Speech Quality Improvement of TV-Sets for Hearing-Impaired Older Adults

Arianna Astolfi, Fabrizio Riente, *Member, IEEE*, Andrea Albera, Louena Shtrepi, Leonardo Scopece, Roberto Albera, and Marco Masoero

***Abstract*—Flat TV-sets sacrifice space for built-in loudspeakers and this implies a degradation of the speech quality, especially for hearing-impaired older adults. Many elderly people report having difficulty understanding speech on TV, even with hearing aids. In the study, a Transfer Function (TF) that improves the listening experience for the above categories of listeners is validated for Rai, the Italian TV broadcasting company. A Digital Audio Optimizer dynamically equalizes the sound level, in real-time, before the transmission to the broadcasting tower. It implements a TF that amplifies the frequency range between 1 kHz to 4 kHz, which is more important for speech intelligibility, with a particular boosting of the one-third octave band of 4 kHz. Subjective tests of the proposed TF have been carried out in a laboratory in compliance with the standard ITU-R BS.1116-3 with 31 hearing-impaired older adults and a commercial flat-screen TV-set. Results showed a statistically significant slight improvement in the perceived audio quality of 24.3%, over three genres, i.e., "Speech", "Singing and Music" and "Sport". This work follows a previous study on 72 normal-hearing subjects where the same methodology was applied, which involved three commercial flat-screen TV-sets and the same genres. The perceived audio quality was judged improved by the 25.3% on average.**

***Index Terms*—Speech quality, enhanced audio quality, broadcasting, flat screen TV-set, subjective tests, hearing-impaired listeners, older adults.**

I. Introduction

IN A FORMER study [1] a Transfer Function (TF) was applied to the audio signal of the TV station before the transmission to the broadcasting tower, which modifies, in real-time, its frequency spectrum. The TF boosted the audio signal in the frequency range between 1 kHz and 4 kHz. Results showed that across three different TV-set models, the perceived improvement in the audio quality compared to the non-processed signal was 25.3% on average for normal hearing subjects, among three genres: "Speech", "Singing and Music" and "Sport". The tests were based on TV tracks from which audio samples from the three genres were extracted. The study was required by Rai (Radiotelevisione italiana), the public TV broadcasting company of Italy, and the aim was to improve the speech quality of commercial (low-cost) TV-sets that do not implement speech enhancement, neither through sound bars nor by the internal TV settings. This is the case, i.e., of TV-sets owned by people that cannot afford the expense of a sound bar or that are not skilled enough to cope with the audio settings of their own TV-set.

This work deals with the improvement of the speech quality of flat screen TV-sets on the platform of Rai for hearing-impaired older adults. The request by Rai was to improve the listening quality by at least 20%. The study extends the work published in [1], which considered normal-hearing listeners.

The paper is organized as follows. Section II provides the physiological needs of hearing-impaired older adults and the state of the art on devices or systems for improving the listening experience of these subjects. Section III includes the recap of the previous setup. The methodology adopted for the study, which includes the TV-set selection and the subjective tests, is described in Section IV. In Section V the results are provided, while the last section concludes the paper and features future perspectives.

II. Background

Age-related hearing loss (ARHL), or presbycusis, is the most common cause of hearing loss and its prevalence approximately doubles every decade of life from the second through the seventh decade [2]. Indeed, the sensitivity of human hearing is well known to decrease with age and the impairment of hearing develops more rapidly for sound at high frequencies than at low frequencies. ARHL is a complex degenerative disease whose development is multifactorial, involving both intrinsic (e.g., genetic predisposition) and extrinsic factors (e.g., environmental noise exposure) acting on the inner ear throughout life and leading to impairments in cochlear transduction of acoustic signals [3]. Moreover, the magnitude of this effect varies considerably among individuals.

Due to ARHL, at the same time as the worldwide spread of electronic devices and televisions, there is increasing attention towards the quality of the audio broadcast: many subjects, especially the elderly, report having difficulty understanding

Manuscript received 9 August 2022; revised 24 February 2023; accepted 28 February 2023. Date of publication 22 March 2023; date of current version 7 June 2023. *(Corresponding author: Fabrizio Riente.)*

Arianna Astolfi, Louena Shtrepi, and Marco Masoero are with the Department of Energy "Galileo Ferraris," Politecnico di Torino, 10129 Turin, Italy.

Fabrizio Riente is with the Department of Electronics and Telecommunications Engineering, Politecnico di Torino, 10129 Turin, Italy (e-mail: fabrizio.riente@polito.it).

Andrea Albera and Roberto Albera are with the Division of Otorhinolaryngology, Department of Surgical Sciences, University of Turin, 10124 Turin, Italy.

Leonardo Scopece is with the Management of Rai Gold, Rai, 10138 Turin, Italy.

This article has supplementary downloadable material available at https://doi.org/10.1109/TBC.2023.3254150, provided by the authors.

Digital Object Identifier 10.1109/TBC.2023.3254150

television and complain about the volume of the "background noise" compared to the dialogues [4], [5].

A. Physiological Needs of Hearing-Impaired Older Adults

The technique of measuring hearing thresholds is standardized (American National Standards Institute, ANSI) and thresholds are expressed in decibel hearing level (dBHL), whereby zero dBHL at a given frequency is defined as the average threshold in 18-year-old people with history free of otological disease. Although the prevalence of presbycusis is highly variable depending on the pure-tone averaged frequencies and the classification system used [6], approximately half of adults in their seventh decade shows hearing loss that is severe enough to affect communication [7]; similarly, speech intelligibility decreases progressively with age-related hearing loss [8], both in quiet and with high noise levels [9]. The relationship between speech quality and speech intelligibility was studied by Preminger and Van Tasell [10], which proved that speech quality is strictly related to speech intelligibility and when speech intelligibility declines speech quality declines in a similar way. Speech quality measures included intelligibility, pleasantness, loudness, effort and total impression. In the case of very high speech intelligibility, near 100%, speech quality varies across individuals in an unpredictable ways. As far as music listening quality from hearing impaired is concerned, it may also relate to speech intelligibility since music often includes lyrics, but it also relates to instrumental components of songs [11]. Concerning the importance of the frequency distribution for speech intelligibility, Preminger and Van Tasell demonstrated the relationship between intelligibility and sound quality perceived in the elderly with hearing aids, highlighting how the subjects reported maximum pleasantness as the low-frequency band decreased [12], while French and Steinberg showed that decreasing the cutoff frequency of a low-pass filter from 7 kHz to 2.85 kHz decreased the percentage of correctly identified syllables presented in quiet from 98 to 82% [13]. Other authors reported the influence of frequency components above 3 kHz on speech intelligibility and sound quality for people with mild-to-moderate high- frequency sensorineural hearing loss such, as in the ARHL, whose amplification would allow to increase the understanding of speech in background noise [14], [15]. Scollie et al. [16] underlined the importance of high frequency audibility, in particular between 0.5 kHz and 3 kHz, for the prescription of hearing aids fitting. Keidser et al. [17] found that the second generation of prescription procedures from the Australian National Acoustic Laboratories (NAL) for fitting wide range compression instruments for hearing aids, prescribes relatively more gain across low and high frequencies and less gain across mid frequencies than its predecessor of 1999.

