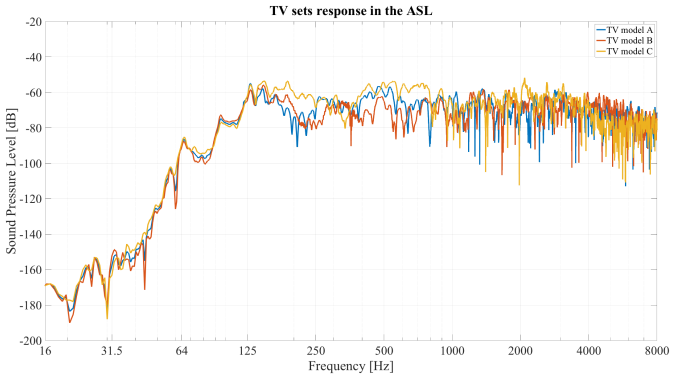


Fig. 5: Impulse response of the three TV-sets in ASL.

C. Audio Space Lab

The Audio Space Lab (ASL) is the room where the subjective and the objective tests were conducted. It is a small, well sound insulated, parallelepipedal room, with volume of 33.3 m^3 (l=5.44 m x w=2.52 m x h=2.43 m). The sound pressure level in the small ASL is affected by the frequency modes up to the Schroeder frequency [19], which is about 134 Hz. Measurements have been carried out in order to detect the most energetic modes in the room. Particularly, a car subwoofer V-6TA 100 W was used as a broadband noise source placed at the room's corners, while a microphone Shoeps CMC 5U MK6, paired to the TASCAM US-144 sound card, was alternatively positioned at the room corners, in the listening position and in other 3 points 10 cm around it. The standard ITU-R BS.1116 [17] specifies that the following dimension ratios should be satisfied to ensure a reasonably uniform distribution of the low-frequency eigentones of the room:

$$1.1 \cdot \frac{w}{h} \leq \frac{l}{h} \leq 4.5 \cdot \frac{w}{h} - 4 \tag{1}$$

where l is the length [m], w the width [m] and h the height [m]. The Eq. 1 becomes:

$$1.1 \cdot \frac{2.5}{2.4} \leq \frac{5.4}{2.4} \leq 4.5 \cdot \frac{2.5}{2.4} - 4 \tag{2}$$

$$1.14 \leq 2.25 \leq 0.68 \tag{3}$$

The criterion is not fully satisfied as shown in equation 3. However, the ITU standard specifies that if the test room cannot completely fulfill these ratios, at least other requirements on the sound field conditions should be fulfilled. Particularly, the background noise, the reverberation time and the listening position requirements have been fulfilled, given the specificity of our test setup, in which the loudspeakers are embedded in the TV-set and the TV-set location has been chosen close to a wall, as in a living room.

Table II shows the natural room resonance frequencies, or normal modes, measured in the ASL in the lowest frequency range, up to the Schroeder frequency. In the lowest frequency range the sound pressure level shows market spatial variations. Results showed the presence of axial modes in the range below 90 Hz that severely affect the measurements inside the room, but which are not in the TV-sets frequency range, that is from ∼95 Hz upward. The location of the listening point was at the center of the room at ∼2 m from the TV-set along l and the position proved to be affected by nodes and anti-nodes. Axial modes are the most energetic and dangerous modes and it was checked that for the one at 127.8 Hz along the x-axis

TABLE II: Normal modes of the Audio Space Lab, average in the listening position. Integers $(\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z})$ identify the mode as axial, tangential, or oblique [20].

Mode	$(\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z})$	**Average in listening position**
Axial	(1,0,0)	32.3 Hz
Axial	(2,0,0)	56.5 Hz
Axial	(0,1,0)	64.6 Hz
Axial	(0,0,1)	72.6 Hz
Tangential	(1,1,0)	–
Tangential	(1,0,1)	–
Tangential	(2,1,0)	91.5 Hz
Axial	(3,0,0)	–
Tangential	(2,0,1)	94.8 Hz
Tangential	(0,1,1)	–
Oblique	(1,1,1)	101.6 Hz
Tangential	(3,1,0)	116.4 Hz
Oblique	(2,1,1)	–
Tangential	(3,0,1)	123.8 Hz
Axial	(4,0,0)	127.8 Hz
Axial	(0,2,0)	–
Oblique	(3,1,1)	–
Tangential	(1,2,0)	138.6 Hz

the listener was not placed in correspondence of either a sound pressure node or anti-node.

In order to complete the acoustical analysis, the reverberation time of the room was measured and the standard ITU-R BS.1116 was used as reference: it states that the average value of the reverberation time T_m measured over the frequency range 200 Hz to 4 kHz should be:

$$T_m = 0.25 \left(\frac{V}{V_0} \right)^{1/3} \tag{4}$$

where V is the volume of the room (m^3) and V_0 is the reference volume of 100 m^3. The tolerances to be applied to T_m over the frequency range 250 Hz to 4 kHz are given in Fig. 6. Results show that the octave-band reverberation time complies with the standard requirements.

The continuous background noise (produced by an air condi-

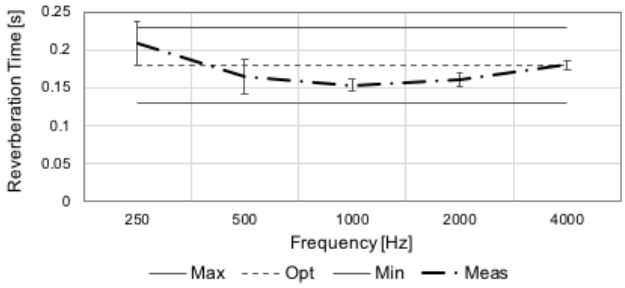


Fig. 6: Reverberation time in the Audio Space Lab and tolerances according to the standard [17].

tioning system, internal equipment or other external sources), measured in the listening position at a height of 1.2 m above the floor should preferably fall between NR 10 and NR 15 and should not be perceptibly impulsive, cyclical or tonal in nature. Fig. 7 shows the measured octave band background noise levels that complies with the requirements of the standard [17] apart from the highest octave bands from 2 kHz to 8 kHz.

