A Sound Activity Detector Embedded Low-Power MEMS Microphone Readout Interface for Speech Recognition

Youngtae Yang
Dept. of Eletrical and Computer Engineering
Seoul National University
Seoul, Korea
youngtae.yang@analog.snu.ac.kr

Jun Soo Cho
Inter-university Semiconductor Research Center
Seoul National University
Seoul, Korea
junsoo.cho@analog.snu.ac.kr

Byunggyu Lee
Dept. of Eletrical and Computer Engineering
Seoul National University
Seoul, Korea
byunggyu.lee@analog.snu.ac.kr

Suhwan Kim
Dept. of Eletrical and Computer Engineering
Seoul National University
Seoul, Korea
suhwan@snu.ac.kr

Abstract— This paper presents a sound activity detector embedded low-power MEMS microphone readout interface for speech recognition. The proposed readout interface exploits the sound activity detector to automatically switch to active/standby mode depending on whether a sound activity is present or not. Since voice recognition applications are mostly in standby mode, standby power consumption is of greater importance than active power consumption. Our readout interface consumes only 14 μA in standby mode, which can significantly extend battery usage time. Also, a fast wake-up feature is provided. In active mode, the readout interface converts the microphone input signal to a high-resolution digital signal. The proposed circuit is fabricated in a 0.18 μm CMOS process. The measurement is performed using a differential piezo MEMS transducer. It achieves A-weighted signal-to-noise ratio of 62.6 dBA and a dynamic range of 104.5 dB.

Keywords— *Speech recognition, MEMS, microphone, readout interface, sound activity detector, standby mode*

I. INTRODUCTION

Speech recognition is becoming more important with the rapid growth of IoT technologies and applications. In the future, many smart devices will be controlled by voice. The process of general speech recognition is as follows. First, the sound is converted into digital data by the microphone, and the processor performs algorithms such as deep-learning to identify the sound. If the quality of the converted digital signal is low, speech recognition accuracy is degraded. Therefore, a microphone that generates high-quality signals with low noise is required. In addition, these microphones are typically used in battery-powered devices, so power consumption should be low.

Unfortunately, there exists a trade-off between performance and power consumption. It is difficult to design a microphone that satisfies both high performance and low power. In general, speech recognition applications spend most of the time waiting for a voice [1]. The actual time to perform speech recognition is only a fraction of the total operating time. Therefore, average power consumption can be drastically reduced by activating the microphone using a wake-up circuit only when a voice is present. Thus, reducing power consumption in standby mode is more important than reducing power consumption in active mode. In addition, a fast wake-up feature is required. If the

wake-up time is too long, the beginning portion of a voice input can remain undetected.

High-performance microphones have been reported in [2]–[4]. However, they should always be in active mode since a wake-up circuit is not integrated. Other previous studies have implemented a voice activity detector (VAD) for a wake-up feature. Previous works have implemented VADs with low-power LNAs and filter banks [5], [6]. However, they only provide wake-up functionality through voice activity detection and do not function as readout interfaces because analog-to-digital converters (ADC) are not integrated. Other studies have implemented VADs by using low power readout interfaces (8-bits ADC) and voice detection algorithms [7], [8]. However, they cannot obtain high-quality digitized voice data due to low ADC resolution. Therefore, they are not suitable for high-level speech recognition, such as understanding the meaning of words.

This paper proposes a low power, high-performance microphone readout interface for speech recognition applications that dramatically reduces overall power consumption by exploiting a sound active detector (SAD). The SAD always monitors the signal of the MEMS transducer and automatically determines whether to operate in active or standby mode. In active mode, the main part of the readout interface circuit converts the signal from the MEMS transducer into high-quality digital data for speech recognition. The proposed readout interface is designed for differential piezo MEMS transducers and demonstrated its performance and operation through measurements.

