

# 装着者の声だけを騒音環境でも利用可能にするフレキシブルマイク

平城 裕隆<sup>\*†</sup> 金澤 周介<sup>†</sup> 三浦 貴大<sup>†</sup> 吉田 学<sup>†</sup> 持丸 正明<sup>†</sup> 暦本 純一<sup>\*‡</sup>

**概要.** オンラインでの音声対話や音声入力を用いたインターフェースは広く利用されている。しかし、騒音環境や複数の利用者がある環境では利用が難しい。これまではマルチモーダルなセンシングやイヤホン型デバイスによる入力などが提案されてきたが、GPUがない環境での推論が難しいことや特定の方向のノイズに弱いことが課題であった。これを改善するマスク型のマイクが提案されているが、マスクは表情を伝達できない上に衛生上の問題がある。本研究ではフレキシブルなマイクを用いた耳掛け型のデバイスを提案する。フレキシブルマイクの曲率による性能と利用可能な範囲を評価し、曲げることにより入力可能な性能が上がる他、顔を隠すことなく利用可能であることを示した。

## 1 はじめに

音声入力は、オンライン会議やスマートデバイスへの入力など、多くのアプリケーションで重要な役割を果たしている。しかし、騒がしい環境でユーザーの声のみを正確に取得することは、従来のマイクロホンにとって依然として大きな課題である。

これまでソフトウェアベースや深層学習による手法が提案されてきたが、実世界の環境での利用や、競合する周囲の話者の音声を正しく区別することに課題が残されている。

本研究ではフレキシブルな導電布によるダイアフラムを用いた耳掛け型のマイクである ClothTalk を提案する。ClothTalk はノイズの角度依存性がなく、リアルタイムにハンズフリーで GPU を用いずに音声強調が可能であるウェアラブルデバイスである。

## 2 関連研究

これまで様々なアプローチでノイズ除去が試みられている。UltraSpeech[2] は超音波による皮膚の変形をマルチモーダルに計測することで音声を強調するが、スマートフォンに向かって話す必要があり、ハンズフリーな利用が難しい。

Clearbuds[1] は左右のイヤホンでビームフォーミングを用いて話者の音声を強調する。しかし、イヤホン型マイクは正面からの音声と周囲の音声を区別が難しく、特に複数の話者が同時に発話する環境での利用に課題がある。

EarSE[3] は超音波センサをヘッドセットに取り付け、発話時の皮膚や舌の変形を計測することでハン

ズフリーな入力を実現した。しかし Transformer による音声強調の処理に GPU が必要なため、スマートフォンやタブレットでの利用が難しい。また、マスク型マイク [4] は騒音環境下で効果的だが、表情伝達を妨げ、衛生上の問題がある。

## 3 提案手法

本研究では、導電布のダイアフラムからなるフレキシブルエレクトレットコンデンサマイクロフォン (ECM) である ClothTalk を提案する (図 1)。デバイスは振動を取得するダイアフラムと周辺回路から構成される。

ダイアフラムは、厚さ 200  $\mu\text{m}$  の銀メッキ導電性布地を 2 枚使用し、間に PFA フィルムを挟んでコンデンサを形成する。最外層にも PFA フィルムを被覆し、呼吸による擾乱を防止する。

デバイスは耳掛け型として設計され、イヤーフックとダイアフラム間は湾曲可能な軟鋼線で接続されている。音声信号は AD コンバータでデジタル変換され、I2S 形式で出力される。総重量約 20g と軽量で、長時間の装着が可能である。

## 4 結果と考察

マイクの周波数応答特性を評価するため、ダミーヘッドを使用した測定を行った。評価は曲率と長さ、および垂直方向の角度の 2 つの観点から実施した。

曲率については、半径 50 mm、半径 150 mm、平面の 3 種類の構成で測定を行った。長さは 10cm  $\times$  1cm と 5cm  $\times$  1cm の 2 種類を評価した。結果として、曲率を持つダイアフラムは平面と比較して優れた音声入力感度を示し、その効果は R5、R15、平面の順に高いことが明らかとなった。また、10cm のダイアフラムが 5cm のものより優れた性能を示し、長さの増加が音声入力の指向性向上に寄与することが示唆された。

Copyright is held by the author(s). This paper is non-refereed and non-archival. Hence it may later appear in any journals, conferences, symposia, etc.

