UNIVERSITAT POLITÈCNICA
DE CATALUNYA

PhD Thesis

A multi-microphone approach to speech processing in a smart-room environment

Author: Alberto Abad Gareta

Advisor: Dr. Fco. Javier Hernando Pericás

Speech Processing Group
Department of Signal Theory and Communications
Universitat Politècnica de Catalunya

Barcelona, February 2007

A mis padres y hermana,

Abstract

Recent advances in computer technology and speech and language processing have made possible that some new ways of person-machine communication and computer assistance to human activities start to appear feasible. Concretely, the interest on the development of new challenging applications in indoor environments equipped with multiple multimodal sensors, also known as smart-rooms, has considerably grown.

In general, it is well-known that the quality of speech signals captured by microphones that can be located several meters away from the speakers is severely distorted by acoustic noise and room reverberation. In the context of the development of hands-free speech applications in smart-room environments, the use of obtrusive sensors like close-talking microphones is usually not allowed, and consequently, speech technologies must operate on the basis of distant-talking recordings. In such conditions, speech technologies that usually perform reasonably well in free of noise and reverberation environments show a dramatically drop of performance.

In this thesis, the use of a multi-microphone approach to solve the problems introduced by far-field microphones in speech applications deployed in smart-rooms is investigated. Concretely, microphone array processing is investigated as a possible way to take advantage of the multi-microphone availability in order to obtain enhanced speech signals. Microphone array beamforming permits targeting concrete desired spatial directions while others are rejected, by means of the appropriate combination of the signals impinging a microphone array.

A new robust beamforming scheme that integrates an adaptive beamformer and a Wiener post-filter in a single stage is proposed for speech enhancement. Experimental results show that the proposed beamformer is an appropriate solution for high noise environments and that it is preferable to conventional post-filtering of the output of an adaptive beamformer. However, the beamformer introduces some distortion to the speech signal that can affect its usefulness for speech recognition applications, particularly in low noise conditions.

Then, the use of microphone arrays for specific speech recognition purposes in smart-room environments is investigated. It is shown that conventional microphone array based speech recognition, consisting on two independent stages, does not provide a significant improvement with respect to single microphone approaches, especially if the recognizer is adapted to the actual acoustic environmental conditions. In the thesis, it is pointed out that speech recognition needs to incorporate information about microphone array beamformers, or otherwise, beamformers need to incorporate speech recognition information. Concretely, it is proposed to use microphone array beamformed data for acoustic model construction in order to take more benefit from microphone arrays. The result obtained with the proposed adaptation scheme with beamformed enrollment data shows a remarkable improvement in a speaker dependent recognition system, while only a limited enhancement is achieved in a speaker independent recognition system, partially due to

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.