B. State of the Art on TV Listening Experience of Hearing-Impaired Older Adults

Many hearing-impaired people report having difficulty understanding speech on TV, even with hearing aids [18]. In fact, the listening difficulties are still encountered by about 40% of the hearing aid users [19]. Listening to media is one of the main listening activities carried out by hearing aid users aged 65+ years old, about 30% of their daily time [20]. Among the compensation strategies there is the increase of the TV or hearing aid volume, the use of closed captioning and a direct TV audio streaming to headphones or to recent hearing aids with optional wireless connection [19]. The use of TV adapters, which are augmentative TV-listening devices to be connected to the television, significantly improve the speech recognition [21]. TV-adapters allow audio information from the television to be sent digitally to the hearing aids via Bluetooth. An improved Signal-to-Noise Ratio (SNR) is obtained with the usage of Frequency Modulation (FM) systems [22]. Anyway, these devices cannot provide enhanced SNR with respect to the broadcast mix. They improve the SNR when there is noise in the listening room, which is not the case in the scenario addressed here (at least in the typical living room of hearing-impaired listeners).

Most televisions have a number of audio settings that can help improving the listening experience. In fact, old adults experiencing high frequency hearing loss may try to compensate with EQ pre-sets that automatically lower the bass and the lower mid range frequencies, while boosting the upper mid range and higher frequencies. Some TVs are outfitted with Bluetooth that send the sound straight to a pair of wireless headphones, which have a speech-enhancement mode that boosts the dialogue while lowering the background noise. There are also special headphones designed to enhance TV sounds for those with hearing loss, which also boosts the frequencies common to dialogue.

Recently, the speech enhancement for hearing-impaired listeners has been improved thanks to a system based on sound source separation and recurrent neural networks (RNN). The separation of speech from background signals and the remixing at a higher signal-to-noise ratio (SNR) brought to a reduction of the listening effort and an increase in the perceived sound quality on real TV-broadcast material [23]. Such approaches are potentially more powerful than a simple equalization because they can actually provide a SNR enhancement. However, still many issues remain uncovered such as the inability to process the signal in real-time, the large amount of training data required and the difficulties in detecting noise in voice-over-voice condition.

The British Broadcasting Corporation (BBC) published a best-practice guidance for program makers with the aim to improve TV speech clarity in the production chain [24]. The document is intended to improve access to BBC contents and services for people with hearing loss. In particular, one of the main point of the guidance is: "Unclear speech, unfamiliar or strong accents, background noise and background music can all affect intelligibility. Audibility can be particularly compromised when more than one of these issues combine" [24].

An important aspect related to speech intelligibility is the loudspeaker position. Shirley found that a center loudspeaker in front of the listener significantly improves speech clarity compared with down-mixed stereo presentation

via two-loudspeakers, for both aided and unaided listeners [5], [25]. The majority of modern flat TV-sets have small built-in loudspeakers mounted either down-firing or rear-firing, i.e., they are directed below or behind the TV-sets. This implicates that, compared to a configuration with the loudspeakers in front of the listener, reverberation is enhanced for listeners seated outside the critical distance, i.e., in the reverberant sound field. In this condition the direct component is reduced and thus speech understanding difficulties from hearing aid listeners increase [26].

III. Previous Setup and Subjective Tests

Subjective surveys have been performed in the Audio Space Lab (ASL) at Politecnico di Torino, which is a small parallelepipedal room (l = 5.44 m × w = 2.52 m × h = 2.43 m) with a volume of 33 m^3. It is well insulated and compliant with the ITU-R BS.1116-3 for listening tests, in which the loudspeakers are embedded in the TVs and the TV has been positioned close to a wall [27]. In particular, the reverberation time, the background noise and the listening position requirements have been satisfied [1]. The ITU-R BS.1116-3 (Methods for the subjective assessment of small impairments in audio systems) recommendation is intended for use in the assessment of systems that introduce small impairments as to be undetectable without rigorous control of the experimental conditions and appropriate statistical analysis [27].

A. Implementation of the Audio Processing System

In a previous work, a methodology has been proposed to improve speech quality in flat commercial TVs [1]. A Digital Audio Optimizer (DAO) dynamically equalizes the sound levels, in real-time, boosting the audio signals towards a flat frequency spectrum without increasing the loudness level. It implements a Transfer Function (TF), named *Heavy* in [1], which is applied to the frequency spectrum of the audio signal from the TV station before the transmission to the broadcasting tower. Fig. 1a, shows the schematic representation of the implemented system, while Fig. 1b shows the schematic representation of the audio processing system within the DAO.

The application of the TF is carried out dynamically by means of a "spectral signature", which represents a reference curve. The "spectral signature" acts as a dynamic multi-band filter that boosts/reduces spectral parts of the signal dynamically [28]. An example of "spectral signature" is shown in Fig. 2. The maximum gain is identified by the size of the white spheres in Fig. 2 and its corresponding value is reported in Tab. I. The spectrum of the input signal is compared to the reference curve and if the signal level in each band is above a certain gate threshold it triggers the equalizer. In fact, to prevent the amplification of noise in a band (especially buzz), a gate threshold is set. If the energy within the band is lower than the threshold, no amplification will take place. Fig. 3 shows the difference between the input and the processed signal in the case of a speech excerpt. A row of colored round circles under the frequency bands indicates if the relative gate is activated or not, yellow or green respectively. The gray circle indicates that the band is switched off. Every single

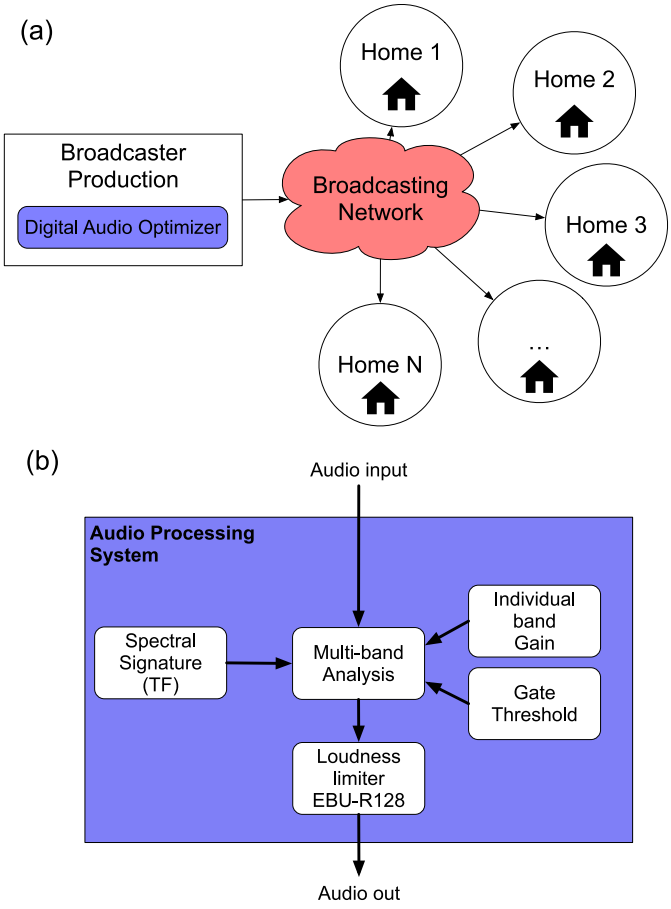

Fig. 1. (a) Schematic representation of the implemented system; (b) Representation of the audio processing system within the Digital Audio Optimizer.

