SSL for Circular Arrays of Mics

Yong Rui, Dinei Florêncio, Warren Lam, and Jinyan Su

Microsoft Research

ABSTRACT

Circular arrays are of particular interest for a number of scenarios, particularly because they can be placed in the center of the sources. That improves the sound capture due to the reduced distance. It also helps on the direction estimation, not only because of the reduced distance, but also because it increases the angle differences. Nevertheless, most research on circular arrays focused on the case of omni-directional microphones. In this paper we present a new algorithm for sound source localization developed specifically for directional microphones. Results obtained from real meeting room setups show a typical error of less than 3 degrees.

General model:

We model signal as reverb and additive noise:

$$x_i(n) = a_i s(n-D) + h_i(n) * s(n) + n_i(n)$$

Two traditional SSL algorithms:

- Steered Beam SSL:

$$p^* = \arg\max_l \left(\sum_{m=1}^{M} x_m(t - \tau_m^l) \right)^2$$

- ML 1-TDOA:

$$p^* = \arg\max_l \left\{ \sum_{r=1}^{M} \sum_{s \neq r}^{M} \left| W_{rs}(f) X_r(f) X_s^*(f) e^{-j2\pi f(\tau_r - \tau_s)} \right|^2 \right\}$$

$$W_{MLR} = \frac{|X_r||X_s|}{2q|X_r|^2|X_s|^2 + (1-q)|N_r|^2|X_r|^2 + |N_s|^2|X_s|^2}$$

A hybrid weighting function

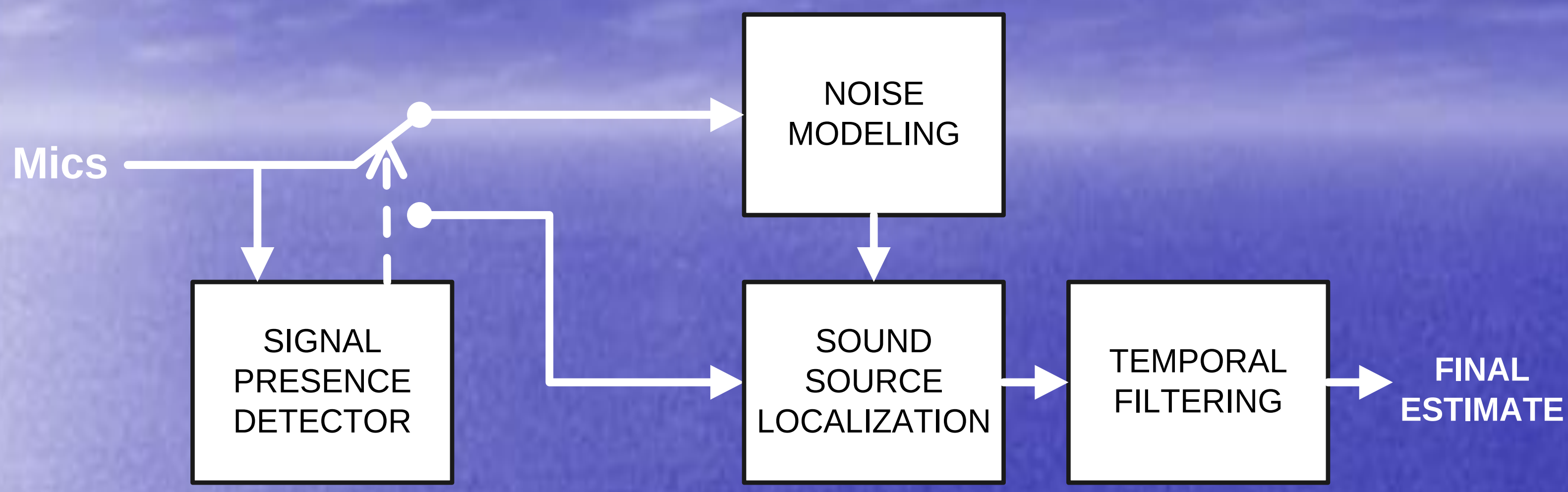
- Introducing a separable weighting function allows reduced complexity:

$$p^* = \arg\max_{l} \left\{ \sum_{r=1}^{M} \sum_{s \neq r}^{M} \left| W_r X_r W_s^* X_s^* e^{-j2\pi f(\tau_r - \tau_s)} \right|^2 \right\}$$

