OPEN ACCESS Freely available online
A. Ragano, E. Benetos and A. Hines, "Automatic Quality Assessment of Digitized and Restored Sound Archives"
J. Audio Eng. Soc., vol. 70, no. 4, pp. 252–270, (2022 April).
DOI: https://doi.org/10.17743/jaes.2022.0002

Automatic Quality Assessment of Digitized and Restored Sound Archives

ALESSANDRO RAGANO,[1,2] **EMMANOUIL BENETOS,**[3,4] **AND ANDREW HINES,**[1,2] *AES Member*
(alessandro.ragano@ucdconnect.ie) (emmanouil.benetos@qmul.ac.uk) (andrew.hines@ucd.ie)

[1]*University College Dublin, School of Computer Science, Ireland*
[2]*Insight Centre for Data Analytics, Ireland*
[3]*Queen Mary University of London, School of Electronic Engineering and Computer Science, UK*
[4]*The Alan Turing Institute, UK*

Archiving digital audio is conducted to preserve and make records accessible. However techniques for assessing the quality of experience (QoE) of sound archives are usually neglected. This paper presents a framework to assess the QoE of sound archives in an automatic fashion. The QoE influence factors, stakeholders, and audio archive degradations are described, and the above concepts are explored through a case study on the NASA Apollo audio archive. Each component of the framework is described in the audio archive life cycle based on digitization, restoration, and consumption. Insights and real-world examples are provided on why digitized and restored audio archives benefit from QoE assessment techniques similar to other multimedia applications, such as video calling and streaming services. The reasons why stakeholders, such as archivists, broadcasters, or public listeners, would benefit from the proposed framework are also provided.

0 INTRODUCTION

On a day-to-day basis, individuals and business organizations create new information, which takes many forms: text documents, sound recordings, videos, and photographs. In order to preserve this material and make it accessible, all state governments and many local governments, universities, businesses, broadcasters, libraries and historical societies maintain archives.[1]

Since the late nineteenth century, audio has been captured using different recording techniques and on different carriers, i.e., the physical media format where the audio signal has been recorded (e.g., wax cylinders, vinyl discs, and digital files). The number of audio items increases day after day, especially thanks to advances in consumer electronics that allow everyone to record audio at any place at any time. Also the recorded material contains varied content, such as speech, music, and general sounds. Consequently, a typical collection held by organizations is heterogeneous because it contains material recorded on different carriers and various source signals.

[1]National Museum of American History, "What are Archives?" http://americanhistory.si.edu/archives/about/what-are-archives.

The British Library keeps a collection of more than 6.5 million recordings of speech, music, wildlife, and the environment held on more than 1.5 million physical items in more than 40 different formats [1, 2], while a more recent survey estimated that United States organizations hold more than 570 million audio recordings, 250 million of which are considered preservation-worthy [3]. Such organizations face several challenges, like deciding which material to preserve, how to preserve it, and how to make it usable [4].

In 1989, most of the organizations agreed to adopt a new way of archiving material: instead of preserving the carrier, which shows clear problems, they opted for safeguarding the content, and the digitization process seemed the proper way for doing this [5]. In 2004, the International Association of Sound and Audiovisual Archives (IASA) reported in the United Nations Educational, Scientific, and Cultural Organization's report Information For All Programme that different reasons pushed the community to use digital technology [6]. Some institutions wanted to provide the collection to a wider audience through the use of the Internet. Other institutions considered digitization as a safe way to provide access to delicate documents. A third group considered digitization as the only way for preserving the in-

Fig. 1. Paper structure summarized. Each circle is a topic related to the audio archive quality of experience (QoE). For each subsection, the name and number are shown in brackets indicating where the subsection can be found.

formation for future generations because some documents cannot last forever.

In some scenarios, after the digitization stage, an audio document could also be subjected to a restoration process. This is required when the content is highly damaged or when organizations decide to provide a different audio quality that better satisfies a certain audience.

Despite the huge effort made for the preservation of sound archives, little has been done toward assessing the audio quality of sound archives subjected to digitization and restoration using an automated approach. In their previous work [7], the authors introduced the concept of the audio archive life cycle based on digitization, restoration, and consumption and argued that in several scenarios of the three life cycle stages, a quality assessment is missing. It was shown how the Quality of Experience (QoE) framework can be adapted to measure audio quality using an approach that is user-centric and automatic. Thanks to the QoE framework, originally designed for multimedia applications, the stakeholders and QoE influence factors [8] have been identified at the three stages of the audio archive life cycle. Future research opportunities were identified that are now covered here. Therefore this paper is an extension of the authors' previous work [7] and aims to:

1) Describe in detail the QoE influence factors that the authors believe should be investigated to predict the QoE of audio archives;
2) Identify the QoE influence factors for each stakeholder;
3) Identify the degradations of the physical carriers and ones caused by the processing stages and how they are perceived;
4) Detect the limitations of the current quality assessment methods used on records that are digitized and restored;
5) Show a case study that discusses the adaptability of the QoE framework in a real-world scenario and how automatic quality assessment can be used.

In particular, real-world examples taken from the NASA Apollo audio archive are shown as a case study that reinforces the idea that using automatic quality assessment on sound archives is feasible and needed by the audio archive community. An overview of the topics covered in this paper is shown in Fig. 1.

1 MOTIVATIONS

Using only technical properties to assess the quality of multimedia applications has shown strong limitations [8]. In many applications, ranging from telecommunications to gaming, user and context factors strongly affect the users' QoE [9–14]. Similarly this paper shows that evaluating the QoE in audio archives requires a user-centric approach, i.e., quality is automatically assessed from the user's perspective instead of only using the system's perspective [12].

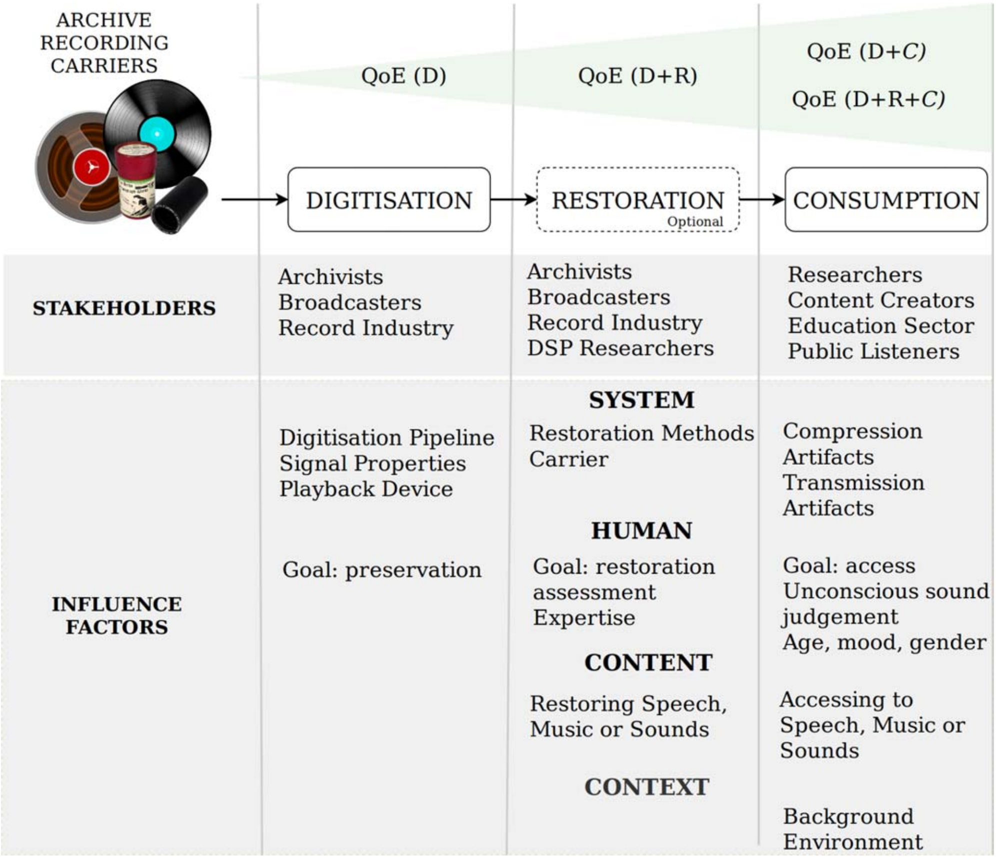


Fig. 2. The audio archive life cycle from the quality of experience (QoE) perspective [7].

The audio archive life cycle proposed in [7] is shown in Fig. 2. The concept of quality changes is argued to be based on the life cycle stage and within the same life cycle depending on the stakeholders. Here some examples are described to motivate why the QoE framework was proposed. These aspects are addressed in more detail in the next sections.

1.1 Digitization

Digitization typically requires that the fidelity of an audio document should be the same as the original copy. For this reason, strict guidelines are applied during the digitization [15, 16]. Quality assessment for the digitization stage is more often called quality control (QC) and it is divided into human-intensive QC and machine-based QC [17]. In this scenario, archivists' goal is to select records that are preservation-worthy and make a digital copy with no errors.

1.2 Restoration

For restoration, similar stakeholders were found (Fig. 2), but they have different quality expectations. Both researchers who need computer-based methods to evaluate audio enhancement techniques and broadcasters and record labels that have to decide how they want to restore their material can be found [18]. Organizations such as broadcasters and record labels often rely on individual judgments without following standards [19], but errors and ethical concerns might arise [20]. Also evaluating large collections with individual judgments is not achievable, considering the high number of records that organizations might restore.

1.3 Consumption

The consumption stage is related to archive accessibility and involves different stakeholders. Here the content quality is one of the most important factors that affects QoE. Successful digitization that adheres to the highest QC standards might not be sufficient for users who explore an audio archive since their QoE might be mostly driven by the perceived quality of the content. Most of the research conducted to improve audio archive accessibility covers related fields such as music information retrieval [21], mobile applications [22], and speech technologies [23], but nothing has been done in terms of the QoE of audio documents.

The examples above show how the concept of quality might have a different meaning based on the context the authors refer to. This means that evaluating the perceived audio quality is not trivial and new user-centric approaches should be taken into account. The proposed life cycle has been designed by adapting the concept of QoE that is defined by the research community as the "degree of delight of the user of a service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users' personality and current state" [8]. QoE differs from other approaches because it is user-centric instead of being system-centric.