II. SYSTEM OVERVIEW

The architecture of the proposed microphone readout interface is shown in Fig. 1 (a). The input signal from a differential MEMS transducer is buffered by a low-noise source follower (SF) and converted to a digital signal by a delta-sigma ADC ($\Sigma\Delta$-ADC). A source follower is widely used as a MEMS interface circuit due to its high input impedance and simple structure. A delta-sigma ADC is suitable for MEMS microphone system because it can achieve high performance through oversampling. For better power efficiency, amplifying the output signal of the SF with a programmable gain amplifier
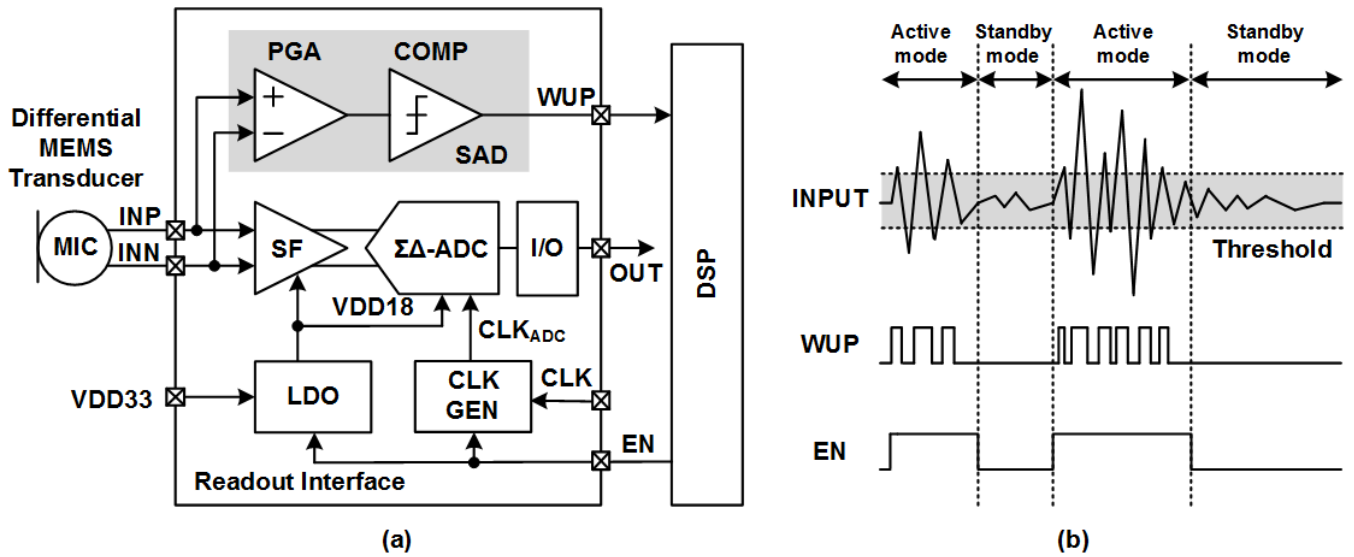
Fig. 1. (a) Proposed MEMS microphone system architecture. (b) Operation waveform of the proposed sound activity detector.

(PGA) may help to reduce the noise burden of the ΣΔ-ADC. However, if a large input signal is applied, the output signal would be clipped due to the supply voltage limit, which causes a significant distortion [4]. In this design, no gain amplifier is used to achieve a wide dynamic range. In addition, since the circuit connected to a differential MEMS transducer, it could be designed to take advantage of a differential structure. As a result, common-mode noise is rejected and excellent noise characteristics can be achieved. Also, a low dropout regulator (LDO) is integrated to provide a stable power supply despite external power supply fluctuation and to enable efficient power control of the circuit. The LDO regulates a 3.3V external supply (VDD33) and provides a regulated 1.8 V supply voltage (VDD18) to the internal circuitry. The external clock is provided to the clock generator, which generates the internal clock for

ADC. The clock gating technique was applied to disable the clock generator.

The active/standby mode switching process of the proposed microphone readout interface is described in Fig. 1 (b). The SAD amplifies the signal from the MEMS transducer and compares the magnitude of the signal with the threshold value to detect a potential sound activity. When a sound activity occurs, the SAD generates the spike-like output (WUP). The output signal is sent to the digital signal processor (DSP) and the DSP finally determines whether the received signal is an actual sound activity. At this time, the DSP performs only a simple sound activity detection algorithm, not a complicate speech recognition algorithm, so it can operate with minimal power consumption. If the DSP determines that sound activity is actually occurring, the enable signal (EN) activates the read interface and the DSP. Otherwise, they remain in standby mode. The power management of the readout interface is performed by enabling or disabling the LDO and the clock generator. The LDO and the clock generator are controlled by the enable signal (EN). In the disable state, the LDO output voltage (VDD18) becomes 0 V and the internal clock is stopped, minimizing the power consumption of the internal circuit.

III. CIRCUIT DESIGN

A. Sound Activity Detector

The most important design consideration of SAD is power consumption and wake-up time. Fortunately, SAD only detects potential sound activity, The noise characteristics of the SAD are not critical. Therefore, it can be implemented with a simple structure rather than complex structures in previous works. This enables fast wake-up and minimal power consumption. Fig. 2 shows the schematic of the designed SAD. It is shown as single-ended for simplicity. The SAD consists of SF, PGA, and comparator. SF buffers the input signal, and PGA amplifies the signal. The gain of the PGA is adjustable from 40 dB to 52 dB through 4-bit control code (GC[3:0]). The amplified signal is compared with the high threshold voltage (V_H) and the low threshold voltage (V_L) by two comparators, respectively. The
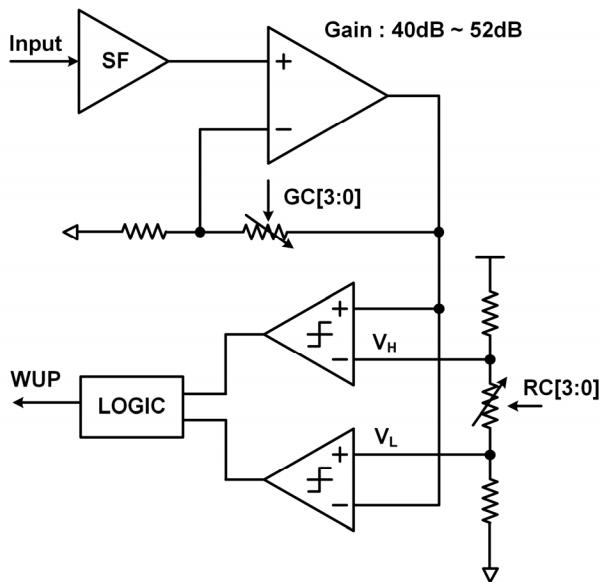

Fig. 2. Sound activity detector schematic

two threshold voltages can be tuned by the control code (RC[3:0]). The combination logic combines the two comparator outputs and generates the WUP signal. The total current consumption of SAD is only 10 μA.

B. Low-Noise Source Follower

The source follower that senses the output of the MEMS transducer is shown in Fig. 3. The MEMS transducer can be modeled as capacitors (C_{MIC}), and acoustic signals (S_{IN}) are converted to an electrical output signal (V_{IN}) by the MEMS transducer (C_{MIC}). The SF buffers the MEMS transducer output signal and drives the input capacitor of the ΣΔ-ADC. The SF also operates as a level-shifter, making the output common-mode voltage half of the power supply voltage. The input DC voltage of the SF is biased to ground by using back-to-back diodes as bias resistors (R_B). The C_P is the parasitic capacitance at node V_{IN}, generated by the input capacitance of SF and the bonding wire between MEMS and SF. The transfer function between S_{IN} and V_{IN} can be expressed as follows.

$$\frac{V_{IN}(s)}{S_{IN}(s)} = \frac{sR_B C_{MIC}}{1 + sR_B(C_{MIC} + C_P)} \qquad (1)$$

The transfer function is a high pass filter, and its cut-off frequency should be less than 20 Hz to avoid filtering out the audio band signal.

$$f_{cutoff} = \frac{1}{2\pi R_B(C_{MIC} + C_P)} < 20 Hz \qquad (2)$$

The value of C_{MIC} is determined by the connected MEMS transducer characteristic and cannot be arbitrarily set. Thus, a huge bias resistor (R_B) is required to prevent loss of low-frequency signals. The sum of C_{MIC} and C_P is usually around 1 pF, and the required resistance is over 10 GΩ. Also, as the resistance value increases, the noise due to the resistance decreases [9]. Back-to-back diodes are used instead of conventional resistors to implement large resistance. The gain of the transfer function can be expressed as follows.
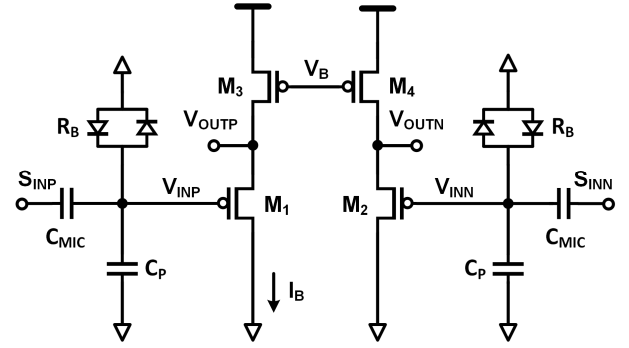

Fig. 3. Low-noise source follower schematic

$$Gain = \frac{C_{MIC}}{C_{MIC} + C_P} \qquad (3)$$

The smaller the C_P, the higher the output gain can be obtained, which is advantageous to achieve a high SNR. The noise components of SF are mainly flicker noise and thermal noise. To reduce flicker noise, input devices (M_1, M_2) must be large enough. However, this causes an increase in parasitic capacitance, resulting in the output gain reduction. In other words, there exists a trade-off between input magnitude and flicker noise. The size of the input devices is optimized to maximize the SNR. Thermal noise is determined by the bias current (I_B). The bias current was set to 225 μA for each pair, designed to meet both drive capability and thermal noise budget. The bias noise of V_B is suppressed by the differential structure.

C. Delta-Sigma Modulator

The schematic of the single-ended version of the delta-sigma modulator (ΣΔM) is shown in Fig. 4. A 3rd order feed-forward 12-level structure with an oversampling ratio of 256 is chosen to fulfill the target noise specification. The coefficients of the modulator were optimized through MATLAB behavior simulation. The modulator is implemented as a switched
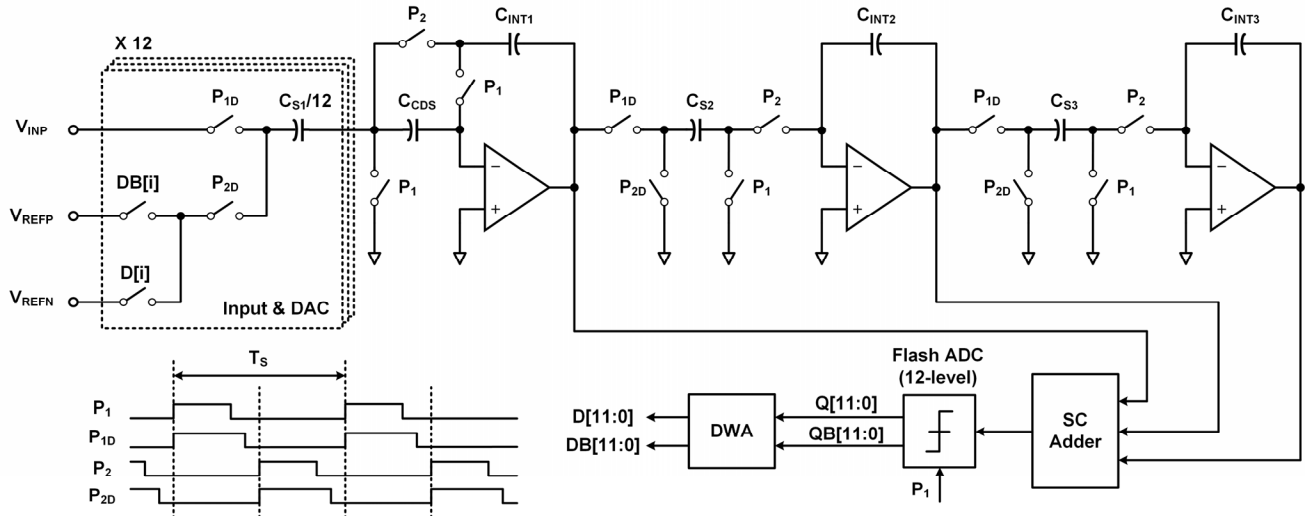

