Adapting the Quality of Experience Framework for Audio Archive Evaluation

Alessandro Ragano[1,2], Emmanouil Benetos[3,4], and Andrew Hines[1,2]
[1] *School of Computer Science, University College Dublin, Ireland* [2] *The Insight Centre for Data Analytics, Ireland*
[3] *School of EECS, Queen Mary University of London, UK* [4] *The Alan Turing Institute, UK*
alessandro.ragano@ucdconnect.ie, emmanouil.benetos@qmul.ac.uk, andrew.hines@ucd.ie

***Abstract*—Perceived quality of historical audio material that is subjected to digitisation and restoration is typically evaluated by individual judgements or with inappropriate objective quality models. This paper presents a Quality of Experience (QoE) framework for predicting perceived audio quality of sound archives. The approach consists in adapting concepts used in QoE evaluation to digital audio archives. Limitations of current objective quality models employed in audio archives are provided and reasons why a QoE-based framework can overcome these limitations are discussed. This paper shows that applying a QoE framework to audio archives is feasible and it helps to identify the stages, stakeholders and models for a QoE centric approach.**

***Index Terms*—Quality of Experience, Digital Audio Archives, Digital Audio Restoration**

I. Introduction

Organizations that collect sound recordings are digitising their material because the analog storage medium where sound has been recorded, referred to as carriers, degrade over time [1]. It is estimated that organizations have only left 15 to 20 years before carriers irreversibly degrade [1]. The urgency is compounded by the quantity of material yet to be digitised, as the British Library alone keeps more than 6.5 million of sound recordings [2]. Unlike an analog carrier, a digital representation can be preserved and cloned without degrading. When a physical carrier material is highly damaged, organizations may decide to restore sound quality by applying Digital Signal Processing (DSP) techniques for improving the listening experience. One crucial aspect that is poorly investigated in this domain concerns assessing the perceived quality of audio documents that have been subjected to digitisation and/or restoration. Audio quality metrics that are typically used in applications such as perceptual audio coding [3] and speech enhancement [4] are not appropriate for archive material for a variety of reasons: (1) different quality expectations of the stakeholders, (2) a reference signal cannot be defined, (3) heterogeneity of audio archives (e.g., multiple carriers and content types), (4) they were developed for different audio impairments. Indeed, most organizations rely on personal and individual judgements when evaluating digitised/restored audio documents rather than using audio quality metrics. The main problem with this approach is that mistakes can occur in the process and quality assessment is not systematic. Each archive audio document has a unique sonic fingerprint and evaluating all of them manually is not feasible.

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289. EB is supported by RAEng Research Fellowship RF/128 and a Turing Fellowship. doi: 10.1109/QoMEX.2019.8743302

In this paper we discuss reasons how the Quality of Experience (QoE) [5] framework can be adapted for assessing perceived audio quality in archive material. We explore the potential of a QoE-based framework compared to the current audio assessment methodologies employed when digitising/restoring archive material. Evaluating QoE concerns assessing different aspects of an application and in this context we only refer to audio QoE. Applying a QoE framework to fields beyond telecommunications and traditional multimedia applications has been shown to be a successful approach, for example to quality of life [6] and gaming [7]. To the authors' knowledge, no prior research has addressed audio archives from a QoE perspective.

II. QoE in the Audio Archive Lifecycle

A white paper published by the QUALINET EU COST network [5] defines QoE 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". The key characteristic of QoE consists of being user-centric instead of system-centric. Indeed, quality is considered as "the outcome of an individual's comparison and judgement process" [5] between perceived quality features and desired quality features. In order to understand both desired and perceived quality features of individuals, key *Influence Factors* (IFs) need to be defined, particularly human IFs, context IFs and system IFs.

The status of digital audio files depends on the processing stages involved in the archive lifecycle. Different stakeholders are involved in providing audio archive services or using the service, e.g. to preserve material, to provide access to audio documents and to consume the service outputs. As a consequence, several QoE models can be defined by means of IFs and quality expectations. Particularly, we identify three stages in the archive lifecycle from a QoE perspective (Figure

