STUDY OF THE PERCEPTUAL AND SEMANTIC DIVERGENCE OF DIGITAL AUDIO PROCESSED BY RESTORATION ALGORITHMS

Sonia Cenceschi
Polytechnic University of Milan
cenceschi@perizieaudiocencor.it

Giorgio Klauer
Music Conservatory of Padua
giorgio.klauer@conservatoriopollini.it

ABSTRACT

This research integrates sensory and scientific instruments to analyze the relationship between subjective evaluations of digitally restored audio and its computer extracted perceptual descriptors. Statistical methods have been used to compare the displacement of three types of remediated content in subspaces obtained by data expressed both by individuals and by feature extraction algorithms.

Qualitative demands in audio restoration are tightly connected to the information embedded in remediated content: it is crucial the awareness that every choice is re-balancing it and affecting its reception. Listeners in their turn don't do an acousmatic reduction of auditory information but recode it interleaving contextual and aesthetic approaches, according to their sensitivity and being influenced by their cultural background.

Thanks to the analysis of the displacement in subspaces related to the descriptive characteristics with greater variability, the semantic divergence resulting from the operations of improving the quality of sound was interpreted and a predictive model aimed at their optimization was assumed.

1. RESTORATION BETWEEN ARCHIVE, CONSERVATION AND REMEDIATION

The practice of audio restoration grounds on a solid methodological basis derived mostly from cultural heritage maintenance. Audio restoration concerns firstly the production of a conservative digitized copy and the implicated technical and cultural issues. In this view its object is the overall audio content, which embodies several layers of information about the mediated content, the acoustic context, the techniques of the recording and the history of the carrier where the audio is stored in. Guidelines on how to handle the digitization procedure are issued mostly from archive science [1, 2] and should be as neutral as possible in respect to its intrinsic complexity.

The digitization of degraded or degradable audio documents is surrounded by a whole set of opportunities related to dissemination, once the digitized medium becomes the subject of conservation of cultural heritage. In this context the study of the complexity of information regressively influence the strategy of restoration, leading to categories of approach like documental, recovery, sociological, and aesthetic, which scruple to lose the information which are respectively considered as ancillary [3, 4, 5, 6].

The practices of audio restoration have been finally considered in the view of the theory of mediation [7]. Referring to the concept of remediation as the representation of one medium in another [8], restoration techniques may be conceived as the key to unleash the potential of audio as a medium, in a framework where the conservation of cultural heritage has to face with increasing functional demands [1].

1.1 Re-evaluating the Aesthetic Content

Despite maintaining substantial theoretical similarities with restoration in general, the role of auditory perception, coupled with the substitution of a tangible object with the digitized copy, makes audio restoration very prone to subjective assessments: if on the one hand restoration aims at recovering and enhancing the medium by acting on certain features, on the other hand this does not necessarily imply optimal improvements in terms of psychology of auditory perception.

Different restoration approaches fear not only the loss of information, but even the mismatch between purposes and aesthetic judgement. This is implicit in the use of restoration algorithms in order to emphasize the qualities of the medium, which have the collateral effect of projecting it in a listening dimension which is divorced from history.

From the point of view of audio restoration, one may argue that the applying of techniques should anyway take care of aesthetic considerations [2]. Rather, the categories of modern audio restoration may benefit from the correlation of computational analyses of the audio medium both from an object-oriented and a listener-centered approach.

The need to determine aesthetic appreciation preventing this to remain an issue of hermeneutics led audio restoration being finally addressed with object-oriented methods of systematic musicology which tries to explain musical

[1] It should be pointed out that a rigid division between a truly archival perspective and several functional, or even commercial approaches in audio restoration is somehow unrealistic: the common practice admits the intervention on local as well as global flaws as later described, albeit in different proportions, and every approach benefits from the same research results in digital signal processing. Recent developments in the creation of conservative copies through optical digitization [9, 10, 11], once regularly implemented, may help revealing goals and needs of the different individuals and groups making decision rules.

[2] This question should not be confused, whether the specialized restorer may be an artist besides a technical expert or a scientist, which is not a debate among scholars and practitioners.

information and its processing by humans, formalizing computational paradigms.

Research in music information retrieval oriented to the analysis of digitally processed audio is very robust. It largely relies on theories and techniques of numerical and statistical analysis and is easily expandable to different genres, virtually covering any audio document.

When related to remediated content, music information retrieval is mainly aimed at cataloguing, or indexing the emotional potential of music. Literature addressing the correlation of the specific processing of restoration with subjective assessments is conversely sparse. If the validation of the techniques of noise reduction lies on a solid experimental basis, the study of the resulting perceptual qualities remains limited to liking [12].

This may be due to the segregation of competencies between scientists and the actors of audio restoration, though notable exceptions are known. Moreover, a predictive model as regards the subjective evaluation of audio restoration is not yet felt as urgent. Furthermore, there is no shared view about what subjective assessments should be categorized.

Believing that audio restoration may benefit from research in qualitative issues, even retrospectively, an exploratory study is proposed.

The ultimate purposes of the study, which would deserve more in-depth case studies on singular issues, are firstly to detect the possible correlation between global operations in digital audio restoration and relevant global effects in auditory perception; secondly, to understand which set of qualitative subjective assessments about the audible content and which set of low-level perceptual features for the computational description may be pertinent; third, if the implementation of multivariate statistics may be the lattice for linking subjective and objective domain in order to single out factors characterizing possible audio restoration strategies from a qualitative point of view.

2. SOME TECHNIQUES IN DIGITAL AUDIO RESTORATION

Particular attention in audio restoration is devoted to the historical context and the technological acquisition and production processes. It is central to discern fingerprints of the original acoustic environment from those due to technical choices and mechanical alterations occurred over time. The purpose is to act on intentional factors firstly, such as ancient and non-neutral noise reduction techniques, and on unintentional factors secondly, caused by technical deficiencies or objective lack of capabilities by humans. An additional level is carrier-dependent and related to its physical degradation: considering that the A/D conversion is the main factor paying in the degradation of the content, the conservative approach tends to operate on the so-called 'original' materials.

2.1 Local Flaws

Sound defects are divided into two categories, temporally localized and affecting the document in its entirety [13, 14]. Localized flaws are impulsive noises (clicks) extended over the entire audible spectrum caused by mechanical agents such as dust, scratches, cuts, or electronic like digital syncloss. Increasing in importance these flaws are renamed scratches, crackle or pop. Models for their removal are the additive ones, suitable for defects of minimal duration and consisting of the local removal of samples and adjacent interpolation; the replacement methods, which span some samples more and implement a reconstructive process. The correction of local flaws is performed before acting on global flaws.

2.2 Global Flaws

Global flaws are divided into background noise, pitch variations and non-linear distortions. The first category is a constant problem being related to any device and has various and complex etiology: the acoustic environment of acquisition, physical degradation, noise of electric motors, and device electrical circuitry. The result is often a stationary noise spread over the entire audible spectrum, so particular subtractive techniques must be applied, considering the phenomenon of masking so as to minimize the spectral area of intervention.

Pitch variations can occur in a long-term form (wow) or a short-term one (flutter). Wow is caused e.g. by non-perfectly concentric grooves engraved into the disk, and relative speed changes due to eccentricity. In this case dynamic resampling is applied, after modelization of the mechanical process generating the error. Non-linear distortions of the signal are caused by more factors and are particularly difficult to correct. Examples are saturation (clipping), harmonic distortion, aliasing, and quantization error. The only way to mitigate these effects is reverse engineering.

3. DOCUMENTS

The mediated contents chosen in this study belong to different categories: mixed vocal, instrumental and electronic music. Criteria were the belonging to a specific set of rare editions of the early twentieth century; a medium degree of physical degradation; a moderate presence of ancillary information; the heterogeneity of emotive potential; the heterogeneity of sound capturing method bound to the musical performance.

For the experimentation one excerpt lasting about one minute was taken from three single-channel digitized media:

guit_01 Solo instrumental music. Medium: digitization of "Goodbye Juna" played by guitarist Cosimo di Ceglie, Columbia (CQ1530)

tale_01 Musical tale - italian spoken voice alternating with instrumental ensemble. Medium: digitization of "La rana ambiziosa" (The Ambitious Frog), Durium - The voice of the empire (Milan 1936-37, AC 583 P007)

elec_01 Analog electroacoustic music. Medium: digitization of a magnetic tape containing the mix of a 40-minute improvisation by the group ArkeSinth (G. De Poli,

M. Sambin, A. Vidolin), unpublished, DEI/CSC, University of Padua, 1972

The choice of media *elec_01* here described has therefore additional counter-proof purposes [3] .

Two different restoration interventions were done using a well-known professional software implementing high-quality algorithms according to audio restoration best practice [4] . The first step for restoration starting from media *tale_01* and *guit_01* was:

_02/03 impulsive noise removal (declipper, threshold -2.8 dB) [5]

_02/03 analog recordings periodic impulsive noise removal (multiband declicker) [6]

Following algorithms were performed:

_02 background noise removal resulting in -4 dB noise level (1.5 sec learning time, 100 ms window, 6 iterations) [7]

_03 background noise removal resulting in -6.5 dB noise level (manual selection of 2 sec background noise, 50 ms window, 10 iterations) [8]

Media *_02* and *_03* resulted with an increase of the tonal signal of +2 dB. For medium *tale_03* an additional tonal gain was performed [9] .

The final step was the equalization performed by listening, looking for a clearer sound. In media *_02* this step was barely audible, whereas in media *_03* this resulted in an increase of brilliance.

Background noise removal in media *elec_02* and *elec_03* was instead very light due to the quality of the recording (manual selection of background noise, 3 iterations in *elec_02* and 1 in *elec_03*). In the equalization of media *elec_* the low frequency range was slightly boosted (+2 dB); in medium *elec_03* the mid range was boosted too.

No final normalization and further processing were performed.

Media *_02* should represent a moderate restoration with a slight enhancement of intrinsic audio qualities, whereas media *_03* represent a less accurate restoration for an average level of attention. The procedure highlighted very well the different features of the media and made versions *_01*, *_02* and *_03* easily distinguishable.

One more version *_04* was done averaging the parameters used in *_02* and *_03* and finalizing with a minimal equalization boosting the mid-frequency range. Media *_04* had the only purpose of the listeners getting accustomed at the beginning of the auditory test, as later described.

4. PROCEDURE

The characterization of the media subjected to audio restoration algorithms was investigated by confronting the maps obtained through multivariate statistical procedures applied to two sets of data got from direct subjective assessments and from computer extracted perceptual descriptors, respectively.

The subjective assessments expressed by a population of listeners have been predefined as concepts related to auditory sensations arising from the audio features themselves and from possible implicit ancillary information.

The audio was analyzed by extracting spectral-based low level perceptual features according to music information retrieval best practice.

A classical metric multidimensional scaling was performed on the similarity matrices derived from both fields of data. This allowed to interpret the displacements of the media as effects of restoration both in respect to subjective judgement and their audible content as an object.

This exploratory study concerned three types of mediated content and three levels of restoration - thus 9 objects - considering this as an appropriate set compared to the size of the interviewed population, the size of the descriptors and the use of classic metric MDS.

4.1 Subjective Assessments

4.1.1 Population

Where the semantics of the mediated audio content plays a fair degree of importance, the reliability of the verbal behaviour in the interviewed population becomes indispensable for the purpose of qualitative assessment accuracy [15]. For this reason, rather than a faithful representation of the universe of listeners, the population descended by non-probabilistic choices based on explicit criteria of the researchers.

The population was defined as a set of experienced musicians in the field of electroacoustic music and sound engineering. The population was equally divided between undergraduate students, graduate students and professionals/academics and consisted of 41 subjects with an age between 30 and 40 years [10] .

4.1.2 Questionnaire

The questionnaire concerned ten complex concepts related to: terms and metaphors used in sound engineering; metaphors of groupings of perceptual features; terms emerging as issues in audio restoration.

[3] No DSP was applied during or after digitization. Media were digitized using dedicated equipment consisting of a restored CESA gramophone (1934), a Nordmende preamplifier and a Tascam 24 bit converter at a 24 kHz sampling rate. Medium *elec_01* was already digitized by the "Sound and Music Computing Lab" at the University of Padua with professional equipment.

[4] The audio restoration processing was done using the software iZotope RX 3. Algorithms are described at `http://help.izotope.com/docs/izotope-rx3-help.pdf`

[5] This algorithm repairs both digital and analog clipping artifacts resulting when A/D converters are pushed too hard or magnetic tape is over-saturated.

[6] Multiband declicker is an impulsive noise removal RX command. It is suited to remove noises due to imperfections in the rotating support.

[7] The learning time option in RX denoiser is an adaptative model which allows the denoise module to adjust its noise profile based on changes in the incoming audio windowing. This option works better when the noise is constantly evolving. The algorithm was applied to the full spectrum.

[8] Noise was reselected at each iteration.

[9] The de-construct RX command is specifically designed for the tonal and noise audio adjustments. RX recognizes the harmonic part of a portion of audio with a predetermined sensitivity, and allows the user to act on the gain of this or of the remaining noisy part.

[10] The subjects spoke italian language as their mother tongue.

These concepts were chosen as independent from each other as possible.

Though assessments change according to the set of operations performed in restoring, since the purpose was enhancing the quality of the mediated content some of them became in fact interdependent.

The subjects gave a numerical rating (0-10), being provided with a verbal description of the range as follows.

- Depth: sonic perspective, from flat (0) to immersive (10)
- Brightness: from dark (0) to overexposed (10)
- Softness: from gross/rigid (0) to harmonious and smooth (10)
- Historical context, from divorced from history (0) to totally coincident with the historical context (10)
- Figures and lead lines: from totally texture-integrated (0) to totally background-detached (10)
- Assumed performance environment: from very small (0) to very large (10)
- Sharpness of detail: from confused (0) to excessively rich in detail (10)
- Bandwidth: from closed and narrow (0) to open and wide (10)
- Background cleanliness: from dirt and unintelligible (0) to focused with no background at all (10)
- Liking: from no liking (0) to complete liking (10)

4.1.3 Data Collection

Each media *_01*, *_02* and *_03* was presented 10 times during a test lasting 90-120 minutes, allowing the subjects to do one different assessment at a time. During the test the media were presented in random order preventing consecutive runs of the same audio. Also the assessments were prescribed in random order.

The test was introduced by the media *_04*, for which no assessment was required. Pauses during the test were allowed. For the high reliability of the subjects, ear fatigue was considered as a uniform error.

The test was done onto an open source CMS accessible through http. The subjects listened to the media through their private studio equipment, with the recommendation of using professional closed monitoring headphones.

Media were encoded in mp3 format (384 kbps, LAME). Due to bandwidth limitations, compression was chosen after verifying the minimal difference between perceptual descriptors extracted from uncompressed and compressed audio and observing quite no difference in MDS displacements. Mp3 encoding was therefore considered a negligible factor as the specific listener's equipment.

The collected data were then arranged and averaged, getting ten mean scores for each medium.

4.2 Perceptual Features

For the computerized analysis of audio content the following FFT-based perceptual features were extracted. The set comprises generic descriptors used in music information retrieval with a stress on restoration-related features [11] .

This had also the purpose of verifying the descriptors against the three types of mediated content in the specific context of restoration, in the view of a research on this particular issue.

- Mel-frequency cepstral coefficients (13 values)
- Chromagram (mean of 12 values)
- Pitch (Hz) [16]
- Confidence in pitch estimation (0-1)
- Roughness [17]
- Brightness (50. spectral percentile, Hz)
- Loudness (sone) [18]
- Spectral flatness [12]
- Spectral ripple (three values: total bandwidth, 0-2kHz, 2-10kHz) [13]
- Spectral variance [14]
- Spectral curve [15]
- Spectral roll-off (80. and 95. spectral percentile, Hz)
- Attack proximity [16]

The data were globally normalized according to the quantity expressed by each listed descriptor, and averaged in time (arithmetic mean) to obtain a set of 27 values for each medium.

The purpose of comparing computer-extracted perceptual features with the mean subjective assessments led to averaging them in time, because of their highly homogeneous behaviour. Considering that the subjects expressed an overall assessment for each medium, the averaging of the features' course was regarded as an approximation of the coding done by mind by the listeners.

4.3 MDS Comparison

Though multivariate statistical procedures are widely implemented in the field of music informatics and notably in music information retrieval, using multidimensional scaling as a primary tool for creating maps in the specific context of audio restoration is not so widespread.

MDS is known as a method of representing subjective assessments in respect to synthesized and processed musical sounds since a long time [20] and is now widely used in music information retrieval to correlate perceptual features and qualitative phenomena, notably emotional content [21].

Choosing an appropriate model for the purpose of this study, the mutual interdependence of perceptual features and qualitative assessments was considered as a good premise to visualize the structure of the relationships between objective and subjective world, implementing among the several existing MDS algorithms the simple classic metric [22].

[11] The algorithms used are developed by the community supporting the open-source scsynth software. The literature which the algorithms are based on is cited in the SuperCollider class documentation (MFCC, Chromagram, Tartini, SensoryDissonance, Brightness, SpecCentroid, Loudness, SpecFlatness, FFTCrest, FFTSpread, FFTSlope, SpecPcile, Onsets). The algorithms were performed with standard settings and window size of 2048 samples.

[12] Ratio between power spectrum geometric mean and its arithmetic mean.

[13] Ratio between the peak of the power spectrum and a list of squared magnitudes in the spectral band.

[14] Power-weighted deviation with respect to the centroid.

[15] Slope of the linear correlation line derived from the spectral magnitudes.

[16] Low-pass filtered onset detection. Onset detection algorithm in [19].

MDS was applied from both the similarity matrices computing the euclidean distance of the 9 media in the 10-dimensional subjective assessment space and in the 27-dimensional perceptual feature space in order to single out possible matches in the displacements caused by the restoration processes.

The significance of the dimensions of these maps was deduced by analyzing the variance of the data from which the similarity matrices were obtained. By interpreting the meaningfulness of the dimensions and correlating the media displacements in the subjective and the objective world, some hypotheses were drawn about audio restoration conveniences.

5. RESULTS

Through the application of classic MDS to subjective and objective descriptors, two separate subspaces were obtained. In the subjective subspace the first eigenvalue reflected the 80% of total variance while the second was already down to 13%. In the objective subspace the first eigenvalue reflected 65% of variance and the second 14%. So MDS gave two main eigenvalues for both the sets of data.

Some residual variance is reflected by eigenvectors 3 and 4 in the objective subspace. This value is however very low and the two-dimensional representation adequately describes the data and allows their interpretation [17] .

As regards the map obtained from perceptual features (Figure 1), spectral curve, loudness, spectral ripple in the low frequency range and confidence in pitch estimation contribute to the most variance reflected by the first eigenvalue. The estimated pitch is instead of the least importance. The first dimension was called *tone*.

The variance reflected by the second eigenvalue is bound to the spectral ripple particularly in the high frequency range, the highest percentile and onset detection. The second dimension was defined as *clarity*.

Analyzing the map in Figure 2, background cleanliness and the sharpness of detail are the most correlated determinations in the first dimension, which has been named *neatness*. Confronting the displacements of the media in the two maps, quality seems to decrease with lower loudness, emphasized spectral curve and increase of spectral ripple in the low-mid range.

The second dimension is determined particularly by subjective assessments on historical context, brightness, and overall liking. Correlated with features such as attack proximity, spectral roll-off and spectral ripple in the high frequency range, gloomy media seem to be associated with the idea of old-fashioned and result in a lower liking, whereas brighter media appear to please especially if they are not historically determined. The second dimension has been named *abstraction*.

[17] A metric MDS algorithm was also applied, showing an acceptable stress for a two-dimensional representation as well (stress in the subjective subspace was much lower). The maps of metric MDS confirmed tha considerations here given, but the second dimension of the objective subspace was more difficult to interpret and to confront with that of the subjective subspace. The use of metric MDS seemed to make the balance of subjective assessments with perceptual features more urgent.

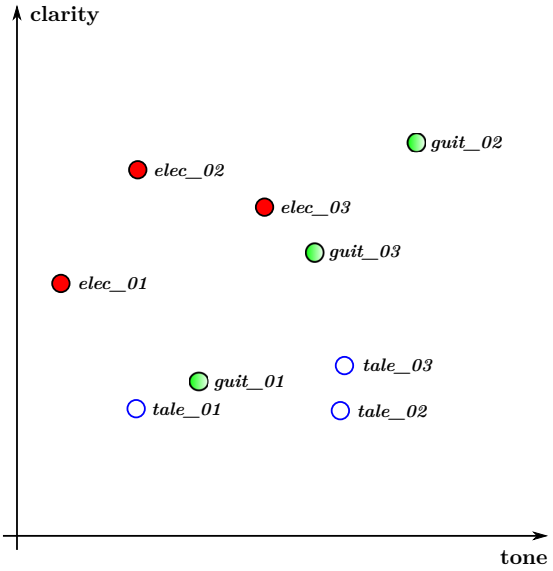

Figure 1. Map obtained from perceptual features.

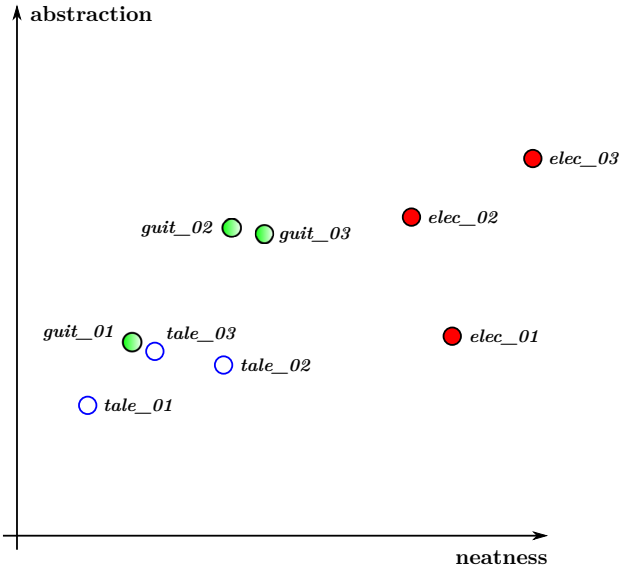

Figure 2. Map obtained from subjective assessments.

6. DISCUSSION

The maps obtained from the subjective and objective sets of data are comparable. Both maps show a clear division by type though medium *guit_01* is located among the media *tale_*, and though the media *guit_* are more spread in the subjective subspace.

The restoration seems to cause a spread of media *_02* and *_03* in respect to media *_01*; the orientation is globally the same in the two dimensions, almost without exceptions.

The position of medium *guit_01* close to the media *tale_* is justified by the same type of physical degradation. Conversely, the group of electronic music media are clearly separated and characterized by a greater degree of *neatness* and *abstraction* (subjective subspace), and *tone* and *clarity* (objective subspace).

In detail, the spread of media *_02* and *_03* shows divergences for each mediated content although almost the same

restoration algorithms were applied.

In the case of the media *guit_* restorations seem to equally satisfy the listener's expectation, though in the objective subspace they are clearly separated. Conversely, media *tale_02* and *tale_03* as an object are not so different as in the listener's subjective point of view.

The position of *guit_02* and *guit_03* in the subjective subspace may point out that particularly in instrumental music the level of intervention should be moderate.

tale_03 is very close to *tale_02* and moved slightly upwards along the dimension *abstraction*. The reason lies in the greater liking by listeners, which contributes to the variance and results particularly high for this medium. The increasing of the tonal signal in *tale_03* had no significant result for the listeners; it is rather the brilliant equalization that, enhancing the intelligibility of the speech, concurs to remove the medium from its historical connotation.

In the case of *elec_* media, restored versions are considered as 'better' than the original, even though the electronic music excerpt is regarded as having much higher quality than the others. The appreciation of electronic music is certainly not assimilable to the traditional and instrumental [23]. The better displacement of *elec_* in both subjective and objective subspaces may point out that electronic music media require specific digital restoration procedures, for which several case studies exist [24], and maybe that the media were over-processed and therefore transformed, rather than restored.

Merging the considerations and observing the displacements of all media, one may suppose that a moderate and detailed intervention represents the best compromise between the listener's qualitative demands and the need to affect the mediated contents as lightly as possible. It remains an open question, therefore, what is the applicable limit for the considered content, and whether this limit is related to cultural conditioning.

7. CONCLUSIONS AND FUTURE WORK

The perceptual and semantic divergence of digital audio processed by restoration algorithms is presented here as an issue related to the practice of remediation. Since methodological studies aimed to confront the objective and receptive information of restored media are not numerous, an exploratory study was proposed to examine the displacements of media in maps obtained from subjectively and objectively issued sets of data.

By predefining and confronting direct subjective assessments with computer extracted descriptors, with the help of multivariate analysis, direct relationships between maps have been identified.

The most interesting findings are related to the different way digital restoration algorithms affect the mediated content in respect to subjective evaluation.

Particularly, the divergent displacements of the media show that restoration algorithms should be used with a content-sensitive approach alongside the information-related one (which depends on the context where the remediation takes place). In this respect, the interpretation of the dimensions of the MDS, though closely linked to the sensitivity of the researcher, may help to establish operational paradigms, to be confronted to existing guidelines.

Further study to ensure scientific consistency is implied. Particularly, from a larger set of case studies could finally ensue a set of constraints characterizing operations belonging to specialized users and a general framework concerning a set of 'sensitive' audio restoration principles, finally leading to predictive models, in the view of combining both functional demands and conservative.

