Sub-band Domain Multi-Hypothesis Acoustic Echo Canceler Based Acoustic Scene Analysis

Link to Paper

Benjamin J. Southwell*, Yin-Lee Ho, David Gunawan
Dolby Laboratories, Sydney Australia
* corresponding author: benjamin.southwell@dolby.com

Abstract

This paper introduces a novel approach for acoustic scene analysis by exploiting an ensemble of statistics extracted from a sub-band domain multi-hypothesis acoustic echo canceler (SDMH-AEC). A well-designed SDMH-AEC employs multiple adaptive filtering strategies with potentially complementary behaviors during convergence, perturbations, and steady-state conditions. By aggregating statistics across the sub-bands, we derive a feature vector that exhibits strong discriminative power for distinguishing different acoustic events and estimating acoustic parameters. The complementary nature of the SDMH-AEC filters provides a rich source of information that can be extracted at insignificant cost for acoustic scene analysis tasks. We demonstrate the effectiveness of the proposed approach experimentally with real data containing double-talk, echo path change and events where the full-duplex device is physically moved. The extracted features enable acoustic scene analysis using existing echo cancellation algorithms and techniques

Sub-band Domain Multi-Hypothesis Cancelers

In this paper, we utilize a sub-band domain multi-hypothesis AEC (SDMH-AEC), depicted in Figure 1, that copies coefficients between two active adaptive filters referred to as the main and shadow filter. These are designed to be complimentary; The shadow filter uses a variable step size (VSS) [1] normalized least mean squares (NLMS) [2]. The VSS algorithm is $\mu_{shadow} = \min\left(\frac{p}{r}, 0.5\right)$ where p is the predicted echo power and r is the residual power. The main filter utilizes a proportionate NLMS[3] adaption strategy with an update rate of $\mu_{main} = 0.5$. The control heuristics block will copy the coefficients from either the main or the shadow into the other should one produce a residual that is at least 10dB lower than the other. The SDMH-AEC also selects which residual, or microphone, signal to output based on which has the lowest power level.