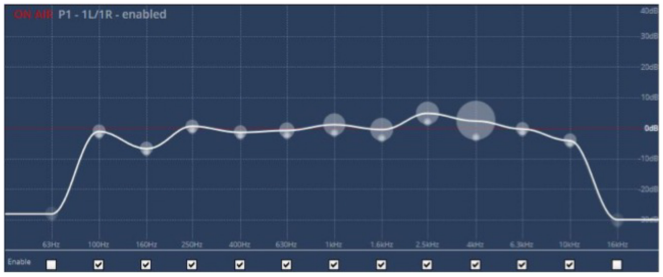

Fig. 2. Frequency spectrum of the filter with the spectral signature represented by the solid white line and the maximum gain (represented by white circles) for each one-third octave band center frequency for the *Heavy* filter [1]. The center frequencies of the one-third octave bands on the x-axis are from 63 Hz to 16 kHz, taken one every two.

band can have an independent gain that limits the amplification/attenuation. While the audio signal is processed, the video stream is delayed by an interval of about 6 ms, which is the time required by the audio processor to equalize the signal. Finally, the processed audio is sent to a loudness limiter block, which is compliant with the EBU-R128 standard [29].

B. Transfer Function for Normal-Hearing Listeners

The aim of the previous work was to enhance the speech quality when listening to TV programs and tested on three

TABLE I
SETTINGS OF THE DIGITAL AUDIO OPTIMIZER FOR EACH ONE-THIRD OCTAVE BAND TAKEN ONE EVERY TWO OF THE *Heavy* FILTER

Frequency [Hz]	Max gain *Heavy* [dB]	Signature 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

commercial TVs, whose frequency responses have been characterized both in the anechoic chamber and the ASL. It has been shown that above 2 kHz the sound pressure level decreases with a slope from −7 dB/oct to −15 dB/oct, in the ASL, for the three TVs. Audio excerpts have been chosen from TV tracks divided in the genres "Speech", "Singing and Music" and "Sport". The tracks consisted in 10s long audio-video excerpts taken from TV programs. All the three genres show a sound pressure level decrease in the frequency bands higher than 600 Hz with a slope of about −20 dB/oct.

The definition of the TF started from the above considerations concerning the decrease of the sound pressure level in the frequency spectrum for both the three TVs and the three genres. Furthermore, the TF has been conceived starting from the consideration that the frequency range which is the most important for speech intelligibility is 0.5 kHz-4 kHz [30], [31], [32] and that the human ear is most sensitive in the range from 3 kHz to 4 kHz [33]. Thus, the TF enhances the energy starting from the one-third octave band 1 kHz to obtain a flat frequency spectrum in the range 100 Hz-4 kHz. Fig. 2 shows the spectral signature of the filter *Heavy* which tends towards a flat spectrum, with the maximum gain for each one-third octave band center frequency that can be applied to the input signal to match the signature. A total of 13 programmable frequency bands are configured, which are the center frequencies of the one-third octave band taken one every two. Table I shows the numerical values for each frequency band of the spectral signature used to obtain a flat spectrum and the maximum gain according to the *Heavy* filter, which boosts the 4 kHz up to 12 dB. Fig. 3 shows the difference between the input signal and the processed audio signal with the transfer function in the case of a speech excerpt.

C. Subjective Tests for Normal-Hearing Listeners

In the previous study, subjective tests were conducted in ASL with the setup shown in Fig. 4. A workstation, which outputs the audio source, is connected to the DAO through an

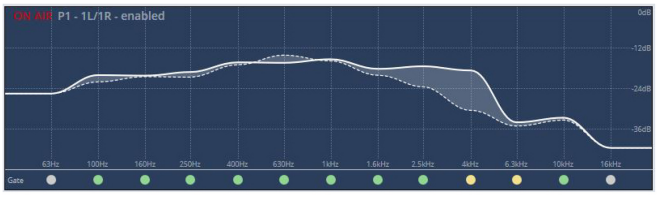

Fig. 3. The window illustrates the difference between the input signal (dashed line) and the processed audio signal with the *Heavy* TF (solid line) in the case of a speech excerpt. The row of colored round circles under the frequency bands indicates if the relative gate is activated or not, in yellow or green respectively. The center frequencies of the one-third octave bands on the x-axis are from 63 Hz to 16 kHz, taken one every two.

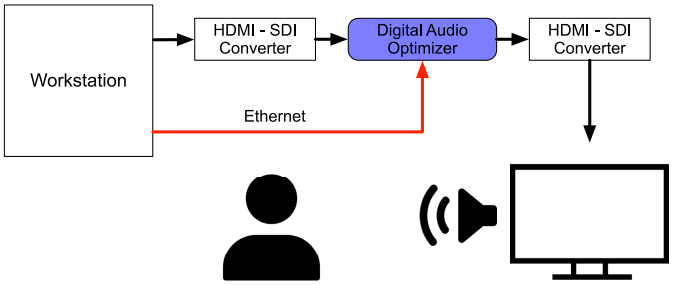

Fig. 4. Setup for running the subjective tests.

HDMI to SDI converter. The audio processor applies the TF and after the HDMI to SDI converter the audio is sent to the TV. The subject, after listening to the audio tracks with and without the TF, performs the subjective test that is handled by the workstation. Indeed, the workstation sends commands to the DAO through an Ethernet connection to activate/deactivate the TF, exploiting the Ember+ control protocol. Ember+ is an initiative of the Lawo Group [34], which makes openly available this communication protocol.

The subjective test was conducted with a total of 72 normal-hearing subjects aged between 21 and 53 years through the "double-blind triple-stimulus with hidden reference" test. It is based on the standard ITU-R BS.1116-3 [27], which is used to evaluate small impairments in audio signals, where the impairment scale was changed from five to seven grades and from single-pole to bipolar. The method implies that there are three stimuli (A, B and C) where the known reference is always A, while the hidden reference and the filtered signal are randomly assigned to B and C depending on the trial. Each subject is asked to assess the impairments on "B" compared to "A", and "C" compared to "A", according to a continuous seven-grade impairment scale as shown in Tab. II. One of the stimuli, "B" or "C", should be imperceptibly different from stimulus "A"; the other one may reveal impairments.

