Antonio Rodà[*] - Alessandro Russo[*] - Sergio Canazza[*] - Piero Cosi[**]

AUDIO DOCUMENTS RESTORATION AS A DOCUMENTARY SOURCE IN THE LINGUISTIC RESEARCH: COMPARISON OF INSTRUMENTS

Abstract. – In the last years, always more speech audio archives digitized their historic corpora with the aim of preserving them and exploiting the advantages offered by digital signal processing techniques, such as formant analysis, automatic transcription, content-based information retrieval. Many of these applications are less effective when the signal to noise ratio (SNR) decreases, as it often happens in field-recorded linguistic corpora, due to the use of non-professional portable analogic devices and environmental noises. Audio restoration tools can therefore be very useful to increase SNR and improve subsequent analyses. At the same time, restoration algorithms should be carefully chosen and tuned to avoid the distortion of essential characteristics, such as formant position and energy. The paper describes, with examples taken from on field-recorded sound excerpts, several algorithms, able to cover different audio restoration categories. The algorithms were then evaluated by means of an automatic speech recognition tool. Results showed that the noise reduction algorithms can improve the phoneme recognition task carried out with the Kaldi toolkit.

1. Introduction

In the last years, archives and cultural institutions have started to be fully aware of the risks arising from unsuitable handling of sound documents in the remediation process, *e.g.*, during the transfer from analogue to new media for preservation and/or distribution and/or restoration purposes. This is also the case of archives which didn't receive adequate attention so far, in particular in the fields of ethno-musicology and speech (dialect, in particular), often intersected between them. These archives are often collected by the owners, in poor condition, using non-professional carriers. Nevertheless,

[*] CSC-Sound and Music Computing Group - Dept. of Information Engineering - University of Padova.

[**] National Research Council - ISTC - Via Martiri della Libertà, 2 - 35117 Padova.

these archives preserve recordings almost always in a unique copy, such as important documents for different research fields, including ethno-musicology, linguistics, and sociology. The value of these archives is evident by the gradual appreciation of the audio recordings by the international digital libraries. Several research projects have been funded by the European Community in the framework of preservation and digital curation[1] (*e.g.*, IST, Culture2000, Mediaplus, Interreg, eContent, Creative Culture, FP5-6-7), but very few of these have considered speech or ethno-music archives (*e.g.*, European project in Culture 2000 framework POFADEAM: Preservation and On-line Fruition of the Audio Documents from the European Archives of Ethnic Music). In addition, the multidisciplinary research that concerns the preservation of cultural musical heritage – which includes preservation and restoration of audio documents – fits perfectly in the Sound and Music Computing roadmap [2], as defined by the scientific community.

The growing interest in audio documents is motivated by the fact that the sound recordings are characterized by a shorter life expectancy than other cultural heritage goods (measured in years or decades, not centuries such as for sculpture and mosaics) and without concrete actions they are doomed to disappear. The urgency is heightened by the fact that the problems introduced by this type of sound files requires the definition of innovative approaches, specifically designed for the needs of these archives. In ethnic music and linguistic documents, a written notation generally doesn't exist (as in modern music Western tradition): the recordings are the only memory of those repertoires.

For the reason stated above, between the wide variety of audio recordings, speech documents represent a very interesting case study also from the preservative point of view. In addition, the preservation of this material is not limited to long-term preservation strategies, but rather includes all actions aimed to allow and improve the access to sound documents by experts and general users for study, research, reinterpretation, entertainment and more. Of course, there is a different situation in musical repertoires, such as clas-

[1] Digital curation [18] is the process of establishing and developing long term repositories of digital assets for current and future reference by researchers, scientists, and historians, and scholars generally.

sical, rock/pop, or jazz, where the record companies have already taken the necessary measures to protect the audio documents, particularly those of high commercial value.

This paper presents a preservation methodology (Section 2) designed for speech archives, some restoration tools developed by the authors (Section 3). They are used to process some audio files shared by the conference scientific committee (Section 4). Section 5 compares the results of automatic speech recognition carried out before and after the restoration, in order to observe if (and how much) the speech enhancement process has actually improved the phonemes recognition.

2. Preservation Methodology

Different approaches to the preservation of sound documents came to light during the debate carried out by the archival community over the last forty years about remediation process [5]. These approaches distinguish themselves for their different degree of accuracy to the integrity of the original document.

Depending on the approach adopted, the signal processing stages (after the A/D process) will be different. The authors are strongly convinced that a preservation intervention is preliminary to any other action. Regardless of the final use for which the recording is made, the quality of the starting digital material is critical, because is the only thing that can maintain a strict continuity between the original document and all copies made (low-quality extracts for online streaming, medium quality versions for local access, high-quality versions for scholars request, . . .).

This section presents the preservation methodology defined by the authors on the basis of their experience in national and international research projects. In [5] is described in detail a general procedure including all the steps from A/D process to the restoration and the storage of the audio documents. In the following, the authors will discuss this issue, focusing on speech recordings.