Evaluating QoE can be done both using listening tests and quality metrics, but user and context perspectives are also taken into account. Mapping a QoE framework outside of telecommunication services has been done before in fields such as quality of life [24], multiplayer online role-playing-games [14], and soundbar technology [25]. As reported in [8], the definition of QoE was explicitly thought to be applied in new domains. Also QoE embraces several application domains compared to other concepts such as quality of service, which is instead designed to be used only in telecommunication services [12]. To the authors' knowledge, no such work in the past has addressed audio archives from a QoE perspective.

1.4 Cultural Heritage

This study is also motivated by the increasing number of ideas and projects proposed in the archive field, which involve the usage of automatic approaches. New digital techniques have a positive impact on both preservation and access to cultural heritage. Sound archives are of fundamental importance for the cultural heritage because they allow the past to be heard. Not all the records can be regarded as heritage. Records that are generally considered as having historical value constitute heritage [26].

Assessing the QoE can be crucial to the cultural heritage and historical fidelity of sound archives as the media must pass through different processing stages: digitization, restoration (optional), and consumption. Inappropriate modifications of sound archives could invalidate the belonging of some recordings to the cultural heritage, similar to the debates regarding colorization of black and white photographs [27]. In particular some organizations have been exploring the usage of machine learning to uncover cultural heritage, make use of archive material, and help archive preservation. Broadcasters proposed to use machine learning to discover material in digital collections, e.g., through the usage of automatic speech recognition in audio and automatic subject detection in video.[2] Libraries have been using the vast amount of text documents in archives to discover new methods for analyzing historical sources using machine learning.[3] More machine learning–based approaches have been used for improving the archival processes, enriching metadata, handling sensitive information, extracting automatic context, and proposing new collection methods to archive documents in the future [28].

2 INFLUENCE FACTORS IN AUDIO ARCHIVES

This section defines the sound archive QoE *Influence Factors* (IFs) [8], which are the key points that characterize the QoE formation process. An IF is defined as "any characteristic of a user, system, service, application, or context whose actual state or setting may have an influence on the Quality of Experience for the user" [8]. IFs may interrelate and are divided into three classes:

- *System IFs* refer to "properties and characteristics that determine the technically produced quality of an application or service. They are related to media capture, coding, transmission, storage, rendering, and reproduction/display, as well as to the communication of information itself from content production to user" [8].
- A *Human IF* is "any variant or invariant property or characteristic of a human user. The characteristic can describe the demographic and socio-economic background, the physical and mental constitution, or the user's emotional state" [8].
- *Context IF* are "factors that embrace any situational property to describe the user's environment in terms of physical, temporal, social, economic, task, and technical characteristics" [8].

In Table 1, IFs that are involved in establishing the QoE related to audio archives are described. It was decided to include a fourth IF, i.e., *Content IF* because user expectations and their QoE are related to the content of the archived material as discussed below.

2.1 Content

Content is considered an IF in the same way as context, human, and system IFs. The content is divided into two types of IFs. The first one is the audio intended by the creator, e.g., a speaker or an artist. The second type of content concerns audio degradations.

When developing a QoE metric (i.e., a machine-based metric suited to evaluate sound quality without human intervention) or a subjective study to measure audio quality in audio archives, the content has to be taken into account since the intended audio could be a mix of speech, music, and sounds. Current objective quality metrics are not designed to work for various applications and various source signals [29–33]. However sound archives show diverse content across the recordings or in the same recording. For example, radio programs include both speech and music in the same audio file [34], which makes a quality evaluation more complicated.

Subjective studies, such as listening tests, might show other issues since it is not common to find speech and music mixed in the same session. It is reasonable to think that possible biases might be introduced when mixing different content in a listening test session, although this aspect has not been investigated yet in the state of the art. Depending on the content, some factors might arise. For example, participants might be influenced by speech intelligibility when asked to judge speech quality or by music preference when judging the sound quality of a musical piece [35]. It is not known how these issues might affect the participants' responses when mixed content is used.

Content is also characterized by the degradations that affect the intended content. Several degradations can be found mainly because of the different formats in which sound was recorded. More generally the degradations can be seen as the result of all possible artifacts faced during the audio

[2]https://www.bbc.co.uk/rd/blog/2018-10-artificial-intelligence-archive-television-bbc4.

[3]https://www.bl.uk/projects/living-with-machines#.

Table 1. Quality of experience influence factors in audio archives.

IF	Type	Example
Content	Source	The intended content
	Degradation	Global, local
System	Signal properties	Frequency bandwidth, dynamic range, number of channels, encoding
	Device	Recording, playback, listening equipment, resonance and quality of devices
	Carrier	Grooved, magnetic, optical, carrier age
	Restoration	Analog, computer-based, human-based
	Digitization	Guidelines, independent
	Network	Web-based application, transmission
	Media	Compression
Human	High level	Stakeholder needs and goals, skills, expertise in using devices and carriers prior experience, music genre preferences, knowledge, expectations, attention, personality, interest, motivation, emotions, purpose of use
	Low level	Gender, age, mood, mental states
Context	Economic	Cost of digitization and restoration
	Physical environment	Room acoustic response, environment background noise
	Social situation	Group or alone

archive life cycle, i.e., recording situation, recording device, carrier, playback device, potential mistakes, and digitization. When information on the degradations is given, this can be used as a feature for creating a quality metric. Also the degradations are important for preparing the development/test data of the quality metric, whether the data are sampled from real sound archives or synthetically generated.

2.2 System

Audio recording and storage technology have been constantly changing over the years, making system IFs a crucial component for determining the final quality. An audio item might be subjected to many processing stages, especially if the document is used in the consumption stage. System IFs that could be investigated and employed to estimate QoE are reported in Table 1. The authors believe that the QoE of some system IFs might be more obvious, e.g., it is known that more recent carriers do not show some disturbances that were typically found in the earliest carriers, and so the latter might produce a higher QoE. Other factors are less obvious and are poorly investigated, e.g., audio restoration approaches have not been assessed in terms of various archive content and are not always evaluated through listening tests [19, 36].

A critical evaluation of the existing methods might help the usage of automatic restoration techniques that are still not employed. If automatic restoration is applied to large collections, a QoE objective metric might be very useful such that an organization can understand on which records the restoration is failing. Automatic restoration might help to restore very large archives. It must be noted that automatic restoration might generate more degradations than a manual approach. This is because automatic restoration procedures would need a degradation recognition [37] step in some cases, which might fail. When the recognition system fails, more distortions could be generated. For example, Brandt et al. [38] discuss the audio quality when click restoration is applied to regions where clicks are absent and propose an improved version of a click detector. Devices and carriers are also of fundamental importance for the QoE because they introduce non-linear distortions [39, 40].

2.3 Human

Human IFs are defined as "any information that the perceiver brings to a situation" [41]. Human factors are usually divided into high-level (behavioral) and low-level (physiological) factors where the former are considered as one of the most difficult IFs to measure [42]. The stakeholder's goal, here identified as a high-level influence factor, widely affects the QoE formation process. For instance, archivists want audio fidelity [43], broadcasters are interested in improving dialogue intelligibility from old movies, etc. Another human IF is the prior experience given that organizations are usually characterized by expert people who have a preconceived idea of the sound quality. Expertise in using devices and carriers is a crucial component when it comes to handling carriers such as lacquer discs and tapes [44] and must be taken into account. The purpose of use, which might be hard to estimate, is included. When the purpose of the use is not known, information on the stakeholders could be used. The section below describes how each stakeholder might be driven by a different purpose of use.

2.4 Context

Context IFs are divided into economic, environmental, and social situations. Economic factors include the cost of digitization and restoration of large archives. Budget constraints might introduce some limitations for improving QoE. Physical environment and social context are studied in other QoE estimation studies [10, 25], and similar approaches could be adapted to the audio archive domain to study the influence of the context on the audio QoE in this domain.

3 STAKEHOLDER INFLUENCE FACTORS

The scope of this section is to define the IFs of stakeholders, i.e., the factors that influence their QoE formation process. The audio archive IFs were introduced in the previous section; now the IFs are mapped to each stakeholder. Most of the audio material in the form of archives has been accumulated by three main creative sectors: the record industry, broadcasters, and academic and cultural bodies (e.g., archivists) [5]. Instead stakeholders who access audio archives include researchers, general public listeners, content creators, and the education sector. This paper discusses IFs related to the following stakeholders: archivists, content creators, public listeners, and researchers. It must be noted that here it is not discussed who or which organization is going to provide the service produced thanks to the adapted QoE framework. Instead the focus is on the users who might need the service so that the QoE IFs can be identified regardless of who is the service provider.

3.1 Archivists

Archivists are mostly interested in QC that could be seen conceptually closer to the quality of service, i.e., more emphasis is given to the system IFs rather than human or context IFs. However, beyond the preservation of sound archives, archivists are also interested in creating the access copy to make the archive usable. The preservation copy must strictly contain the content of the original copy, while the access copy could be restored and/or compressed to make it usable. Therefore the digitization of the old recordings is a crucial system IF. Especially, archivists want to make sure that no alteration has been introduced after digitizing the signal, i.e., its authenticity must be guaranteed [5, 16].

QC is divided into human-intensive and machine-based techniques on which software checks the presence of mistakes. The two approaches are commonly employed together since the current technology representing the machine-based approach is not able to fully solve all of the digitization problems [17]. According to Casey [17], the following situations are not handled correctly by the current machine-based approaches: 1) field recordings on audiotapes show a change in speed and reversed audio, and 2) some metadata errors occur. The above-mentioned issues are all solved by human-intensive approaches.

For these reasons, most of the system IFs are included as shown in Table 2. The digitization mistakes could be evaluated at either the digitization or consumption stage since it is known that mistakes, such as selecting the wrong speed, would have strong implications on the timbre of the instruments or voice and the color of the performance [45]. However the authors argue that it would be much better to intervene at the digitization stage instead of the consumption stage (see Fig. 2). This is because preventing these mistakes is what the archivists want and typically an organization does not want to digitize the same copy several times. For example, a tape speed detector that works on the copy digitized with the wrong speed might help archivists to prevent the above-mentioned mistakes and speed up the digitization, which is an important contribution given that analog carriers degrade over time. As a consequence, removing these mistakes from the digitization could improve the QoE of the stakeholders belonging to the restoration or consumption stages as shown in Fig. 2, given that other stages depend on the digitization.

The goal of the access copy, instead, is to make the material usable. Therefore it can be restored for improving the listening experience, compressed for saving network bandwidth, and since the material is provided via a web application, transmission artifacts are also involved. Human IFs include the goals of these organizations: on the one hand, archivists are interested in preserving the material, but on the other hand, they are concerned about how to make it usable. In both cases, archivists will judge sound using a critical analysis approach that was included as a prior experience factor.