The test was implemented through a dedicated software application, displayed full-screen on the TV, to facilitate the user interaction. The detailed description can be found in [1]. The tests are based on the "Subjective Difference Grade" (SDG), which is the difference between the evaluation of the filtered signal and the reference signal as shown in Equation (1):

$$SDG = Evaluation_{\text{signal}_{\text{fdt}}} - Evaluation_{\text{hidden ref}} \quad (1)$$

TABLE II
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

The SDG range is from -3 to $+3$, where $+3$ corresponds to a $+100\%$ speech improvement, while -3 to a -100% worsening. Equation (2) shows the SDG in percentage:

$$SDG_{\%} = SDG \cdot \frac{100}{3}[\%]. \tag{2}$$

Results showed improvement in the audio quality of 25.3% on average, over the three TVs and the three genres.

IV. CURRENT STUDY METHODOLOGY

A. Transfer Function for Hearing-Impaired Listeners

The aim of this study was to test the *Heavy* filter in listening with ARHL. The filter amplifies the high frequencies with a maximum emphasis at 4 kHz, and several studies show that frequencies above 3 kHz are important for the perception of speech for hearing-impaired listeners, especially when background sounds are present [14], [35], [36]. The TF *Heavy* of the former study can be applied to hearing-impaired listeners based on the following considerations: i) studies of human physiology report a maximum resonance of the external auditory canal (EAC) approximately between 2.5 kHz and 4 kHz [37]; ii) the most important frequency range for speech intelligibility is from 0.5 kHz to 4 kHz [30], [31], [32]; iii) the importance of medium frequencies on hearing is confirmed by Italian legislation, in particular by the National Institute for Insurance against Accidents at Work (INAIL), which applies a biological damage of 25% and 35% in the case of hearing loss at 1 kHz and 2 kHz respectively, while only 5% is attributed when hearing loss hits the 4 kHz frequency [38], even if it is underlined in the literature as the audiometric threshold at 4 kHz and perhaps 6 kHz should be taken into account when assessing noise-induced hearing loss in a medico-legal context [14], [39]; iv) preferences for high-frequency amplification (up to 9 kHz) is also revealed for hearing aid wearers with mild-to-moderate hearing loss [16], [17], [40].

B. TV Selection

Among the three commercial TVs used in the previous work [1] with normal-hearing subjects, the TV chosen for this work was the model A. Its main characteristics are ultra-HD 4K display with 3840x2160 pixels, 55" display size, Dolby Digital audio decoder, 2.0 ch loudspeakers with 20 W power facing downward.

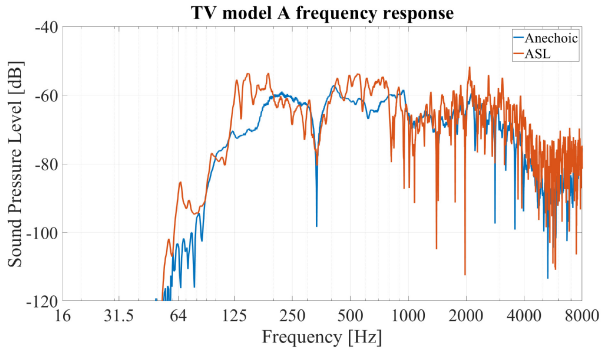

Fig. 5. Frequency response of the model A TV-set in the anechoic chamber and in the ASL.

To select the adopted TV, the obtained results from previous work were taken into account. TV model A was the one that showed the highest average audio quality improvement in the subjective tests. The selected TV was characterized in terms of frequency response. Two measurements were carried out using a convolution technique with an exponential sweep signal from 50 Hz to 20 kHz. The first measure was conducted in an anechoic chamber to analyze the response without the influence of any reflection, the second one was conducted in the ASL to evaluate how the TV would behave in the room for subjective test, which is also a typical indoor listening environment. The two obtained responses are plotted in Fig. 5. In the anechoic chamber the response is flat from 100 Hz up to 3 kHz, with a 40 dB drop at 1.5 kHz. Above 3 kHz the response decays with a rate of -14 dB/oct. In the ASL the response is generally less flat and more jagged, above 3 kHz it decays with a rate of -7 dB/oct.

C. TV Tracks

In this work we used the same TV tracks selected for the previous study with normal-hearing subjects [1]. The tracks consist of 30 short audio-video samples, 10s long, provided by Rai and extracted from programs aired between 2017 and 2019. The tracks are divided into the three genres "Speech", "Singing and Music" and "Sport". Speech includes movies, news, TV fiction. Sport event commentaries fall into the Sports genre. In particular, basket, soccer, cycling and volley. In "Singing and Music" there are some singing excerpts from musical contests, i.e., Sanremo Festival and TV music-show. According to the ITU-R BS.1116-3 [27], the duration of each video should range between 10 s and 25 s and for each genre there are at least 5 tracks. In particular, 18 for "Speech" (12 among fictions and films, 6 news), 5 for "Singing and Music" and 7 for "Sport" event commentaries (2 soccer, 2 cycling, 2 volley, 1 basket). All the selected excerpts contained speech or singing content with very low or nearly absent environmental noise. They are provided in the supplementary material. The frequency content of each track was computed and the average spectrum was extracted for each genre. Their comparison is shown in Fig. 6. The average spectra have a similar trend starting from 600 Hz with a slope of about -10 dB/oct for "Speech" and "Sport" and slightly less steep for

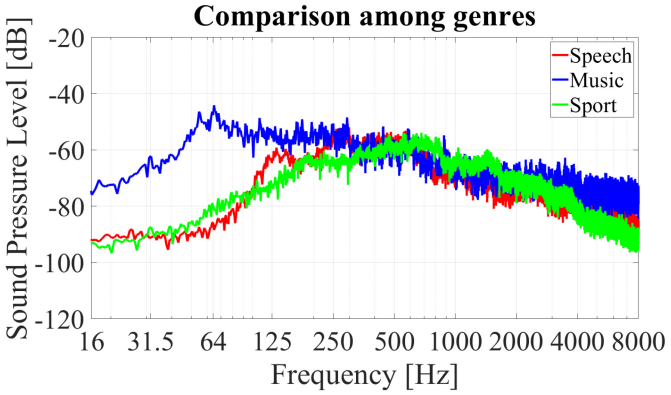

Fig. 6. Overlapping of the average frequency spectra for each genre. The genre "Singing and Music" has been shortened in "Music."

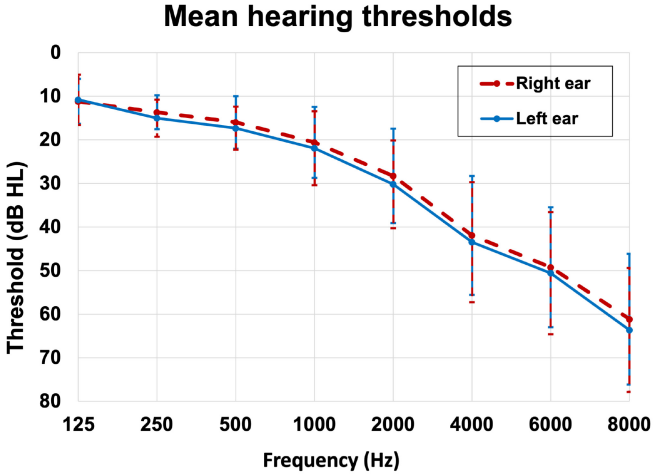

Fig. 7. Mean hearing threshold of the 26 elderly subjects with ARHL.

