mmProcess: Phase-Based Speech Reconstruction from mmWave Radar

Hyeongjun Choi
Korea University
tiger8135@korea.ac.kr

Young Eun Kwon
Korea University
equation@korea.ac.kr

Ji Won Yoon
Korea University
jiwon_yoon@korea.ac.kr

***Abstract*—This paper presents mmProcess, a novel phase-based approach for speech reconstruction using millimeter-wave (mmWave) technology, offering an alternative to existing Doppler-based and deep learning-dependent methods. By leveraging the phase variations in mmWave signals, mmProcess enables precise detection of fine vibrations caused by sound, facilitating accurate speech reconstruction without the need for large training datasets, prior knowledge, or complex neural networks. This eliminates the limitations of deep learning approaches, such as degraded performance with unseen languages and the significant time and cost required for system development.**

mmProcess combines advanced signal processing techniques, including range processing, phase unwrapping, and noise filtering, to transform raw mmWave radar data into high-fidelity speech signals. Experimental evaluations validate the effectiveness of the method, demonstrating its capability to operate in challenging scenarios while maintaining adaptability and cost efficiency.

I. Introduction

The proliferation of millimeter-wave (mmWave) technology has revolutionized wireless communication and sensing applications, offering high data rates, low latency, and improved spatial resolution. Beyond its traditional use in radar systems and high-frequency communication [1], mmWave signals have shown significant potential for non-traditional applications, such as gesture recognition [2], health monitoring [3], [4], and even audio reconstruction [5].

In this paper, we introduce mmProcess, a novel phase-based approach for speech reconstruction utilizing mmWave technology, offering an alternative to existing mmWave-based Doppler methodologies [6]. The proposed method enables the detection of finer vibrations, facilitating accurate speech reconstruction without reliance on deep learning models or prior knowledge. In contrast to existing Doppler-based techniques which employ deep learning, our approach addresses critical limitations, such as degraded performance when reconstructing speech in languages that are not included in the training data, and the substantial time and resource investment required for system development. By eliminating the dependence on deep learning, the proposed method provides a more adaptable, cost-effective, and efficient solution for mmWave-based speech reconstruction. The overall is in Fig. 1.

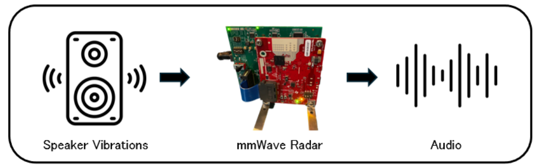

Fig. 1. Overall diagram of mmProcess

The key contributions of this paper are as follows:

- We propose a novel phase-based approach for speech reconstruction using mmWave technology.
- The methodology does not require deep learning, making it adaptable to new languages while reducing system development costs.
- Experimental evaluations demonstrate the effectiveness of our approach in detecting fine vibrations and reconstructing speech signals.

II. Background

A. mmWave

mmWave, a subset of electromagnetic waves, operates within the frequency range of approximately 30 GHz to 300 GHz, corresponding to wavelengths between 1 mm and 10 mm. The short wavelength of mmWave enables the utilization of high-frequency signals, facilitating compact antenna designs and highly precise signal processing. Moreover, its broad frequency bandwidth supports exceptionally high data transmission rates, making it particularly well-suited for advanced wireless communication technologies such as 5G and Wi-Fi 6E.

Despite its advantages, mmWave signals exhibit significant propagation challenges due to their high frequency. They are characterized by strong linear propagation and are unable to easily penetrate physical obstacles such as walls and trees, leading to signal reflection and absorption. Additionally, substantial signal attenuation occurs in the air, resulting in a limited transmission range. To overcome these limitations, techniques such as beamforming and the use of repeaters are commonly employed to enhance signal coverage and maintain communication quality.

