12.1 3D Ultrasonic Gesture Recognition

Richard J. Przybyla¹, Hao-Yen Tang¹, Stefon E. Shelton², David A. Horsley², Bernhard E. Boser¹

¹University of California, Berkeley, CA, ²University of California, Davis, CA

Optical 3D imagers for gesture recognition suffer from large size and high power consumption. Their performance depends on ambient illumination and they generally cannot operate in sunlight. These factors have prevented widespread adoption of gesture interfaces in energy- and volume-limited environments such as tablets and smartphones. Wearable mobile devices, too small to incorporate a touchscreen more than a few fingers wide, would benefit from a small, low-power gestural interface. Gesture recognition using sound is an attractive alternative to overcome these difficulties due to the potential for chip-scale size, low power consumption, and ambient light insensitivity. Using pulse-echo time-of-flight, MEMS ultrasonic rangers work over distances of up to a meter and achieve sub-mm ranging accuracy [1,2]. Using a 2-dimensional array of transducers, objects can be localized in 3 dimensions.

This paper presents an ultrasonic 3D gesture-recognition system that uses a custom transducer chip and an ASIC to sense the location of targets such as hands. The system block diagram is shown in Fig. 12.1.1. Targets are localized using pulse-echo time-of-flight methods. Each of the 10 transceiver channels interfaces with a MEMS transducer, and each includes a transmitter and a readout circuit. Echoes from off-axis targets arrive with different phase shifts for each element in the array. The off-chip digital beamformer realigns the signal phase to maximize the SNR and determine target location.

The 450µm diameter piezoelectric micromachined ultrasound transducers (pMUTs) used in this work are made up of a 2.2µm thick AIN/Mo/AIN/AI stack deposited on a Si wafer and released with a back-side through-wafer etch. The bottom electrode is continuous, while each pMUT has a top electrode lithographically defined to actuate the trampoline mode. Each pMUT can transmit and receive sound waves, and is operated at its resonance of 217kHz ± 2kHz with a bandwidth of 12kHz. The impedance of the transducers is dominated by the 10pF transducer capacitance, and the motional resistance at resonance is ~2.4M\Omega. The resonant frequencies of the pMUTs vary due to fabrication, temperature, and packaging stress, so online frequency tracking is used to maintain maximum SNR during operation.

Two pMUTs are used for transmission and seven for reception as illustrated in Fig. 12.1.1. The receive array is 3.5 wavelengths wide in the x-angle axis, allowing targets separated by more than 15° to be distinguished. In the y-angle axis the array is only 0.16 wavelengths wide, sufficient to determine the y-angle to the target by measuring the average phase difference along the y axis of the array. The center element of the receive array and the element 900µm above it are used to launch a $138\mu s \approx 24$ mm long pulse of sound into the environment. The transmit configuration illuminates a wide field of view, permitting the capture of an entire scene in a single measurement. Applications requiring better target resolution or greater maximum range can also use transmit beamforming at the expense of reduced measurement rate.

Each cycle begins with the launch of an acoustic pulse. Figure 12.1.2 shows the schematic of a single channel. High-voltage level shifters actuate the S_{Tx} transmit switches, setting the transducer's bottom electrode to 16V to permit bi-polar actuation of the transducer. The transmitter then excites the transducer with a $32V_{pp}$ square wave for 30 cycles at the transmit frequency f_{Tx} which is locked to $1/16^{\rm m}$ of the sampling frequency f_s . At the end of the transmit phase, the mechanical energy stored in the inertia of the pMUT dissipates and the pMUT rings down at its natural frequency. The S_{RX} receiver isolation switches are turned on, and a resistor converts the ringdown current to a voltage that is subsequently amplified and digitized by the receiver normally. The ringdown signal is then I/Q demodulated with f_{TX} . The slope of the phase signal during the ringdown indicates the frequency offset and is used to update the f_s and f_{TX} used in the next measurement. Figure 12.1.3 shows the offset measured by the frequency autotuning loop as it is enabled. An initial 57kHz offset frequency is nulled to 1kHz within 30 measurement cycles.