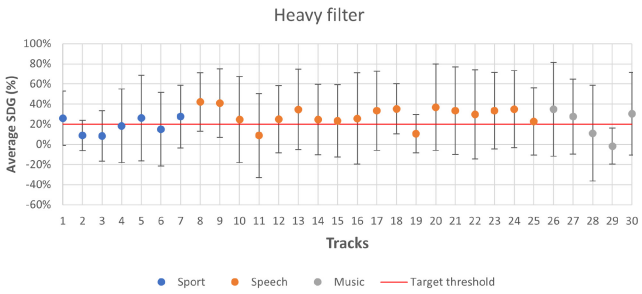

Fig. 8. Average Subjective Difference Grade for each track as a result of the subjective tests with 26 hearing-impaired listeners. The genre Singing and Music has been shortened in Music.

"Singing and Music". This was taken into account in shaping the Transfer Function since all genres could benefit if signal processing to enhance speech intelligibility is applied in this range of frequencies.

D. Selection of the Subjects

Thirty-one elderly listeners between the ages of 62 and 85 years (mean = 71.7 years; standard deviation 5.9 years) participated in the study testing the *Heavy* TF. Twenty-six subjects participated in the main experiment, while 5 subjects were involved in a preliminary pilot test. Pure-tone thresholds were measured using TDH-39 headphones and a clinical audiometer (Triangle, Inventis Srl, Padova, Italy) in a sound-attenuated booth at octave and interoctave frequencies from 250 Hz to 8 kHz. All the subjects demonstrated to have a high-frequency sensorineural hearing loss, although mild in some, consistent with ARHL. On average, listeners had symmetrical mild to moderate hearing loss at high frequencies. Figure 7 shows the mean hearing threshold of the 26 elderly subjects with ARHL. Otoscopy was normal in all subjects, and none reported having had significant ear disease in their medical history nor chronic undue noise exposure or having used potentially ototoxic drugs. Participants with a clinically relevant conductive hearing loss (10 dB or above in the air-bone gap) were excluded.

E. Subjective Tests in ASL

A preliminary test was carried out with 5 subjects aged between 65 and 75 years with ARHL, non involved in the study, to try the test set-up and the instructions for this category of older adults. The 5 subjects showed a compatible hearing loss with the 26 subjects involved in the main test. The preliminary test gave satisfactory results in term of comprehension and usability of the tools, and thus the main test was started. The investigations were conducted in the ASL. We adopted the same experimental setup described in [1]. The position of each participant was at 2 m from the TV sitting on a chair. The subject was presented only the audio part of the excerpt during the test, the only video component was the user interface on the TV upon which the subjects made the subjective evaluations. The instructions for the test were presented by the test conductor. The subjects were asked to rate the speech quality, in terms of speech clarity, of each excerpt evaluating the improvement or deterioration. The complete test was composed by two parts:

- *Training (15 min):* in this part the interface of the test was presented to the subject by the conductor that was inside the room. Three excerpts were presented to each listener. The subject was asked to adjust the TV volume at his/her comfort level, which remained fixed during the main test.
- *Main test (30 min):* the subject was left alone in the room and was asked to listen to the 30 excerpts and evaluate them.

It was decided to use only the audio part of the excerpt during the tests in order to avoid distractions coming from the visual part. The subjective tests were conducted according to the procedure described in Section III. In order to cope with the visual impairment of the older subjects the user-interface was set full-screen on the TV at start up. For each subject a single SDG was computed as the average of the SDG referred to each track, for each genre.

V. Results and Discussions

Figure 8 shows the average SDG for each track subdivided among "Speech", "Singing and Music" and "Sport". It can be observed that the majority of the tracks have an average SDG

higher than 20%, which is the lowest limit requested by Rai in terms of overall improvement of the perceived audio quality, over the different genres. The average SDG for "Speech", "Singing and Music" and "Sport is 29% (std. dev. 9.2%), 21% (std. dev. 15.3%) and 19% (std. dev. 8.2%), respectively. The values and the standard deviations are similar to the ones obtained with normal-hearing listeners, i.e., "Speech" 32% (std. dev. 8.9%), "Singing and Music" 28.6% (std. dev. 18.2%) and "Sport" 30.1% (std. dev. 13.6%)". The high values of the standard deviations are due to the intrinsic variability of the subjective tests [1]. The overall average SDG across the genres is 24.3% (std. dev. 10.9%), that is in compliance with the 20% target. This result is in agreement with the average SDG obtained with normal-hearing subjects on the same TV set in the former study that was 29.9% (std. dev. 11.5%). A possible explanation of the lower averaged SDG for "Sport" and "Singing and Music" compared to "Speech" is the slightly lower SNR in the audio excerpt for these two genres, which intrinsically include noise or music overlapped to speech and sing.

The conversion from a discrete 7-points scale to outputs presented as percentage was carried out to comply with the Rai request of reaching at least 20% of improvement in the perceived audio quality with the application of the TF. The SDG is mapped here onto a percentage scale by setting the highest possible rating ("Highly improved") to +100% and the lowest possible rating ("Highly worse") to −100%. Since the mapping is linear, the range of 200% is separated into 6 equal steps and one step corresponds to 33.3%. In other words, an improvement of 33.3% would correspond to subjects consistently assessing the test signal to be "slightly improved". The data obtained here hardly ever reach this value, except for a few individual tracks, with mean values in the low 20$^{\text{th}}$ies. This suggests that we are looking to a less than a "slight improvement", on average, which is in line with algorithmic approaches that provides SNR enhancement [23]. Anyway, the aim of the study was to improve speech quality from flat TV-sets in speech excerpts with very low noise. This is the reason why algorithms based on SNR enhancement have not been taken into account. Many issues still are not fully settled for audio processing based on SNR enhancement in the broadcasting field. In particular, the source separation and subsequent remixing at a higher SNR relative to the original mix [23] is one of the most promising and recently proposed method for improving the perceived speech quality and reducing the listening effort in TV broadcasting. This method reaches the goal of providing easier speech perception while preserving the original sound atmosphere as much as possible. Results show that the method is able to reduce the listening effort by 2 points out of 13 on the listening effort scale, for common background noise conditions for "Music", "Sport" and "Environment", based on the quality of the estimation of the ambient noise from the speech pauses. In the case of voice-over-voice conditions the method does not work because there is a background voice in the speech pauses and the algorithm does not classify it as background noise. The main issues of this method are the choice of the best architecture that performs the speech separation in real-time with accurate trained models, and which do not require a specific action of the listener.

TABLE III
PERCENTAGE OF INDIVIDUALS THAT EVALUATED POSITIVELY AND NEGATIVELY THE TF FOR EACH GENRE WITH NORMAL HEARING AND HEARING IMPAIRED SUBJECTS, WITH RESPECTIVELY 1932 AND 603 TOTAL EVALUATIONS

		Hearing impaired	Normal hearing
Positive	Speech	70.2%	42.9%
	Sing. & Music	10.1%	11.6%
	Sport	13.1%	15.3%
Negative	Speech	17.1%	17.8%
	Sing. & Music	7.5%	5.0%
	Sport	10.6%	7.4%

As far as the slight improvement of the *Heavy* TF obtained here for the hearing-impaired listeners, which boosts the amplification of the 4 kHz one-third octave band as largely described in Section IV, the range 0.5 kHz to 4 kHz is the most important for speech intelligibility, and it is well known that the human ear is most sensitive in the range from 3 kHz to 4 kHz [33]. These can be the reasons why the *Heavy* filter has reached this improvement, which is comparable to normal-hearing listeners.