Improvements in the procedure here presented will consist primarily in the isolation of the issues related to cultural factors, using several known and unknown spoken/sung languages, several instruments and musical excerpts belonging to mixed instrumental-electroacoustic repertoire.

One more improvement is the identification, through several statistical procedures, of the subjective assessments which could determine a match in the variance expressed by MDS dimensions in both the subjective and objective field.

8. REFERENCES

[1] T. C. IASA, *The safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. International Association of Sound and Audiovisual Archives, 2005.

[2] ——, *Guidelines on the Production and Preservation of Digital Audio Objects*. International Association of Sound and Audiovisual Archives, 2009.

[3] G. Brock-Nannestad, "The objective basis for the production of high quality transfers from pre-1925 sound recordings," *AES Preprint*, vol. 4610, 1997.

[4] ——, "The requestor decides - the fundamental ethical issues when dealing with sound recordings," in *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, 1998, pp. 157–158.

[5] D. Schüller, "The ethics of preservation, restoration, and re-issues of historical sound recordings," *Journal of the Audio Engineering Society*, vol. 39, no. 12, pp. 1014–1016, 1991.

[6] W. Storm, "A proposal for the establishment of international re-recording standards," *Association for Recorded Sound Collections Journal*, vol. 15, no. 2-3, pp. 26–27, 1983.

[7] S. Canazza and M. C. T. Monti, *Ri-mediazione dei documenti sonori*. Udine: Forum, 2006.

[8] J. Bolter and R. Grusin, *Remediation: Understanding New Media*. MIT, 1999.

[9] S. Cavaglieri, O. Johnsen, and F. Bapst, "Optical retrieval and storage of analog sound recordings," in *Proceedings of the AES 20th International Conference*, Budapest, 2001.

[10] V. Fadeyev and C. Haber, "Reconstruction of mechanically recorded sound," Lawrence Berkeley National Laboratory, Tech. Rep., 2003.

[11] S. Canazza, G. Ferrin, and L. Snidaro, "Photos of ghosts: Photos of grooves and holes, supporting tracks separation," in *Proceedings of the XVII Colloquium on Musical Informatics*, Venice, 2008, pp. 171–176.

[12] R. Krolikowski and A. Czyzewski, "Perceptual denoising of audio," in *Ri-mediazione dei documenti sonori*. Forum, 2006, pp. 421–447.

[13] S. Godsill and P. Rayner, *Digital Audio restoration: A Statistical Model Based Approach*. Springer, 1998.

[14] S. Canazza, "Tecniche di filtraggio per il restauro audio: modelli a confronto," in *Ri-mediazione dei documenti sonori*. Forum, 2006, pp. 259–338.

[15] D. Bogdanov, J. Serrà, N. Wack, and P. Herrera, "From low-level to high-level: Comparative study of music similarity measures," in *IEEE International Symposium on Multimedia. Workshop on Advances in Music Information Research*, 1999.

[16] P. McLeod and G. Wyvill, "A smarter way to find pitch," in *Proceedings of the International Computer Music Conference*, Barcelona, 2005, pp. 138–141.

[17] W. Sethares, "Consonance based spectral mappings," *Computer Music Journal*, vol. 22, no. 1, pp. 56–72, 1998.

[18] B. Moore, B. Glasberg, and M. Stone, "Use of a loudness model for hearing aid fitting: A general method for deriving initial fittings for hearing aids with multichannel compression," *British Journal of Audiology*, vol. 33, no. 4, pp. 241–258, 1999.

[19] D. Stowell and M. Plumbley, "Adaptive whitening for improved real-time audio onset detection," in *Proceedings of the International Computer Music Conference*, Copenhagen, 2007.

[20] J. Grey, "An exploration of musical timbre," Ph.D. dissertation, Stanford University, 1975.

[21] P. Saari, T. Eerola, and O. Lartillot, "Generalizability and simplicity as criteria in feature selection: Application to mood classification in music," in *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, no. 6, 2011.

[22] I. Borg and P. Groenen, *Modern Multidimensional Scaling. Theory and Applications*. Springer, 1997.

[23] A. Camci, "Cognitive approach to electronic music: Theoretical and experiment-based perspectives," in *Proceedings of the International Computer Music Conference*, Ljubljana, 2012.

[24] S. Canazza, "Restauro di continuo di bruno maderna: un approccio basato sul filtro di kalman esteso," in *Ri-mediazione dei documenti sonori*. Forum, 2006, pp. 449–457.