IMPORTANCE SAMPLING PARTICLE FILTER FOR ROBUST ACOUSTIC SOURCE LOCALISATION AND TRACKING IN REVERBERANT ENVIRONMENTS

Eric A. Lehmann and *Robert C. Williamson*

National ICT Australia
216 Northbourne Avenue, Braddon ACT 2612, Australia
Email: {Eric.Lehmann,Bob.Williamson}@nicta.com.au

The concept of *acoustic source localisation and tracking* (ASLT) plays an important role in many practical speech acquisition systems. Exact knowledge of the speaker position is usually the key to acquiring clean speech using e.g. beamforming or equalisation. Multipath sound propagation in practical environments however constitutes a major challenge to overcome for any array-based tracker. The performance of methods used traditionally for this purpose, such as steered beamforming (SBF) and time delay estimation (TDE), are heavily influenced by reverberation and background noise. Recently, the concept of *particle filtering* (PF) was proposed as a new approach to this problem [1, 2, 3]. This method relies on a Bayesian filtering principle which can be summarised as follows.

Let the *state variable* $\boldsymbol{\mathcal{X}}_k$ represent the position and velocity of the acoustic source, $\boldsymbol{\mathcal{X}}_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]^{\mathrm{T}}$. At regular time intervals ($k = 1, 2, \ldots$), the microphone array delivers a collection of frames of audio signal, one per sensor, which can be denoted by the *observation variable* $\boldsymbol{\mathcal{Y}}_k$. With $\boldsymbol{\mathcal{Y}}_{1:k}$ representing the concatenation of all the measurements up to time k, Bayesian filtering provides a recursive derivation of the *posterior* density $p(\boldsymbol{\mathcal{X}}_k | \boldsymbol{\mathcal{Y}}_{1:k})$. Important statistical information about the target position can be derived from this density, including an estimate $\widehat{\boldsymbol{\mathcal{X}}}_k$ of the target position.

Particle filtering is an approximation technique that implements the Bayesian filtering recursion by means of a sample-based representation of the posterior density, and it is able to deal with nonlinear and non-Gaussian problems. Because a PF delivers location estimates based on a series of past acoustic measurements rather than the current observation only, this method is also more efficient at dealing with the spurious effects of acoustic reverberation than traditional ASLT algorithms.

Previous research [2, 3] made use of the basic *bootstrap* particle filter, introduced in [4]. The conceptual simplicity of this algorithm leads to straightforward practical implementations and moderate computational requirements. However, this PF suffers from a major drawback: during each iteration, the particles are relocated in the state space without knowledge of the current observations. The PF might hence omit some important regions of the state space when searching for the target, which mainly precludes the PF from reinitialising after a target disappears or becomes occluded for a short period of time. Despite showing promising results, this algorithm consequently still lacks some important characteristics necessary for a smooth operation in practical scenarios, such as the automatic detection of new targets and the ability to recover from track loss.

In the current research, we developed a particle filtering method based on the more general concept of *importance sampling* (IS), in which particles are generated during each iteration on the basis of *both* the particle set at the previous time step and the current measurement. The critical design step of this technique consists in choosing a suitable *importance function*, from which the particles can be easily sampled. In previous literature, this importance function has been implemented to take advantage of measurements from auxiliary sensors [5], or to draw on information obtained from two different measurement processes derived from the same raw data [6]. The latter approach is applied here to the ASLT problem definition.

Experimental results have shown that SBF has improved tracking performance compared to other TDE-based methods [2]. Hence, the importance function is implemented here on the basis of delay-and-sum beamforming (DSB) results, computed for a small range of low frequencies: $f \in [100, 400\text{Hz}]$. Likelihood weights are then assigned to each particle using a *pseudo-likelihood* function (as introduced in [2]) built from DSB results obtained for a wider range of frequencies: $f \in [300, 3000\text{Hz}]$. The SBF output computed over a small range of low frequencies is known to provide coarse localisation information, whereas results obtained for high frequencies are more accurate but also suffer from spatial aliasing. Combining these two acoustic measurements in the global statistical framework of particle filtering leads to improved tracking and localisation results. The

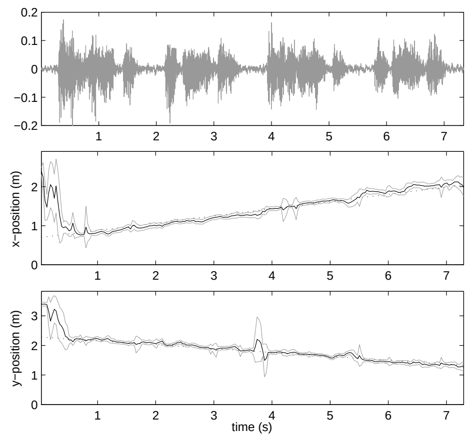

Fig. 1. *Top plot:* signal recorded with one array sensor. *Bottom plots:* true source position (dotted line), source location estimate (solid lines) and lines representing +/− one standard deviation of the particle set (grey lines).

