

A PROTOCOL FOR THE PRESERVATION OF SPEECH DOCUMENTS ARCHIVES: TOWARDS THE DIGITAL CURATION OF THE "ITALIAN DIALECTS MAP"

Canazza Sergio¹, Silvia Calamai², Pier Marco Bertinetto³, De Dominicis Amedeo⁴

¹University of Padova, Dept. of Information Engineering, Padova, Italy - canazza@dei.unipd.it; ²University of Siena, Siena, Italy - calamai@unisi.it; ³Scuola Normale Superiore, Pisa, Italy - bertinetto@sns.it; ⁴University of Tuscia, Italy - dedomini@unitus.it

Keywords: Speech Audio Archives, Linguistic Analysis, Dialectology, Preservation Audio Documents.

Despite the youthfulness of the audio recording technologies, the preservation of audio documents is already becoming an urgent issue to prevent the loss of precious cultural heritage recordings. These documents have a shorter life span compared to other cultural heritage materials and their preservation introduce novel problems requiring original approaches. In particular, the speech audio documents are in danger of disappearing, because of the poor quality of the material on which they were recorded and the rapid evolution of the recording formats. Moreover, the audio speech recordings were often made with non-professional systems or in fieldwork circumstances (i.e. with a low quality), and in particular cases even risk being forgotten, due to widespread lack of well-established archiving procedures. This has not been the destiny of music repertoires of wide interest (classical music, rock). In these cases, the recording companies have re-recorded most of the audio documents, particularly those of high commercial value. The same has not happened to speech audio documents, often the only testimonial of disappeared oral cultures. Their preservation requires procedures, detailed for every type of carrier, able to preserve the recorded audio at quality level suitable for linguistic analysis. In the past, lack of philological and engineering accuracy and care to source consistency has already produced several dramatic mistakes in the A/D transfer of speech audio archives, producing yet more sources featuring disfiguring transmission errors.

In several international audio archives projects, the present authors defined and put it into practice a quality control for the task of audio documents active preservation. In particular, they are applying it in the project, partially supported by *Regione Toscana*, focussed on *Carta dei Dialetti Italiani*. The *Carta dei Dialetti Italiani* archive is a neglected open reels speech archive containing both answers to phonetic and morphological questionnaires and the oral performance of the *Parable of the Loss Child*. It has been collected in a relevant number of Italian villages by linguists and dialectologists. This important fieldwork, offering a unique database of Italian dialects from the Sixties and Seventies, sank into oblivion because of the death of the founder, Oronzo Parlangeli, and subsequent financial and organizational difficulties. While we are still tracing back the reels of other Italian regions, we succeeded in finding almost all the reels referring to Tuscan fieldwork, together with all the related reports and notes written by the équipe of linguists co-ordinated by Gabriella Giacomelli.

In order i) to choose the optimum re-recording system, ii) to verify the correctness of the preservative copies and iii) to validate the de-noise processing in the access copies – an experiment was performed: a speaker recognition system, developed by some of the present authors, is used to observe the deterioration of the speaker's features as the re-recording process changes. The results that will be presented confirm the goodness of our protocol and the effectiveness of the audio restoration system used.