CCIT Report # 469 February 2004

On the Use of Spatial-Temporal Information in Speaker Localization Applications

Tsvi Gregory Dvorkind and Sharon Gannot

Abstract

Speaker localization based on a microphone array and using a dual step approach is addressed. The first stage, which is not the main concern of this paper, is comprised of estimating the *time difference of arrival* between the speech signal received by each microphones pairs. These readings are then used by the second stage for the actual localization. The speaker's smooth trajectory is used for improving the current position estimate. Two localization schemes, which exploit the temporal information, are presented. The first is the *Extended Kalman filter*. The second is a recursive form of the Gauss method. Experimental study as well as approximate analytical evaluation supports the potential of the proposed methods.

Keywords

Talker Localizer; Extended Kalman filter; Gauss method; Cramér-Rao Lower Bound

I. Introduction and Problem Formulation

Determining the spatial position of a speaker finds a growing interest in video conference scenario where automated camera steering and tracking are required. In this work we address approaches for determining speaker position which are comprised of two stages. In the first stage, the TDOA is estimated using spatially separated microphone pairs (e.g. [1],[2], [3], and [4]). In the second stage, these readings are used for the actual localization (e.g. [5], [6] and [7]). These methods exploit the spatial information obtained by different microphone pairs, but do not exploit the temporal information available from adjoint speaker position estimates. This information is relevant for the current position estimate, due to the speaker smooth trajectory. A shorter preliminary conference version of these ideas has been published in [8].

Consider an array of $M+1$ microphones, placed at the Cartesian coordinates $\underline{m}_i \triangleq [x_i, y_i, z_i]^T$; $i = 0, \ldots, M$ where $\underline{m}_0 = [0,0,0]^T$ is the reference microphone, placed at the axes origin and $(\cdot)^T$ stands for the transpose operation. Define the source coordinate at time instance t by $\underline{s}(t) \triangleq [x_s(t), y_s(t), z_s(t)]^T$. Each of the M microphones, combined with the reference microphone, is used at time instance t to produce a TDOA measurement $\tau_i(t)$; $i = 1 \ldots M$. Denote the i-th **range difference** measurement by $r_i(t) = c\tau_i(t)$, where c

T. G. Dvorkind is with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel (email: dvorkind@tx.technion.ac.il).

S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel (email: gannot@eng.biu.ac.il)
Corresponding author: S. Gannot