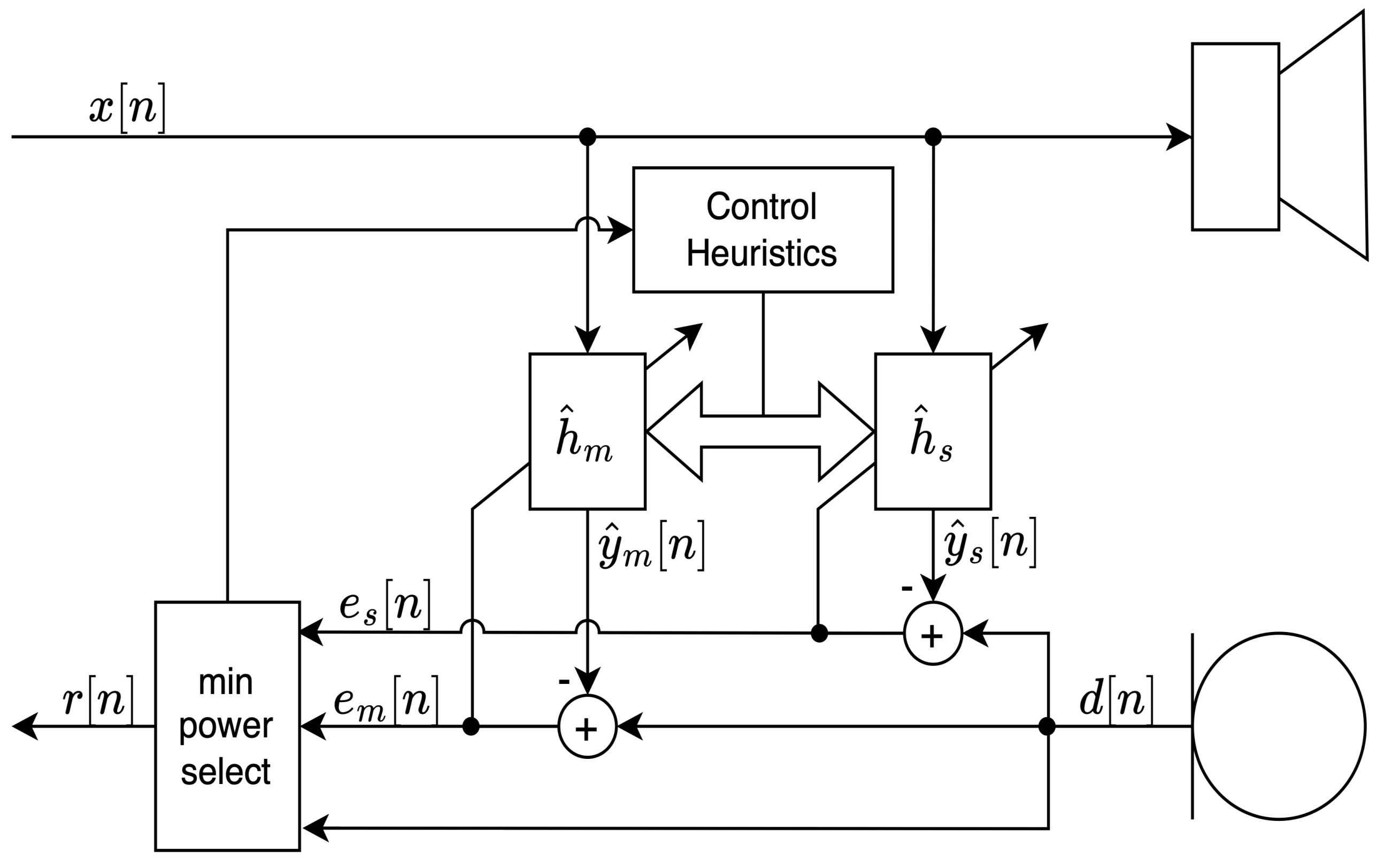

Fig. 1. A Multi-hypothesis AEC. In the sub-band scheme, this represents the processing that is done on a single sub-band where the analysis and synthesis transforms are not shown.

Extracting Statistics

The SDMH-AEC operates on all critically sampled sub-bands obtained using a modified discrete Fourier transform [4] where the sample rate is 48kHz and the block size is 512 samples. However, for the purposes of extracting statistics, we only process the first 100 sub-bands, i.e., from 0 to 4687.5Hz. We aggregate the main, P_m, shadow, P_s, and microphone, P_d, probabilities. These are simply the counts, after being normalized and smoothed, over the ensemble of sub-bands of which residual was selected by the minimum power select block, showing in Figure 1. We also aggregate the counts of the main update and shadow update probabilities. These are the normalized counts of sub-bands where the shadow filter coefficients were copied into the main filter (thus updating it) or vice versa.

In Figure 2, steady-state operation cancelling spectrally rich music with no perturbation is shown. We can see that the main and shadow filter both have a probability of approximately 0.5. This is characteristic of the filters themselves and the operating conditions. The microphone probability is close to zero and we observe an insignificant amount of main and shadow update events.