A. Statistical Analysis and Comparison With Normal Hearing Subjects

In order to deepen the significance of the results, given only the less than "slight" improvement in the sound quality that we obtained both from the normal hearing and the hearing impaired subjects, we carried out statistical analysis on both the subject categories based on discrete data. Fig. 9 shows the occurrences of the SDG scores from −3 to +3 according to the impairment scale reported in Table II for the three genres "Speech", "Singing and Music" and "Sport" in the case of normal hearing and hearing impaired subjects. It is shown how the most recurrent grade is 1 (Slightly improved) for both the categories of subjects. In Tab. III, we reported in the percentage of individuals that positively (SDG greater than 0) and negatively (SDG lower than +1) evaluated the TF in the current and the former study. The evaluation preferences in the current study are in agreement with the results obtained with normal hearing subjects. The one-sample Wilcoxon signed-rank test [41], [42], right-tailed, is used to determine whether the median of the sample is higher than a theoretical value, which in our case is the SDG equal to "0" on the subjective scale, a value that we specified based on our expectation ("0" means imperceptible difference). Unlike the one-sample t-test, the one-sample Wilcoxon test is a non-parametric test, meaning that it does not require the assumption of normality of the data. The p-value lower than the significance level alpha equal to 0.05 reject the null hypothesis of the $median \leq 0$ and the alternate hypothesis that data comes from a distribution with $median > 0$ is accepted. Table IV shows the p-value for the right-tailed one-sample Wilcoxon signed-rank test related to SDGs for each genre and tracks, for the normal hearing subjects tested in a previous study [1], and the hearing impaired

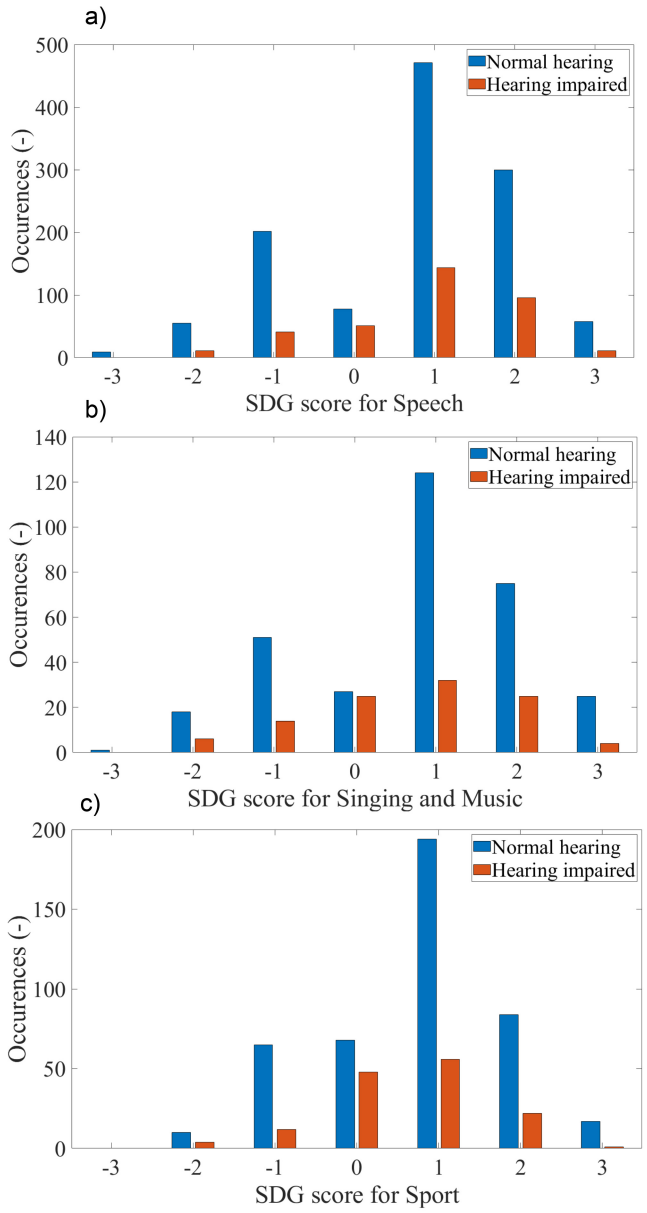

Fig. 9. Occurrences of the SDG scores for: a) Speech; b) Singing and Music; c) Sport in the case of normal hearing and hearing impaired subjects.

subjects tested in the present one. As far as normal hearing subjects are concerned, 29/30 p-values are ≤ 0.05 while in the case of hearing impaired subjects 26/30. The main differences are on the "Singing and Music" and "Sport" tracks. These results mean that subjects have consistently judged the tracks with the implemented transfer function improved compared to the tracks without the transfer function. Most of the medians reported in Table IV are either 1 or 2, which means "Slightly Improved" or "Improved", respectively, according to Table II and as shown in Fig. 9.

VI. Conclusion

In this study, the perceived audio quality improvement of one commercial TV-set has been positively verified after the implementation of a Transfer Function that modifies the audio signal in real-time. This has been verified for subjects with age-related hearing loss. The TF enhances the frequency spectrum of the audio signal from the Italian radio and TV broadcasting company Rai before the transmission to the broadcasting tower. It amplifies the frequency range between 1 kHz to 4 kHz, which is the most important for speech intelligibility, with a particular boosting of the one-third octave band of 4 kHz. Subjective tests improved perceived audio quality that: i) met the 20% improvement set by Rai and ii) was in agreement to a similar study run with normal hearing listeners. The statistical analysis revealed that both the categories of subjects judged "Slightly improved" the perceived audio quality of the tracks with the applied transfer function.

TABLE IV
Right-Tailed One-Sample Wilcoxon Signed-Rank Test p-Values and Median for the Normal-Hearing Subjects (Indicated With Subscript 1) and Hearing-Impaired Listeners (Indicated With Subscript 2)

Genres	# Track	p-value$_1$	p-value$_2$	me$_1$	me$_1$
Speech	1	**0**	**0.007**	1	1
	2	**0**	**0.05**	1	1
	3	**0.001**	**0.014**	1	1
	4	**0**	**0.007**	1	1
	5	**0.001**	**0.007**	1	1
	6	**0.002**	**0.006**	1	1
	7	**0**	**0.001**	1	1
	8	**0.022**	**0.001**	1	1
	9	**0**	**0.003**	1	1
	10	**0**	**0**	1	1
	11	**0**	**0.02**	1	0
	12	**0.033**	**0.17**	1	1
	13	**0**	**0.001**	2	1
	14	**0**	**0.002**	1	1
	15	**0**	**0.01**	1	1
	16	**0**	**0**	1	1
	17	**0**	**0.002**	2	1
	18	**0**	**0**	1	1
Singing & Music	19	**0**	0.174	1	0
	20	**0**	**0.002**	2	2
	21	**0.048**	**0.002**	1	1
	22	**0**	**0.003**	1	1
	23	**0**	0.813	1	0
Sport	24	**0**	**0.03**	1	1
	25	**0**	**0.006**	1	1
	26	**0**	0.06	1	0
	27	**0**	**0.001**	1	1
	28	**0**	**0**	1	1
	29	**0.002**	**0.016**	0	0
	30	0.466	0.227	0	0

Acknowledgment

The authors would like to thank Lorenzo Meli, Caterina Maria Boffa and Gabriele Perrone who have participated in the data acquisition, measuring of hearing function and the data elaboration.