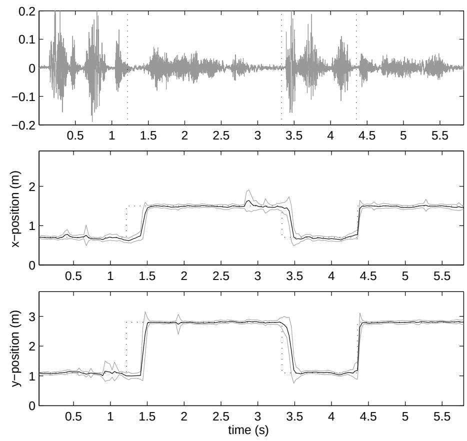

Fig. 2. *Top plot:* signal recorded with one array sensor. Vertical dotted lines denote a change of speaker. *Bottom plots:* tracking results in x and y-coordinates. Dotted lines represent the position of the active source.

importance sampling approach also provides the algorithm with the ability to *reinitialise*, allowing the resulting tracker to automatically recover from total track losses and detect new targets entering the acoustic scene.

Practical tests were conducted with the IS algorithm developed in this research. The experimental setup was based on an enclosure measuring $2.9\text{m} \times 3.8\text{m} \times 2.7\text{m}$, and fitted with a planar array of 8 omnidirectional microphones organised as one pair on each wall. The number of filter particles was 30. An example of the tracking results achieved with the IS algorithm is depicted in Fig. 1, showing the estimated source position versus time. The multi-channel audio data used in this example was recorded in a real office room with reverberation time $T_{60} = 0.39$s. The results depicted in Fig. 2 were obtained with a scenario where two speakers take part in an alternating conversation. The simulation was carried out using the image method to generate signals originating from two different locations in the room, with a reverberation time $T_{60} = 0.35$s. This result demonstrates the reinitialisation capabilities of the IS method which automatically switches between talkers as they take turns. Finally, the IS method was implemented on a 1.7GHz computer, used in conjunction with a 16 microphone array set up in a $3.5\text{m} \times 4.5\text{m} \times 2.7\text{m}$ office room with reverberation time $T_{60} = 0.5$s. This implementation delivers real-time location estimates and demonstrates the robustness of the IS algorithm when localising sources and tracking fast target motions. Movie recordings of the implementation results can be found at http://rsise.anu.edu.au/~eric/tracking.html.

Speaker localisation and tracking are complicated array processing applications, made especially challenging by complex reverberation effects and the discontinuous nature of speech signals, among others. In this research, we have developed a particle filter using an importance sampling approach, which proved to drastically improve the performance of traditional acoustic source localisation methods. Also, the resulting algorithm was found to be a much better candidate than bootstrap-only methods for practical implementations.

References

[1] J. Vermaak and A. Blake, "Nonlinear filtering for speaker tracking in noisy and reverberant environments," in *Proc. IEEE ICASSP*, Salt Lake City, USA, May 2001, vol. 5, pp. 3021–3024.

[2] D. Ward, E. Lehmann, and R. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Trans. Speech and Audio Proc.*, vol. 11, no. 6, pp. 826–836, November 2003.

[3] D. Ward and R. Williamson, "Particle filter beamforming for acoustic source localization in a reverberant environment," in *Proc. IEEE ICASSP*, Orlando, USA, May 2002, vol. 2, pp. 1777–1780.

[4] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE Proc. Radar and Sig. Proc.*, vol. 140, no. 2, pp. 107–113, April 1993.

[5] J. Vermaak, M. Gangnet, A. Blake, and P. Pérez, "Sequential Monte Carlo fusion of sound and vision for speaker tracking," in *Proc. IEEE Intl. Conf. on Computer Vision*, Vancouver, Canada, July 2001, vol. 1, pp. 741–746.

[6] M. Isard and A. Blake, "ICONDENSATION: unifying low-level and high-level tracking in a stochastic framework," in *Proc. European Conf. on Computer Vision*, Freiburg, Germany, June 1998, vol. 1, pp. 893–908.