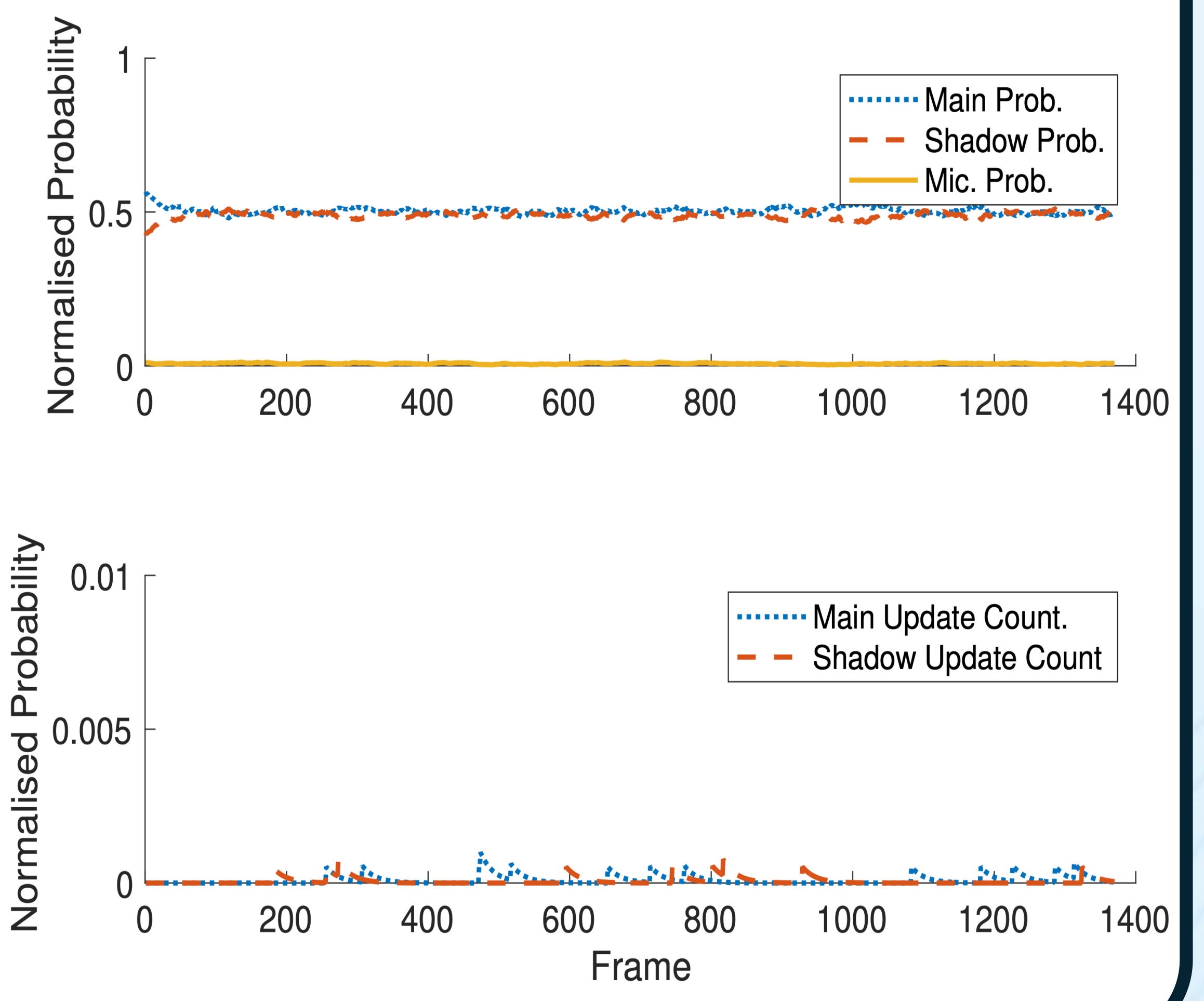

Fig 2. Elements of the statistics vector steady-state operation with no perturbation

References

[1] R. H. Kwong and E. W. Johnston, “A Variable Step Size LMS Algorithm,” IEEE Transactions on Signal Processing, vol. 40, no. 7, pp. 1633–1642, 1992.

[2] S. Haykin, Adaptive filter theory, 5th ed. Prentice Hall, 2014. Upper Saddle River, NJ:

[3] D. L. Duttweiler, “Proportionate normalized least-mean-squares adaptation in echo cancelers,” IEEE Transactions on Speech and Audio Processing, vol. 8, no. 5, pp. 508–517, 2000.

[4] T. Karp and N. J. Fliege, “Modified DFT filter banks with perfect reconstruction,” IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, vol. 46, no. 11, pp. 1404–1414, 1999.

Experimental Results

In Figure 3, Figure 4 and Figure 5, the statistics vectors extracted from real-world recordings are depicted where double talk, an extrinsic echo path change, and an intrinsic echo path change events occur. Due to the complimentary nature of the main and shadow filtering strategies combined with the control heuristics, there are distinct signatures associated with each class of events.

Double talk events, such as that shown in Figure 3, result in the main filter mis adapting, cancelling some of the double talk and causing a momentary peak in P_m. Not long after, the misadaption of the main causes the shadow filter to become the better filter, resulting in P_s dominating. The shadow filter coefficients are copied into the main leading to the system reconverging to steady state after the event.