Fig. 4. Simplified schematic of the delta-sigma modulator

capacitor to realize accurate filter coefficients and to prevent performance degradation due to clock jitter. Non-overlapping clocks are employed to avoid side effects such as charge injection. A feed-forward topology is used to reduce the operational transconductance amplifiers (OTA) output swing. Summation of the feed-forward signal is performed using a conventional switched-capacitor adder. The summation signal is converted to a 12-bit thermometer code by a 12-level flash ADC and fed back to the first integrator through 12 capacitive DACs. A data-weighted averaging (DWA) is used to prevent non-linearity of DAC due to capacitor mismatch.

In order to satisfy the thermal noise specification (kT/C noise), the sampling capacitance of the 1st integrator (C_{S1}) is designed to be 15 pF [10]. The first OTA consumes significant current to drive large loading capacitors. Since the output swing is small, the current consumption can be optimized using a 1st-stage folded-cascode OTA and its current consumption is 1.08 mA. The noise of the second and third integrators is filtered by the first integrator, which enables to relax the capacitor size and OTA consumption current. The sampling capacitors for the second and third integrators are 0.2 pF and 0.1 pF, respectively, and the dissipation currents are both 84 µA. In addition, the correlated double sampling (CDS) technique is applied to remove flicker noise and the CDS operation is performed through the C_{CDS}.

The output of the modulator (D[11:0]) is transmitted to the decimator. The decimator converts the modulator output into a 24-bit audio band signal by out-band signal filtering and down-sampling. The decimator is designed by using a cascaded integrator-comb (CIC) filter and a finite impulse response (FIR) filter. The 24-bit output signal of the decimator is transmitted to the outside as a digital bitstream of the I²S and S/PDIF standard.

D. Low Dropout Regulator

The LDO is integrated for stable power supply and efficient power management. The schematic of the LDO is shown in Fig. 5. The LDO generates the internal 1.8V voltage (VDD18) from the external 3.3V supply (VDD33). A reference voltage of 1.2V (V_{REF}) is given from the bandgap reference (BGR). The resistance ratio is set to 2:1 to generate an internal supply voltage of 1.8V. An external capacitor is used to stabilize the LDO and
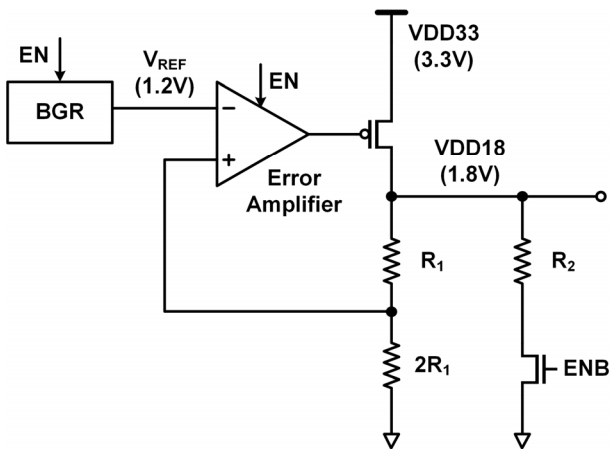
reduce the ripple of the output voltage. As mentioned before, the LDO is activated or deactivated by the SAD. When the enable signal (EN) becomes low and deactivated, the internal supply voltage is pulled down to ground via R_2. At the same time, the error amplifier and BGR were also deactivated to minimize the total current consumed by the LDO.

IV. EXPERIMENTAL RESULTS

The proposed readout interface is fabricated in a 0.18 µm CMOS process. Microphotograph is shown in Fig. 6 and the die size is 3.066 mm². The SAD occupies an area of only 0.12 mm². Fig. 7 shows the behavior of the SAD and its corresponding output response of the LDO when the word "hello" is pronounced in the silence state. In response to the voice, the SAD generates spike-like signals (WUP), and the DSP transmits
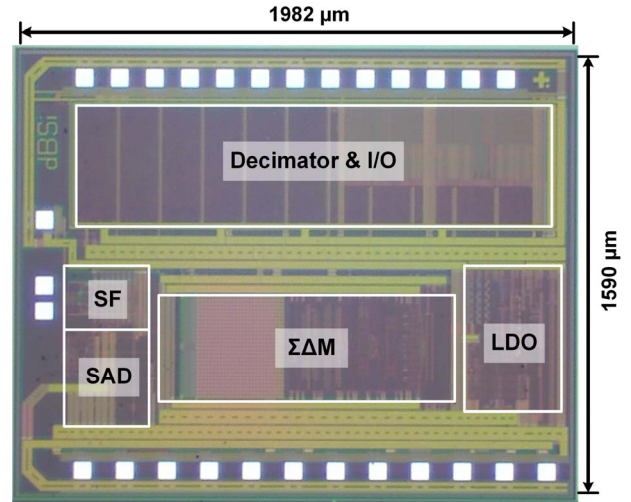
Fig. 6. Readout interface microphotograph
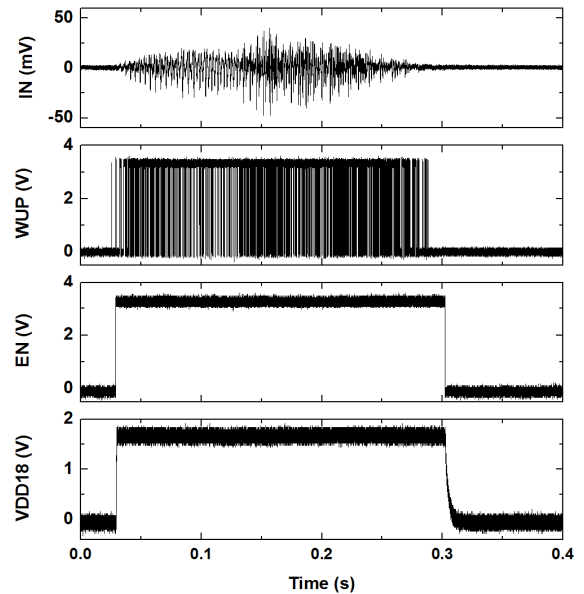
Fig. 5. Low dropout regulator schematic

Fig. 7. Measured response of the readout interface to the voice

the enable signal (EN) to the readout interface. As a result, the output voltage of the LDO rises from 0 V to 1.8 V and the readout interface becomes active. As shown in Figure 7, the LDO is enabled as soon as the sound activity occurs. The wake-up time from when sound activity occurs to when the supply voltage stabilizes at 1.8V is 0.76 ms. The power on settling time of the internal circuit is less than 100us, which is negligible compared to the wake-up time of the LDO. When it becomes silent again, the LDO output voltage fall to 0 V and the readout interface operates in standby. In standby mode, the total measured current consumption of the readout interface including SAD and other circuitry is only 14 μA.

The measurement of readout interface including MEMS transducer was conducted in an anechoic box for accurate measurements. Audio precision (AP2722) drives the speaker to generate the input and captures the digital output bitstream of the readout interface. Fig. 8 shows the power spectral density (32768 points) of the readout interface including the MEMS
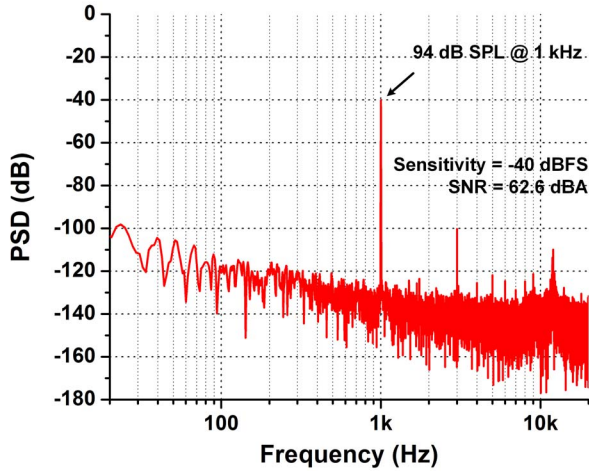
transducer within the audio band when acoustic input of 1 kHz 94 dB SPL (1 Pa) is applied. The influence of flicker noise of SF appears at low frequency. The third harmonic distortion is from the MEMS transducer. A resonance peak due to the resonance frequency of the MEMS transducer is observed at 12 kHz. From the measured output signal magnitude, the sensitivity of our microphone is -40 dBFS/Pa and its measured A-weighted SNR is 62.6 dBA.

The electrical measurement of the readout interface also performed. For electrical measurements, the electrical inputs are applied directly to the readout interface. When there is no input, the readout interface shows the total in-band noise power of -108 dBFS which correspond to 68 dBA performance at 94 dB SPL. Fig. 9 plots the A-weighted SNR and SNDR versus input amplitude. The measured dynamic range is 104.5 dB. The SNDR decreases from -20 dBFS (= 114 dB SPL) due to the non-linear characteristic of SF. Considering that the 114 dB SPL is fairly loud, the drop in SNDR is unimportant. Note that the
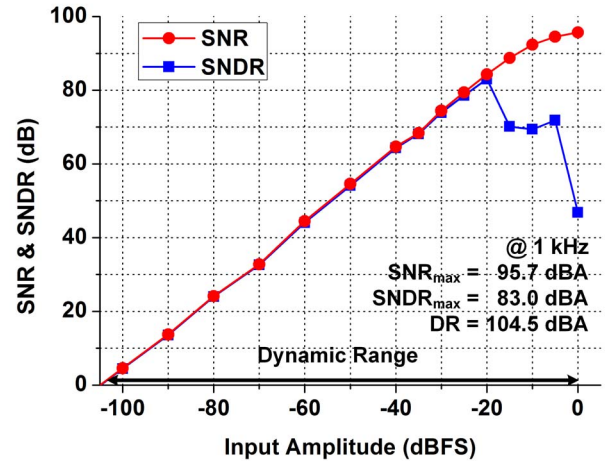

Fig. 8. Measured power spectrum density of the readout interface with MEMS transducer (32768 FFT points)

Fig. 9. Measured A-weighted SNR/SNDR versus input amplitude (electrical measurements)

TABLE I.   COMPARISON WITH PREVIOUS DIGITAL MEMS MICROPHONES

Reference	This work	ESSCIRC [2]	ICECS [3]	ISSCC [4]
MEMS device type	Differential Piezoelectric	Single-Ended Capacitive	Single-Ended Capacitive	Differential Capacitive
Process	0.18 μm	0.18 μm	0.25 μm	0.13 μm
Supply voltage	3.3 V	1.8 V	1.8 V	1.8 V
Sensitivity @ 1 Pa	-40 dBFS	-42 dBFS	-26 dBFS	-46 dBFS
SNR @ 1 Pa	62.6 dBA	63 dBA	63 dBA	67 dBA
Dynamic Range	**104.5 dBA**	80 dBA	83 dBA	113 dBA
Active Current	2.26 mA	0.46 mA	0.47 mA	1.2 mA
Standby Current	**14 μA**	N/A	N/A	N/A
Average Current (standby ratio = 0.9)	**0.24 mA**	0.46 mA	0.47 mA	1.2 mA
Wake-up time	**0.76 ms**	N/A	N/A	N/A

distinction in the electrical measurement at -40 dBFS (= 94 dB SPL) is negligible. In other words, the nonlinearity of SF does not matter because the non-linearity of the MEMS transducer is dominant when the signal is large enough. Table 1 compares the performance summary of this work with previous digital MEMS microphones. Our readout interface and [4] provide a much wider dynamic range (over 100 dBA) compared to [2], [3]. The active power consumption, excluding the digital interface, is 2.26 mA, slightly higher than [4]. Among those papers, however, only our circuit supports standby mode. Thus the average power consumption is lower than other previous works assuming that the activation time is 10% of the total operating time (standby ratio = 0.9).