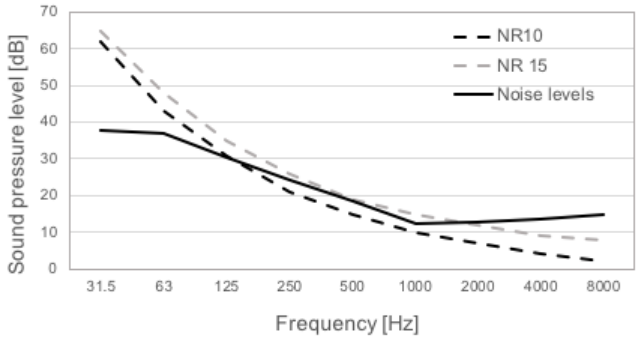


Fig. 7: Background noise level in the Audio Space Lab.

D. Video Tracks Selection

The broadcast material used to test the effectiveness of the method was provided by Rai and refers to various programs distributed between 2017 and 2019. The video were divided into three genres: Speech (movies, news, TV fictions, documentaries), Music (musical contests) and Sport (sport events and running commentaries). Priority must be given to the Speech genre, since the goal is to improve speech quality. However, the proposed TF must not worsen speech quality in the other genres as well. Moreover, norm ITU-R BS.1116 [17] states that there must be a minimum of 5 excerpts per genre and their duration must be in the range 10 s to 25 s.
All things considered, 30 excerpts in total were selected, each 10 s long: 18 for Speech, 5 for Music and 7 for Sport.
In order to find out a common enhancement strategy, a preliminary analysis in the frequency domain was carried out on each genre to identify possible similarities. Using software Adobe Audition 3.0 the spectrum of each excerpt was computed and compared with the others belonging to the same genre; eventually the average spectrum per genre was calculated and graphed in Fig. 8.

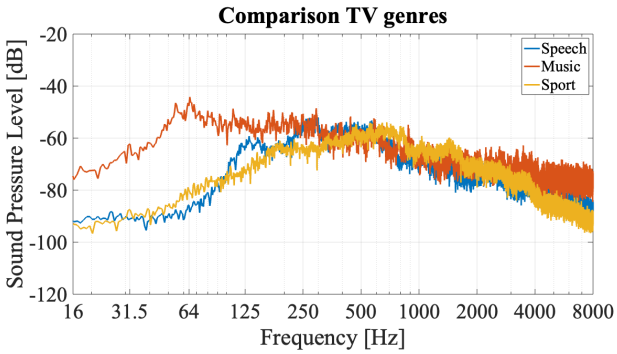


Fig. 8: Average spectra for each genre superimposed.

All the three series present similar behaviour in the frequency bands higher than 600 Hz: this suggests that, if a strategy concerning only this range is adopted, all three genres would benefit and intelligibility would be improved.

E. Transfer Function Definition

The definition of the transfer function started from four considerations: i) the most important frequency range for speech intelligibility is from 0.5 kHz to 4 kHz [13] [14] [15], ii) the human ear is most sensitive in the range from 3 kHz to 4 kHz [21], iii) as shown in Fig. 5 the frequency behaviour of the flat TV-sets decreases in the frequency range higher than 2 kHz, iv) the frequency domain of the audio tracks depicted in Fig. 8 shows a similar behavior in the frequency bands higher than about 600 Hz, with a slope of about $-20\,\mathrm{dB}/oct$.
Starting from these considerations, two transfer functions have been initially proposed, which have been named *Heavy* and *Normal*. It was firstly decided to enhance the energy starting from 1 kHz in order to obtain a flat frequency spectrum in the wide range from 100 Hz to about 4 kHz (see the normalized frequency spectrum in Tab. III). This has been done because of the decrease of the sound pressure level above 2 kHz for each TV-set and above 600 Hz for the three genres. Indeed, a single TF for the three genres has been proposed. Secondly, with the aim to dynamically obtain a flat frequency spectrum, a maximum gain/attenuation to every frequency band was applied, as shown in Tab. III. In particular, for the frequency bands that are most important for speech intelligibility, i.e. from 630 Hz to 4 kHz available in the DAO, a larger max gain was defined. Furthermore, according to the literature, the human ear is most sensitive in the frequency range around 4 kHz. The two proposed TFs differ only in the maximum gain at 4 kHz, where the *Normal* filter gain is 6 dB and the *Heavy* filter gain is 12 dB.
A dynamic equalizer was preferred over a static one to avoid constant amplification of certain frequency bands that are already in line with the expected flat spectrum. A total of 13 programmable frequency bands has been configured. These 13 bands are the one-third octave band center frequencies taken one every two. A maximum gain/attenuation has been set for every band. It was also possible to set a threshold above which the band is triggered. The configuration for every frequency

TABLE III: Digital Audio Optimizer settings for each frequency band of the *Heavy* filter.

Frequency [Hz]	Max gain [dB]	Normalized gain [dB]
100	2	-1.1
160	1.6	-6.7
250	1.5	0.6
400	2.2	-1.4
630	3.7	-0.8
1000	5.8	1.1
1600	6.1	-0.5
2500	6.1	4.8
4000	12.0	2.3
6300	1.1	-0.3
10000	0.7	-4.1

band of the *Heavy* filter is reported in Tab. III.
The metering window reported in Fig. 9 illustrates the effect of the transfer function applied to the input signal in the case of a speech excerpt. The solid line represents the processed signal.