Echo path changes, such as those shown in Figure 4 and Figure 5, are characterized by P_m dominating the P_s for the duration of the perturbation. Reconvergence of the system occurs after the shadow filter is updated with the main coefficients. Repositioning the device is distinct to an extrinsic echo path change due to the noise introduced into the microphone when the device is touched and then placed back on a surface.

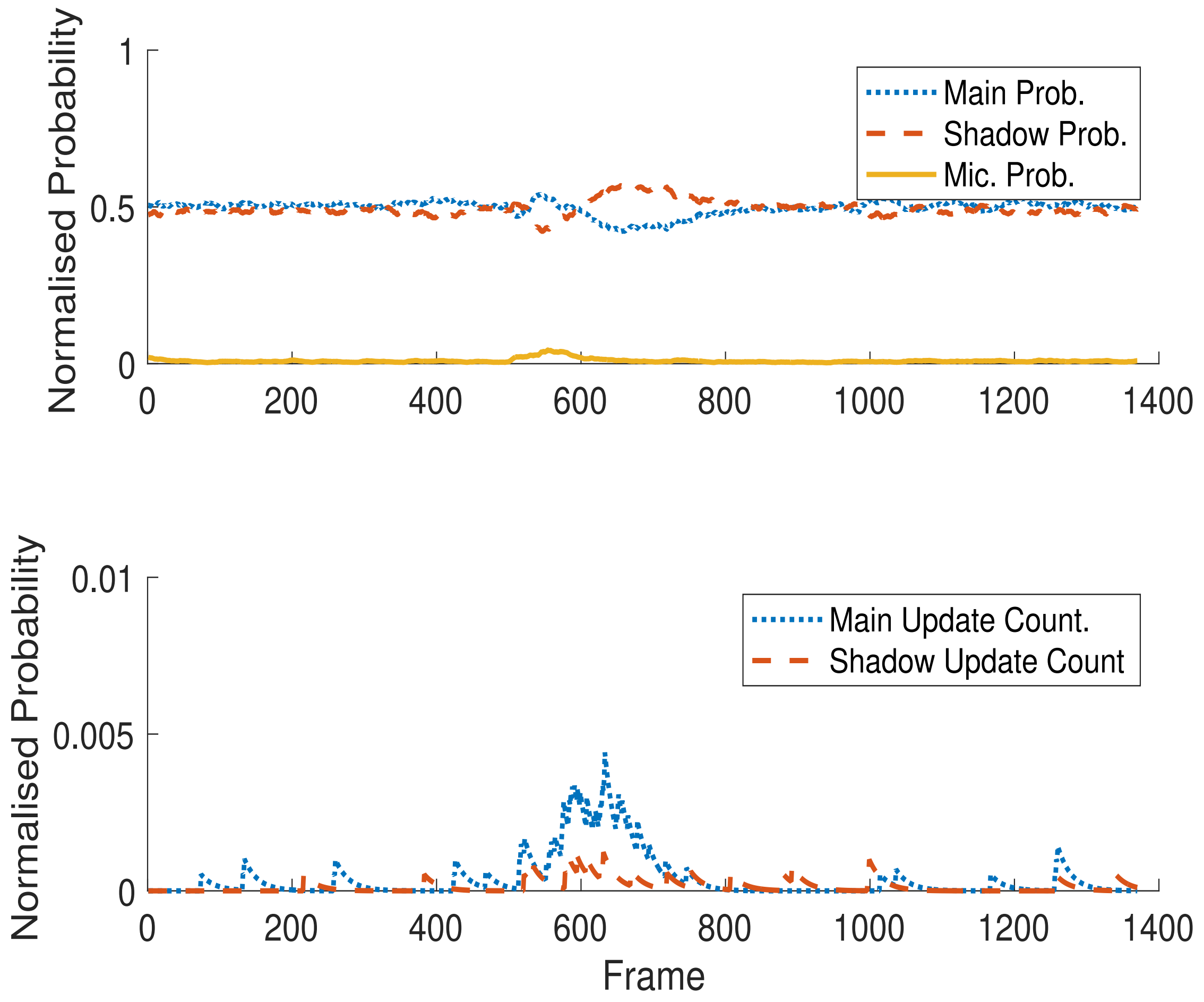

Fig. 3. Elements of the statistics vector during a double-talk event

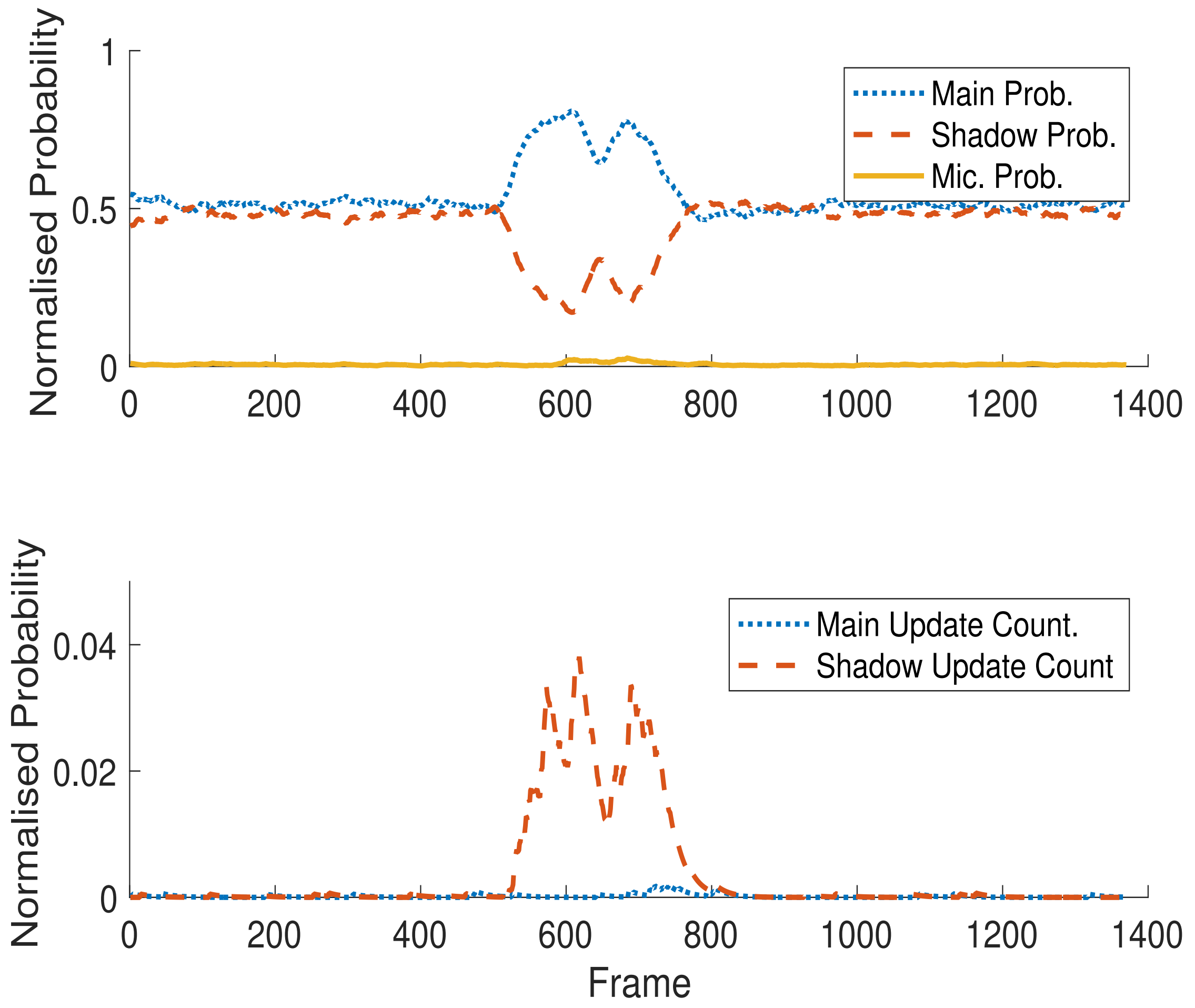

Fig. 5. Elements of the statistics vector when the device is repositioned

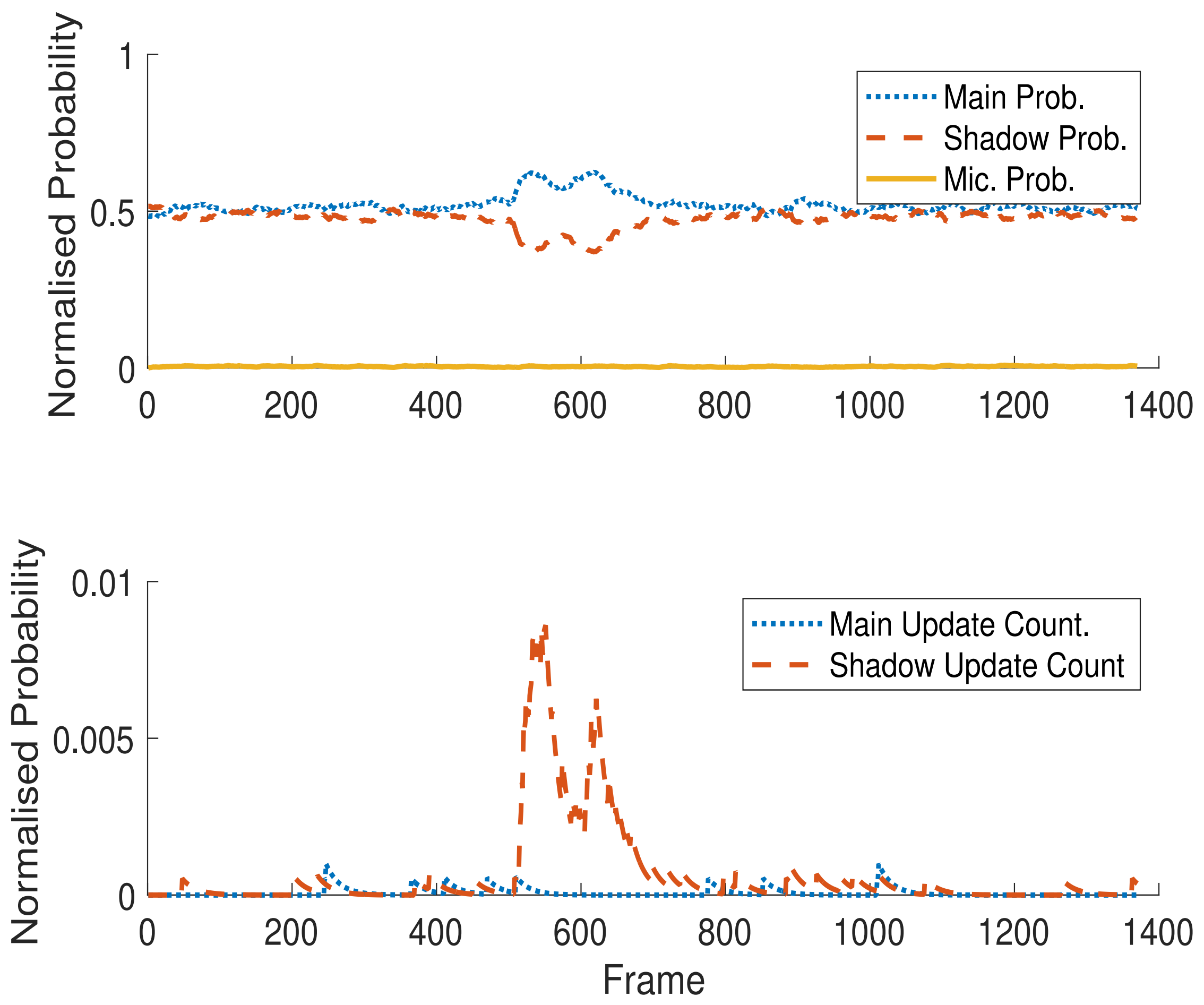

Fig. 4. Elements of the statistics vector during an echo path change

Acoustic State Estimation & Future Work

In Figure 6, the two-component t-SNE projection of the statistics vectors obtained from 184 real-world acoustic events are shown. Event classes are indicated by the color of the point. By inspection we can see that this preliminary investigation suggests that these features are suitable for acoustic state estimation due to the separability we can see.