After 86μ s, the ringdown signal has decayed sufficiently for the S_{ring} switch to be opened, beginning the processing of received echoes. At this point, the signal from the transducer is integrated on the transducer's capacitance, and the front-end measures a voltage that is proportional to the displacement of the transducer's membrane.

The front-end amplifier consists of an open-loop current-reuse OTA with both NMOS and PMOS differential pairs biased near subthreshold for current efficiency. The front-end current is integrated onto the integrating capacitor of the second stage, which also makes up an integrator in the first of two switched capacitor resonators. Although the second stage is a switched capacitor integrator, the front-end current is processed in a continuous-time fashion before it is sampled at the output of the second integrator. As a result, the second integrator acts as an anti-aliasing filter for the wideband noise generated by the front-end and prevents this noise, the dominant noise source in the receiver architecture, from being aliased into the band of interest.

The signal then passes through a second switched-capacitor resonator and is quantized by a comparator. The high in-band gain provided by the 4th-order bandpass filter shapes the wideband quantization noise to be away from the signal at f_{TX} . The SC resonators are designed to resonate at 1/16 of the sampling frequency f_{s} , which is locked to the transducer's resonance by the ringdown autotuning circuit. This centers the bandpass $\Delta\Sigma$'s noise notch on the signal at f_{TX} .

The output of each $\Delta\Sigma$ ADC is I/Q demodulated, filtered, and downsampled offchip. A digital beamformer [3] processes the received signals to maximize the receive SNR and determine the x-angle location of the target. This process can be repeated in the orthogonal angle axis to implement 3D beamforming; in this work we forgo 3D beamforming since the tiny y-axis aperture does not provide any y-axis resolution.

Thermal noise in the front-end amplifier and the thermal motion of air limit the minimum detectable echo. The input-referred noise of the amplifier is $11nV/\sqrt{Hz}$, and the noise voltage of the transducer is $6nV/\sqrt{Hz}$ at resonance. Figure 12.1.4 shows the measured signal-to-noise ratio vs. range for a $127mm \times 181mm$ flat rectangular target. Figure 12.1.4 also shows the rms error in the range and direction measurement. Amplitude noise in the received signal limits the accuracy of the time-of-flight estimate. Figure 12.1.5 shows the output of the digital beamformer from a single measurement, which captures the echoes from a user's hands and head as he poses as shown. The system tracks objects between 45mm to 1m away and over an angular range of $\pm 45^\circ$. Echoes from targets at a range of 1m return after 5.8ms, and this sets the maximum measurement rate of the system at 172 frames per second (fps).

Figure 12.1.7 shows a micrograph of the readout IC, which is fabricated in a 0.18 μm CMOS process with 32V transistors. For a 1m maximum range, the system presented here uses 13.6 μJ per measurement. At 30 fps, the receive power consumption is 335 μW and the transmit power consumption is 66 μW . The energy consumption scales roughly linearly with maximum range. For a maximum range of 0.3m, the energy per frame is reduced to <0.5 μJ per channel per frame. Single-element range measurements can be conducted at 10 fps using only 5 μW .

Figure 12.1.6 compares the performance of this system to an earlier MEMS ultrasonic 1D rangefinder [4] and two recent optical 3D rangers [5,6]. This ultrasonic 3D rangefinder offers dramatically reduced energy consumption compared to optical methods while permitting 3D target tracking. The energy consumption trades off with performance, permitting continuous operation in even tiny mobile devices. These characteristics enable energy-efficient gestural interfaces in applications such as smartphones and tablets, and permit gestural user interfaces in tiny mobile devices too small to accommodate a conventional touchscreen.

References:

[1] R. Przybyla, *et al.*, "In-air Ultrasonic Rangefinding and Angle Estimation using an Array of AIN Micromachined Transducers," in *Proc. Hilton Head Workshop*, pp. 50-53, 2012.