Finally, the cost of digitization affects the QoE. Large resources are needed to carry out a digitization process. These organizations face the problem of choosing what audio documents have to be digitized because of budget and time constraints.

3.2 Content Creators

The term content creators refers to individuals who use archive material as the source input of their work. This includes stakeholders who adapt or modify content, such as composers, musicians, sound artists, video makers, and artistic producers. It also includes stakeholders who use archive audio to create content without modifying it, such as journalists and broadcasters. Content creators' QoE formation process is primarily affected by the content, and this makes content IFs one crucial influencing factor. In this paper's scenario, the authors do not consider the content as explained in [41], where it is part of the system IFs. Given that an audio archive contains different content, it is considered an inherent characteristic that can or cannot satisfy the QoE of an individual.

Content creators will probably choose to use an audio document even if it shows low audio quality (e.g., perceivable loud background noise). However, if they are given similar content tracks, they will be interested in finding the one that has the best audio quality. This scenario might occur in classical music archives where several versions of the same composition are very common in the same audio archive.[4]

Content creators might benefit from knowing the audio quality of audio records since digital audio archives are vast and exhibit many low-quality recordings. The authors believe that this group will judge sound using technical skills and prior experience, and they make a final judgment using a critical approach. Therefore human IFs are included in their QoE formation process. Generally it is not obvious that a specific degradation is unwanted, but it depends on the user needs. Indeed carriers (system IFs) and degradations (content IFs) play a key role given that on one hand,

[4] https://sounds.bl.uk/Information/About/.

Table 2. Influence factors of the audio archive stakeholders.

	Archivists	Content Creators	Public Listeners	Researchers
Content		Source Degradations	Source Degradations	Source Degradations
System	Digitization Guidelines Recording Devices Playback Devices Carriers, Restoration	Signal Properties Carriers Network Compression	Recording Devices Playback Devices Restoration Recording Devices	Digitization Guidelines Playback Devices Carriers
Human	Goal Prior Experience	Prior Experience Skills, Interest Purpose of Use	Sound Preconception Goal, Purpose of Use Motivation, Interest, Age	Motivation Purpose of Use Skills
Context	Cost	Physical Environment	Physical Environment Social situation	Cost

they may be inappropriate for their artistic product, while on the other hand, they might be preferred for emulating a sound of the past. Content creators can also be influenced by the environment where they work, and therefore room acoustic response and background noise are considered as possible factors that may introduce bias in the QoE formation process. These factors are classified as a physical context IF.

3.3 Public Listeners

Accessing audio archives is a service that concerns public listeners and business organizations. This group desires to access audio archives for different reasons: curiosity, looking for ancestors' witnesses, creation of amateur content, and entertainment. It can be considered that public listeners do not have the same skills as professionals, and therefore a conscious judgment is hard to be identified.

Content IFs are the most important QoE influence factors. Public listeners would use audio archives similarly to content creators, i.e., they are looking for content. Unlike content creators, quality could be measured using a lower discrimination power since they are considered non-experts. Similarly to content creators, public listeners might want to search for the best quality recordings in an archive. The average public listener has a preconception of sound (human IF), which can be considered as unconscious and depends on several factors concerning their life lived. Their goal is also a crucial human IF: if they are looking for ancestors' witnesses, they could be less interested in enhanced quality, even though they could prefer a restored version if it is available. On the other hand, when their goal is about entertainment, the enhanced signal is desired. Familiarity with the degradations also matters since public listeners are not experts. It has been shown that listeners' abilities to learn the details of an unfamiliar sound change after periods of training [46].

Age is also meaningful when comes to creating a sound preconception, given that technology has been changing over the years. Age not only matters from a physiologic perspective [47] but also from an experience perspective, i.e., the distinction between a digital and analog native. It must be noted that the unconscious preconception can be modified if an external factor, e.g., from an expert person, is introduced and make them reflect. An external factor can also be found in content IF, i.e., degradations. A certain degradation could make them change their sound judgment. Therefore, from a technical perspective, restoration is as meaningful as transmission and compression since they access the material mainly using web applications. Generally the authors believe that in this category, the comfort of listening supersedes the original sound of the analog carriers; however it depends on the goal. Among the system IFs, the listening equipment is also included where a wide quality range that goes from low-quality loudspeakers (e.g., smartphones) to a high fidelity audio system (e.g., audiophiles and home audio enthusiasts) is expected.

3.4 Researchers

Researchers from both the private and public sectors need to access audio archives for their own research goals. Examples of researchers include speech scientists, wildlife biologists, linguists, historians, computer scientists in the computational audio sector, musicologists, and digital humanists. It is assumed that researchers need the original copy or a restored one based on the research goal. Therefore both digitization and restoration are crucial for them because they can invalidate the data employed in a research outcome.

Recording equipment is also crucial. For instance, many documents used for linguist research purposes have been recorded without caring about the recording devices and environment conditions [48]. Carriers, digitization (system IFs), and degradations (content IF) concern profiles like musicologists where sound evaluation and analysis drastically change depending on the various versions of a musical work [19]. Similar to content creators, researchers would consciously evaluate the QoE, but they will be driven by different goals (human IF). The recording conditions of the sources is extremely important for such data. For instance, animal sounds are usually recorded in hostile environments where capturing the scene faces several issues. Examples include getting the microphone close enough to the animal, reducing extraneous sources like wind noise produced by the outdoor environment, and taking specialist advice for some type of wildlife recording where ultrasonics should be captured, as in the case of bats (system IFs) [49].

4 AUDIO ARCHIVE DEGRADATIONS

A digital audio archive is composed of material derived from various formats. This peculiar characteristic makes QoE evaluation a complicated challenge. This chapter provides a taxonomy of the most common degradations that can be found in audio archives. The taxonomy can be useful to (1) study what and how carriers and degradations affect audio QoE, which help to understand the causes; (2) propose the usage of degradations (e.g., metadata) as features for developing an objective metric; (3) provide useful information for automatic restoration that might require a degradation recognition stage; and (4) help the creation of datasets for sound archives, which should include archive degradations, either if data are sampled from real-world sound collections or synthetically generated.

Degradations can be grouped into local degradations, where only subsets of samples are corrupted, and global degradations, which concern the whole waveform. Local degradations include clicks and low-frequency noise pulses. Clicks are more common in grooved material and less frequent in magnetic tapes. Clicks are typically perceived as "tick" or "crackle" noises. However clicks could also occur in optical discs or digital files because of digitization mistakes, such as concealed digital errors and timing problems. Low-frequency noise pulses are typical of grooved material, such as shellac discs (i.e., so-called 78-rpm discs), because of large scratches and breakages in the surface of the discs [50]. The perceived effect can be described as "pop" or "thump."

Among the global degradations, broadband noise, hum, pitch variation defects, and distortions are found. Broadband noise or hiss is typical of every recording and storage system; indeed the term "hiss" is typically used for describing everyday electrical noise. Even though hiss is caused by circuitry, it can also be due to ambient noise captured from the recording environment [36].

Hum is a particular noise that depends on power line interference, which commonly occurs during recording or digitizing and therefore concerns every carrier. It is perceived as a low-pitch "whirring." Pitch variation defects are present in every analog carrier for different reasons [51]. In gramophone records (wax cylinders and shellac discs), pitch variation defects arise because of the variation of the rotational speed of the recording medium during either the recording or playback stage. In the case of magnetic tapes, pitch variation defects can be encountered because of irregular stretches during playback or storage. The perceived effect is an overall pitch variation that was not intended by the person who recorded the audio. Depending on the speed variation, it is called "wow" (slow variation) or "flutter" (rapid variation) [36]. The last degradation class is clip, which is a type of distortion. Clipped audio can be caused by groove wall deformation, recording mistakes, or digitization mistakes. It consists of truncating the waveform when the maximum allowed value is exceeded. Clipped audio is usually perceived as a typical distortion of the signal.

Table 3 shows an overview of the relationships between degradations and carriers and the way they are perceived by listeners. Regarding the perceived quality of degradations, little has been done on the state of the art. Understanding if some degradations show higher quality than others might help to create some labels when listening tests cannot be conducted. To the best of the authors' knowledge, there is only one paper that analyzed the perceptual difference of degradations in old recordings [52]. The authors concluded that impulsive noises (i.e., local degradations) are the most annoying disturbances and that their removal introduces audible signal quality loss. However hiss removal methods are also well known for introducing musical noise artifacts [53, 54] that are easily perceivable. Therefore more studies are needed for evaluating the perceived quality of audio archive degradations. More degradations could be generated by other restoration techniques, such as automatic restoration methods [38, 55], machine learning–based restoration techniques [56–60], audio inpainting [61, 62] and sound source modeling–based approach [63].

5 QOE EVALUATION

Since IF has been established, how to assess the sound archive QoE needs to be discussed. Quality assessment approaches are typically divided into two categories: subjective and objective methods. This section first shortly reviews the two approaches and then discusses the challenges of the current methodologies in sound archives.

5.1 Subjective Audio Quality Assessment

Subjective quality assessment methods can be divided into two groups. Listening tests are conducted in a controlled environment (i.e., a lab) or with crowd-sourcing approaches where the experimenter is not physically present with the participants. Within the context of audio quality evaluation, lab-based listening tests represent the most reliable way of assessing audio quality. The idea is to gather information from human assessors in a dedicated room, who assign scores after being exposed to several stimuli of various quality levels. Quality scores are collected through rating scales, which normally are either categorical or continuous. Once many participants rate a stimulus, the mean opinion score (MOS) is used to define the audio quality of that stimulus. Even though lab-based listening tests are the most reliable approach to assess audio quality, several issues can be encountered, such as expensive cost, time-consumption, demographic limitation, artificial environment, wrong formulation of the questions made to the participants, etc. Examples of listening test protocols provided by the International Telecommunication Union (ITU) include the ITU-R BS.1534, commonly called the MUlti Stimulus test with Hidden Reference and Anchor [64], which is suitable for intermediate audio quality in a broadband scenario (i.e., 20 Hz–20 kHz) and requires expert listeners, and the ITU-T P.800 [65], which suits speech transmission quality.

Crowd-sourcing in QoE evaluation consists of presenting and assessing different stimuli using a web interface instead of the laboratory setting [41]. Examples of crowd-sourcing

Table 3. Overview of degradations in audio recordings.