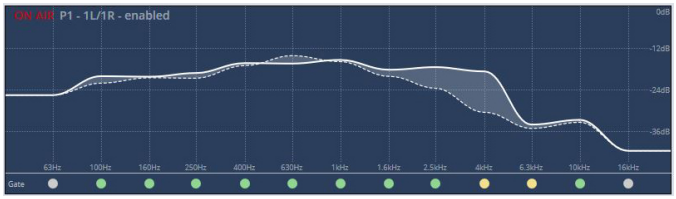


Fig. 9: Metering window, illustrating the difference between the input (dotted line) and the output (solid line) signal in the case of a speech excerpt.

F. Subjective Tests

The subjective evaluations were carried out according to the procedure described in the standard ITU-R BS.1116 [17], which is specifically intended as a guide for the assessment of sound quality where small impairments are considered. It was used as a reference to define the experiment variables, that is, the experimental room conditions, the participant selection, and the audio excerpts. The main examination consisted in providing 30 audio excerpts to each participant and adopting the double-blind triple-stimulus with hidden reference method. This phase was conducted in the ASL. The overall test consisted in three steps:

- hearing test in the anechoic chamber (15 minutes)
- training test in the Audio Space Lab (5 minutes)
- main test in the Audio Space Lab (30 minutes).

A preliminary test was conducted in which 4 volunteers were subjected to the *Normal* filter and 4 to the *Heavy* one. All the volunteers used the same audio material. Results shown in Tab IV proved the latest to be more effective and so the following tests were carried out with the *Heavy* filter only. In the following, the procedure for the subjective tests administration is described.

TABLE IV: Results of the preliminary test. Four volunteers for each transfer function and three TV-sets were considered. The acronym SDG stand for Subjective Difference Grade.

Normal		**Heavy**	
Volunteer	Average SDG (%)	Volunteer	Average SDG (%)
Test_1	22.6	Test_5	11.4
Test_2	7.4	Test_6	35.8
Test_3	29.8	Test_7	19.5
Test_4	16.6	Test_8	34.9

1) Selection of subjects: The total number of subjects that participated on volunteering basis was 72. In particular, 20 participants performed the *Heavy* filter test and 4 were involved in the *Normal* filter test for each TV model (see Tab V).
The participants' age ranged between 21 to 53 years. The test participation was organized using a Doodle survey that has been sent to selected groups of people. Particularly, young students and professors from Politecnico di Torino were chosen. The ITU standard describes the characteristics that each participant should fulfil in order to be considered valid for the subjective test. It is required that each listener is able to perceive relatively small differences of the audio signal to produce a quantitative estimate of the impairments. Moreover, a rejection technique has to be introduced either before (pre-screening) or after (post-screening) the test. Rejection means omitting the whole test of a particular subject. The adopted pre-screening procedure consisted in a hearing test implemented in the anechoic chamber, in order to minimize as much as possible the influence of the environmental noise. As far as the post-screening is concerned, subjects with evaluations strongly different from the mean results or with an evident inability to make correct assessments, were identified as outliers and were removed from the dataset. Indeed, with subjective evaluations outliers are problematic because they may not be originated by the mental process under scrutiny or may not reflect the ability under examination. Thus, in case of subjective measures the measurand is rather uncertain and the outliers removal allows to reduce its variability in order to provide a more robust statistical model [22]. This has been done with the Cook's distance for the outliers' removal [22] [23].

TABLE V: Participants involved during the test.

TV-set	**Filter**	**Number of listeners**
TV model A	Heavy	20
	Normal	4
TV model B	Heavy	20
	Normal	4
TV model C	Heavy	20
	Normal	4
Total		**72**

2) Hearing test in the anechoic chamber: During this phase each participant was provided of an iPad mini with "uHear

Kiosk" App [24][25] and Sennheiser HD 650 headphones. The test is individual and requires the person to be alone inside the anechoic chamber. The testing procedure consists in a series of pure tones at different frequencies alternatively sent to each ear. The duration is approximately 10 minutes and aims at testing the hearing ability in the range 0.5 kHz to 6 kHz. When the user detects a sound, he/she must press the Heard it button in the screen. After the test the app shows the hearing conditions of both ears.

3) Test method: The main test consisted in providing 30 audio excerpts using the double-blind triple-stimulus with hidden reference, described in the standard ITU-R BS.1116 [17]. The name implies that there are three stimuli (A, B, C) in total. The test is individual: one subject at a time is involved and it is required to be alone inside the room in order to avoid any bias coming from the presence of the test conductor. The known reference is always available as stimulus A. The hidden reference and the filtered signal are simultaneously available and randomly assigned to stimuli B and C. The assignation changes for every track. In other words, one of the stimuli, B or C, should be indiscernible from stimulus A. Any perceived difference between the reference and the other stimulus must be interpreted as an impairment. The impairment scale was defined according to the ITU standard, as shown in Tab. VI. The analysis of the tests is based on the Subjective Difference Grade (SDG), which is defined as the difference between the evaluation of the processed signal and the reference as reported in Equation 5.

$$SDG = Evalutation_{\text{signal}_{\text{fdt}}} - Evaluation_{\text{hidden ref}} \quad (5)$$

The possible values range from -3 to $+3$, where $+3$ refers to a $+100\%$ speech improvement, whereas -3 to a -100% worsening. The SDG in percentage is obtained according to Equation 6.

$$SDG = SDG \cdot \frac{100}{3} \quad [\%] \quad (6)$$

4) Training and Main tests in the Audio Space Lab: The training and the main test have been performed in the Audio Space Lab (ASL), where each volunteer was positioned in front of a TV-set, sitting on a chair which was located at a distance of 2 m from the TV itself. The procedure of the test was presented to the subjects in a verbal and written form by the test conductor. The instruction given to the subject was to evaluate the improvement or the worsening of the audio quality mostly referred to the speech clarity.
The two test phases consist of:

- Training phase (15 min): in this phase three excerpts were presented to the tester, one for each genre; they were different from the ones in the main test. The interface was shown by the test conductor and the user was instructed on how to operate it; during the whole training phase the test conductor had to be present inside the room to help the user. In this phase it was possible to freely adjust the TV volume, which was then fixed in order to be constant during the main test.
- Main test (30 min): in this phase the volunteer was left alone in the room and had to evaluate 30 excerpts using a software with an interface identical to the one of the training phase. The duration of the main test was approximately 30 minutes.