The ensemble statistics that can be extracted from metadata within an SDMH-AEC come at insignificant cost. Nevertheless, they have strong discriminative power as they are derived from a multitude of diverse adaptive filtering algorithms and VSS schemes. Future work will investigate the use of more complex feature vectors as inputs to both classifiers and regressors for the purposes of detecting and estimating acoustic state.

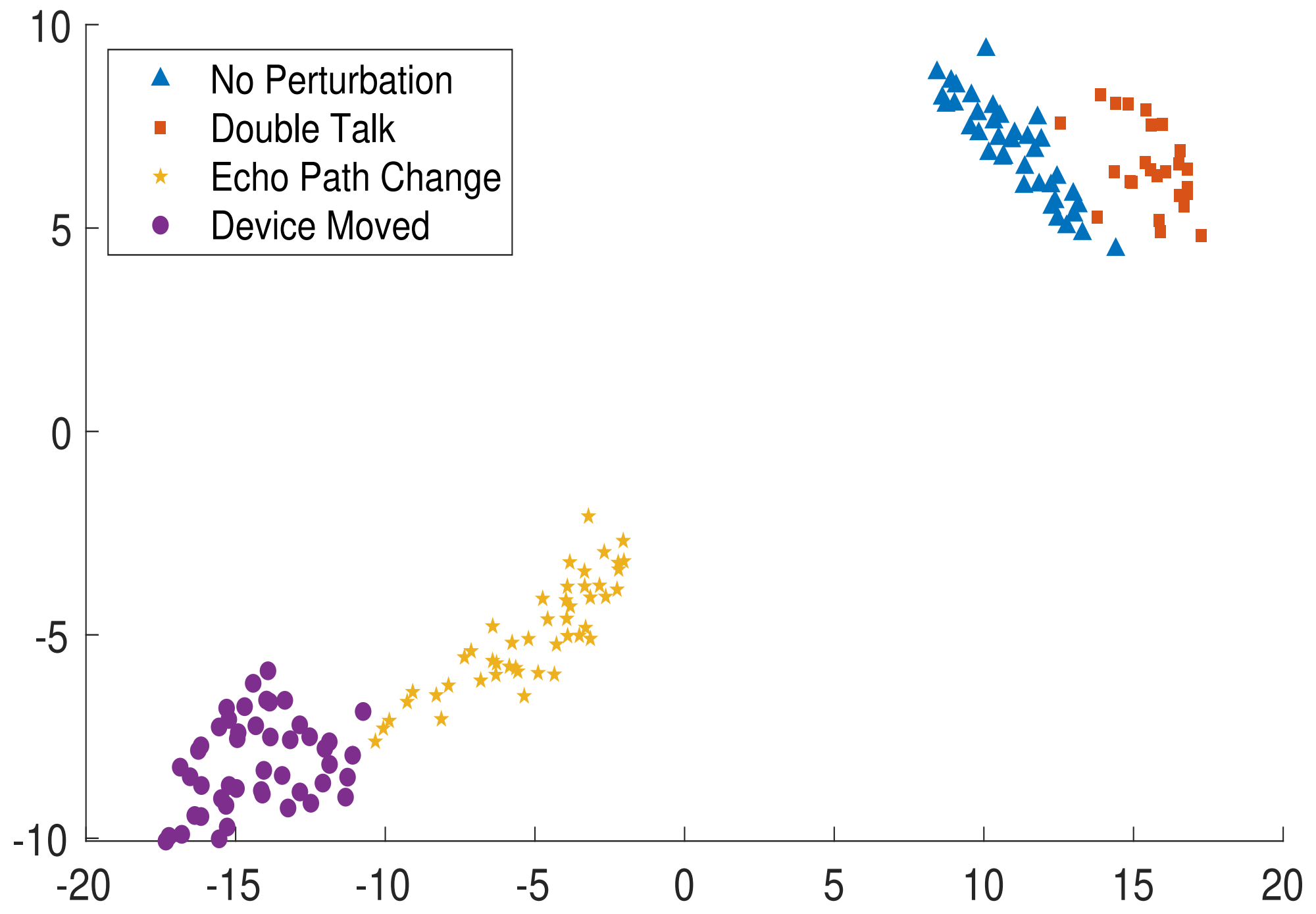

Fig. 6. The two-component t-SNE projection of the processed statistics vector extracted from 184 real world acoustic events