	Carrier	Perception	Cause
Clicks	Grooved material, optical, digital	Ticks, crackles	Dirt, dust, granularity, small scratches, error concealment
Low-frequency noise pulses	Grooved material	Pop, thump	Large scratches, breakages
Broadband noise	All	Hiss	Electric circuits, recording environment
Hum	All	Whirring	Power line interference
Pitch variation defects (wow, flutter)	Grooved material, magnetic tapes	Overall pitch variation	Rotational speed variations, eccentricity, irregular stretches
Clip	Grooved material, optical, digital	Distortion	Groove deformation, exceeding maximum range

methods are given by Chen et al. [66] and Keimel et al. [67]. In both frameworks, different stimuli are provided to the participants, whose goal is to assign scores to each stimulus. Crowd-sourcing for QoE evaluation shows the following advantages: being less time-consuming than lab-based listening tests, having a real-life environment, having a wide demography, and being cheap. However the participants are mostly motivated by financial reasons, and they can cheat in order to get paid more; transmission artifacts must be taken into account; and there are time constraints due to the participant fatigue [41, 68]. Recent studies have shown that crowd-sourcing listening tests can produce similar MOS values to lab-based listening tests at the expense of having to mitigate some issues such as checking participant reliability, fatigue, and environment suitability and using more participants to get the same statistical power [69, 68].

5.2 Objective Metrics

The disadvantages of listening tests pushed researchers to develop computational methods that automatically predict audio quality from the input signal. Three major categories of objective measurements are found: *full reference*, *no-reference*, and *reduced reference* or *parametric* [32, 33]. In full reference metrics, the noisy signal is compared with a reference signal, usually the clean version of the noisy signal. No-reference means that only the noisy signal is used for predicting the quality, and reduced reference refers to the use of only some information about the reference signal or some parameters.

The main problem with using objective metrics is that they do not always reflect the real perceived quality, especially when used for applications that do not represent the target application they were created for. Examples of full reference metrics for audio/music include PEAQ [70], which is suitable for broadband audio with intermediate impairments, and ViSQOLAudio [71] developed for music compressed with low bitrate codecs. Full-reference metrics developed for speech quality include PESQ [72], which is suitable for narrowband speech with small impairments, and POLQA [73]. Instead, an example of a no-reference method is ITU-T P.563 [74], which is designed for narrowband speech quality. An example of a parametric metric is the E-model [75] used to estimate speech quality in a conversational context [33].

More recently, several deep learning-based approaches for no-reference speech quality assessment [76–79] and music quality assessment [80] emerged. Some of these approaches show higher performance than full-reference metrics in some datasets, which suggests that deep learning can be a powerful approach to deal with speech quality assessment. Even for contemporary media archives (e.g., YouTube or SoundCloud) where the original recording may be digital, no reference objective models may required. Because digital metadata only records the latest coding information, relying on digital coding metadata to quantify quality may be an issue if media is re-encoded [81]. For a review of objective metrics for audio quality the reader is referred to [29–33].

6 QOE EVALUATION IN AUDIO ARCHIVES: CHALLENGES AND LIMITATIONS

Several challenges arise when assessing audio archive QoE. This section goes through each life cycle stage to detect the corresponding QoE assessment challenge.

6.1 Digitization

The digitization stage (Sec. 1.1) suits naturally a reduced-reference metric that is based on predicting quality based on parameters. A quality metric would check if the digitization parameters are correct. An example could be checking the metadata file as discussed above. Other kinds of metrics could be no-reference during the digitization stage. For instance, a quality metric that detects the severity of degradations produced by digitization mistakes might help archivists to find when digitization failed. This issue is typically found in digital audio tapes [82]. A high-priority challenge is that archive data exhibiting digitization mistakes are mostly available from institutions such as national libraries or broadcasters. Therefore a collaboration between some of the archive institutions and academic research is needed. A second problem concerns the annotations that should be provided by these institutions.

6.2 Restoration

Restoration (Sec. 1.2) might benefit from both full-reference and no-reference metrics. Full-reference metrics might consider the signal before the restoration and the restored one, while no-reference metrics work on the noisy signal only.

Restoration is a crucial stage for determining the final quality of a restored digital audio archive. Organizations restore material through the usage of Digital Audio Workstations, which include tools such as clip removal, hiss reduction, click removal, and so on. The mechanism that underlies these tools is built on top of digital audio restoration techniques [36, 83, 84]. The authors are interested in how restoration is evaluated. Digital restoration techniques were assessed using full-reference metrics [85], listening tests, or objective metrics that are not based on human auditory perception [86, 87]. However full-reference metrics using a clean signal are not appropriate since the clean signal is missing in real-world sound archives and listening tests are costly. Sound engineers use their listening skills for evaluating restored audio documents. One problem with this approach is the strong subjectivity of the result and large collections cannot be restored manually. In this context, the authors believe that quality evaluation procedures can help both audio processing researchers and sound engineers for improving QoE.

6.3 Consumption

The consumption stage (Sec. 1.3) requires a no-reference metric since no other signal is available. The main challenge for the consumption stage concerns the heterogeneity of the content that constitutes a sound archive. Libraries organize content in the form of web pages, and they usually divide archives into collections. Collections could represent anything: musical genre, period, programs from a specific broadcaster, a particular carrier, an ethnic group, a specific composer, a specific performer, and so on. For this reason, an objective quality metric should be able to deal with all the various content. This means that datasets including both speech and music should be used to create and test an objective quality metric. Music exhibits even more challenges because sound archives include many genres and music from several cultures. As discussed above, current objective metrics are not designed to work with such heterogeneous content.

6.4 Common Challenges

Given the promising performance of recent deep learning approaches [76–79], the authors suggest that deep learning warrants further attention and application to develop a QoE metric for a sound archive. It requires machine learning expertise and the collection of data and annotations. The latter can be collected allowing participants to rate MOS as previously mentioned. Given the novelty of the problem, no ground truth is available in the literature, and it needs to be found using auditory experiments. Other IFs can be extracted using questionnaires and/or surveys.

Here is an overview of the main challenges required for using this approach:

- Identifying the stakeholders,
- Collecting the data either synthetically or from real-world recordings,
- Defining and conducting a listening test for obtaining reliable ground truth,
- Generating synthetic data for pretraining in the case that collecting the ground truth is not feasible,
- Defining procedure for obtaining IF, and
- Evaluating machine learning models.

It should be noted that the description level is very general. Also several aspects could change compared to what was mentioned above. For instance, in case of real audio archives not being available, one could think to create artificial data. As a consequence, a reference signal is available and different procedures can be applied. In such a case, the fact that artificial data are considered must be taken into account when evaluating quality. Does it match real-world scenarios?

Also it is quite hard to define a reference signal even if it can be obtained. For instance, in applications such as perceptual audio coding, the reference signal is the audio file before compression. However, in digital audio archives, it cannot always be assumed that there is the desired signal that corresponds to the content creator's intention and an undesired signal that is represented by all the technical stages of the processing chain that caused noise-like components. As mentioned earlier, everything depends on the stakeholder who judges sound quality. A typical approach used in audio restoration consists of artificially creating the carrier degradation and adding it to a clean signal [38, 83, 61]. Then full reference measures are used for evaluating performance. This means that an assumption has been made, i.e., the desired signal is the one that contains audio that was intended to be recorded only. However this assumption does not apply in all the scenarios, and a QoE-based evaluation can fill this gap.

7 CASE STUDY

This section presents a case study to illustrate with examples the topics discussed above. The authors use the NASA audio collection that has also gained attention in the audio and speech processing community with the release of the Fearless Steps challenge [88]. Additionally, the NASA audio collection suits the goal of showing a case study given that it represents many IFs experienced in audio archives.

The NASA audio collection includes archival recordings from certain NASA missions (Apollo, Gemini, and Mercury). The Apollo audio recordings include conversations between astronauts, crew members, and backroom staff at the NASA Mission Control Center (MCC) during the Apollo missions. Audio recordings are divided into different categories: onboard, commentary, technical air-to-ground, MCC recordings, and pre and post-mission recordings. Onboard recordings include all the conversations between the astronauts on the lunar module and the Command Service Module. Commentary and technical air-to-ground recordings include conversations between astronauts and the capsule communicator. Commentary recordings overlap with technical air-to-ground recordings with the public affairs officers' comments overdubbed. MCC audio includes conversations between the staff members like communications between flight controllers and back-room specialists. Pre and post-mission recordings include interviews and press conferences.

The broad acoustic scenario of this audio collection poses many challenges for speech processing tasks, audio quality prediction included. The Apollo collection includes recordings that are full of silence, almost undetectable speech, and variable signal-to-noise ratio. Also the usage of different voice channel implementations that change according to the mission status makes quality evaluation more difficult [89].

7.1 Stakeholders and Influence Factors

The stakeholders of the Apollo collection span the entire audio archive life cycle from digitization to consumption. This section discusses the stakeholders and influence factors for each life cycle stage.

7.1.1 Digitization

Digitization of the Apollo collection, and more generally the whole NASA audio collection, is a time-consuming operation. The Apollo mission only exhibits ~100,000 h of recordings stored in more than 700 analog tapes, with 30 tracks per tape [90]. It has been estimated that it would take 11 years to digitize the whole Apollo collection, so a new technique able to decode all 30 tracks simultaneously has been proposed [90]. Evaluating digitization mistakes with a manual approach is infeasible in this context. Stakeholders such as NASA would benefit from having a quality metric as discussed in SEC. 3.1. The goal is to detect potential digitization mistakes that might occur during the digitization process [91] and also quantify the severity of the degradations caused by digitization. For the organizations that preserve the Apollo collection, some of the IFs that were assigned to archivists in Table 2 are identified (digitization guidelines, carriers, goal, and digitization cost).

7.1.2 Restoration

Restoration concerns stakeholders such as audio processing researchers and sound engineers. Restoring the Apollo collection is not feasible with a manual approach. Manual approaches could include both types of researchers applying a speech enhancement algorithm or sound engineers applying plugin settings. In both cases, mistakes will occur since the content requires different settings based on the degradations that occur in that particular region. Therefore more attention should be given to automatic restoration [38, 55, 60, 57, 56] and to finding quality assessment techniques for automatically restored recordings. The IFs detected at this stage are mostly focused on system and context IFs and include source, degradations, carriers, devices, cost of restoration, and restoration methods. Predicting the quality of restored records could help sound engineers to inform their performance and be used to detect where automatic restoration might fail.