REFERENCES

[1] A. Astolfi, F. Riente, L. Shtrepi, A. Carullo, L. Scopece, and M. Masoero, "Speech quality improvement of commercial flat screen TV-sets," *IEEE Trans. Broadcast.*, vol. 67, no. 3, pp. 685–695, Sep. 2021.

[2] F. R. Lin, R. Thorpe, S. Gordon-Salant, and L. Ferrucci, "Hearing loss prevalence and risk factors among older adults in the United States," *J. Gerontol. A, Biol. Sci. Med. Sci.*, vol. 66A, no. 5, pp. 582–590, 2011.

[3] T. Yamasoba, F. R. Lin, S. Someya, A. Kashio, T. Sakamoto, and K. Kondo, "Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways," *Hearing Res.*, vol. 303, pp. 30–38, Sep. 2013.

[4] H. Fuchs and D. Oetting, "Advanced clean audio solution: Dialogue enhancement," *SMPTE Motion Imag. J.*, vol. 123, no. 5, pp. 23–27, Jul. 2014.

[5] B. Shirley and R. Oldfield, "Clean audio for TV broadcast: An object-based approach for hearing-impaired viewers," *J. Audio Eng. Soc.*, vol. 63, pp. 245–256, Apr. 2015.

[6] A. Rodríguez-Valiente, Ó. Álvarez-Montero, C. Górriz-Gil, and J. R. García-Berrocal, "Prevalence of presbycusis in an otologically normal population," *Acta Otorrinolaringologica Espanola*, vol. 71, no. 3, pp. 175–180, 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/31506162/

[7] Y. Agrawal, E. A. Platz, and J. K. Niparko, "Prevalence of hearing loss and differences by demographic characteristics among U.S. adults: Data from the national health and nutrition examination survey, 1999–2004," *Arch. Internal Med.*, vol. 168, no. 14, pp. 1522–1530, Jul. 2008. [Online]. Available: https://doi.org/10.1001/archinte.168.14.1522

[8] *Acoustics—Statistical Distribution of Hearing Thresholds Related to Age and Gender*, Standard ISO 7029:2017, Int. Org. Stand., Geneva, Switzerland, 2017. [Online]. Available: https://www.iso.org/standard/42916.html

[9] R. Plomp and A. M. Mimpen, "Speech–reception threshold for sentences as a function of age and noise level," *J. Acoust. Soc. Amer.*, vol. 66, no. 5, pp. 1333–1342, 1979. [Online]. Available: https://doi.org/10.1121/1.383554

[10] J. E. Preminger and D. J. Van Tasell, "Quantifying the relation between speech quality and speech intelligibility," *J. Speech Lang. Hearing Res.*, vol. 38, no. 3, pp. 714–725, 1995. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/jshr.3803.714

[11] N. Condit-Schultz and D. Huron, "Catching the lyrics: Intelligibility in twelve song genres," *Music Percept.*, vol. 32, no. 5, pp. 470–483, Jun. 2015. [Online]. Available: https://doi.org/10.1525/mp.2015.32.5.470

[12] J. E. Preminger and D. J. Van Tasell, "Measurement of speech quality as a tool to optimize the fitting of a hearing aid," *J. Speech Hearing Res.*, vol. 38, no. 3, pp. 726–736, 1995. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/7674663/

[13] N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," *J. Acoust. Soc. Amer.*, vol. 19, no. 1, pp. 90–119, 1947. [Online]. Available: https://doi.org/10.1121/1.1916407

[14] B. C. J. Moore, "A review of the perceptual effects of hearing loss for frequencies above 3 kHz," *Int. J. Audiol.*, vol. 55, no. 12, pp. 707–714, 2016. [Online]. Available: https://doi.org/10.1080/14992027.2016.1204565

[15] D. A. Vickers, B. C. J. Moore, and T. Baer, "Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies," *J. Acoust. Soc. Amer.*, vol. 110, no. 2, pp. 1164–1175, 2001.

[16] S. Scollie et al., "The desired sensation level multistage input/output algorithm," *Trends Amplif.*, vol. 9, no. 4, pp. 159–197, 2005. [Online]. Available: https://doi.org/10.1177/108471380500900403

[17] G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer, "The NAL-NL2 prescription procedure," *Audiol. Res.*, vol. 1, no. 1, p. e24, 2011. [Online]. Available: https://www.mdpi.com/2039-4349/1/1/e24

[18] M. S. Sommers, "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," *J. Acoust. Soc. Amer.*, vol. 101, no. 4, pp. 2278–2288, 1997.

[19] O. Strelcyk and G. Singh, "TV listening and hearing aids," *PLoS One*, vol. 13, no. 6, pp. 1–21, Jun. 2018. [Online]. Available: https://doi.org/10.1371/journal.pone.0200083

[20] S. S. Hasan, O. Chipara, Y.-H. Wu, and N. Aksan, "Evaluating auditory contexts and their impacts on hearing aid outcomes with mobile phones," in *Proc. 8th Int. Conf. Pervasive Comput. Technol. Healthc.*, 2014, pp. 126–133. [Online]. Available: https://doi.org/10.4108/icst.pervasivehealth.2014.254952

[21] M. L. Sjolander, M. Bergmann, and L. B. Hansen. "Improving TV listening for hearing aid users." Oct. 2009. [Online]. Available: https://www.hearingreview.com/inside-hearing/research/improving-tv-listening-for-hearing-aid-users

[22] M. S. Lewis, C. C. Crandell, M. Valente, and J. E. Horn, "Speech perception in noise: Directional microphones versus frequency modulation (FM) systems," *J. Amer. Acad. Audiol.*, vol. 15, no. 6, pp. 426–439, 2004.