Workshop on Security and Privacy in Standardized IoT (SDIoTSec) 2025
24 February 2025, San Diego, CA, USA
ISBN 979-8-9919276-6-6
https://dx.doi.org/10.14722/sdiotsec.2025.23xxx
www.ndss-symposium.org

These unique characteristics have positioned mmWave as a pivotal technology across a range of applications. Its capacity for ultra-fast data transmission and low-latency communication has made it a cornerstone in 5G networks and Wi-Fi 6E/7 systems. Furthermore, mmWave radar is highly reliable in adverse weather conditions, enabling high-resolution obstacle detection and distance measurement, which are critical for autonomous driving technologies. Beyond communication and sensing, mmWave has also been adopted for non-invasive monitoring of physiological signals, including heart rate and respiratory rate, further expanding its applicability in healthcare and monitoring systems.

B. Speech extraction

Chenhan Xu et al. introduced WaveEar technology, which uses mmWave to detect skin vibrations near the vocal cords and restore voice [5]. Chao Wang et al. proposed mmPhone, a technology that can eavesdrop on voice even in soundproof environments by utilizing piezoelectric effects and mmWave [7]. Wei-Han Chen et al. presented a technology that reconstructs speech by detecting minute vibrations induced on surrounding objects by sound using mmWave [8]. Chao Wang et al. and Suryoday Basak et al. proposed mmEve and mmSpy, respectively, which enable remote eavesdropping on speech emitted from smartphone earpiece mode using mmWave technology [9], [10]. Pengfei Hu et al. proposed MILLIEAR, a system that uses mmWave technology to detect minute vibrations of a speaker and reconstruct entire conversations [11]. Pengfei Hu et al. also introduced mmEcho, which utilizes mmWave radar and signal processing to detect minute vibrations of specific objects and reconstruct speech without relying on machine learning or large datasets [6]. Unlike their work, which performed peak detection on the results of range FFT and further processed it with Doppler FFT, this paper enhances performance by leveraging phase changes instead. Yiwen Feng et al. introduced mmEavesdropper, which amplifies mmWave signals using beamforming and Chirp-Z transform to reconstruct speech [12].

III. METHODOLOGY

We propose a comprehensive methodology for reconstructing speech signals from mmWave radar data. The proposed approach spans from raw data acquisition to the generation of refined audio signals, leveraging phase information embedded in mmWave signals to achieve speech reconstruction.

A. Data Acquisition

To reconstruct speech using mmWave technology, the initial step involves detecting vibrations induced by sound on objects within the sensing range of the mmWave radar. This process leverages mmWave Studio, a specialized software tool provided by the radar manufacturer, to configure the radar system and capture raw data [13]. The software is operated on a Windows-based laptop, where all necessary settings are applied, and the resulting data is stored for further analysis. While this manual process is effective, it is both time-intensive and operationally cumbersome, particularly when frequent adjustments to the radar configuration are required. To address these limitations, an automated workflow has been implemented using Lua scripting, enabling efficient and consistent configuration of the radar system. This approach not only reduces the time and effort required for setup but also minimizes potential human errors, ensuring reliability and repeatability in data acquisition for mmWave-based speech reconstruction.

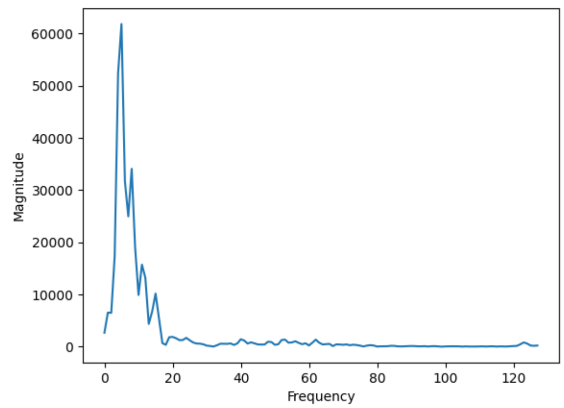

Fig. 2. Frequency spectrum obtained from range FFT

During the data collection process, the distance between the mmWave radar and the speaker was set to approximately 25 cm. The radar configuration included a total of 300 frames, with each frame consisting of 128 ADC samples. The system utilized all three transmit (TX) antennas and four receive (RX) antennas to maximize data acquisition capability. Each frame contained 384 chirps, calculated as the product of the three TX antennas and 128 loops per frame. The raw data was stored in binary format (.bin) for subsequent processing and analysis.