When an authoritative version of an original documents disappears – because lost, stolen, accidentally destroyed, or corrupted for physical degradation – it's necessary to find a new document replacing the first as authoritative reference. The master copy of an audio document is exactly designed for this purpose. As defined by International Association of Sound and Audiovisual Archives (IASA), the master (or archival) copy is “the artefact designated to be stored and maintained as the preservation master”. Such a designation means that the item is used only under Exceptional Circumstances” [13]. Its purpose is to preserve the documentary unit and its bibliographical equivalents are the facsimile and the diplomatic copy. The restoration is permitted only for the repair/optimization of the physical carrier. The documentation accompanying the preservation master is defined by Schüller [17], who specifies that all the compensation and the processing applied during the remediation process must be “based on the capacity for precise counteraction” (which implies the reversibility of each operation and, consequently, the ability to go back to the original characteristics). For the master copy (logical) structure, see [5].

The actions starting from the assessment of the document condition to the time when the document is ready to be re-archived are included on the remediation process (see [5]). In general, a restoration laboratory can process multiple documents at once, although some activities require careful and constant control by the operator: in particular, the monitoring of the signal transfer aims to document any corruptions and other relevant events (related to the audio signal and not to the content). In the case in which the operator knows the (presumed) content of the recording, s/he can discern more clearly the noise produced by the recording process from non-intentional alterations resulting from the equipment of reproduction. The determination of the causes of a local corruption (clicks, pops, crackles, bumps) is not a trivial task. And the fact that the artifact is to be attributed to the playing equipment implies that the signal transfer must be repeated from the beginning. This kind of recognition can be performed only if the operator monitors the signal transfer during the whole process (including the silent parts). A lack of documentation of disturbances in the digital signal invalidate the reliability of the preservation master: actually, when a sound file will be analysed in future, and the original media will be lost or otherwise inaccessible, it would be impossible to determine the source of the noise. Monitoring, with the purposes described above, is effective only if the monitored audio is not taken directly from the playback device, but is first processed through the A/D conversion (*i.e.*, the

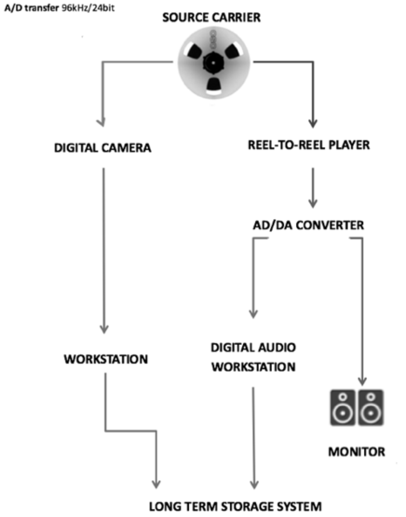


Fig. 1. – A generic signal flow scheme proposed by the authors.

signal should be redirected from digital workstation, via D/A converter, to the speakers or headphones input). In fig. 1 is reported a resuming signal flow scheme proposed by the authors.

2.1. *First and Second Level Access Copies.*

The aims of the digitization projects by archival institutions are: (i) the possibility to grant the access to a larger user base; (ii) the desire to promote new kind of access, and (iii) the documents preservation. If the ultimate goal is the documents preservation, the work process ends when the preservation copies are created. If the goal is the access, other kinds of documents must be produced. A crucial step missing between the conservative and, for example, audio files downloaded from an internet user, is the description of documents content. At this level, a change from a documentary approach to a content oriented one is required. The cataloguer sees in the magnetic tape (wax cylinder, phonographic disc, cassette, . . .), not a sound recording featuring technical audio parameters, but the music, the interview, the conference (with the same gap that separates a manuscripts restorer from the one who studies the text, essay, paper contained in the manuscript). The *first level access copies* are versions generally of lower quality than the preservation copies, and they are used by cataloguers to accomplish the analysis interpretation step. The main function of the first level access copies is to allow the access to the preservation master content, which must be "used only under exceptional circumstances" [14], and which are usually stored in different locations with respect to the archive with slow access (for example, tape drives for long-term storage). In the first level access copies, the length, the tracks number (and their order) should remain unchanged from the preservative master. At this level, some processing is permitted on the audio signal: for example, de-hiss and de-noise filtering for improving the Signal-to-Noise Ratio (SNR) could be useful in order to increase the intelligibility (speech enhancement) to help cataloguers.

The *second level access copies* are digital audio files generated during the cataloguing process. Their length depends on the content. The second level access copies are the containers of sound events (abstract entities) identified by cataloguers during her/his analysis and interpretation of content, organizing the sound material that was provided them in unmediated manner with respect of the content (first level access copies). For examples: (i) a work in mp3 format, from which the silent parts of the tape have been eliminated, and which has been divided into tracks corresponding to the parts of an interview; (ii) a collection of songs in a ethno-musical research. The result of the cataloguer work is a collection of digital resources of variable duration, which can vary in size and quality depending on the flexibility of the protocol adopted by the preservation project. In any case, the flexibility is permitted by the fact that a preservation master exists, and that it is possible to go back to it by means of the information provided by cataloguers during the step of creation of access copies.