TABLE VI: Impairment scale for the "double-blind triple-stimulus with hidden reference test".

Impairment	Grade
Highly Improved	3
Improved	2
Slightly Improved	1
Imperceptible	0
Slightly worse	-1
Worse	-2
Highly worse	-3

Only the audio of the excerpts was presented to the listeners in order to avoid that the visual part could distract the the participants from the main task. Fig. 10 shows a volunteer

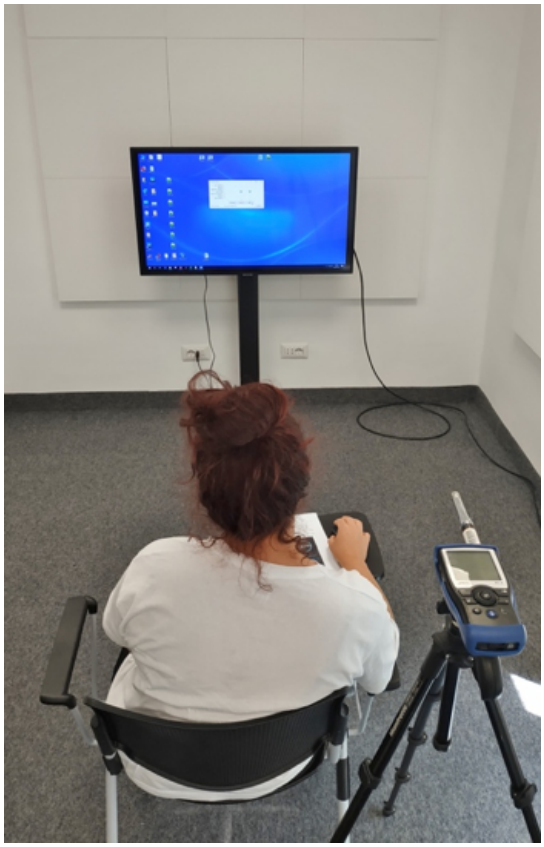


Fig. 10: Training phase of the subjective test inside the ASL.

during the training phase, holding a wireless mouse in her right hand in order to autonomously operate the interface. In order to carry out the test a GUI (Graphic User Interface) was developed in C++. Two similar programs were created, one for each phase of the subjective test inside the ASL. In the first screen, as shown in detail in Fig. 11, the user was asked to insert some personal data that were relevant for the evaluations. Upon entering these data, the main interface was presented to the user (Fig. 12). In the bottom left corner, the number of evaluated tracks is displayed. The 30 tracks

proposed to the user were randomly ordered at every run of the application. Three buttons were present in the main interface: Ref is for reproducing the reference, original, unaltered signal; B and C are the two signals that had to be compared to the other one. Evaluations were done by dragging up or down the cursor related to each signal. To avoid an excessive duration of the evaluation session, the Ref signal could be reproduced up to five times: if exceeded, a pop-up window appeared asking the user to make the evaluation. Once the two signals had been evaluated, the user could click on the Next button and proceed to the next track. After the last track, the test program automatically generated a report file containing the evaluations for the signals B or C, as well as the ID of the track that was referred to as the hidden reference.

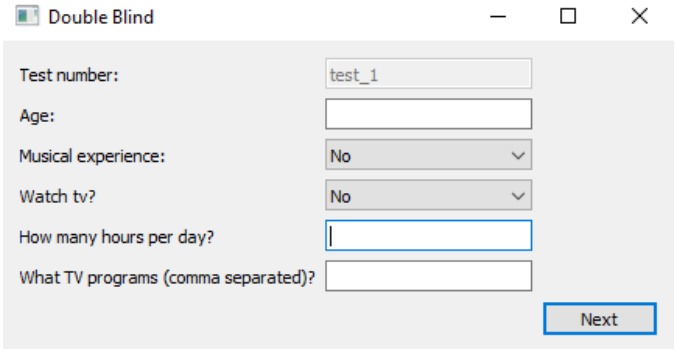


Fig. 11: First screen of the test software.

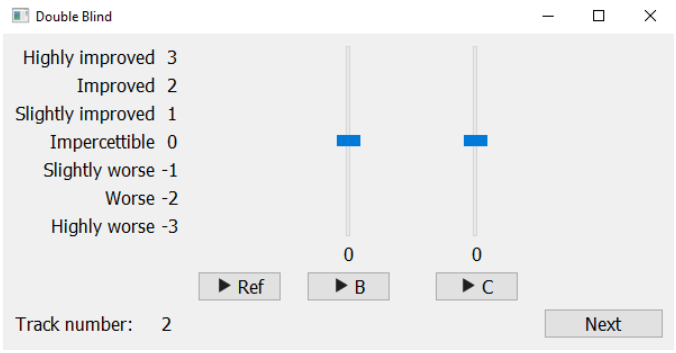


Fig. 12: Main interface of the GUI of the test software.

IV. Results

A. Subjective outcomes

The analysis of variance (ANOVA) showed significant differences in the subjective outcomes between TV models (*p-value* 0.002), genres (*p-value* 0.028) and filter-genre combination (*p-value* 0.028), and therefore the results are firstly shown in the following for each TV model and genre.

Figure 13 shows the mean values of the SDGs (vertical axes) obtained for the TV model A divided per track (horizontal axis). The 20% goal has been highlighted in red. As it can be clearly seen, every track is above the threshold, so every track shows an improvement, and only 4 tracks are below the 20% threshold. The highest standard deviation across the tracks is 12%. Similarly, in Fig. 14, TV model B obtained an improvement in each track and only 8 tracks are below the 20% goal. The highest standard deviation across the tracks is 13%. Differently from the previous ones, results of TV model C in Fig. 15 show that 5 tracks are below the 0% threshold, (speech is worsened in these cases) and the majority of tracks is below the 20% goal. The highest standard deviation across the tracks is 15%. The results showed an average audio quality

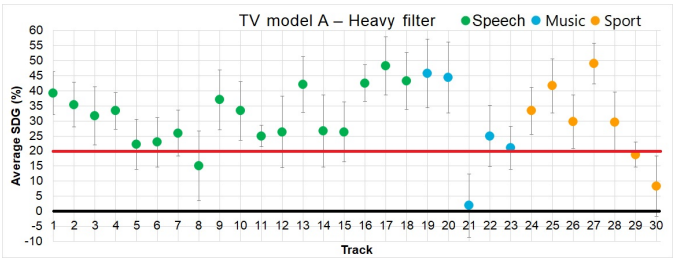


Fig. 13: TV model A per track.

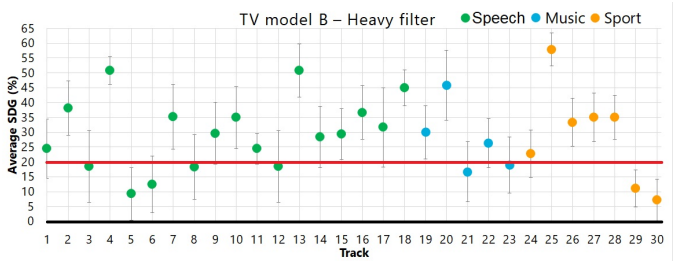


Fig. 14: TV model B per track.

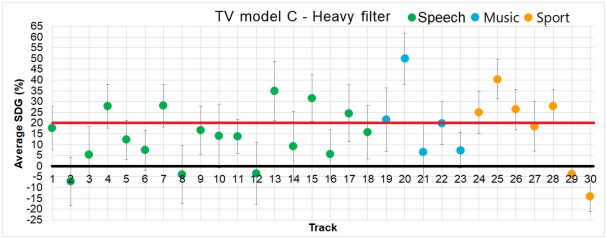


Fig. 15: TV model C per track.

improvement equal to 32% and 30% for TV model A and B respectively. TV model C shows a lower 14% improvement, possibly due to the drop of the sound pressure level around 5 kHz, highlighted in Fig. 5. Results per genre, averaged on the three TV-sets, are reported in the following:

- Speech: 25.5%
- Music: 27.0%
- Sport: 23.6%

All genres showed an improvement above the 20% goal. Overall, the perceived improvement in the audio quality compared to non-processed signal, was 25.3% (standard deviation 14.6%) on average among the three TV-set models and the three genres.

B. Objective index

After the subjective tests, a set of acoustical and psychoacoustic parameters was estimated in order to correlate the results of the subjective test with an objective index. The aim was to obtain the subjective outcomes starting from an objective measure.
The Objective Index (*OI*), similarly to the SDG, ranges from -3 to +3 and is based on the difference between the parameters related to the audio signal recorded at 2 m from the TV-set with the filter (signal post) and without the filter (signal pre). Particularly, the Sharpness, the percentile levels difference $LA_{10} - LA_{90}$, the Booming Index and the Fluctuation Strength have been chosen for these analyses. The Sharpness is a psychoacoustic parameter that refers to the high frequency content of an audio signal, which is important for speech intelligibility; $LA_{10} - LA_{90}$ expresses the amplitude variability of the signal, which is related to the speech-to-noise-ratio; the Booming Index refers to the low frequency content of the signal; the Fluctuation Strength is related to the perceived fluctuation of a modulated signal.

1) Sharpness (S): it is expressed in $acum$, where 1 $acum$ is related to the noise at sharp band centred at 1 kHz with width equal to the critical band [26].

$$S = 0.11 \frac{\int_0^{24Bark} N' g(z) dz}{\int_0^{24Bark} N' dz} \quad [acum] \tag{7}$$

where N is the loudness and $g(z)$ is a weight function, which is equal to 1 up to 16 Bark and then increases exponentially with respect to frequency. The computation of the loudness is based on norm DIN 45692 [27].
The Objective Index for the sharpness (OI_S) is computed as:

$$OI_S = \frac{S_{\text{post}} - S_{\text{pre}}}{S_{\text{pre}}} \cdot 100 \tag{8}$$

where S_{pre} is referred to the signal without filter and S_{post} to the filtered signal.

2) LA10 - LA90: LA90 expresses the sound pressure level exceeded for the 90% of the measurement time. It is used to identify the background noise level of a sound signal.
Similarly, *LA10* expresses the sound pressure level exceeded for the 10% of the measurement time. It is used to identify the highest levels of a sound signal.
The relative Objective Index (OI_L) for the difference *LA10-LA90* is expressed as:

$$OI_L = \frac{(LA_{10} - LA_{90})_{\text{post}} - (LA_{10} - LA_{90})_{\text{pre}}}{(LA_{10} - LA_{90})_{\text{pre}}} \cdot 100 \tag{9}$$

where $(LA_{10} - LA_{90})_{\text{pre}}$ refers to the signal without filter and $(LA_{10} - LA_{90})_{\text{post}}$ to the signal with the filter.

3) Booming Index (BI): the Booming Index is obtained from the power summation of the weighted 1/3-octave band levels of the sound signal and the ratio of loudness relative to the low frequency (0-280 Hz) range, (N_{low}), to the overall loudness, (N_{all}) [28], as reported in Eq.10 [29].

$$BI = 10 \cdot \log \left(\sum_{i=1}^{28} 10^{\frac{(B(Fc_i) - T(Fc_i)) \cdot W(Fc_i)}{10}} \right) \cdot \frac{N_{low}}{N_{all}} \tag{10}$$

Where, Fc_i is the *i*th 1/3-octave band center frequency (i = 128, Fc_1 = 25 Hz, Fc_{28} = 12.5 kHz), $B(Fc_i)$ the band level at Fc_i, $T(Fc_i)$ the threshold level in quiet at Fc_i [26], and $W(Fc_i)$ the weighing function obtained from subjective test at Fc_i, as reported in [30], which reveal that sounds below 200 Hz have a major effect on the booming sensation.
The Objective Index for the Booming Index (OI_B) is computed as:

$$OI_B = \frac{BI_{\text{post}} - BI_{\text{pre}}}{BI_{\text{pre}}} \cdot 100 \tag{11}$$

where BI_{pre} is referred to the signal without filter and BI_{post} to the filtered signal.

4) Fluctuation Strength (F): the Fluctuation Strength is used to describe the degree of perceived fluctuation of a modulated sound signal. It applies to low modulation frequencies up to 20 Hz and considers both the modulation depth and the modulation frequency. Zwicker [26] defined that a 60 dB and 1 kHz tone, 100% amplitude-modulated at 4 Hz, produces a perceived fluctuation strength of 1 *vacil*. It is calculated according to Eq. 12, proposed in [31].

$$F = cal \sum_{n=1}^{75} f_i (k_{i-2} \cdot k_i)^2 \quad [vacil] \tag{12}$$

The value of the calibration factor (cal) is set to produce a fluctuation strength of 1 *vacil* if a 100% amplitude-modulated 1 kHz tone at 60 dB Sound Pressure Level (SPL) is applied to the input with a modulation frequency of 4 Hz. The f_i term is the specific fluctuation strength in the *i*th channel. The 75 channels are obtained through the Equivalent Rectangular Bandwidth (ERB) auditory filter bank [32]. The symbols k_{i-2} and k_i are the cross correlation coefficients between the filtered signals in the *i*th channel in the time domain ($h_{BP,i}(t)$); i.e. $h_{BP,i-2}(t)$ and $h_{BP,i}(t)$ determine k_{i-2}, while $h_{BP,i}(t)$ and $h_{BP,i+2}(t)$ determine k_i.
The Objective Index for the Fluctuation Strength (OI_F) is computed as:

$$OI_F = \frac{F_{\text{post}} - F_{\text{pre}}}{F_{\text{pre}}} \cdot 100 \tag{13}$$

where F_{pre} is referred to the signal without filter and F_{post} to the filtered signal.

C. Correlations between Subjective Evaluations and Objective Indexes

This section shows the results of the correlations between the objective parameters and the average SDG. As a first step, outliers were removed from the objective data. Outliers are those values whose Cooks distance is greater than $\frac{4}{n}$, where n is the total number of observations. Up to 4 outliers were removed for the Speech genre, depending on the regression

between the SDG and the four psychoacoustic parameters. Up to 2 outliers were removed for the Music genre, up to 3 for the Sport genre and up to 8 for all the genres together.
As shown in Tab. VII, the SDG can be estimated from OI_L, OI_S, OI_B and OI_F for each genre with standard error that ranges between 10% and 14%. The best prediction models were obtained with multiple regressions for Speech and with single regression for Music and Sport. The coefficient R^2, which measures the goodness of adaptation of the regression model to the data set, is not higher than 0.5 for all the genres and this could be due to the difference between TV models, highlighted by the ANOVA. For all the genres the Booming Index appears to be the most suitable psychoacoustic parameter for estimating the subjective evaluations.

TABLE VII: Regressions between SDG and Objective Indexes.

Genre	**Sign.**	$\mathbf{R^2}$	**St. err.** (%)	**SDG** (%)
Speech 1	0.001	0.25	10.6	$25.27 - 0.24OI_L - 0.41OI_B$
Speech 2	0.013	0.18	10.7	$32.63 - 0.33OI_L - 0.35OI_S$
Speech 3	0.003	0.16	12.2	$26.05 - 0.4OI_B$
Music 1	0.085	0.36	13.7	$9.24 + 0.3OI_L + 0.45OI_S$
Music 2	0.005	0.52	9.6	$22.26 - 0.72OI_F$
Sport 1	0.022	0.4	10.2	$20.62 + 0.37OI_S - 0.44OI_B$
Sport 2	0.054	0.21	11.3	$17.17 + 0.46OI_S$
Sport 3	0.001	0.45	12.9	$29.85 - 0.67OI_B$
All Genres 1	0.000	0.14	12.6	$26.42 - 0.37OI_B$
All Genres 2	0.001	0.16	12.3	$25.5 - 0.11OI_L - 0.37OI_B$

V. Conclusions

The improvement of the perceived audio quality of flat screen TV-sets has been successfully tested in this work, following the implementation of a Transfer Function which modifies in real-time the frequency spectrum of the audio signal from the Italian radio and TV broadcasting company Rai before the transmission to the broadcasting tower. Particularly, the TF boosts the frequency range that is more important for speech intelligibility, i.e. from $1\,\mathrm{kHz}$ to $4\,\mathrm{kHz}$. Subjective evaluations of the proposed TF have shown the perceived improvement in the audio quality compared to non-processed signal that was 25.3% on average among three commercial TV-set models and three genres, i.e. Speech, Singing and Music and Sport. The regression formula allows the estimation of the subjective outcomes for each genre and for all the genres together from objective indexes based on the acoustical parameters Sharpness, $LA_{10} - LA_{90}$, Booming Index and Fluctuation Strength, with standard errors between 10% and 14%. This error is lower than the standard deviation of the SDG and it is mainly due to the different behaviour of the TV model C, which does not match the results obtained for the models A and B. Nevertheless, the regression formulas show that additional subjective tests could be performed to develop more accurate relationships with the objective indexes.

Acknowledgment

The authors would like to thank Giorgio Fatale from Rai Direzione Qualit e Pianificazione for his contribution in the project conception and development.
Finally, we would like to acknowledge Giacomo Barbero and Viviana Cennamo that have participated in the data acquisition and elaboration.

References

[1] H. Behrends, W. Bradinal, and C. Heinsberger, "Loudspeaker systems for flat television sets," in *Audio Engineering Society Convention 123*, Oct 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14359

[2] G. Cermak, M. Pinson, and S. Wolf, "The relationship among video quality, screen resolution, and bit rate," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 258–262, 2011.

[3] R. Shmueli, O. Hadar, R. Huber, M. Maltz, and M. Huber, "Effects of an encoding scheme on perceived video quality transmitted over lossy internet protocol networks," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 628–640, 2008.

[4] J. T. Geiger, P. Grosche, and Y. L. Parodi, "Dialogue enhancement of stereo sound," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 869–873.

[5] F. Rumsey, "hearing enhancement," *journal of the audio engineering society*, vol. 57, no. 5, pp. 353–359, may 2009.

[6] A. Nakamura, N. Seiyama, A. Imai, T. Takagi, and E. Miyasaka, "A new approach to compensate degeneration of speech intelligibility for elderly listeners-development of a portable real time speech rate conversion system," *IEEE Transactions on Broadcasting*, vol. 42, no. 3, pp. 285–293, 1996.

[7] C. Uhle, O. Hellmuth, and J. Weigel, "speech enhancement of movie sound," *journal of the audio engineering society*, october 2008.

[8] H. Fuchs, S. Tuff, and C. Bustad, *Dialogue Enhancement - Technology and Experiments*, 01 2012.

[9] P. Mapp, "Intelligibility of cinema & tv sound dialogue," in *Audio Engineering Society Convention 141*, Sep 2016. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=18436

[10] R. Niederjohn and J. Grotelueschen, "The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 277–282, 1976.

[11] ——, "Speech intelligibility enhancement in a power generating noise environment," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 4, pp. 378–380, 1978.

[12] M. Torcoli, J. Herre, H. Fuchs, J. Paulus, and C. Uhle, "The adjustment/satisfaction test (a/st) for the evaluation of personalization in broadcast services and its application to dialogue enhancement," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 524–538, 2018.

[13] H. J. Steeneken and T. Houtgast, "Mutual dependence of the octave-band weights in predicting speech intelligibility," *Speech Communication*, vol. 28, no. 2, pp. 109 – 123, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167639399000072

[14] I. C. Secretary, "Sound system equipment - part 16: Objective rating of speech intelligibility by speech transmission index," International Electrotechnical Commission, Geneva, CH, Standard IEC 60268-16, 2020. [Online]. Available: https://webstore.iec.ch/publication/26771

[15] A. S. of America, "Methods for calculation of the speech intelligibility index," American National Standards Institute, New York, NY, Standard ANSI/ASA S3.5-1997, 2017. [Online]. Available: https://webstore.ansi.org/standards/asa/ansiasas31997r2017

[16] E. B. Union, "Loudness normalisation and permitted maximum level of audio signals," https://tech.ebu.ch/docs/r/r128.pdf.

[17] "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," International Telecommunication Union, Geneva, CH, Recommendation ITU-R BS.1116-3, 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

[18] "Lawo group - ember+ protocol," https://github.com/Lawo/ember-plus/wiki, accessed: 2010-09-30.

[19] H. Kuttruff, *Room Acoustics, Fifth Edition*. Taylor & Francis, 2009. [Online]. Available: https://books.google.it/books?id=X4BJ9ImKYOsC

[20] F. Everest and K. Pohlmann, *Master Handbook of Acoustics*. McGraw-Hill Education, 2009. [Online]. Available: https://books.google.it/books?id=6tiJ1cwnwxoC

[21] "Acoustics normal equal-loudness-level contours," International Organization for Standard, Geneva, CH, Standard ISO 226:2003, 2003. [Online]. Available: https://www.iso.org/standard/34222.html

[22] D. Cousineau and S. Chartier, "Outliers detection and treatment: a review." *International Journal of Psychological Research*, vol. 3, no. 1, pp. 58–67, Jun. 2010. [Online]. Available: https://revistas.usb.edu.co/index.php/IJPR/article/view/844

[23] IBM Corp., "Ibm spss statistics for windows." [Online]. Available: https://www.ibm.com/it-it/analytics/spss-statistics-software

[24] J. Szudek, A. Ostevik, P. Dziegielewski, J. Robinson-Anagor, N. Gomaa, W. Hodgetts, and A. Ho, "Can uhear me now? validation of an ipod-based hearing loss screening test," *Journal of otolaryngology - head & neck surgery = Le Journal d'oto-rhino-laryngologie et de chirurgie cervico-faciale*, vol. 41 Suppl 1, pp. S78–84, 04 2012.

[25] J. C. Wang, S. Zupancic, C. Ray, J. Cordero, and J. C. Demke, "Hearing test app useful for initial screening, original research shows," *The Hearing Journal*, vol. 67, no. 10, 2014. [Online]. Available: https://journals.lww.com/thehearingjournal/Fulltext/2014/10000/Hearing_Test_App_Useful_for_Initial_Screening,.8.aspx

[26] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, ser. Springer series in information sciences. Springer Berlin Heidelberg, 2007. [Online]. Available: https://books.google.it/books?id=eGcfn9ddRhcC

[27] "Measurement technique for the simulation of the auditory sensation of sharpness," German Institute for Standardisation, Berlin, Recommendation DIN 45692, 2009. [Online]. Available: https://infostore.saiglobal.com/en-us/standards/din-45692-2009-455777_saig_din_din_1026948

[28] "Procedure for calculating loudness level and loudness," German Institute for Standardisation, Berlin, Standard DIN 45631, 1991-03. [Online]. Available: https://www.din.de/en/getting-involved/standards-committees/nals/publications/wdc-beuth:din21:1666531

[29] T. H. S. Hatano, "Booming index as a measure for evaluating booming sensation," in *2000 29th International Congress and Exhibition on Noise Control Engineering*, Aug 2000, pp. 869–873.