The mmWave radar system transmits signals from the TX antennas, which are reflected by target objects and received by the RX antennas. These reflections are converted into Analog-to-Digital Converter (ADC) data, stored in binary format. Each frame consists of multiple chirps recorded across multiple TX and RX antennas.

The raw ADC data is processed using dedicated hardware and software tools, such as DCA1000EVM, to organize it into a structured data cube. The data cube represents dimensions of frames, chirps, antennas, and ADC samples, enabling efficient processing during subsequent stages. To address inconsistencies in frame sizes, padding is applied to ensure uniformity.

B. Range Processing

The organized data cube undergoes range processing to transform time-domain signals into range spectra. This is achieved by applying a one-dimensional window function, such as the Blackman window, to enhance resolution and suppress noise. The transformation is mathematically expressed as:

$$X(f) = \sum_{n=0}^{N-1} x(n) \cdot w(n) \cdot e^{-j2\pi fn}, \quad (1)$$

where $x(n)$ is the time-domain signal, $w(n)$ is the window function, and N is the total number of samples.

The result of range processing is shown in Fig. 2. From this result, peak detection is performed to identify peaks that correspond to vibrations caused by speech. The identified peaks are then used for subsequent phase analysis.

C. Phase Difference Extraction

The next step involves extracting phase differences between consecutive frames. For each chirp and RX antenna, the magnitude spectrum $|X(f)|$ is computed, and a peak detection algorithm identifies the most significant frequencies. The phase information at these frequencies is given by:

$$\phi(f) = \arg(X(f)). \quad (2)$$

The phase difference between the current and previous frames is calculated as:

$$\Delta\phi(f) = \phi_{\text{curr}}(f) - \phi_{\text{prev}}(f). \quad (3)$$

To handle discontinuities, phase unwrapping is applied, expressed as:

$$\Delta\phi_{\text{unwrapped}}(f) = \Delta\phi(f) + 2\pi k, \quad k \in \mathbb{Z}, |\Delta\phi(f)| < \pi. \quad (4)$$

Phase unwrapping is necessary to address the inherent periodicity of phase values, which are typically constrained within the range $-\pi$ to π. Without unwrapping, these periodic constraints can result in discontinuities or "wrap-around" effects when the true phase difference exceeds this range. Phase unwrapping corrects these discontinuities by ensuring the phase values remain continuous over time, thereby restoring the actual phase difference. This is particularly critical for accurate speech signal reconstruction, as it allows for precise tracking of temporal changes in the mmWave signals.

By resolving these discontinuities, phase unwrapping ensures that the extracted phase differences represent the true temporal variations, enabling reliable and high-quality reconstruction of the speech signal.

D. Audio Signal Construction

The unwrapped phase differences are aggregated across all chirps and RX antennas to form a raw audio signal. This signal is normalized to maintain consistent amplitude levels, described as:

$$x_{\text{normalized}}(t) = \frac{x(t)}{\max(|x(t)|)}. \quad (5)$$

Normalization ensures that the signal is suitable for further processing.

E. Signal Filtering

We applied multi-step filtering to enhance the quality of the reconstructed audio signal. The approach includes three filtering techniques: bandpass filtering to selectively allow frequencies within the range of human speech (80–2000 Hz) to pass while attenuating frequencies outside this range, Wiener filtering to reduce noise using statistical properties of the signal, and median filtering to smooth the signal irregularities and eliminate high frequency spikes.

F. Audio Signal Output

The filtered signal is scaled to 16-bit integer values and saved in WAV format. The scaling process is described as:

$$x_{\text{scaled}}(t) = x_{\text{normalized}}(t) \cdot 32767. \quad (6)$$

Scaling is necessary to map the normalized signal, which typically ranges from $[-1, 1]$, to the 16-bit integer range $[-32768, 32767]$ used by the WAV file format. The value 32767 represents the maximum positive value in this range, ensuring the signal fits within the standard 16-bit integer representation while preserving its dynamic range.

G. Conclusion

Our methodology integrates mmWave technology with signal processing techniques to achieve reliable speech reconstruction without relying on deep learning. By leveraging phase information, this approach overcomes the limitations of deep learning-dependent methods, such as the need for large datasets, high computational cost, and poor generalization to unseen languages. Instead, it provides a robust and efficient framework for non-contact speech capture and reconstruction.

