Audio Surveillance Tapping Transcription Quality Estimation Reliability

Clara Borrelli, Paolo Bestagini, Fabio Antonacci, Augusto Sarti, Stefano Tubaro

Dipartimento di Elettronica, Informazione e Bioingegneria

Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Email: {clara.borrelli, paolo.bestagini, fabio.antonacci, augusto.sarti, stefano.tubaro}@polimi.it

Abstract—The abstract goes here.

I. Introduction

Thanks to the recent advances in technology, audio communication systems are becoming more and more pervasive and diverse. In fact, nowadays a conversation can happen not only face-to-face but also on a variety of channels, like phone calls, VoIP or voice messages on instant messaging platforms.

At the same time the ability of monitoring environmental, digital or phone communications has become an urgent necessity for national security issues. These technologies can effectively assist law enforcement agencies in foiling terrorist attacks or revealing harmful intents. The aforementioned heterogeneity of the communication channels has produced in this sense both advantages and drawbacks.

On one side, information gathering has become easier and ubiquitous. Regarding environmental monitoring, audio surveillance devices are getting cheaper and easier to deploy. For these reasons the number of control spots can be increased, ensuring higher coverage, for example, in large public spaces. Concerning digital or phone wire taps, ...

On the other side, a blind and massive data collection can results in huge databases whose analysis might be infeasible. Moreover the audio results can be affected by several type of noises which

- Audio communication systems are becoming more pervasive and diverse (e.g., phone calls, voip, voice messages, etc.).
- For national security reason (e.g., to avoid terroristic attacks, uncover malicious things, etc.), the ability of monitoring communications through phone tapping or environmental monitoring in several situations is an urgent necessity.
- Fortunately, audio surveillance devices are increasingly cheaper and easier to deploy.
- However, a blind and massive data collection might result in huge databases whose analysis might become unfeasible.
- It is therefore needed to develop methods that automatically enable to speed the analysis of these huge corpora of collected audio excerpts.

- In this paper we propose a framework to evaluate the quality of noisy speech recordings estimating the likelihood of obtaining a reliable transcript.
- Two applications: i) to process large databases with automatic speech-to-text transcriptors; ii) to provide a quality feedback to any investigator deploying an audio surveillance system.
- The main rationale is to use ...
- Results obtained on a lots of data show that ...

[1]

II. Problem statement and background

In this section bla bla ...

A. Problem formulation

Given an audio excerpt $x(t)$ composed by ... containing some speech ... we want to estimate how much it is useful for an analyst. This is done by estimating a score indicating how likely is it possible to correctly understand the speech.

B. Background

Ci sono tanti metodi in tante aree che si occupano di stima dell'intelligibilita'. Citiamono un po' e facciamo capire che sono tipicamente cuciti su scenari applicativi per noi non rilevanti (e.g., acustica di stanza, transmissioni telefoniche, etc.).

III. Method

Cappello introduttivo in cui si spiega che il framework e' composto da tre elementi

- La scelta di una misura di qualita' oggettiva (Jaccard, Coseno su vettori di parole)
- L'estrazione di descrittori (MFCC, etc.)
- Classificazione / regressione

A. Objective quality measure

B. Feature extraction

C. Decision taking

IV. Results

In this section bla bla ...

A. Dataset

Da dove siamo partiti, tipi di rumore, SNR, trascrittore.

B. Setup

Tipi di classificatore / regressore usati. Divisione train, validation e test (o forse in dataset).

C. Numerical analysis

risultati di classificazione / regressione con metodi diversi, e feature diverse.

V. CONCLUSIONS

It worked!

ACKNOWLEDGMENT

REFERENCES

[1] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, 2013.