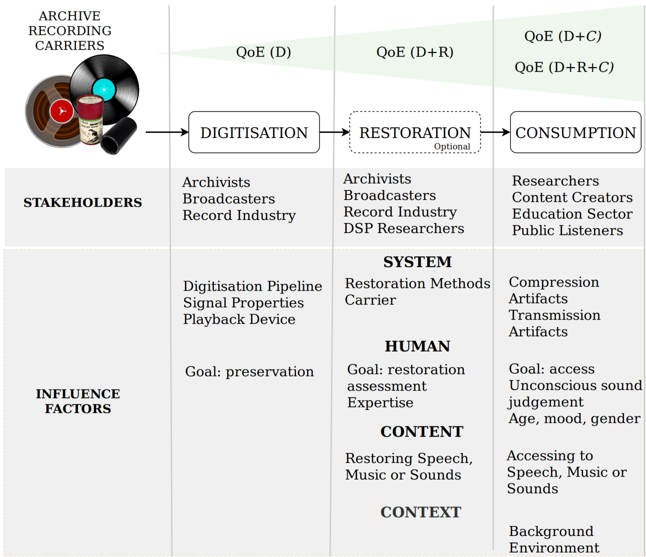

Fig. 1. Audio archive lifecycle. 3 stages are represented: (D)igitisation, (R)estoration and (C)onsumption. Each stage is affected by different Influence Factors and a different QoE model can be developed. Each QoE model takes into account the previous stages (D, D+R, D+C, D+R+C).

1): digitisation, restoration and consumption. Digitisation is generally conducted by archivists, broadcasters and the record industry as they keep most of the archive material. QoE of digitisation focuses on system IFs and on the preservation goal i.e., reproducing digital copies as close as possible to the analog versions by following digitisation guidelines [8]. The same organizations generally make restoration decisions. Restoration is needed when material is objectively highly damaged (archivists) or when a different listening experience wants to be provided (record industry, broadcasters). Restoration is generally conducted by skilled and experienced people that are aware of the desired expectations in terms of audio quality (human IF). This stage also concerns DSP researchers that develop new DSP-based restoration techniques and would benefit from audio quality metrics that assess digital audio restoration methods. The quality perspective of this stakeholder is also affected by content IFs, considered separated from system IFs given that content influences quality expectations. For instance, evaluating restored speech is not the same as evaluating restored music. The last stage concerns consumption i.e., accessing the material. QoE depends on the target users we have in different scenarios. Researchers from different fields access audio archives for conducting experiments. In certain scenarios a restoration could invalidate a research hypothesis while if the same content is provided to public listeners for entertainment purposes, they would desire enhanced audio QoE. It must be noted that each stage also depends on the previous ones. Indeed, public listeners benefit from having restored material that should be in line with their expectations. Therefore, assessing their QoE means also considering how this material has been restored and if degradations of analog carriers are treated in a proper way. The consumption stage also requires the definition of content IFs as inherent characteristics that affects quality expectations. For instance, given two historic musical works where the first has a louder background noise than the second, it does not mean that the latter gives higher QoE since the content of the former is preferred. Typically, a digital restoration method that is considered as technically 100% successful from an engineering perspective does not imply the same QoE for all stakeholders.

III. QoE Evaluation

Audio QoE is typically evaluated with subjective and objective methods. Subjective methods include listening tests where participants rate the quality for a variety of stimuli. Objective methods are computer-based measures that emulate this process to predict how audio quality for a given stimulus would be rated. Listening tests are time consuming and expensive and are an impractical approach considering the quantity of archived recordings. Objective methods can be classified in signal-based models and parametric models. Signal-based models output scores using the signal and they can be divided in full-reference models (a noisy signal is compared with a clean desired signal) and reference-free models (quality is estimated based only on evaluation of the noisy signal). Parametric models employ technical specifications of the processing systems without relying on the signal. We believe that different quality models may be better suited to each stage of the archive lifecycle. Parametric models suit digitisation since guidelines employed by stakeholders provide technical specifications which can be monitored for guaranteeing high quality digital copies. Conversely, restoration and consumption will need to rely on reference-free models. Indeed, defining a reference signal in audio archives is not feasible. Even if we assume that the desired signal is represented by the audio that was intentionally recorded while everything else represents the undesired signal, we are not given this separation in real scenarios. A solution that is typically used in this case consists in artificially creating noisy signals by simulating carrier degradations. This solution does not guarantee a reliable evaluation since real archives cannot be assessed. Also, defining a desired signal in archive material is ambiguous. For instance, some musicologists desire a faithful reproduction of the musical work that must include degradations of the old carriers. Noise-like components do not always have to be treated as undesired components since this depends on the stakeholder. In this case we believe that a data-driven approach can be developed for predicting quality. Ground truth quality labels can be collected allowing participants to rate scores through auditory experiments. This approach, displayed in Fig. 2, has the following steps: (1) identify the stakeholder, (2) define and conduct listening tests to obtain reliable ground truth, (3) identify the archive material in line with the stakeholder, (4) define the procedure for obtaining IFs and quality features, (5) evaluate machine learning models for regression tasks. Under certain circumstances, digital audio restoration can be also assessed using full-reference models.