The dichotomy between carrier and content (*i.e.*, between artefact and information) distinguishes audio recordings by other cultural heritage, such as sculptures and paintings: in these cases, preservation and restoration actions are directed to the object representing the heritage, that can't be separated from its physical expression [4]. Conversely, this separation can be applied to the audio recordings (the re-mediation process, described in this paper, per-forms an abstraction of the content, in this way the content can survive in new containers), with the result that, in presence of a preservation master, (potentially) infinite number of different restorations (*i.e.*, interpretations) can be performed without affecting directly the source document. The second level access copies are a logic re-organization of sound material, but they are also the type of document that incorporates the results of the various restoration proposals.

3. Audio Restoration: Techniques and Tools

The speech audio documents are usually recorded in non-professional carriers by means of amateur recording system. Thus, for their appropriate fruition and/or for a suitable use of automatic speech recognition techniques, it is often useful to process the audio signals by means of audio restoration algorithms, during the creation of second level access copies.

The audio restoration algorithms can be divided into, at least, three categories [8]:

1. frequency-domain methods, such as various forms of non-casual Wiener filtering or spectral subtraction schemes and recent algorithms that attempt to incorporate knowledge of the human auditory system; these methods use little *a priori* information;

2. time-domain restoration by signal models such as Extended Kalman Filtering (EKF): in these methods, a lot of *a priori* information is required in order to estimate the statistical description of the audio events;

3. restoration by source models: only *a priori* information is used.

The advantage of frequency-domain methods is that they are straightforward and easy to implement. However, the limitations are as follows: musical noise (short sinusoids randomly distributed over time and frequency) is unavoidable; the results depend on a good noise estimation. Restoration by source model is limited to very few cases (*e.g.* only monophonic recordings) and it is not generalizable. The EKF is able, in principle, to simultaneously solve the problems of filtering, parameter tracking and elimination of the outliers, but it is very sensitive to parameter setting (*e.g.*, the order model; the length of the signal vector, the length of the initial training segment in the bootstrap procedure, the adaptation speed, the threshold for detection of impulsive noise).

The authors developed innovative algorithms, using the Virtual Studio Technology (VST) plug-in architecture developed by Steinberg. The algorithms are (for a detailed description, from a computer science point of view, see [6]):

- CREAK (Canazza REstoration Audio - extended Kalman filter): a de-noise and de-click system based on Extended Kalman Filter [7], dedicated to the restoration of audio signal re-recorded from discs coated with shellac, the most common material used in the early production of the 78 RPM recordings, typically affected by low SNR, clicks, pops, crackle. At medium/high SNR, the performance of such filter is (at least) comparable to that of other standard methods like STSA (see below), nevertheless its plain use does not guarantee the best results. One reason for this is that the non-stationarity of audio signals leads to errors in parameter and tracking noise filtering, especially during fast transients. In order to achieve maximum performance from the EKF, it is essential to optimize its implementation.

- CMSR (Canazza-Mian Suppression Rule): a de-noise algorithm based on STSA (Short Time Spectral Attenuation), dedicated to the restoration of audio signal re-recorded from wax and amberol cylinders and shellac discs: low SNR. Adaptations of the spectrally-based techniques applied in speech processing have received most attention. They are based upon the well-known spectral weighting [3], in which individual spectral components are weighted according to expected noise and signal components. Such techniques can be viewed as finite block-size approximations to frequency domain Wiener filtering. As a result of these approximations (necessary to follow the time-varying nature of the useful signal) undesirable distortions can occur, the most notable being known as “musical noise” in which statistical fluctuations in the frequency components of noise lead to random tonal artefacts in the processed signal. Various techniques have been applied to mask or eliminate these distortions. CMSR adopts the Ephraim and Malah suppression rule [12].

- PAR (Perceptual Audio Restoration): a de-hisser based on perceptual algorithm for reel-to-reel tapes and cassettes: high SNR. The "perceptual" processing task requires to transform the audio signal from an "outer" to an "inner" representation, that is to resort to a representation that takes into account how the human ear perceives the sound. The combination of the psychoacoustic model and frequency-domain algorithms, permits to define a promising restoration methodology [9].

Some files, chosen among the set shared by the conference scientific committee, are restored, using CREAK, PAR (CMSR war discarded because it is specifically designed for wax cylinders) and a commercial software tool (iZotope RX, based on a STSA algorithm able avoid the "musical noise" artefact, designed to repair, enhance, and restore audio signals, improving sound quality and clarity), with different setups. In particular, files named "oro" and "maiale" (8 s long, wave format) were chosen. No further information about the way the files were recorded are known by the authors.