- Assumption of constant reverb and noise leads to separable weighting function:

$$W_m(f) = \frac{1}{q\,|X_m(f)| + (1-q)\,|N_m(f)|}$$

System Level Diagram



- SSL run only frames where speech is detected;
- Temporal filtering based on particle filtering;
- Noise modeling used to update the weighting function;

The Phase Problem

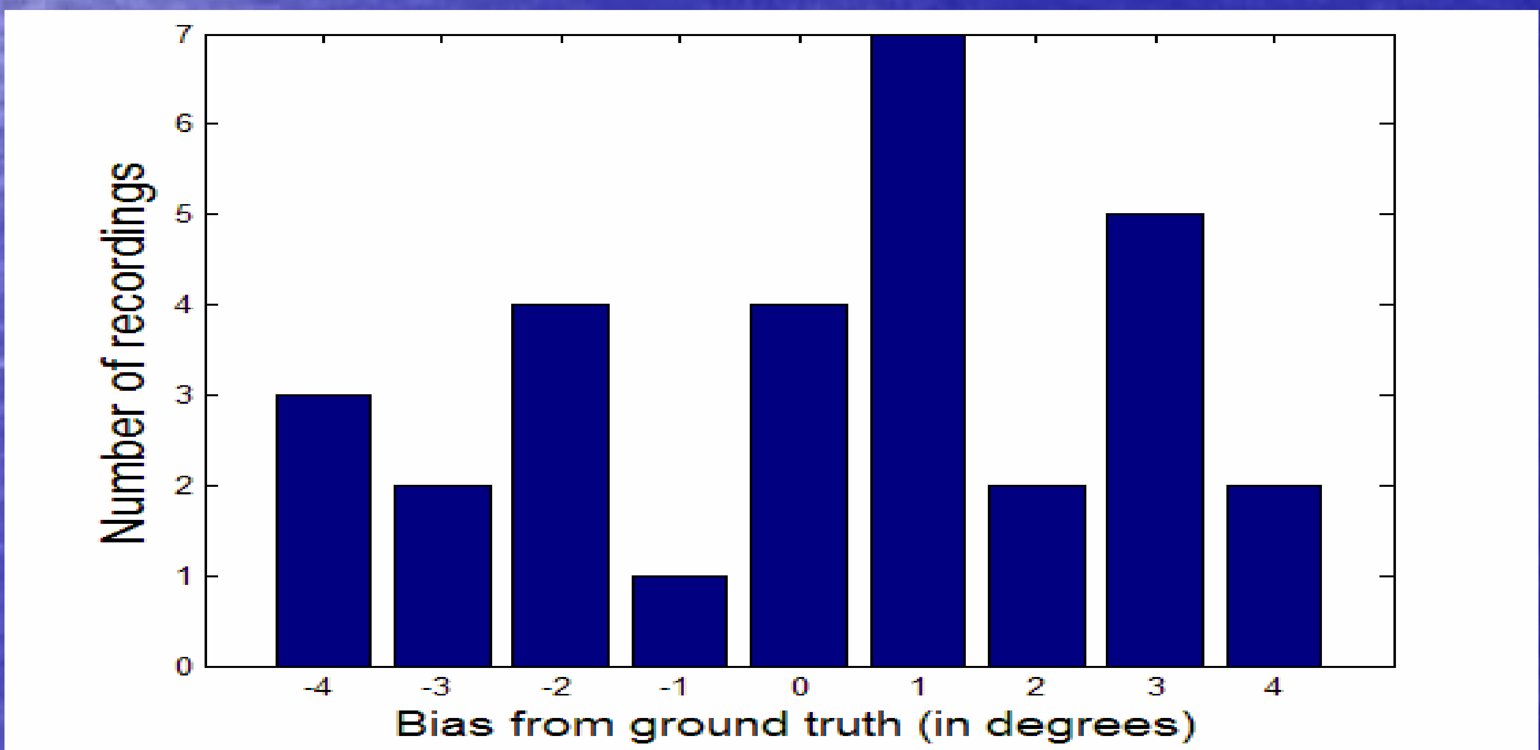
- Typical directional mics have strong phase response variance for non-frontal incidence;
- Phase response variability makes modeling unpractical;
- SOLUTION: use only mics in predictable DOA range, usually extending up to 90º or 120º;

	No Compensation			With Compensation			Use cutoff angle		
	Bias	**Std**	**#Fs**	**Bias**	**Std**	**#Fs**	**Bias**	**Std**	**#Fs**
T1	-1.9	1.2	249	29.8	79.8	195	-3.9	0.6	338
T2	0.0	0.0	298	156.9	21.2	54	0.0	0.0	320
T3	0.7	0.2	205	-146.3	94.6	229	0.8	0.0	282
T4	-0.6	0.6	112	-168.2	5.7	34	-2.6	0.2	429
T5	0.1	0.2	153	159.6	16.5	271	0.0	0.0	422
T6	7.0	29.9	66	160.0	4.0	426	2.5	0.1	419
T7	24.3	52.6	43	114.7	49.3	76	-2.7	0.4	351
T8	-0.2	0.4	161	-151.3	0.7	3	-1.0	0.4	333
T9	-3.6	0.6	92	-132.7	89.7	154	-3.8	0.4	450

Hardware

EXPERIMENTAL RESULTS

- 48 recordings;
- Six directional mics, 14cm radius;
- Rooms 3.6X6m to 5.4X12m;



Bias histogram. On the reference test set (single speaker), all biases were within 4º from the ground truth.

CONCLUSIONS

The proposed algorithm is based on:

- Preprocessing (VAD);
- Hybrid SSL function, incorporating Microphone Selection;
- Post-processing;

Typical errors around 3°.

References:

[1] Yong Rui, Dinei Florencio, Warren Lam, and Jinyan Su, "Sound source localization for circular arrays of directional microphones," in Proc. of *ICASSP*, 2005.

[2] Yong Rui and Dinei Florencio, "Time delay estimation in the presence of correlated noise and reverberation," in ICASSP, 2004.

[3] Yong Rui and Dinei Florencio, "New Direct Approaches to Robust Sound Source Localization," in Proc. of ICME, 2003.

[4] Dinei Florencio and Henrique Malvar, "Multichannel Filtering for Optimum Noise Reduction in Microphone Arrays," in Proc. of ICASSP, 2001.

[5] Bradford Gillespie, Henrique Malvar, and Dinei Florencio, "Speech Dereverberation via Maximum Kurtosis Subband Adaptive Filtering," in Proc. of ICASSP, 2001.

[6] Dinei A. Florencio, "Investigating the Use of Asymmetric Windows in CELP Vocoders," in Proc. of ICASSP, 1993.

[7] Dinei A. Florencio, "On the use of Asymmetric Windows for Reducing the Time Delay in Real-time Spectral Analysis," in Proc. of ICASSP, 1991.

[8] S. Birchfield and D. Gillmor, "Acoustic source direction by hemisphere sampling", Proc. of ICASSP, 2001.

[9] Cutler, R., et al. "Distributed Meetings: A Meeting Capture and Broadcasting System", Proc. of ACM Multimedia 2002.

[10] Kleban, J., Combined acoustic and visual processing for video conferencing systems, MS Thesis, Rutgers University, 2000.

[11] Wang, H., and Chu, P., Voice source localization for automatic camera pointing system in videoconferencing, In Proc. of ICASSP, 1997.

[12] Arlindo F. Conceicao, Jin Li, Dinei A. Florencio, and Fabio Kon, "Is IEEE 802.11 ready for VoIP?," in preparation.

[13] Xun Xu, Li-wei He, Dinei Florencio, and Yong Rui, "PASS: Peer-aware Silence Suppression for Internet Voice Conferences," in preparation.

[14] Amit Chhetri, Jack W. Stokes, and Dinei Florencio, Acoustic Echo Cancelation for High Noise Environments, in preparation.