\* 東京大学

† 産総研

‡ Sony CSL

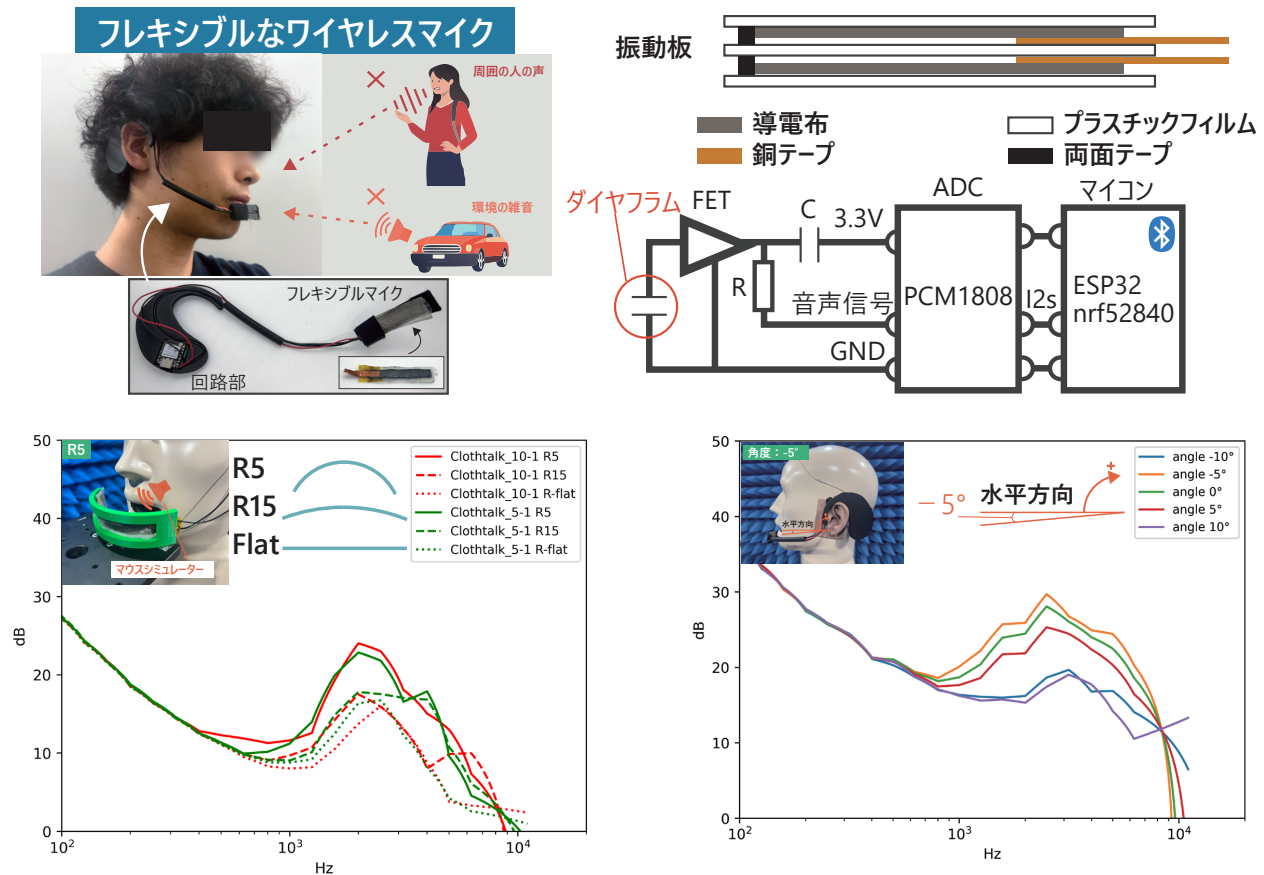


図 1. ワイヤレスマイクの概要 (上図). 導電布からなるフレキシブルなマイクを制作. マイクを曲げた場合の入力の変化 (左下図), 付ける位置を変えた場合の入力の変化 (右下図)

次に、垂直方向の角度による影響を評価した。マイクが口部正面に位置する 0 度を基準とし、 $-10$  度から  $+10$  度の範囲で 5 度間隔で測定を行った。結果として、 $-5$  度から  $+5$  度の範囲では入力が一貫して効果的に維持されることが確認された。 $-5$  度の位置では、マイクは口部から約 2cm 下方に位置し、口の表情を視認できる。しかしながら、 $\pm 10$  度を超える極端な角度においては性能の大幅な低下が観察された。

これらの結果は、フレキシブルなマイクの設計における重要なパラメータとして曲率と長さが重要であること、また実用的な装着位置の範囲を示唆している。今後は、より広範な環境および多様なユーザーシナリオでの評価を通じて ClothTalk の実用化を進める予定である。

## 謝辞

本研究は JST ACT-X グラント番号 JPMJAX23 KG, JST ムーンショット型研究開発事業グラント番号 JPMJMS2012, 国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) 委託業務 (JPNP 23025) の支援を受けたものです。

## 参考文献

- [1] I. Chatterjee, M. Kim, V. Jayaram, S. Gollakota, I. Kemelmacher, S. Patel, and S. M. Seitz. ClearBuds: Wireless Binaural Earbuds for Learning-Based Speech Enhancement. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services, MobiSys '22*, p. 384–396, New York, NY, USA, 2022. Association for Computing Machinery.
- [2] H. Ding, Y. Wang, H. Li, C. Zhao, G. Wang, W. Xi, and J. Zhao. UltraSpeech: Speech Enhancement by Interaction between Ultrasound and Speech. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(3), sep 2022.
- [3] D. Duan, Y. Chen, W. Xu, and T. Li. EarSE: Bringing Robust Speech Enhancement to COTS Headphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(4):1–33, Jan. 2024.
- [4] T. M. M. Y. M. M. J. R. Hirota, Hiraki, Shusuke Kanazawa. WhisperMask: a noise suppressive mask-type microphone for whisper speech. In *The Augmented Humans International Conference (AHs 2024)*, pp. 1–16, Melbourne, VIC, Australia, 2024.