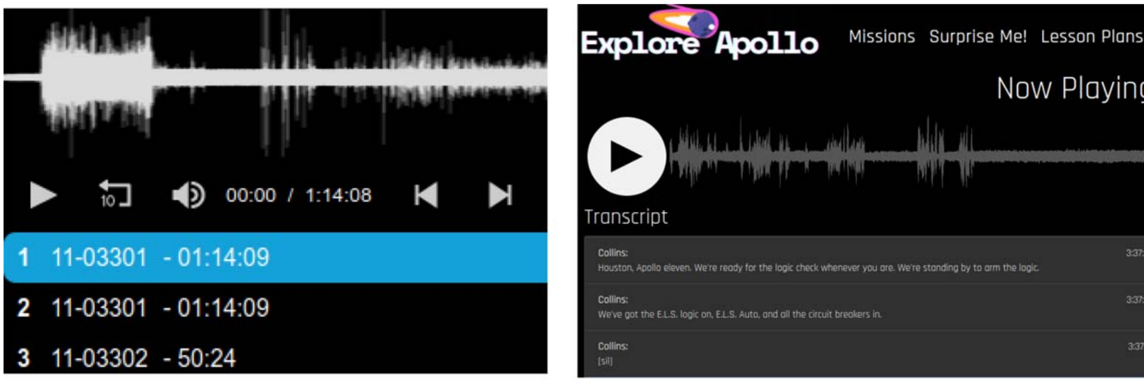
7.1.3 Consumption

For the consumption stage, the purpose of use of public listeners and content creators is discussed. QoE can be seen in this context in terms of how to improve archive accessibility. Currently the NASA audio collection is accessible through the Internet Archive,[5] and some Apollo missions are available on the web applications Explore Apollo[6] and Apollo In Real Time.[7] The first two are shown in Fig. 3. They all provide a different way to explore recordings but none of the cases considers the sound quality.

The Internet Archive provides many long audio files (some last three or four hours) that are not sorted or organized in collections (e.g., by mission status). Also the filenames refer to the tape where NASA stored the recording, which makes archive exploration hard. The web applications provide a mechanism to extract the highlighted moments of the missions or to search speech through text which improves QoE. However no quality labels can be found and accessing the recordings might not be ideal for content creators and public listeners. Indeed the NASA audio collection exhibits parts where audio can be described as annoying (e.g., unintelligible at key points or long periods without active content). As a result, navigating and exploring the archive to find interesting content can easily cause listening fatigue.

Similar scenarios can be found in other archive collections, where some recordings might be very unpleasant, i.e., some of the early gramophone recordings or amateur modern recordings. Users QoE of the Apollo collections will benefit from having audio quality information alongside other content retrieval tools so that they can retrieve the best quality items. Improving archive accessibility is a human IF that might be explored with questionnaires to understand how it affects the QoE and how to improve it. Since consumption is at the end of the archive life cycle, it will benefit from a successful restoration. It must be noted that content IFs play a key role in the Apollo collection. In their previous work [92], the authors showed that clusters of audio features in the Apollo collection overlap with mission status (onboard, air-to-ground, etc.). This means that the mission status could be seen as a useful feature to predict audio quality. Similarly this concept could be extended to archive collections that exhibit different carriers that show clear, distinct time-spectral characteristics, e.g., gramophone recordings are easy to distinguish from tape recordings and Apollo onboard recordings captured in the spaceship are different from commentary recordings. Further listening tests should be conducted to explore this content IF. Also context IFs are included, since public lis-

[5]https://archive.org/details/nasaaudiocollection.
[6]https://app.exploreapollo.org/.
[7]https://apolloinrealtime.org/.



(a) Internet Archive collection (b) Explore Apollo app

Fig. 3. Example of how the Apollo collections are currently delivered. Having access to quality labels could improve content creators' and public listeners' QoE, e.g., the clips could be sorted by quality and the best quality recordings could be retrieved within the same mission status. (a) Internet Archive collection and (b) Explore Apollo app.

teners would have access to the Apollo collection in many different physical environments.

7.2 Audio Degradations

Sec. 4 introduced a taxonomy of the archive degradations, by dividing them into global and local. Degradations could also be divided between those generated from an external source (e.g., environmental noise, hiss, hum) and the ones that are applied to the content (e.g., clip, general non-linear distortions, reduced content frequency bandwidth). In the Apollo collection, many recordings show both characteristics. For example, onboard recordings exhibit loud environmental noise and a reduced frequency bandwidth, while commentary recordings show almost no hiss. When judging the overall quality, it is not clear if the quality judgment is because of the content or degradation type that affects the background. This is also important depending on which quality factor someone is studying. For example, distortions, such as amplitude clipping, might affect the overall quality but preserve speech intelligibility.

Asking participants the quality of separated scales, e.g., speech vs. background, might help to determine the causes that generate high or low quality. Labeling both speech and background quality might also help to discover the best quality recordings of a certain mission status or carrier in a general sound archive. For example, onboard recordings exhibit loud noise that causes extreme fatigue as reported by participants in a listening test survey feedback [92]. This means that finding the audio recordings with low-quality background noise might help to filter out recordings where effort is needed. Archive audio, such as gramophone recordings, could lead to listener fatigue, which could benefit from a similar application.

7.3 QoE Evaluation

This section shows the challenges related to the Apollo collection for the following issues: subjective quality assessment, data collection for a quality model, and development of an objective quality model.

7.3.1 Subjective Quality Assessment

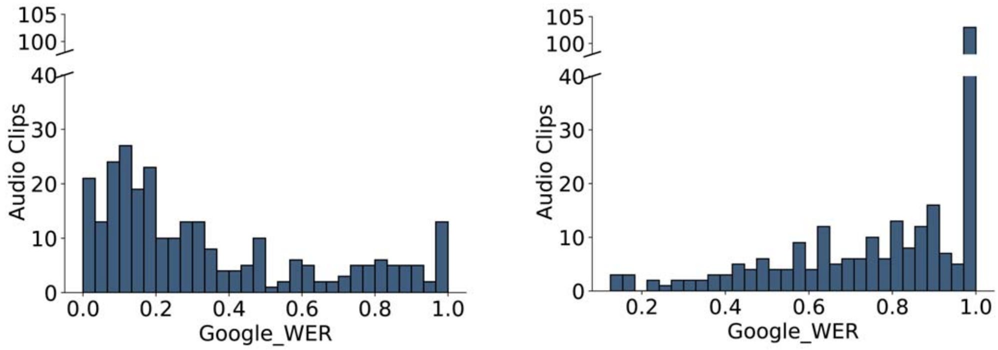
The overall quality of a speech stimulus can be seen as the interaction of many factors: pleasantness, intelligibility, listening effort, and loudness [33]. Evaluating the contribution of each factor in the Apollo collection will help the design of a suitable listening test. In their previous work, the authors have conducted a small-scale listening test to evaluate speech intelligibility, which shows low correlation with some objective metrics [92]. The next study can focus on understanding how intelligibility contributes to the overall quality. The authors assume that sound archives like the Apollo collection might benefit from an intelligibility assessment when it comes to public listeners and content creators. However the authors suggest that more studies are needed to explore the overall quality. New findings could be adapted to other speech archives that represent important events of mankind.

Further listening tests are needed to understand how the mission status contributes to the overall quality or intelligibility, which might help the development of an objective quality metric as discussed earlier.

7.3.2 Data Collection

The development and evaluation of a data-driven quality metric require the collection of a suitable dataset, which is currently missing for sound archives. A dataset can be created through artificially degrading clean recordings or sampling from real-world recordings.

In their previous work [92], the authors proposed an approach to building a dataset for speech quality prediction by sampling real-world recordings from the Apollo collection. Proposing a dataset that is artificially degraded is hard in the case of the Apollo recordings. Techniques for antiquing modern recordings to emulate the quality of the analog carriers have been proposed [93]. However the Apollo col-



(a) Google STT-WER Commentary (b) Google STT-WER Onboard

Fig. 4. Word error rate (WER) of Google speech-to-text (STT) computed from onboard recordings (Apollo 11) and commentary (Apollo 11, 17) [92].

lection shows degradations that go beyond the magnetic tape degradation, which would be hard to emulate. Examples include space-to-ground communication, recordings made on the spacecraft with far-field microphones picking environmental noise that varies over time, and the usage of both head-mounted microphones and far-field microphones [94].

The “in the wild” recording situation could be reproduced by extracting noise-only regions and mixing them to some clean speech. However the non-linear distortions that modify the clean speech signal will not be covered with this approach, and they are not easy to generate. Plus, finding clean speech similar to natural astronaut conversations can be hard. Although the usage of real-world recordings allows mitigation of these issues, the uncontrolled conditions pose other challenges that should be investigated.

1) Sampling clips randomly from the Apollo collection is not ideal for creating a dataset because of the massive presence of repetitive recordings. In particular, it was identified that the MOS distribution of random sampling is skewed [92] and a mechanism to avoid data repetition is needed. This is because otherwise, a data-driven model will be less capable of predicting quality in the entire MOS range.
2) Audio clips need to be annotated through listening tests to collect labels. The listening session should be prepared such that every participant is exposed to all the degradation conditions. This helps to avoid listening test biases. One issue is that without controlling the degradations, it becomes hard to select all the degradation conditions that are present in the archive. This means that some participants risk being exposed to a subset of the degradation conditions, which produces biases.

These two challenges can be found in any audio collection that has real-world recordings. The solutions that the authors proposed are the following:

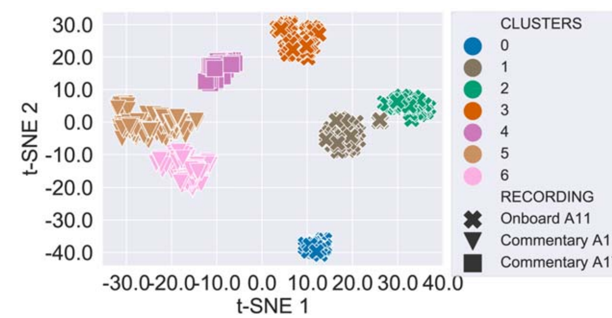


Fig. 5. HDBSCAN clusters of features computed from onboard recordings (Apollo 11) and commentary (Apollo 11, 17) [92]. The feature space is reduced in 2 dimensions using t-distributed stochastic neighbor embedding (t-SNE).

1) To avoid data repetition, it was found that the word-error-rate (WER) of the Google Speech-to-Text (STT) is correlated to subjective labels while other non-intrusive quality metrics such as ITU.T P.563 [74] fail. This means that an STT-WER could be used to detect and then prune audio clips that cause a skewed MOS distribution. The STT-WER distributions of commentary and onboard recordings from the Apollo collection are shown in Fig. 4.
2) To prepare listening tests such that all the degradation conditions are used, the authors propose [92] to find clusters using audio features and the HDBSCAN [95] technique. Distinct clusters are shown in Fig. 5. When preparing the listening test, the experimenter can pick an audio clip uniformly from each cluster to guarantee that all the degradations are covered in a listening test session.