IV. Evaluation

A. Experimental setup

The experimental setup was designed to capture, process, and analyze mmWave radar data using a combination of hardware and software components. For data acquisition, the AWR1843BOOST, a single-chip automotive radar sensor evaluation module operating in the 76-GHz to 81-GHz frequency range, was used as the primary sensor to capture high-resolution radar signals. The DCA1000EVM, a data capture and streaming evaluation module, was employed to enable real-time access to the raw sensor data via an LVDS interface. The ASUS ExpertBook, running on Windows OS with an 11th Gen Intel Core i5-1135G7 processor, 8 GB of RAM, and Python 3.11.7, served as the data acquisition platform, facilitating seamless collection and transfer of radar data from the sensor.

Once the data was collected, processing and analysis were conducted on a high-performance desktop system running Ubuntu 22.04. The desktop, equipped with an 11th Gen Intel Core i9-11900K processor, 64 GB of RAM, and Python 3.10.9, was used to execute the proposed methodology, including range processing, phase extraction, and audio signal reconstruction. For audio playback during data collection, the

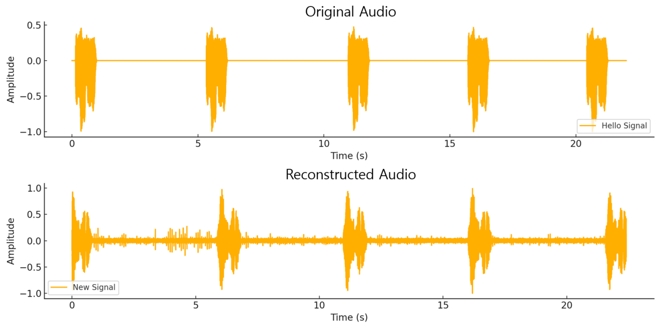

Fig. 3. Comparison of the original and the reconstructed audio

Creative PEBBLE PRO speaker was employed to provide consistent, high-quality reference audio signals within the radar's sensing range.

B. Result

The reconstructed speech signal derived from the preceding results is presented in Figure 3. This figure illustrates the outcome of an experiment conducted at a distance of 25 cm. The Mean Squared Error (MSE) between the original and reconstructed speech signals was calculated to be 0.0197, indicating a moderate level of similarity. To evaluate the robustness of our approach, we conducted additional experiments by playing the same audio at a consistent volume using a speaker and placing the mmWave radar at increasing distances in 25 cm increments to collect data. The experimental results showed that the reconstructed speech signal had an MSE of less than 0.05 compared to the original speech signal at a maximum distance of 1.5 m, which can be considered as practically similar.

V. Discussion

We propose mmProcess, a methodology for reconstructing speech signals using mmWave radar technology without relying on deep learning, and explore its potential applications in eavesdropping. By detecting minute vibrations caused by sound on objects, mmWave radar functions as a tool for acquiring audio data in a non-contact manner, eliminating the need for physical access. This approach addresses the limitations of deep learning, such as the need for large datasets, high computational costs, and reduced performance when processing languages not included in training data. It is particularly advantageous in scenarios where the target is unaware of the sensing equipment or where direct access is infeasible.

The proposed methodology successfully integrates signal processing techniques, including range processing, phase unwrapping, and noise filtering, to reconstruct speech signals without deep learning. Additionally, the automation of data collection and radar configuration significantly improved repeatability and efficiency, reducing setup time while maintaining consistent accuracy—key advantages for practical eavesdropping applications.

However, the accuracy of speech signal reconstruction heavily depends on the characteristics of the vibrating object and environmental factors, such as interference or multipath reflections. Weak vibrations or environments with significant interference may constrain the system's performance, limiting its applicability. Furthermore, the computational resources required by the proposed pipeline may restrict its real-time deployment on low-power edge devices.

Future research could focus on enhancing robustness against environmental factors like interference and multipath reflections while optimizing computational efficiency for real-time applications. Improvements in the signal processing pipeline could further reduce noise and enhance audio quality without integrating machine learning techniques. Expanding experimental validation across diverse real-world scenarios would provide deeper insights into the system's scalability and performance. Such advancements could establish mmWave radar as a key tool for non-contact speech eavesdropping systems.