4. Audio Restoration: Results

4.1. *iZotope RX Audio Restoration Tool.*

Audio files were processed with a -20 dB de-noise filter (iZotope RX, Algorithm Type D) and with a 50 Hz de-hum filter. The intention was to apply a massive removal of the noise floor, extracting the vocal part as much as possible. Although this process might increase the risk of creation of audio artifacts, it also enhances the speech components and allows improved automatic phoneme recognition by the software. The noise profile was extracted from the first second of the record, in a fragment in which vocal parts were not present (fig. 2). In fig. 3 is showed an example of noise profile, extracted from the first second of a not processed audio file. After the processing, the spectrograms show how the spoken part appears more defined and emphasized in comparison to the noise background (fig. 4). Spectrograms were obtained with the free software Sonic Visualizer developed by London University [10]. In the first seconds of the fragment is particularly clear the action of the de-hum and the -20 dB de-noise filter, especially in the frequency range between 100 and 1000 Hz, which appears now less intense. In some cases, the effect of the de-noise algorithm can be also observed along the whole spectrogram as a lighter color band in the frequency range between 1000 and 1700 Hz (fig. 5). In fact, observing the noise profiles it is possible to notice how in this particular frequency range the noise components are more intense. In order to further emphasize the extraction of the spoken parts, a second -20 dB de-noise filter has been applied. The second noise profile was extracted from the residual audio information still present in the first second of the record after the DD (Denoise-Dehum) processing (fig. 6). As shown in fig 7, the second de-noise algorithm caused a significant further enlargement of the lighter color band around 1000 and 1700 Hz, due to the attenuation of middle-high range frequencies and the high amount of information removed. In fact, although the noise background was almost removed in its totality, the filter influence area included also fundamental components of the speech part.

4.2. *Perceptual Audio Restoration*

In comparison to the results obtained after the application of the algorithms developed by Izotope, the application of the de-hisser based process PAR has led to a minor alteration of the original audio files. PAR was applied with three different growing intensities: “Soft”, “Medium” and “High”. As shown in fig. 8, as opposed to the former results, the background floor reduction is now inferior and it slowly increases with an heavier application of the algorithm. In fact, vocal component frequencies are emphasized and they appear more intense in the high-intensity spectrogram, in comparison to the low-intensity one. It is possible to notice that, in this case, there is no massive intensity reduction in the frequency range between 1000 and 1700 Hz, as confirmed by the absence of the lighter colour band like the one that was created after the application of the former commercial algorithm. With

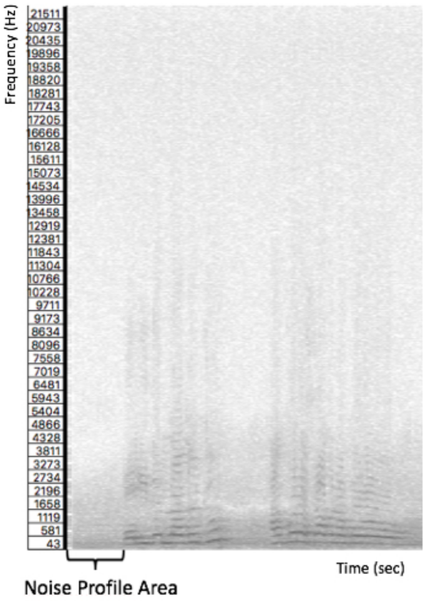


Fig. 2. – A generic noise area chosen as target for the extraction of the noise profile.

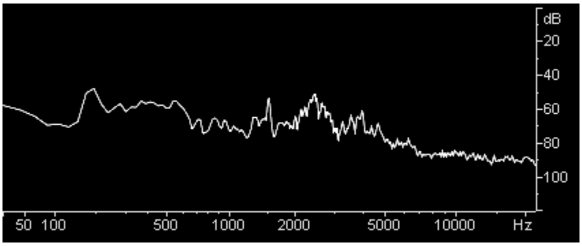


Fig. 3. – Noise profile extracted from “oro” audio file.

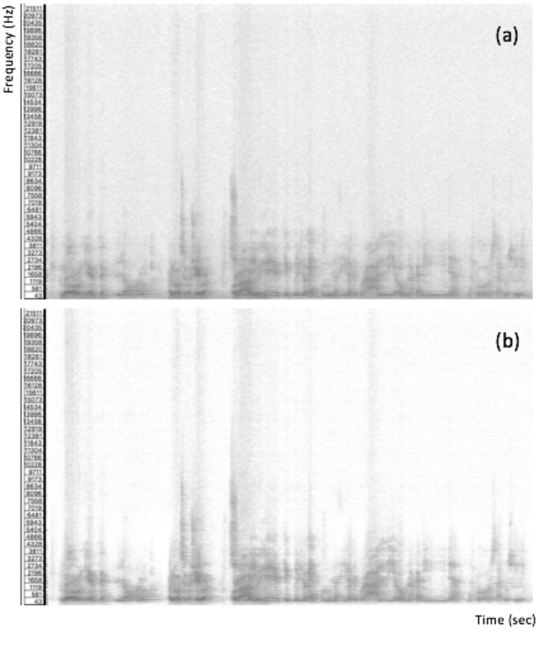


Fig. 4. – Comparison between "oro" audio file not processed (a) and after the application of de-noise and de-hum algorithms (Izotope RX).