7.3.3 Objective Quality Assessment

Given the vast number of recordings, the development of a quality metric can be explored with unsupervised (no labels) or semi-supervised (dataset partially labeled) techniques. In their work [76], the authors proposed an approach based on pretraining with a clustering technique and fine-tuning with a multitask approach that uses labels generated

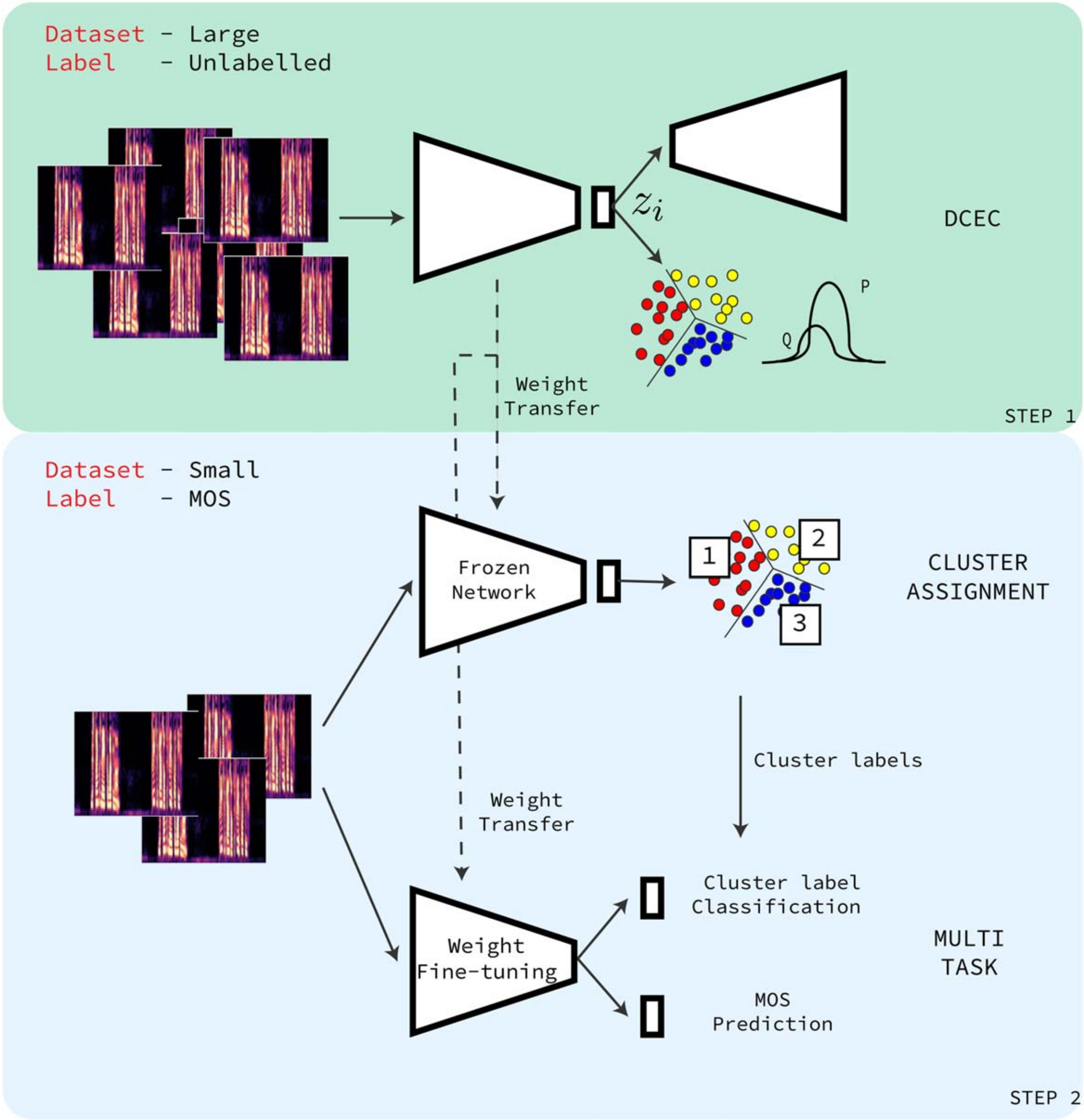


Fig. 6. Deep convolutional embedded clustering (DCEC) for mean opinion score (MOS) prediction plus multi-task with cluster labels [76].

by clusters. The authors' technique is particularly useful for pretraining in an unsupervised fashion. The proposed model is shown in Fig. 6. Thanks to the proposed model, superior performance to some existing non-intrusive metrics can be achieved by using a very small dataset annotated with quality labels.

8 CONCLUSION

Several organizations and individuals collect digital audio archives. Although archive preservation is the most urgent operation, technology for evaluating the QoE of digitization, restoration, and consumption of audio archives is missing, and QoE assessment is not properly conducted. In this paper, the authors proposed to extend the adaptation of the QoE framework to digital audio archives by focusing on five key points:

1) Collecting the potential IFs for audio archive QoE evaluation,
2) Finding the stakeholders and their corresponding QoE IFs,
3) Proposing a taxonomy of the audio archive degradations,
4) Showing the limitations of existing subjective and objective quality assessment techniques, and
5) Using the NASA Apollo collection as a case study to show real-world examples.

The proposed QoE framework is designed to improve research and tackle the above-mentioned challenges related to sound archives. It must be noted that the collected IFs will need to be evaluated in the future. It is expected that some IFs will have more influence on quality than others.

Research methodologies that can be produced for audio archive applications through the usage of the adapted QoE framework are also useful for other applications. Evaluating the QoE of audio applications becomes more urgent given the growth of multimedia content in digital archives. For example, videos corresponding to a different version of the same song (e.g., different live and studio versions) are uploaded to streaming services, and delivering the best quality version is beneficial to improve the QoE. Broadcasters need to automate metadata labeling to detect content that needs to be restored or know which record exhibits an acceptable perceived quality for the users. General consumer access services such as the Internet Archive or web search engines can benefit from labeling quality content to improve their services.

The proposed framework could be further extended to include more stakeholders or more QoE influence factors. With the proposed framework, automatic quality assessment of sound archives is presented as a feasible approach. The goal of this paper is also to encourage the communities of stakeholders to give attention to automatic quality assessment techniques for sound archives.

9 ACKNOWLEDGMENT

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grants 17/RC-PhD/3483 and 17/RC/2289_P2. This work was supported by The Alan Turing Institute under the Engineering and Physical Sciences Research Council Grant EP/N510129/1.

10 REFERENCES

[1] A. Tovell, "Audio Preservation and Access: Overcoming the Challenges," in *Proceedings of the AES International Conference on Audio Archiving, Preservation & Restoration* (2018 Jun.), paper P-8.

[2] F. Rumsey, "Preserving Our Audio Heritage," *J. Audio Eng. Soc.*, vol. 66, no. 10, pp. 852–856 (2018 Oct.).

[3] M. Casey, "Why Media Preservation Can't Wait: The Gathering Storm," *IASA J.*, vol. 44, pp. 14–22 (2015 Jan.).

[4] F. Rumsey, "Will You Be Mine Forever? Audio Archiving, Multitracks, and 90s Digital," *J. Audio Eng. Soc.*, vol. 68, no. 4, pp. 304–307 (2020 Apr.).

[5] D. Schüller, "Preserving the Facts for the Future: Principles and Practices for the Transfer of Analog Audio Documents into the Digital Domain," *J. Audio Eng. Soc.*, vol. 49, no. 7/8, pp. 618–621 (2001 Jul.).

[6] United Nations Educational, Scientific and Cultural Organization, *Information for All Programme (IFAP). Report 2004/2005* (UNESCO, Paris, France, 2006). https://unesdoc.unesco.org/ark:/48223/pf0000147135.

[7] A. Ragano, E. Benetos, and A. Hines, "Adapting the Quality of Experience Framework for Audio Archive Evaluation," in *Proceedings of the 11th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–3 (Berlin, Germany) (2019 Jun.). https://doi.org/10.1109/QoMEX.2019.8743302.

[8] K. Brunnström, S. A. Beker, K. De Moor, et al., "Qualinet White Paper on Definitions of Quality of Experience," https://hal.archives-ouvertes.fr/hal-00977812/document (2013 Mar.).

[9] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the Influence of Rebuffering Interruptions on the User's Quality of Experience During Mobile Video Watching," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 47–61 (2013 Mar.). https://doi.org/10.1109/TBC.2012.2220231.

[10] Y. Zhu, I. Heynderickx, and J. A. Redi, "Understanding the Role of Social Context and User Factors in Video Quality of Experience," *Comput. Hum. Behav.*, vol. 49, pp. 412–426 (2015 Aug.). https://doi.org/10.1016/j.chb.2015.02.054.

[11] N. Murray, G. Muntean, Y. Qiao, S. Brennan, and B. Lee, "Modeling User Quality of Experience of Olfaction-Enhanced Multimedia," *IEEE Trans. Broadcast.*, vol. 64, no. 2, pp. 539–551 (2018 Jun.). https://doi.org/10.1109/TBC.2018.2825297.

[12] M. Varela, L. Skorin-Kapov, and T. Ebrahimi, "Quality of Service Versus Quality of Experience," in S. Möller and A. Raake (Eds.), *Quality of Experience: Advanced Concepts, Applications and Methods*, T-Labs Series in Telecommunication Services, pp. 85–96 (Springer, Cham, Switzerland, 2014). https://doi.org/10.1007/978-3-319-02681-7.

[13] M. Schmitt, J. Redi, D. Bulterman, and P. S. Cesar, "Towards Individual QoE for Multiparty Videoconferencing," *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1781–1795 (2017 Jul.). https://doi.org/10.1109/TMM.2017.2777466.

[14] M. Suznjevic, L. Skorin-Kapov, and M. Matijasevic, "The Impact of User, System, and Context Factors on Gaming QoE: A Case Study Involving MMORPGs," in *Proceedings of the 12th Annual Workshop on Network and Systems Support for Games (NetGames)*, pp. 1–6 (Denver, CO) (2013 Dec.). https://doi.org/10.1109/NetGames.2013.6820606.

[15] IASA Technical Committee, "Handling and Storage of Audio and Video Carriers," *Standard IASA-TC 05* (2014). https://www.iasa-web.org/handling-storage-tc05.

[16] IASA Technical Committee, "Guidelines on the Production and Preservation of Digital Audio Objects," *Standard IASA-TC 04* (2004). www.iasa-web.org/tc04/audio-preservation.

[17] M. Casey, "Quality Control for Media Digitization Projects," *IASA J.*, vol. 1, no. 50, pp. 45–52 (2019 Aug.). https://doi.org/10.35320/ij.v0i50.92.