[23] N. L. Westhausen, R. Huber, H. Baumgartner, R. Sinha, J. Rennies, and B. T. Meyer, "Reduction of subjective listening effort for TV broadcast signals with recurrent neural networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 3541–3550, Nov. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9610991

[24] "Editorial policy guidance: Hearing impaired audiences." BBC. Mar. 2011. [Online]. Available: http://downloads.bbc.co.uk/guidelines/editorialguidelines/pdfs/hearing-impaired.pdf

[25] B. Shirley and L. Ward, "Intelligibility versus comprehension: Understanding quality of accessible next-generation audio broadcast," *Universal Access Inf. Soc.*, vol. 20, no. 4, pp. 691–699, Nov. 2021. [Online]. Available: https://doi.org/10.1007/s10209-020-00741-8

[26] R. W. Harris and M. L. Reitz, "Effects of room reverberation and noise on speech discrimination by the elderly," *Audiology*, vol. 24, no. 5, pp. 319–324, 1985. [Online]. Available: https://www.tandfonline.com/doi/abs/10.3109/00206098509078350

[27] *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, Rec. ITU-R BS.1116-3, Int. Telecommun. Union, Geneva, Switzerland, 2015. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/bs/R-REC-BS.1116-3-201502-I!!PDF-E.pdf

[28] *D*AP4—Digital Audio Processor*, Junger, Berlin, Germany, 2020.

[29] "Loudness normalisation and permitted maximum level of audio signals." European Broadcasting Union. 2020. [Online]. Available: https://tech.ebu.ch/docs/r/r128.pdf

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

[31] *Sound System Equipment—Part 16: Objective Rating of Speech Intelligibility by Speech Transmission Index*, Standard IEC 60268-16, Int. Electrotech. Commission, Geneva, Switzerland, 2020. [Online]. Available: https://webstore.iec.ch/publication/26771

[32] *Methods for Calculation of the Speech Intelligibility Index*, Standard ANSI/ASA S3.5-1997, Amer. Nat. Stand. Inst., New York, NY, USA, 2017. [Online]. Available: https://webstore.ansi.org/standards/asa/ansiasas31997r2017

[33] *Acoustics—Normal Equal-Loudness-Level Contours*, Standard ISO 226:2003, Int. Org. Stand., Geneva, Switzerland, 2003. [Online]. Available: https://www.iso.org/standard/34222.html

[34] "Lawo group—Ember+ protocol." Accessed: Sep. 30, 2010. [Online]. Available: https://github.com/Lawo/ember-plus/wiki

[35] E. Hänsler and G. Schmidt, Eds., *Applications of Adaptive Signal Processing Methods in High-End Hearing Aids*. Berlin, Germany: Springer, 2006, pp. 599–636. [Online]. Available: https://doi.org/10.1007/3-540-33213-8_15

[36] B. W. Y. Hornsby and T. A. Ricketts, "The effects of hearing loss on the contribution of high- and low-frequency speech information to speech understanding. II. Sloping hearing loss," *J. Acoust. Soc. Amer.*, vol. 113, no. 3, pp. 1706–1717, 2003. [Online]. Available: https://doi.org/10.1121/1.1553458

[37] A. P. R. D. Silva, W. Q. Blasca, J. R. P. Lauris, and J. R. M. D. Oliveira, "Correlation between the characteristics of resonance and aging of the external ear," *CoDAS*, vol. 26, no. 2, pp. 112–116, 2014.

[38] "Circolare n. 22 del 7 luglio 1994." INAIL. 1994. [Online]. Available: https://www.inail.it/cs/internet/docs/ci199422.pdf

[39] M. I. Gomez, S.-A. Hwang, L. Sobotova, A. D. Stark, and J. J. May, "A comparison of self-reported hearing loss and audiometry in a cohort of New York farmers," *J. Speech Lang. Hearing Res.*, vol. 44, no. 6, pp. 1201–1208, 2001. [Online]. Available: https://pubs.asha.org/doi/abs/10.1044/1092-4388%282001/093%29

[40] T. A. Ricketts, A. B. Dittberner, and E. E. Johnson, "High-frequency amplification and sound quality in listeners with normal through moderate hearing loss," *J. Speech Lang. Hearing Res.*, vol. 51, no. 1, pp. 160–172, 2008. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18230863/

[41] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945. [Online]. Available: http://www.jstor.org/stable/3001968
[42] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*. New York, NY, USA: Taylor & Francis Group, 2020.

Arianna Astolfi is an Associate Professor of Building Physics with the Department of Energy, Politecnico di Torino, where she is responsible for the Applied Acoustics Laboratory. She has been a Vice-President with the European Acoustical Association since 2022, the Co-Chair of the EAA Technical Committee of Room and Building Acoustics since 2017 and a member of the National Council of the Italian Acoustic Association since 2014. She has been serving the Italian National Unification Body (UNI) since 2016. She is an Associate Editor of *Applied Acoustics* and a member of the editorial board of *Acoustics* and *Building Acoustics*. Her main research interests include classroom acoustics, speech intelligibility, and voice monitoring, but she also works on sound diffusion, acoustical characterization of materials, sound insulation, and soundscape. She is the author of more than 90 peer-reviewed journal papers and hundreds of conference papers; she has three patents and has created two start-ups incubated in the I3P incubator of the Politecnico di Torino.

Fabrizio Riente (Member, IEEE) received the M.Sc. degree *(magna cum laude)* in electronic engineering and the Ph.D. degree from the Politecnico di Torino in 2012 and 2016, respectively. He was a Postdoctoral Research Associate with the Technical University of Munich in 2016. He is currently a Postdoctoral Research Associate with the Politecnico di Torino. His primary research interests are device modeling, circuit design for nano-computing, with particular interest on magnetic QCA. His interests also cover the development of EDA tool for beyond-CMOS technologies, with the main focus on the physical design.

Andrea Albera received the second level master's degree in Otoneurosurgery from the University of Padova. He is currently pursuing the Ph.D. degree in bioengineering and surgical sciences with the Politecnico of Turin. He is an Otorhinolaryngology Physician with the University Hospital "Città della Salute e della Scienza," Turin. He received the University Research Grant from the Department of Neuroscience "Rita Levi Montalcini," University of Turin.

Louena Shtrepi received the university degree in architecture from the Politecnico di Torino and the Politecnico di Milano, the Alta Scuola Politecnica diploma degree, as part of her master degree, in 2010, and the Ph.D. degree in Metrology: Measuring Science and Techniques in 2015. She has been an Assistant Professor with the Department of Energy "Galileo Ferraris," Politecnico di Torino since 2018. 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. She was 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.

Leonardo Scopece was born in Foggia, Italy, in 1955. He received the degree in physics from the University of Turin in 1988. He has been employed with RAI—Radiotelevisione Italiana since July 1978. From 1978 to 1979, he was an Audio Engineer with the Radio Broadcasting Department, RAI Production Centre, Turin. Since 1979, he has been a Researcher with the RAI Centre for Research and Technological Innovation. He has invented and patented a method for shooting and recording audio, processing the signal to obtain virtual repositionable and virtual zoom enabled microphones.

Roberto Albera is a Full Professor of Otorhinolaryngology with the University Hospital "Città della Salute e della Scienza," Turin. He is the Director of the Department of Surgical Sciences, University of Turin. He was a Past President of the Italian Society of Audiology and Phoniatrics.

Marco Masoero received the degree in civil engineering from the Politecnico di Torino and the degree in mechanical and aerospace engineering from Princeton University. He is a Professor with the Department of Energy "Galileo Ferraris," Politecnico di Torino, which he directed for two mandates (1995–1999 and 2012–2015), and an International Faculty Affiliate with the Department of Mechanical and Industrial Engineering, 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."

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement