LOUD: A 1020-Node Microphone Array and Acoustic Beamformer*

Eugene Weinstein[1], Kenneth Steele[2], Anant Agarwal[2,3], James Glass[3]

[1] Courant Institute of Mathematical Sciences
[2] Tilera Corporation
[3] MIT Computer Science and Artificial Intelligence Lab

* Based on work done at MIT CSAIL

Introduction

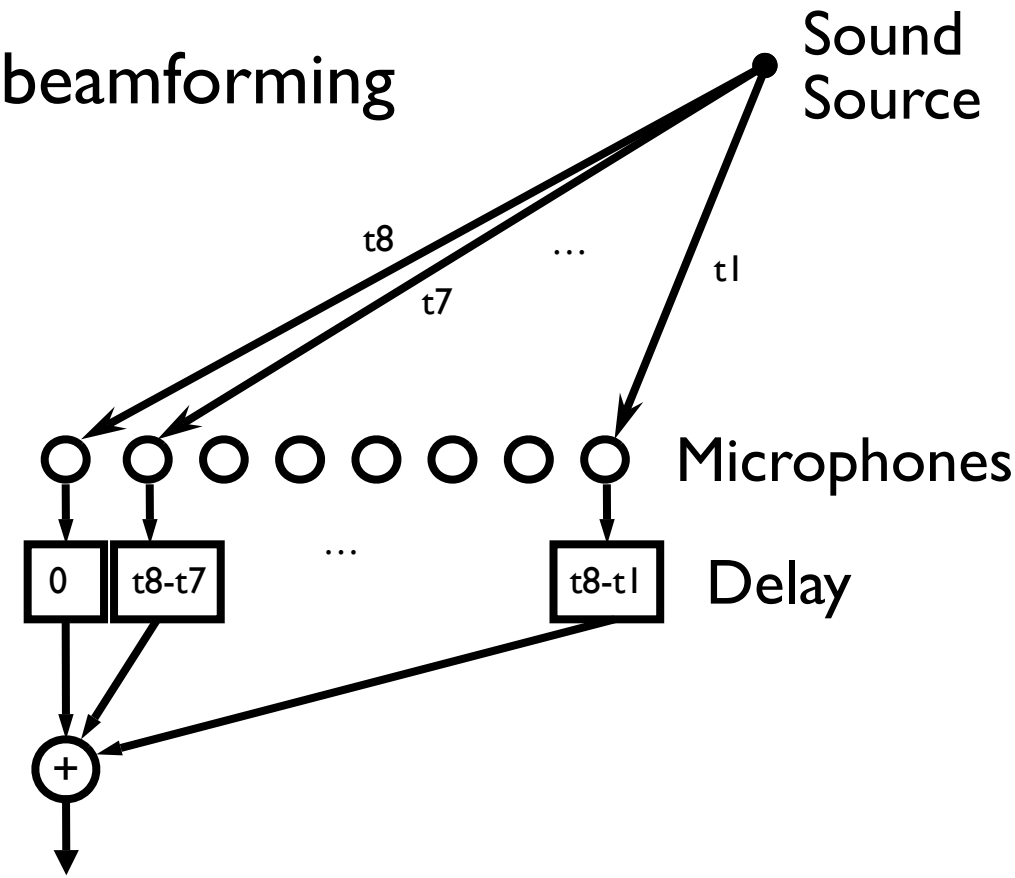
- Recording sound in high-noise settings is difficult
  - e.g., noisy lab or conference room
- Can use close-talking microphones (e.g., lapel mic)
- However, an untethered solution is more natural
- Idea: use software-steerable microphone arrays
  - Isolate and amplify sound using beamforming
  - Target application: speech recognition

Large Microphone Arrays

- Large acOUstic Data (LOUD) array: 1020 microphones
- Microphone array gain increases linearly with the number of microphones
- Past large-array speech recognition experiments scarce
- Processing large quantities of data in real-time is a compelling application for novel computing architectures
  - LOUD generates 400 Mbits/sec
  - We use Raw, a 16-tile parallel architecture

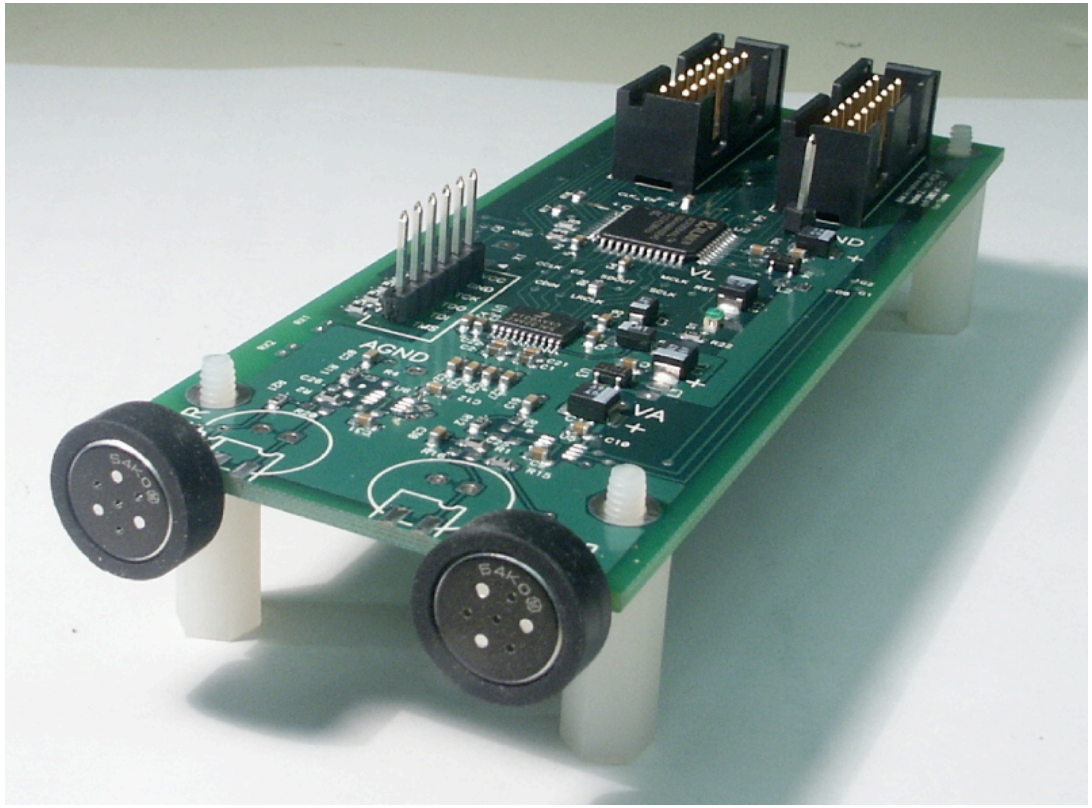
Acoustic Beamforming

- Selectively amplify a sound source at a particular location
- Take advantage of sound propagation through space
- Use simple delay-and-sum beamforming

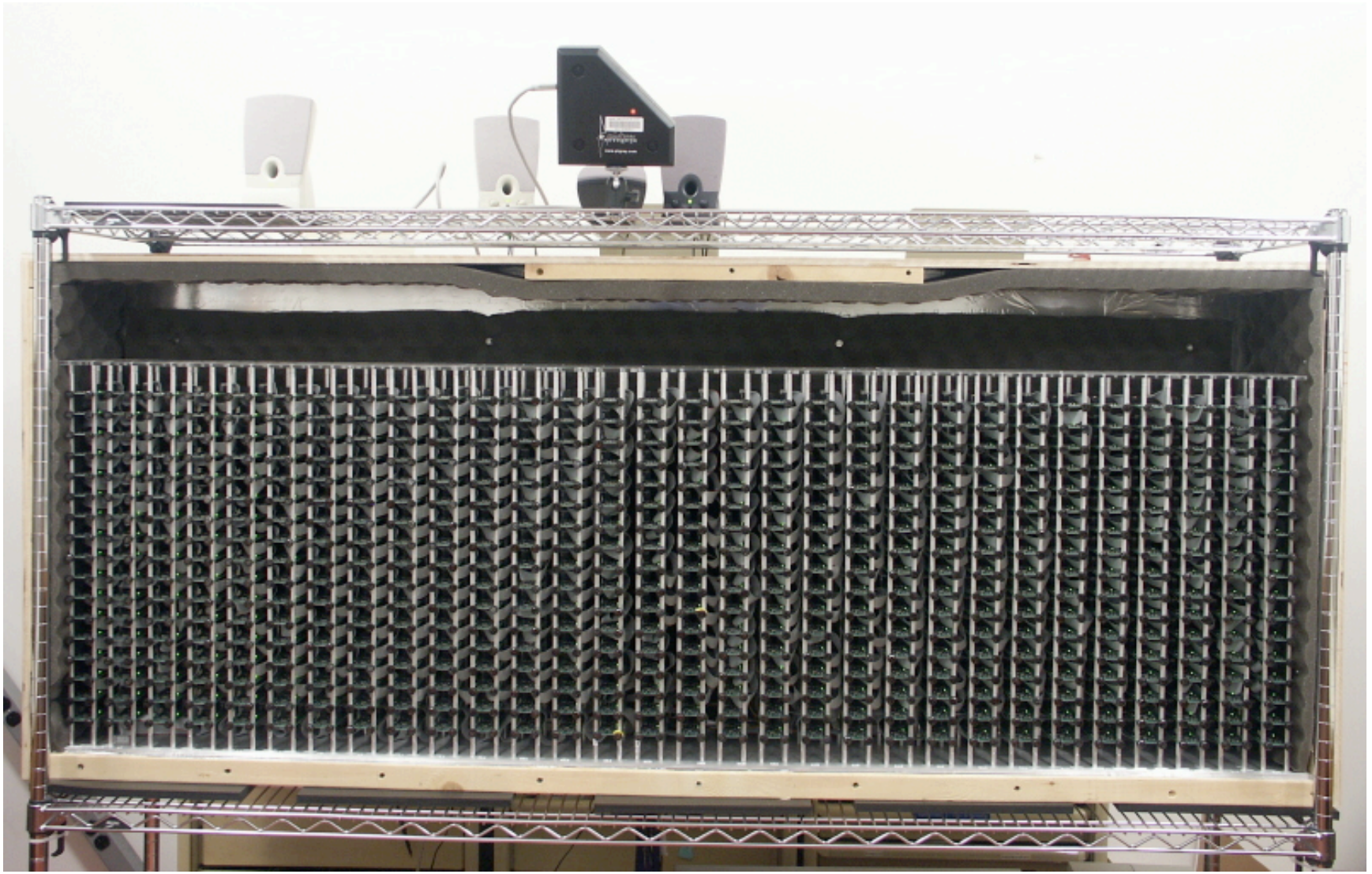


Two-microphone PCB

- On-board A/D converter feeds into CPLD
- Data streamed to CPU using time-division multiplexing



1020-Microphone Array



Microphone Positions

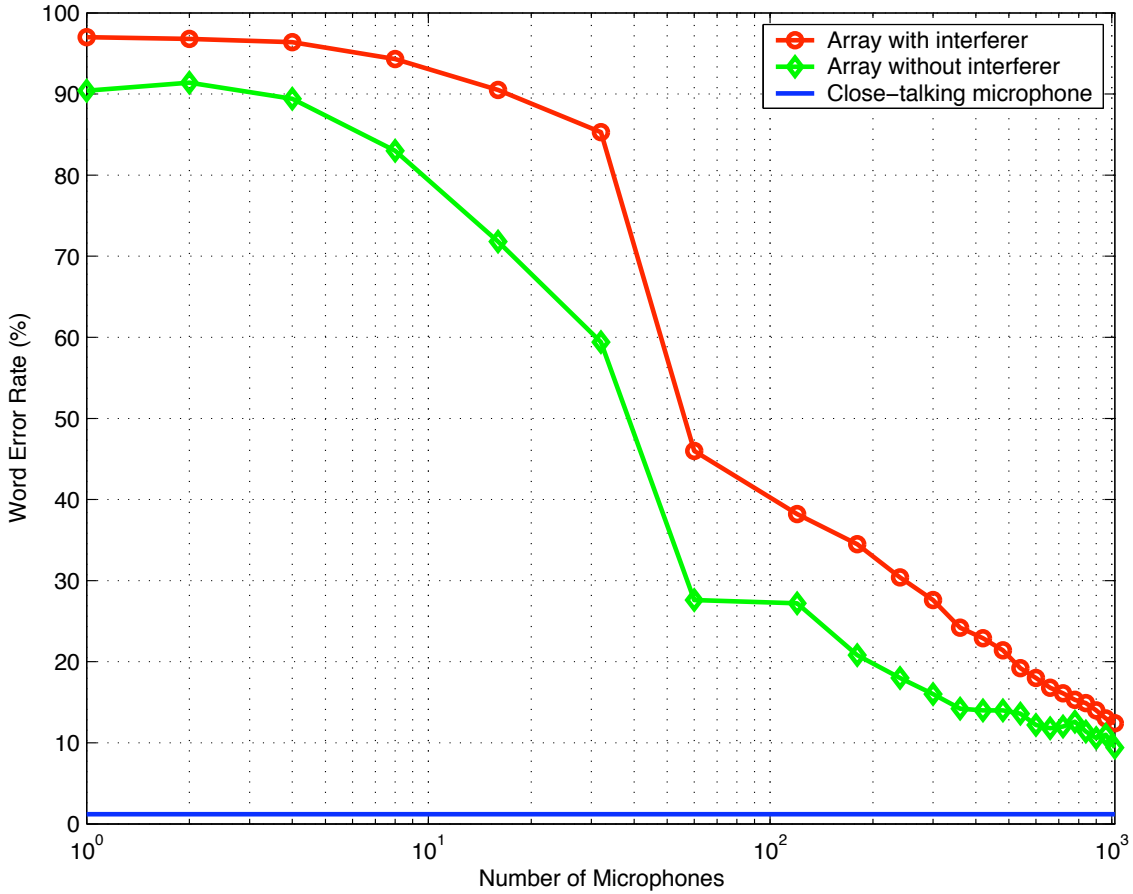
- Automated procedure to calibrate microphone positions
- Play a test audio "chirp" through a speaker
- Record with reference mic at speaker position and at each array mic
- Peak of cross-correlation function between reference, array microphones gives propagation delay
  - Solve for precise array geometry

Experiments

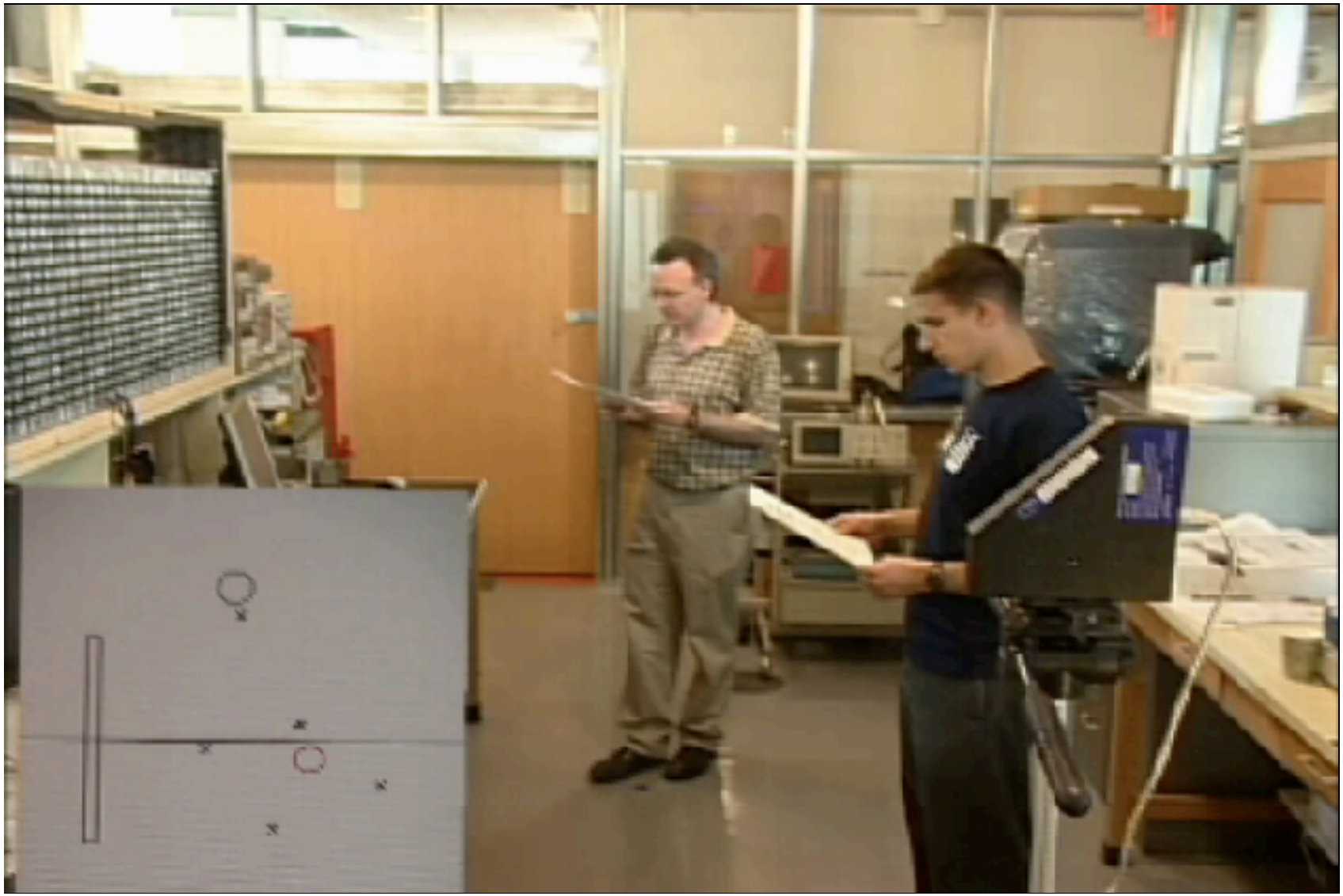
- Setting: extremely noisy hardware lab
  - Subject and "interferer" talking at the same time
- Goal: demonstrate that speech recognition accuracy improves with microphone array size
  - Speaker-independent recognizer for digit strings
- Record 150 utterances with interferer, 110 without
- Baseline: high quality close-talking mic, 80 utterances

Recognition Accuracy

- Word error rate (WER) decreases with array size
- WER drops by 87% (w/ interferer), 91% (no interferer) from one to 1020 mics
- Accuracy approaches close-talking microphone levels!



LOUD Demo



Summary/Future Work

- LOUD allows high-quality untethered recording in very noisy settings
- Speech recognition experiments demonstrate benefit of large arrays
- Future work:
  - Implement more sophisticated beamforming techniques
  - Automatic speaker tracking
  - Conduct more experiments with different geometries, noise settings