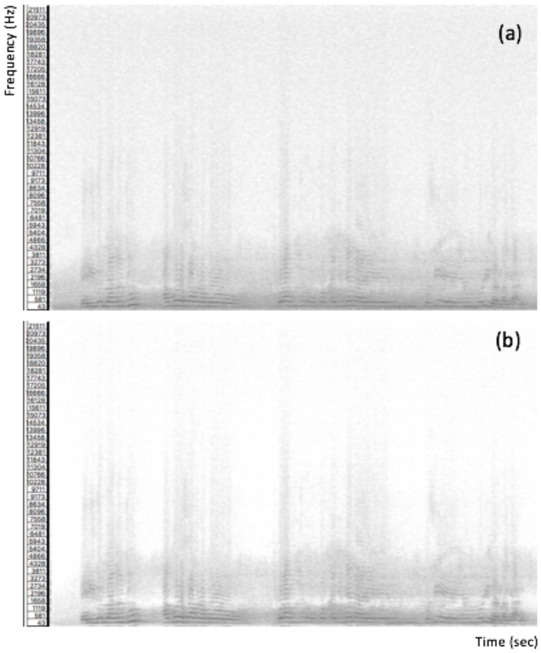


Fig. 5. – Comparison between “maiale” audio file not processed (a) and after the application of de-noise and de-hum algorithms (Izotope RX) (b). It is possible to observe the lighter colour band in the frequency range around 1000-1700 Hz, due to the removal of some fundamental audio components from the speech part.

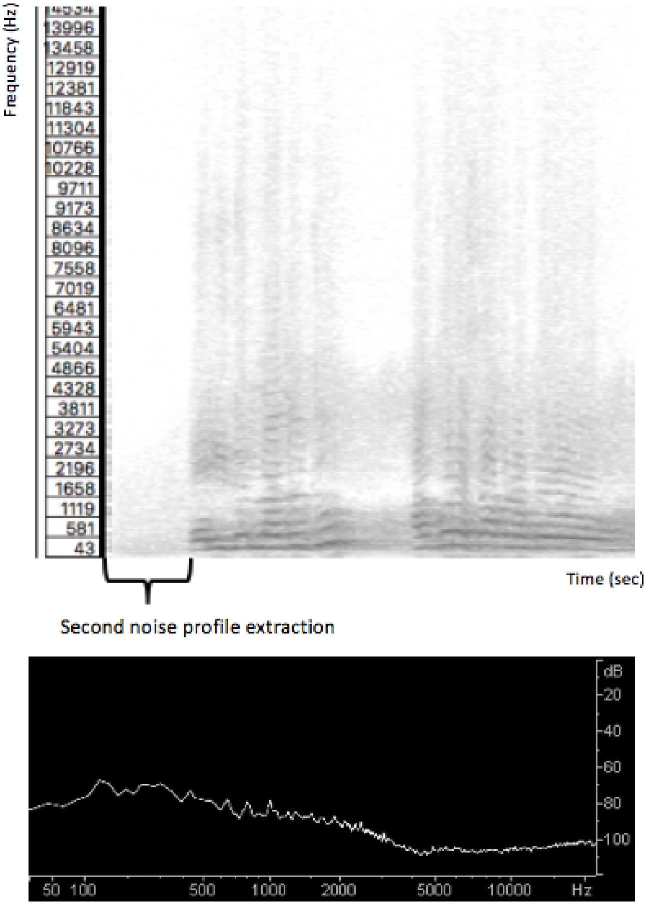


Fig. 6. – Extraction of a second noise profile, used as target to process the file with a second de-noise algorithm (Izotope RX).

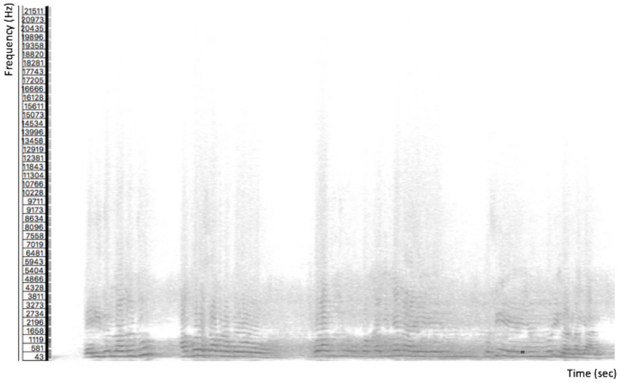


Fig. 7. – Spectrogram of “maiale” audio file after the processing with the first de-noise filter (-20 dB, Izotope RX), de-hum and the second de-noise filter (-20 dB, Izotope RX). It is possible to observe the spread of the lighter colour band due to the massive removal of fundamental audio components.

the application of PAR process, high frequencies (>1.5 kHz) components of the audio signal are also emphasized.