[30] S.-H. Shin, J.-G. Ih, T. Hashimoto, and S. Hatano, "Sound quality evaluation of the booming sensation for passenger cars," *Applied Acoustics*, vol. 70, no. 2, pp. 309–320, 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X08000765

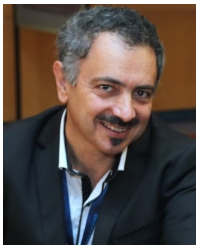
[31] t. zhou, m. zhang, and c. li, "a model for calculating psychoacoustical fluctuation strength," *journal of the audio engineering society*, vol. 63, no. 9, pp. 713–724, september 2015.

[32] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing Research*, vol. 47, no. 1, pp. 103–138, 1990. [Online]. Available: https://www.sciencedirect.com/science/article/pii/037859559090170T

Arianna Astolfi is Associate Professor of Building Physics at the Department of Energy of the Politecnico di Torino, Italy, where she is responsible for the Applied Acoustics Group and Laboratory. She is co-chair of the Technical Committee of Room and Building Acoustics of the European Acoustical Association, member of the National Council of the Italian Acoustic Association, and member of the UK Institute of Acoustics and of the Acoustical Society of America. She regularly organizes Structured Sessions on Room Acoustics for the EAA Conferences and she is usually appointed as Chairman of the Classroom Acoustics session. She serves the UNI committee which is developing technical standards on acoustic requirements for indoor environments such as schools, offices and hospitals. She is member of the editorial board of the journals Acoustics and Building Acoustics, is guest editor of a number of special issues in international journals and author of more than 50 peer-reviewed articles on topics of classroom acoustics, voice monitoring, concert-hall acoustics, soundscape and sound insulation. She has two patents and has created two start-ups incubated in the I3P incubator of the Politecnico di Torino. She has participated in the scientific committees of several conferences in the field of acoustics and building physics and was invited as expert speaker to conferences in the fields of Audiology, Phoniatrics, and Speech Therapy.

F. Riente received his M.Sc. Degree with honors (Magna Cum Laude) in Electronic Engineering in 2012 and the Ph.D. degree in 2016 from the Politecnico di Torino. He was Postdoctoral Research Associate at the Technical University of Munich in 2016. He is currently Postdoctoral Research Associate at the Politecnico di Torino. His primary research interests are device modeling, circuit design for nano-computing, with particular interest on field coupled devices. His interests cover also the development of EDA tool for beyond-CMOS technologies, with the main focus on the physical design and the simulation engine.

Louena Shtrepi is an assistant professor at Politecnico di Torino in the Department of Energy Galileo Ferraris since 2018. She holds a university degree in architecture both from Politecnico di Torino and from Politecnico di Milano. Moreover, as part of her master degree, she obtained the Alta Scuola Politecnica diploma in 2010. She received her PhD degree in 2015 in Metrology: Measuring Science and Techniques, rewarded with the Newman Medal (Newman Student Award Fund and Acoustical Society of America) for excellence in the study of acoustics and its application to architecture. Her research and teaching interests rely on applied acoustics, more specifically in room acoustics and building acoustics. Since 2012, she started working on acoustic materials properties, acoustic simulations and measurement uncertainty. Furthermore, her research aim is to raise awareness about acoustic issues and solutions since the early stages of the design process by involving actively architects and designers. These aspects have been deeply studied in multidisciplinary investigations that involved also subjective perceptual testing. Her research results have been published in highly rated journals and rewarded with several grants at different conferences.



Alessio Carullo was born in Italy in 1966. He received his M.S. degree in Electronic Engineering in 1992 from Politecnico di Torino (Italy) and his Ph.D. degree in Electronic Instrumentation in 1997 from the Universit di Brescia (Italy). He is currently with the Politecnico di Torino, Department of Electronics and Telecommunications, as an Associate Professor of Electrical and Electronic Measurements. He works in the development and characterization of intelligent instrumentation and in the validation of automatic calibration systems.

Leonardo Scopece , born in Foggia (Italy) in 1955. He received his degree in Physics in 1988 from University of Turin. In was employed in RAI - Radiotelevisione Italiana since July 1978. From 1978 to 1979 in RAI Production Centre of Turin as audio engineer inside the Radio Broadcasting department. From 1979 as reasearcher at RAI Centre for Research and Technological Innovation. He have invented and patented a method for shooting and recording audio, processing the signal to obtain virtual repositionable and virtual zoom enabled microphones.

Marco Masoero holds degrees in Civil Engineering from Politecnico di Torino and in Mechanical and Aerospace Engineering from Princeton University. He is professor at Politecnico di Torino in the Department of Energy Galileo Ferraris, which he directed for two mandates (1995-1999 and 2012-2015), and International Faculty Affiliate in the Department of Mechanical and Industrial Engineering of the University of Illinois at Chicago. He teaches graduate courses on the Design of HVAC Systems in the Mechanical Engineering and in the Energy Engineering programs, and on Sound Systems Engineering in the Cinema and Media Engineering program. His present activity mostly deals with Architectural Acoustics, Acoustic Quality of Living and Working Spaces, Noise and Vibration Impact of Transportation Systems. He is the Artistic Director of the concert season Polincontri Classica.