V. CONCLUSION

A readout interface with built-in SAD for speech recognition is presented. The proposed readout interface provides high-quality readout feature and supports mode conversion (active/standby) by exploiting SAD. The SAD detects sound activity and automatically switches the readout interface to active or standby mode. In standby mode, the total current consumption is 14 μA, which significantly reduces average power consumption. In addition, our readout interface supports fast wake-up. The measured wake-up time is 0.76 ms. A low-noise SF and a ΣΔM were implemented to convert sound into high-resolution digital data for speech recognition. The proposed circuit is fabricated in a 0.18 μm CMOS process. The measurement is performed using a differential piezo MEMS transducer, and it achieves A-weighted SNR of 62.6 dBA and dynamic range of 104.5 dB.

ACKNOWLEDGMENT

REFERENCES

[1] F. Höflinger, G. U. Gamm, J. Albesa, and L. M. Reindl, "Smartphone remote control for home automation applications based on acoustic wake-up receivers," in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pp. 1580–1583, May 2014.

[2] S. A. Jawed, D. Cattin, M. Gottardi, N. Massari, A. Baschirotto, A. Simoni, "A 828μW 1.8V 80dB dynamic-range readout interface for a MEMS capacitive microphone," *Proc. on European Solid-State Circuits (ESSCIRC)*, pp.442–445, Sept. 2008.

[3] A. Barbieri, G. Nicollini, "A 470μA Direct Readout Circuit for Electret and MEMS Digital Microphones," *International Conference on Electronics, Circuit, and Systems (ICECS)*, pp. 341–344, Dec. 2013.

[4] M. Cho, et al., "A 1.8V True-Differential 140dB SPL Full-Scale Standard CMOS MEMS Digital Microphone Exhibiting 67dB SNR," *International Solid-State Circuits Conference (ISSCC)*, pp. 166–167, Feb. 2017.

[5] T. Delbruck, T. Koch, R. Berner, and H. Hermansky, "Fully integrated 500μW speech detection wake-up circuit," *International Symposium on Circuits and Systems (ISCAS)*, pp. 2015–2018, May. 2010.

[6] M. Cho, et al., "A 1μW Voice Activity Detector Using Analog Feature Extraction and Digital Deep Neural Network," *International Solid-State Circuits Conference (ISSCC)*, pp. 346–348, Feb. 2018.

[7] S. Jeong, et al., "A 12nW always-on acoustic sensing and object recognition microsystem using frequency-domain feature extraction and SVM classification", *International Solid-State Circuits Conference (ISSCC)*, pp. 362–363, Feb. 2017.

[8] M. Cho, et al., "A 142nW Voice and Acoustic Activity Detection Chip for mm-Scale Sensor Nodes Using Time-Interleaved Mixer-Based Frequency Scanning," *International Solid-State Circuits Conference (ISSCC)*, pp. 278–279, Feb. 2019.

[9] S. Ersoy, R. H. M. van Veldhoven, F. Sebastiano, K. Reimann, and K. A. A. Makinwa, "A 0.25mm2 AC-biased MEMS microphone interface with 58dBA SNR," *International Solid-State Circuits Conference (ISSCC)*, pp. 382–383, Feb. 2013.

[10] R. Schreier, et al., "Design-Oriented Estimation of Thermal Noise in Switched-Capacitor Circuits," *IEEE Trans. Circuits and Systems I, vol. 52,* no. 1, pp. 2358–2368, Nov. 2005.