In general, this process permitted to optimize the frequency attenuation for the speechless parts, in comparison to the vocal ones, less deprived of frequency components.

4.3. *Extended Kalman Filter.*

The application of this particular restoration chain led to less evident results and the processed files result equal or slightly different from the original ones.

Neither the analysis of the spectra profiles showed any relevant differences in comparison to the not processed ones (fig. 9). In fact, the noise components appear slightly less intense in the files processed with the EKF, but almost in a imperceptible way. Generally, EKF best results are obtained with low SNR and for impulsive noises like pop and crackle and not for continuous noise disturbs like the ones that the processed files presented; that’s probably the reason why, in comparison with the iZotope commercial algorithms and the

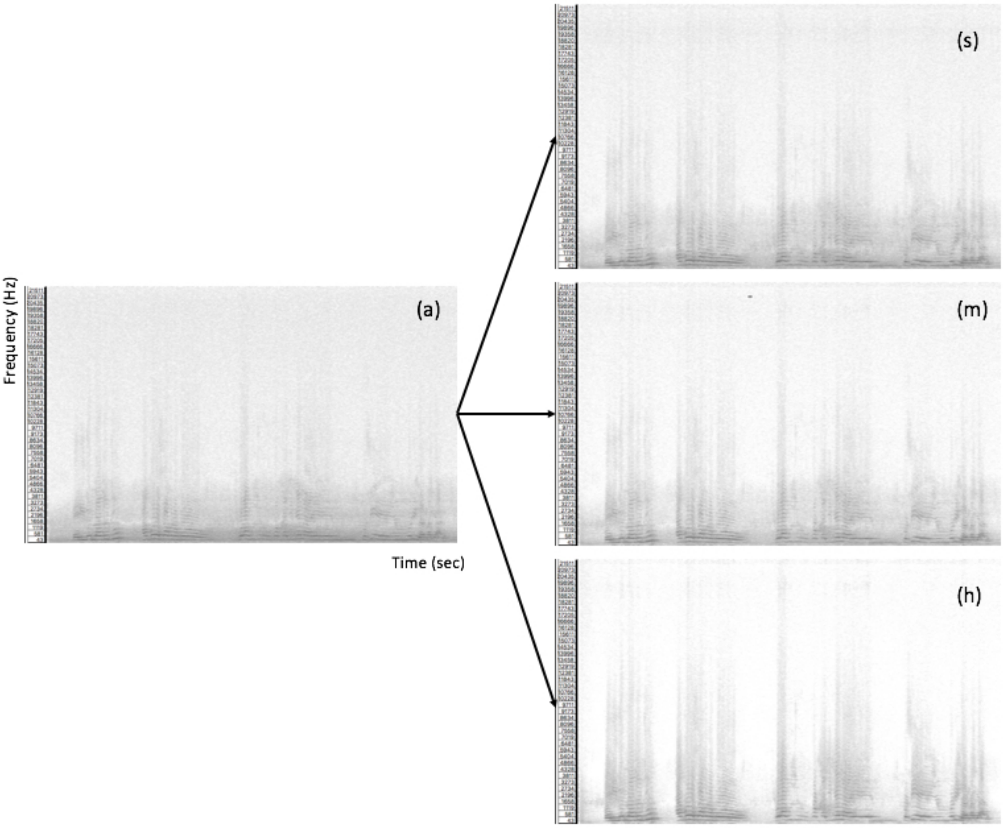


Fig. 8. – Comparison between unprocessed "maiale" audio file (a) and the same file processed with PAR algorithm with three different intensities (s), (m) and (h).

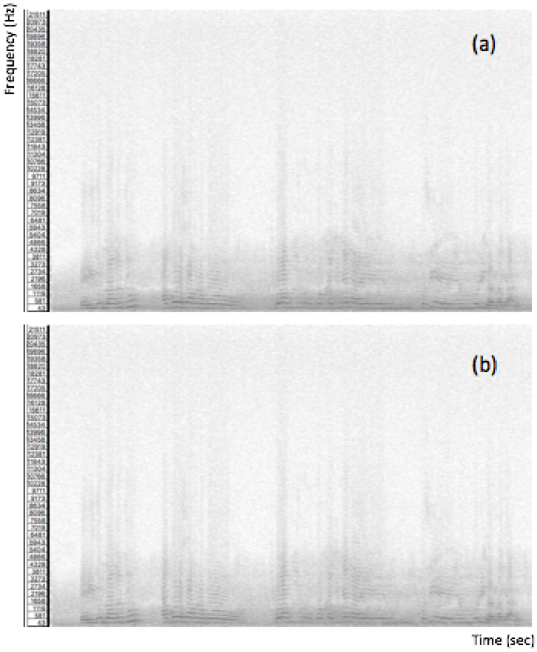


Fig. 9. – Comparison between unprocessed “maiale” audio file (a) and after the application of the EKF (b).