VI. Conclusion

This paper demonstrated the feasibility and effectiveness of reconstructing speech signals from mmWave radar data without relying on deep learning. The proposed methodology, mmProcess, successfully captured and analyzed minute vibrations caused by sound on objects by leveraging the phase information of mmWave signals. The combination of signal processing techniques, including range processing, peak detection, phase unwrapping, and noise filtering, enabled reliable speech signal reconstruction even in challenging environments.

Unlike deep learning-based methods, mmProcess does not require large training datasets or extensive computational resources and overcomes the limitations of reduced performance when processing languages not included in the training data. This makes the methodology highly adaptable for scenarios such as privacy-sensitive environments and remote speech sensing, where traditional microphones or deep learning approaches may be impractical.

This study establishes a strong foundation for advancing mmWave-based audio processing, particularly in eavesdropping applications. Future research could focus on enhancing robustness against environmental factors like interference and multipath reflections, as well as optimizing computational efficiency for real-time processing. Additionally, expanding experimental validation across diverse real-world scenarios could provide deeper insights into the system's scalability and performance.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2024-00460321, Development of Digital Asset Transaction Tracking Technology to Prevent Malicious Financial Conduct in the Digital Asset Market)

REFERENCES

[1] N. K. Mallat, M. Ishtiaq, A. Ur Rehman, and A. Iqbal, "Millimeter-wave in the face of 5g communication potential applications," *IETE journal of research*, vol. 68, no. 4, pp. 2522–2530, 2022.

[2] B. Yan, P. Wang, L. Du, X. Chen, Z. Fang, and Y. Wu, "mmgesture: Semi-supervised gesture recognition system using mmwave radar," *Expert Systems with Applications*, vol. 213, p. 119042, 2023.

[3] J. Wu and N. Dahnoun, "A health monitoring system with posture estimation and heart rate detection based on millimeter-wave radar," *Microprocessors and Microsystems*, vol. 94, p. 104670, 2022.

[4] B. Zhang, B. Jiang, R. Zheng, X. Zhang, J. Li, and Q. Xu, "Pi-vimo: Physiology-inspired robust vital sign monitoring using mmwave radars," *ACM Transactions on Internet of Things*, vol. 4, no. 2, pp. 1–27, 2023.

[5] C. Xu, Z. Li, H. Zhang, A. S. Rathore, H. Li, C. Song, K. Wang, and W. Xu, "Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 14–26.

[6] P. Hu, W. Li, R. Spolaor, and X. Cheng, "mmecho: A mmwave-based acoustic eavesdropping method," in *Proceedings of the ACM Turing Award Celebration Conference-China 2023*, 2023, pp. 138–140.

[7] C. Wang, F. Lin, T. Liu, Z. Liu, Y. Shen, Z. Ba, L. Lu, W. Xu, and K. Ren, "mmphone: Acoustic eavesdropping on loudspeakers via mmwave-characterized piezoelectric effect," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 820–829.

[8] W.-H. Chen and K. Srinivasan, "Acoustic eavesdropping from passive vibrations via mmwave signals," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 4051–4056.

[9] C. Wang, F. Lin, T. Liu, K. Zheng, Z. Wang, Z. Li, M.-C. Huang, W. Xu, and K. Ren, "mmeve: eavesdropping on smartphone's earpiece via cots mmwave device," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 338–351.

[10] S. Basak and M. Gowda, "mmspy: Spying phone calls using mmwave radars," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 1211–1228.

[11] P. Hu, Y. Ma, P. S. Santhalingam, P. H. Pathak, and X. Cheng, "Milliear: Millimeter-wave acoustic eavesdropping with unconstrained vocabulary," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 11–20.

[12] Y. Feng, K. Zhang, C. Wang, L. Xie, J. Ning, and S. Chen, "mmeavesdropper: Signal augmentation-based directional eavesdropping with mmwave radar," in *IEEE INFOCOM 2023-IEEE Conference on Computer Communications*. IEEE, 2023, pp. 1–10.

[13] "mmwave studio," https://www.ti.com/tool/MMWAVE-STUDIO.