[2] R. Przybyla, *et al.*, "A Micromechanical Ultrasonic Distance Sensor With >1 Meter Range," in *Transducers Dig. Tech. Papers*, pp. 2070-2073, 2011.

[3] M. Skolnik, *Introduction to Radar Systems*. 3rd edition, McGraw-Hill, 2001.
[4] C. Kuratli and Q. Huang, "A CMOS Ultrasound Range-Finder Microsystem," *VELP* 4, Colid Citat Circuits and 25, pp. 10, pp. 2005, 2017, Day 2005.

IEEE J. Solid-State Circuits, vol.35, no.12, pp. 2005-2017, Dec. 2000.

[5] W. Kim, *et al.* "A 1.5Mpixel RGBZ CMOS Image Sensor for Simultaneous Color and Range Image Capture," *ISSCC Dig. Tech. Papers*, pp. 392-393, Feb. 2012.

[6] O. Shcherbakova, *et al.*, "3D Camera Based on Linear-Mode Gain-Modulated Avalanche Photodiodes," *ISSCC Dig. Tech. Papers*, pp. 490-491, Feb. 2013.



Figure 12.1.3: Ringdown frequency offset measurement and tuning loop settling behavior.



Figure 12.1.5: Echo from user's hands and head when posing as shown. Color axis shows y-angle position of the targets. Beamformed data is thresholded at 12dB SNR.

Figure 12.1.4: Signal-to-noise ratio and target localization accuracy vs. range for 127mm×181mm flat rectangular target.

This work	[4]	[5]	[6]
Time-of-flight	Phase-shift	Phase-shift	Phase-shift
sound	sound	light	light
AIN pMUT	Si Thermal	Pinned	Avalanche
array	Ultrasound	Photodiode	Photodiode
1.6mm / 10mm	3.6mm / 250mm	850nm /	850nm /
1.00000 / 400000		15m	12m
0.18µm	0.8µm	0.13µm	0.35µm
45mm /	18mm /	200mm /	500m /
1m	0.11m	7m	5m
0.41mm _{rms} @	2.5mm @	7mm _{ms} @	19mm _{rms} @
0.5m	0.1m	1m	2m
90° / 90°	No angle	Not specified	18º /18º
	measurement		
14.8°/ NA	Single target	(480px x	0.28º / 0.28º
	only	360px)	(64px x 64px)
170 fps	50 fps	100 fps	200 fps
13.64µJ/frame	6.5mJ/frame	20mJ/frame	>2.9mJ/frame
	This work Time-of-flight sound AIN pMUT array 1.6mm / 48mm 0.18µm 45mm / 1m 0.41mmms 0.5m 90° / 90° 14.8° / NA 170 fps 13.64µJ/frame	This work [4] Time-of-flight sound Phase-shift sound AlN pMUT Si Thermal Ultrasound AlN pMUT Si Thermal Ultrasound 1.6mm / 48mm 3.6mm / 250mm 0.18µm 0.8µm 45mm / 18mm / 1m 0.11m 0.41mmms@ 2.5mm @ 0.5m 0.1m 90° / 90° No angle measurement 14.8° / NA Single target only 170 fps 50 fps 13.64µJ/frame 6.5mJ/frame	This work [4] [5] Time-of-flight sound Phase-shift Phase-shift AlN pMUT Si Thermal Pinned array Ultrasound Photodiode 1.6mm / 48mm 3.6mm / 250mm 850nm / 15m 0.18µm 0.8µm 0.13µm 45mm / 18mm / 200mm / 1m 0.11m 7m 0.41mm_ms@ 2.5mm @ 7mm _{ms} @ 0.5m 0.1m 1m 90° / 90° No angle measurement Not specified 14.8°/ NA Single target (480px x 360px) 170 fps 50 fps 100 fps 13.64µJ/frame 6.5mJ/frame 20mJ/frame

Figure 12.1.6: Comparison table.

ISSCC 2014 PAPER CONTINUATIONS