PAR, EKF leaded to a less efficient removal of the undesired audio information.

5. Evaluation by means of ASR

Automatic speech recognition is always more used in many applications, from multimodal human-computer interfaces and virtual assistants to automatic transcription and annotation. With the aim of evaluating how the noise reduction process influences the automatic recognition of speech, the original recording and the de-noised versions described in the previous section were analyzed by an acoustic model implemented by means of Kaldi [16], a free and open-source toolkit for speech recognition research. Kaldi includes tools for calculating standard features, such as MFCC and PLP, and supports model-space transformations and projections both speaker-independent, *e.g.* LDA and MLLT/STC, and speaker-specific, such as fMLLR. Moreover, Kaldi supports conventional acoustic models (*i.e.* diagonal GMMs) and Subspace Gaussian Mixture Models (SGMMs). As the excerpts to be evaluated are both in Italian language, we employed the adaptation of Kaldi to Italian detailed in Cosi *et al.* [11]. The Italian FBK APASCI [1] was employed to train the acoustic model. APASCI is an Italian speech database recorded in a silent room with a Sennheiser MKH 416 T microphone; it includes 5290 phonetically rich sentences and 10800 isolated digits, for a total of 58924 word occurrences (2191 different words) and 641 minutes of speech material which was read by 100 Italian speakers (50 male and 50 female).

Table 1 shows the performance of the speech recognition algorithm applied to the excerpts “maiale” (A) and “oro” (B), both the original versions and the processed ones. The results are reported in terms of Phoneme Error Rate (PER), calculated on the basis of the Levenshtein distance [15], and defined as

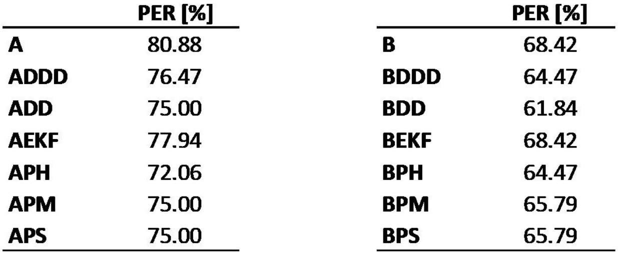
$$PER = \frac{S + D + I}{N}$$

where S is the number of substitutions, D is the number of deletions, I is the number of insertions, and N is the number of phonemes in the sentence. Consequently, a lower value means a better recognition.

Overall, the PER is included between 61.84% and 80.88%. As one might expect, the phonemes of the excerpt “maiale”, that is characterized by a worse sound quality, are less recognizable in comparison to excerpt “oro”. In all the cases but one, the PER is less for the processed versions in comparison to the original recordings, meaning that the noise reduction algorithms improved the automatic phoneme recognition. Only the version processed with the Extended Kalman Filter of the excerpt “oro” obtained the same PER of the original recording: this result could be explained by the fact that the excerpt “oro” was recorded with a quite good sound quality and the application of the EKF, that is known to work better with low SNR [6], resulted in a signal that is very similar to the original one.

Apart EKF, which results are equal or slightly lower than the original recording, the noise reduction algorithms improved the phoneme recognition by

Table 1 – Phoneme Error Rate calculated for both the excerpts “maiale” (A) and “oro” (B) is reported. The original recordings have been processed by different de-noise algorithms (DDD = Denoise+Dehum+Denoise, DD = Denoise+Dehum, EKF = Extended Kalman Filter, P = Perceptual, H = High, M = Medium, S = Small).



	PER [%]		PER [%]
A	80.88	**B**	68.42
ADDD	76.47	**BDDD**	64.47
ADD	75.00	**BDD**	61.84
AEKF	77.94	**BEKF**	68.42
APH	72.06	**BPH**	64.47
APM	75.00	**BPM**	65.79
APS	75.00	**BPS**	65.79

an amount that depends on the type of algorithm and the applied settings. As concern the excerpt “maiale”, the best performance (PER = 72.06%) is obtained by the PAR (Perceptual Audio Restoration) algorithm with the setting High, that implies a higher noise attenuation. Regarding the excerpt “oro”, instead, the commercial tool iZotope obtained the best performance (PER = 61.84%) by applying in sequence a de-noise and a de-hum algorithm (DD). In general, PAR worked better with the setting High that removes more noise, whereas iZotope obtained a better performance with the sequence DD, that removes less noise in comparison to the sequence Denoise-Dehum-Denoise (DDD).