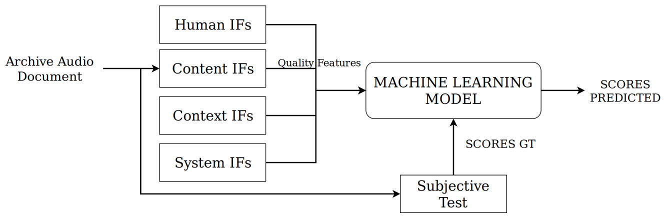

Fig. 2. Data-driven QoE prediction. Human, context and system IFs are given as a metadata and are collected through questionnaires. Ground Truth (GT) is obtained labelling audio with quality ratings.

Indeed, even if a clean signal is not available, a comparison between the signal before and after being restored can be made by means of useful information extracted from the pre-restored signal.

IV. Assessing Restoration

In Table I we list an overview of audio recordings degradations, summarized from [9]. Restoring material means removing these degradations without altering the main signal in order to improve listening experience. Typically sound engineers use tools in digital audio workstations such as declick, dehiss and dehum. Those tools are developed by DSP engineers. Several problems are encountered when evaluating restoration algorithms. In the best case scenario, audio quality metrics developed for other applications such as speech enhancement [4] and audio coding [3] are employed. Those audio quality metrics are unfeasible for audio archives since they are full-reference. In the worst case scenario, papers published during the 1990s [9] report results only through informal listening tests using a single listener or one example of reconstructed signal that does not provide a reliable assessment. Finally, unlike digitisation, no standards are used for restoration. As reported in [8], “audio practice executes the rule: Do as you like” when comes to restoring material for public. Therefore, audio quality metrics that are able to address each degradation accurately can bring benefit to both DSP researchers and sound engineers. DSP researchers can validate outcomes in a more rigorous way, instead of using inappropriate evaluation models and artificial data. Restoration is not only used for repairing audio from degradations. It can be also employed for overcoming limitations of analog carriers. Dynamic range and frequency range of grooved material such as wax cylinder and shellac discs is very poor compared to the today’s formats. Several bandwidth extension methods such as audio inpainting [10] and audio resolution [11] have been developed for enhancing the listening experience but they are not evaluated on old recordings. Those methods need to be evaluated using reference-free models as no reference is available.

TABLE I
Overview of degradations in audio recordings

	Carrier	Perceived as	Cause
Clicks	Grooved material, optical discs, digital files	Ticks, scratches, 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	Grooved material, magnetic tapes	Overall pitch variation	RSV, eccentricity, irregular stretches during playback/recording/storage
Clip	Grooved material, optical discs, digital files	Distortion	Groove deformation, exceeding maximum range

V. Conclusions

This paper describes the importance of evaluating quality in digital audio archives and proposes a QoE framework approach to quality assessment. We show that predicting quality for audio archives will often require a data-driven reference-free approach. In particular when audio documents are restored no standards are employed and audio quality can be easily unsatisfactory. We discussed issues related to the sonic fingerprints of archive material and how a human-based assessment brings to unsatisfactory QoE. No prior research has applied a QoE framework approach to audio archive quality. We have illustrated that it can be applied to audio archives to identify the stages, stakeholders and models for a QoE centric approach.

References

[1] M. Casey, “Why Media Preservation Can’t Wait: the Gathering Storm,” *IASA Journal*, vol. 44, pp. 14–22, 2015.

[2] A. Tovell, “Audio Preservation and Access: Overcoming the Challenges,” in *Audio Engineering Society Conference: 2018 AES International Conference on Audio Archiving, Preservation & Restoration*, 2018.

[3] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerends, and C. Colomes, “PEAQ-The ITU standard for objective measurement of perceived audio quality,” *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000.

[4] 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 *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752, 2001.

[5] European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), “Qualinet white paper on definitions of quality of experience,” 2013.

[6] A. Hines and J. D. Kelleher, “A framework for post-stroke quality of life prediction using structured prediction,” *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017*, 2017.

[7] S. Möller, D. Pommer, J. Beyer, and J. Rake-Revelant, “Factors Influencing Gaming QoE: Lessons Learned from the Evaluation of Cloud Gaming Services,” in *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013)*, pp. 163–167, 2013.

[8] N. Wallaszkovits, “Between standards and arts: Digitisation and restoration of audio material–a balancing act between authenticity and manipulation?,” *Journal of New Music Research*, vol. 47, no. 4, pp. 285–290, 2018.

[9] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*. Springer-Verlag London, 1 ed., 1998.

[10] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval, and M. D. Plumbley, “Audio inpainting,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 3, pp. 922–932, 2012.

[11] V. Kuleshov, S. Z. Enam, and S. Ermon, “Audio Super Resolution using Neural Nets,” *International Conference on Learning Representation (ICLR)*, 2017.