[18] F. Rumsey, "Act Before the Window Closes: Safeguarding Audio Heritage," *J. Audio Eng. Soc.*, vol. 69, no. 7/8, pp. 607–611 (2021 Jun.).

[19] N. Wallaszkovits, "Between Standards and Arts: Digitisation and Restoration of Audio Material – A Balancing Act Between Authenticity and Manipulation?" *J. New Music Res.*, vol. 47, no. 4, pp. 285–290 (2018 Oct.). https://doi.org/10.1080/09298215.2018.1514058.

[20] D. Schüller, "The Ethics of Preservation, Restoration, and Re-Issues of Historical Sound Recordings," *J. Audio Eng. Soc.*, vol. 39, no. 12, pp. 1014–1017 (1991 Dec.).

[21] T. Lidy, C. N.Jr Silla,, O. Cornelis, et al., "On the Suitability of State-of-the-Art Music Information Retrieval Methods for Analyzing, Categorizing and Accessing Non-Western and Ethnic Music Collections," *Signal Process.*, vol. 90, no. 4, pp. 1032–1048 (2010 Apr.). https://doi.org/10.1016/j.sigpro.2009.09.014.

[22] S. Canazza, C. Fantozzi, and N. Pretto, "Accessing Tape Music Documents on Mobile Devices," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 12, no. 1s, paper 20 (2015 Oct.). https://doi.org/10.1145/2808200.

[23] J. Goldman, S. Renals, S. Bird, et al., "Accessing the Spoken Word," *Int. J. Digit. Libr.*, vol. 5, no. 4, pp. 287–298 (2005 Aug.). https://doi.org/10.1007/s00799-004-0101-0.

[24] A. Hines and J. D. Kelleher, "A Framework for Post-Stroke Quality of Life Prediction Using Structured Prediction," in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (Erfurt, Germany) (2017 May). https://doi.org/10.1109/QoMEX.2017.7965672.

[25] T. Walton, *The Quality of Experience of Next Generation Audio: Exploring System, Context and Human Influence Factors*, Ph.D. thesis, Newcastle University, Tyne, UK (2017 Dec.). https://doi.org/10.17634/154300-81.

[26] H. A. Taylor, "The Collective Memory: Archives and Libraries as Heritage," *Archivaria*, vol. 15, pp. 118–130 (1982 Jan.).

[27] L. Watkins, "The Politics of Nostalgia: Colorization, Spectatorship and the Archive," *Comp. Cinema*, vol. 9, no. 17, pp. 123–145 (2021 Dec.). https://doi.org/10.31009/cc.2021.v9.i17.07.

[28] G. Colavizza, T. Blanke, C. Jeurgens, and J. Noordegraaf, "Archives and AI: An Overview of Current Debates and Future Perspectives," *J. Comput. Cult. Herit.*, vol. 15, no. 1, paper 4 (2022 Feb.). https://doi.org/10.1145/3479010.

[29] M. Torcoli, T. Kastner, and J. Herre, "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1530–1541 (2021 Jan.). https://doi.org/10.1109/TASLP.2021.3069302.

[30] D. Campbell, E. Jones, and M. Glavin, "Audio Quality Assessment Techniques—A Review, and Recent Developments," *Signal Process.*, vol. 89, no. 8, pp. 1489–1500 (2009 Aug.). https://doi.org/10.1016/j.sigpro.2009.02.015.

[31] A. W. Rix, J. G. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, "Objective Assessment of Speech and Audio Quality—Technology and Applications," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 1890–1901 (2006 Nov.). https://doi.org/10.1109/TASL.2006.883260.

[32] Z. Akhtar and T. H. Falk, "Audio-Visual Multimedia Quality Assessment: A Comprehensive Survey," *IEEE Access*, vol. 5, pp. 21090–21117 (2017 Sep.). https://doi.org/10.1109/ACCESS.2017.2750918.

[33] S. Möller, W.-Y. Chan, N. Cote, et al., "Speech Quality Estimation: Models and Trends," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 18–28 (2011 Nov.). https://doi.org/10.1109/MSP.2011.942469.

[34] E. Wieser, M. Husinsky, and M. Seidl, "Speech/Music Discrimination in a Large Database of Radio Broadcasts From the Wild," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2134–2138 (Florence, Italy) (2014 May). https://doi.org/10.1109/ICASSP.2014.6853976.

[35] A. Wilson and B. M. Fazenda, "Perception of Audio Quality in Productions of Popular Music," *J. Audio Eng. Soc.*, vol. 64, no. 1/2, pp. 23–34 (2016 Jan.). https://doi.org/10.17743/jaes.2015.0090.

[36] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration* (Springer, London, UK, 1998). https://doi.org/10.1007/978-1-4471-1561-8.

[37] A. Ragano, E. Benetos, and A. Hines, "Audio Impairment Recognition Using a Correlation-Based Feature Representation," in *Proceedings of the 12th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (Athlone, Ireland) (2020 May). https://doi.org/10.1109/QoMEX48832.2020.9123111.

[38] M. Brandt, S. Doclo, T. Gerkmann, and J. Bitzer, "Impulsive Disturbances in Audio Archives: Signal Classification for Automatic Restoration," *J. Audio Eng. Soc.*, vol. 65, no. 10, pp. 826–840 (2017 Oct.). https://doi.org/10.17743/jaes.2017.0032.

[39] D. Preis and H. Polchlopek, "Restoration of Nonlinearly Distorted Magnetic Recordings," *J. Audio Eng. Soc.*, vol. 32, no. 1/2, pp. 26–30 (1984 Feb.).

[40] S. A. White, "Restoration of Nonlinearly Distorted Audio by Histogram Equalization," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 828–832 (1982 Nov.).

[41] S. Möller and A. Raake (Eds.), *Quality of Experience: Advanced Concepts, Applications and Methods*, T-Labs Series in Telecommunication Services, (Springer, Cham, Switzerland, 2014). https://doi.org/10.1007/978-3-319-02681-7.

[42] Z. Akhtar, K. Siddique, A. Rattani, S. L. Lutfi, and T. H. Falk, "Why is Multimedia Quality of Experience Assessment a Challenging Problem?" *IEEE Access*, vol. 7, pp. 117897–117915 (2019 Aug.). https://doi.org/10.1109/ACCESS.2019.2936470.

[43] R. Wright, "Digital Preservation of Audio, Video and Film," *VINE J. Inf. Knowl. Manag. Syst.*, vol. 34, no. 2, pp. 71–76 (2004 Jun.). https://doi.org/10.1108/03055720410550869.

[44] F. Rumsey, "Challenges in Archiving and Restoration," *J. Audio Eng. Soc.*, vol. 64, no. 1/2, pp. 94–97 (2016 Jan.).

[45] F. Lemmers, "Digitizing Sound Archives at Royal Library of Belgium. Challenges and Difficulties Encountered During a Major Digitization Project," *Bibliothek Forschung und Praxis*, vol. 42, no. 2, paper 3209 (2018 Mar.).

[46] M. R. Leek and C. S. Watson, "Learning to Detect Auditory Pattern Components," *J. Acoust. Soc. Am.*, vol. 76, no. 4, pp. 1037–1044 (1984 Oct.).

[47] S. Gordon-Salant, “Hearing Loss and Aging: New Research Findings and Clinical Implications.” *J. Rehabil. Res. Dev.*, vol. 42, no. 4, pp. 9–24 (2005 Jul.). https://doi.org/10.1682/jrrd.2005.01.0006.

[48] D. Nathan, “Sound and Unsound Practices in Documentary Linguistics: Towards an Epistemology for Audio,” in P. K. Austin (Ed.), *Language Documentation and Description*, vol. 7, pp. 262–284 (SOAS, London, UK, 2010).

[49] R. Kettle, “Natural History,” in D. Lance (Ed.), *Sound Archives - A Guide to their Establishment and Development* (International Association of Sound Archives, London, UK, 1983).

[50] P. A. Esquef, L. W. Biscainho, and V. Välimäki, “An Efficient Algorithm for the Restoration of Audio Signals Corrupted With Low-Frequency Pulses,” *J. Audio Eng. Soc.*, vol. 51, no. 6, pp. 502–517 (2003 Jun.).

[51] A. Czyzewski, M. Dziubinski, A. Ciarkowski, et al., â€œNew Algorithms for Wow and Flutter Detection and Compensation in Audio,” presented at the *118th Convention of the Audio Engineering Society* (2005 May), paper 6353.

[52] S. Canazza, G. Coraddu, G. De Poli, and G. A. Mian, “Objective and Subjective Comparison of Audio Restoration Methods,” *J. New Music Res.*, vol. 30, no. 1, pp. 93–102 (2001). https://doi.org/10.1076/jnmr.30.1.93.7120.

[53] M. Niedzwiecki and M. Ciołek, “Elimination of Impulsive Disturbances From Archive Audio Signals Using Bidirectional Processing,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 5, pp. 1046–1059 (2013 May). https://doi.org/10.1109/TASL.2013.2244090.

[54] S. V. Vaseghi and R. Frayling-Cork, “Restoration of Old Gramophone Recordings,” *J. Audio Eng. Soc.*, vol. 40, no. 10, pp. 791–801 (1992 Oct.).

[55] M. Brandt, S. Doclo, and J. Bitzer, “Automatic Noise PSD Estimation for Restoration of Archived Audio,” *J. Audio Eng. Soc.*, vol. 67, no. 1/2, pp. 38–53 (2019 Jan.). https://doi.org/10.17743/jaes.2018.0073.

[56] C. F. Stallmann and A. P. Engelbrecht, “Gramophone Noise Detection and Reconstruction Using Time Delay Artificial Neural Networks,” *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 47, no. 6, pp. 893–905 (2017 Jun.). https://doi.org/10.1109/TSMC.2016.2523927.

[57] A. Porov, E. Oh, K. Choo, et al., “Music Enhancement by a Novel CNN Architecture,” presented at the *145th Convention of the Audio Engineering Society* (2018 Oct.), paper 10036.

[58] A. Czyzewski, “Learning Algorithms for Audio Signal Enhancement, Part 1: Neural Network Implementation for the Removal of Impulse Distortions,” *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 815–831 (1997 Oct.).

[59] A. Czyzewski, “Learning Algorithms for Audio Signal Enhancement, Part 2: Implementation of the Rough-Set Method for the Removal of Hiss,” *J. Audio Eng. Soc.*, vol. 45, no. 11, pp. 931–943 (1997 Nov.).

[60] Y. Li, B. Gfeller, M. Tagliasacchi, and D. Roblek, “Learning to Denoise Historical Music,” in *Proceedings of the 21st International Society for Music Information Retrieval Conference (ISMIR)*, pp. 504–511 (Montreal, Canada) (2020 Oct.).

[61] A. Adler, V. Emiya, M. G. Jafari, et al., “Audio Inpainting,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 3, pp. 922–932 (2012 Mar.). https://doi.org/10.1109/TASL.2011.2168211.

[62] D. Barry, A. Ragano, and A. Hines, “Audio Inpainting Based on Self-Similarity for Sound Source Separation Applications,” in *Proceedings of the IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6 (2020 Sep.). https://doi.org/10.1109/MMSP48831.2020.9287104.

[63] P. A. A. Esquef, V. Välimäki, and M. Karjalainen, “Restoration and Enhancement of Solo Guitar Recordings Based on Sound Source Modeling,” *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 227–236 (2002 Apr.).

[64] International Telecommunication Union, “Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems,” *Recommendation ITU-R BS.1534-1* (2003 Jan.).

[65] International Telecommunication Union, “Methods for Subjective Determination of Transmission Quality,” *Recommendation ITU-R P.800* (1996 Aug.).

[66] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, “Quadrant of Euphoria: A Crowdsourcing Platform for QoE Assessment,” *IEEE Netw.*, vol. 24, no. 2, pp. 28–35 (2010 Apr.). https://doi.org/10.1109/MNET.2010.5430141.

[67] C. Keimel, J. Habigt, C. Horch, and K. Diepold, “QualityCrowd — A Framework for Crowd-Based Quality Evaluation,” in *Proceedings of the Picture Coding Symposium*, pp. 245–248 (Krakow, Poland) (2012 May). https://doi.org/10.1109/PCS.2012.6213338.

[68] T. Hoßfeld, C. Keimel, M. Hirth, et al., “Best Practices for QoE Crowdtesting: QoE Assessment With Crowdsourcing,” *IEEE Trans. Multimed.*, vol. 16, no. 2, pp. 541–558 (2014 Feb.). https://doi.org/10.1109/TMM.2013.2291663.

[69] B. Naderi, R. Z. Jiménez, M. Hirth, S. Möller, F. Metzger, and T. Hoßfeld, “Towards Speech Quality Assessment Using a Crowdsourcing Approach: Evaluation of Standardized Methods,” *Qual. User Exp.*, vol. 6, no. 1, paper 2 (2020 Nov.). https://doi.org/10.1007/s41233-020-00042-1.

[70] T. Thiede, W. C. Treurniet, R. Bitto, et al., “PEAQ—The ITU Standard for Objective Measurement of Perceived Audio Quality,” *J. Audio Eng. Soc.*, vol. 48, no. 1/2, pp. 3–29 (2000 Feb.).

[71] A. Hines, E. Gillen, D. Kelly,Â et al., â€œViSQOLAudio: An Objective Audio Quality Metric for Low Bitrate Codecs,” *J. Acoust. Soc. Am.*, vol. 137, no. 6, pp. EL449–EL455 (2015 Oct.). https://doi.org/10.1121/1.4921674.

[72] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, “Perceptual Evaluation of Speech Quality (PESQ) - A New Method for Speech Quality Assessment of Telephone Networks and Codecs,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*,

pp. 749–752 (Salt Lake City, UT) (2001 May). https://doi.org/10.1109/ICASSP.2001.941023.

[73] J. G. Beerends, C. Schmidmer, J. Berger, et al., "Perceptual Objective Listening Quality Assessment (POLQA), The Third Generation ITU-T Standard for End-to-End Speech Quality Measurement Part I—Temporal Alignment," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 366–384 (2013 Jun.).

[74] L. Malfait, J. Berger, and M. Kastner, "P. 563—The ITU-T Standard for Single-Ended Speech Quality Assessment," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 6, pp. 1924–1934 (2006 Nov.). https://doi.org/10.1109/TASL.2006.883177.

[75] N. O. Johannesson, "The ETSI Computation Model: A Tool for Transmission Planning of Telephone Networks," *IEEE Commun. Mag.*, vol. 35, no. 1, pp. 70–79 (1997 Jan.). https://doi.org/10.1109/35.568213.

[76] A. Ragano, E. Benetos, and A. Hines, "More for Less: Non-Intrusive Speech Quality Assessment With Limited Annotations," in *13th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 103–108 (2021 Jun.). https://doi.org/10.1109/QoMEX51781.2021.9465410.

[77] G. Mittag, B. Naderi, A. Chehadi, and S. Möller, "NISQA: A Deep CNN-Self-Attention Model for Multidimensional Speech Quality Prediction With Crowdsourced Datasets," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2127–2131 (Brno, Czech Republic) (2021 Aug.). https://doi.org/10.21437/Interspeech.2021-299.

[78] J. Serrà, J. Pons, and S. Pascual, "SESQA: Semi-Supervised Learning for Speech Quality Assessment," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 381–385 (Toronto, Canada) (2021 Jun.). https://doi.org/10.1109/ICASSP39728.2021.9414052.

[79] A. R. Avila, H. Gamper, C. Reddy, et al., "Non-Intrusive Speech Quality Assessment Using Neural Networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 631–635 (Brighton, UK) (2019 May). https://doi.org/10.1109/ICASSP.2019.8683175.

[80] K. Kilgour, M. Zuluaga, D. Roblek, and M. Sharifi, "Fréchet Audio Distance: A Reference-Free Metric for Evaluating Music Enhancement Algorithms," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2350–2354 (Graz, Austria) (2019 Sep.). https://doi.org/10.21437/Interspeech.2019-2219.

[81] C. Sloan, N. Harte, D. Kelly, A. C. Kokaram, and A. Hines, "Bitrate Classification of Twice-Encoded Audio Using Objective Quality Features," in *Proceedings of the 8th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6 (Lisbon, Portugal) (2016 Jun.).

[82] M. Casey, "From Violence to Violins: Debugging DAT," *Media Digitization & Preservation Initiative Blog: Latest Thoughts From Inside MDPI*, https://blogs.iu.edu/mdpi/2020/05/18/from-violence-to-violins-debugging-dat/ (2020 May).

[83] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 6, pp. 1109–1121 (1984 Dec.). https://doi.org/10.1109/TASSP.1984.1164453.

[84] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445 (1985 Apr.). https://doi.org/10.1109/TASSP.1985.1164550.

[85] Y. Hu and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," *IEEE Trans. Audio Speech Lang. Process.*, vol. 16, no. 1, pp. 229–238 (2008 Jan.). https://doi.org/10.1109/TASL.2007.911054.

[86] S. J. Godsill and P. J. W. Rayner, "A Bayesian Approach to the Restoration of Degraded Audio Signals," *IEEE Speech Audio Process.*, vol. 3, no. 4, pp. 267–278 (1995 Jul.). https://doi.org/10.1109/89.397091.

[87] S. J. Godsill, P. J. Wolfe, and W. N. Fong, "Statistical Model-Based Approaches to Audio Restoration and Analysis," *J. New Music Res.*, vol. 30, no. 4, pp. 323–338 (2001). https://doi.org/10.1076/jnmr.30.4.323.7489.

[88] J. H. L. Hansen, A. Joglekar, M. C. Shekhar, et al., "The 2019 Inaugural Fearless Steps Challenge: A Giant Leap for Naturalistic Audio," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 1851–1855 (Graz, Austria) (2019 Sep.). https://doi.org/10.21437/Interspeech.2019-2301.

[89] J. H. Dabbs and O. L. Schmidt, "Apollo Experience Report: Voice Communications Techniques and Performance," *NASA Technical Note NASA TN D-6739* (1972 Mar.).

[90] L. Kaushik, A. Sangwan, and J. H. L. Hansen, "Multi-Channel Apollo Mission Speech Transcripts Calibration." in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 2799–2803 (Stockholm, Sweden) (2017 Aug.). https://doi.org/10.21437/Interspeech.2017-1778.

[91] F. Bressan, V. Burini, E. Micheloni, et al., "Reading Tapes Backwards: A Legitimate Approach to Saving Time and Money in Digitization Projects?" *Appl. Sci.*, vol. 11, no. 15, paper 7092 (2021 Jul.). https://doi.org/10.3390/app11157092.

[92] A. Ragano, E. Benetos, and A. Hines, "Development of a Speech Quality Database Under Uncontrolled Conditions," in *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 4616–4620 (2020 Oct.). https://doi.org/10.21437/Interspeech.2020-1899.

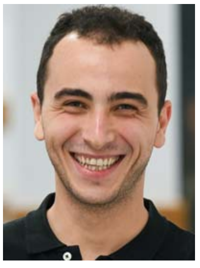
[93] V. Välimäki, S. González, O. Kimmelma, and J. Parviainen, "Digital Audio Antiquing—Signal Processing Methods for Imitating the Sound Quality of Historical Recordings," *J. Audio Eng. Soc.*, vol. 56, no. 3, pp. 115–139 (2008 Mar.).

[94] A. Ziaei, L. Kaushik, A. Sangwan, J. H. L. Hansen, and D. W. Oard, "Speech Activity Detection for NASA Apollo Space Missions: Challenges and Solutions," in *Proceedings of the Annual Con-*

ference of the International Speech Communication Association (INTERSPEECH), pp. 1544–1548 (Singapore) (2014 Sep.). https://doi.org/10.21437/Interspeech.2014-369.

[95] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical Density Based Clustering," *J. Open Source Softw.*, vol. 2, no. 11, paper 205 (2017 Mar.). https://doi.org/10.21105/joss.00205.

THE AUTHORS



Alessandro Ragano

Emmanouil Benetos

Andrew Hines

Alessandro Ragano is a Ph.D. student in the School of Computer Science at University College Dublin. His research focuses on sound quality assessment techniques for audio archives. He obtained an M.Sc. in Computer Science and Engineering from Politecnico di Milano in 2018. His research interests include signal processing and machine learning for speech and audio quality assessment.

•

Emmanouil Benetos is Senior Lecturer with the School of Electronic Engineering and Computer Science, Queen Mary University of London, and Turing Fellow with The Alan Turing Institute. He has authored or co-authored more than 140 peer-reviewed papers spanning several topics in audio and music signal processing. His research focuses on signal processing and machine learning for audio and music signal analysis, applications to music information retrieval, sound scene analysis, and computational musicology.

•

Andrew Hines is an Assistant Professor with the School of Computer Science, University College Dublin, Ireland, where he leads the QxLab research group. He is a member of the Audio Engineering Society (AES) and committee member of the AES Ireland Section. He was facilities chair for the 146th AES Convention in Dublin, Ireland. He is a senior member of the IEEE.