6. Conclusions

The state-of-the-art techniques in audio noise reduction offer interesting tools for improving the quality of audio recordings, as a preliminary task for subsequent linguistic analyses. In this paper, a set of commercial and research tools for audio noise reduction have been described and applied on a couple of on-field recorded excerpt. For each original record, a set of different post-processed versions was obtained. The utilization of the commercial de-noise and de-hum algorithms developed by iZotope was intentionally heavy and the amount of audio information removed was finalized to enhance and emphasize the speech part as more as possible, permitting a best intelligibility of phonemes by the software. In addition to the files processed with the commercial algorithms, other versions were obtained processing the audio material with the Perceptual Audio Restoration (soft, medium and high intensity) and with an Extended Kalman Filter-based restoration algorithm (CREAK).

While EKF did not lead to significant results and the files appeared to be almost identical to the original ones, in PAR processed files the speech components were emphasized, without an excessive alteration of the original record.

The noise reduction algorithms were then evaluated by means of Kaldi, a toolkit for automatic speech recognition. Results showed that, at least for the analyzed cases, the noise reduction algorithms reduce the phoneme error rate, improving the automatic recognition task.

Some limitations of the present paper should be acknowledged. Firstly, the results were obtained with only two short (8 s long) audio excerpts and a more extensive set of audio recordings would be necessary to have more significant results. Secondly, the values of PER obtained with the ASR tool are quite high: for such values, the improvement due to the noise reduction algorithms, with a phoneme error rate however greater than 60%, could not imply a real usefulness for automatic annotation/transcription tasks. The low performances in terms of PER could be explained with the differences between the dataset used to train the acoustic model and the two analyzed excerpts. In particular, the training dataset is composed by Italian utterances recorded with professional equipment in a controlled environment (anechoic room), whereas the excerpts were recorded on-field, with dialect inflections. Specific adaptation of the acoustic model is therefore necessary to improve the recognition task.

References

[1] B. Angelini, F. Brugnara, D. Falavigna, D. Giuliani, R. Gretter, M. Omologo, *A baseline of a speaker independent continuous speech recognizer of Italian*. Proceedings of the 3rd European Conference on Speech Communication and Technology, 1993, 847-850.

[2] N. Bernardini, X. Serra, M. Leman, G. Widmer, *A roadmap for sound and music computing*. The S2S2 Consortium, 2007.

[3] S.F. Boll, A.V. Oppenheim, *Suppression of acoustic noise in speech using spectral subtraction*. IEEE Trans. Acoustics, Speech and Signal Processing, 2(27), 1979, 113-120.

[4] C. Brandi, *Teoria del restauro*. Edizioni di Storia e Letteratura, Roma 1963 (*Theory of Restoration*, trad. Cynthia Rockwell, rev. Dorothy Bell, Firenze, Nardini 2005).

[5] F. Bressan, S. Canazza, *A systemic approach to the preservation of audio documents: Methodology and software tools*. Journal of Electrical and Computer Engineering, 2013.

[6] S. Canazza, *The digital curation of ethnic music audio archives: from preservation to restoration*. International Journal of Digital Libraries, 2/3(12), 2012, 121-135.

[7] S. Canazza, G. De Poli, G.A. Mian, *Restoration of audio documents by means of Extended Kalman Filter*. IEEE Trans on Audio Speech and Language Processing, 6(18), 2010, 1107-1115.

[8] S. Canazza, A. Vidolin, *Preserving electroacoustic music*. Journal of New Music Research, 4(30), 2001, 289-293.

[9] S. Canazza, G. De Poli, S. Maesano, G.A. Mian, *On the performance of a noise reduction technique based on a psychoacoustic model for the restoration of old audio recordings*. Proceedings of Diderot Forum, 1999, 29-35.

[10] C. Cannam, C. Landone, M. Sandler, *Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files*. Proceedings of the ACM Multimedia International Conference, 2010.

[11] P. Cosi, G. Paci, G. Sommavilla, F. Tesser, *Kaldi: yet another ASR toolkit? Experiments on adult and children Italian speech*. Proceedings of 11th Convegno Nazionale Associazione Italiana di Scienze della Voce, 2015, 429-438.

[12] Y. Ephraim, D. Malah, *Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator*. IEEE Trans. Acoustics, Speech and Signal Processing, 21(6), 1984, 1109-1121.

[13] IASA-TC 03, *The Safeguarding of the Audio Heritage: Ethics, Principles and Preservation Strategy*. IASA Technical Committee, 2005.

[14] IASA-TC 04, *Guidelines on the Production and Preservation of Digital Audio Objects*. IASA Technical Committee, 2004.

[15] V.I. Levenshtein *Binary codes capable of correcting deletions, insertions, and reversals*. Soviet Physics Doklady, 10(8), 1966, 707-710.

[16] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, K. Vesely *The Kaldi speech recognition toolkit*. IEEE workshop on automatic speech recognition and understanding, 2011, 404-439.

[17] D. Schüller, *Preserving the facts for the future: Principles and practices for the transfer of analog audio documents into the digital domain*. Journal of Audio Engineering Society, 7/8(49), 2001, 618-621.

[18] Yakel, *Digital curation*. OCLC Systems & Services. International digital library perspectives, 23(4), 2007, 335-340.