# WhisperMask: 騒音環境でも囁き声が利用可能なウェアラブルマイク

平城 裕隆 *† 金澤 周介 † 三浦 貴大 †* 吉田 学 † 持丸 正明 †* 暦本 純一 * ‡

**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いる環境では利用が難しい．これまでは，咽喉マイクや NAM マイクと行った，首元や耳の後ろなどの体の表面に取り付けることでマイクが提案されてきた．しかし，これらのマイクは体の表面に取り付ける必要があり，歩行など別の動作によるノイズが生じる．また，体の中を通る音であるため，囁き声のような呼気によって生じる音を披露ことが難しい．そこで導電性フィルムを用いたコンデンサマイクを提案する．このマイクは自身の声のみを拾うことができ，ピンマイクと比べて騒音環境下でも高い SNR を得ることができた．

## 1 はじめに

音声による対話は人間が持つ基本的なコミュニケーションの手段である．近年は Covid-19 によるオンラインでの音声対話や，スマートフォン・スマートデバイスなどによる音声入力を用いたインタフェースが広く利用されている．しかし，騒音環境では音声認識の精度が低下するため利用が難しいほか，複数の利用者がいる環境では意図しない音声入力が発生する上，話者ごとの音源の分離が必要であり，認識が困難である．自身の声のみを集音するハードウェアとして，これまでは，咽喉マイクや NAM マイク [5] が利用されてきた，首元や耳の後ろなどの体の表面に取り付けることでマイクが提案されてきた．これらは，首の周りや耳の後ろなどの体の表面にマイクを取り付けることで自身が発生した音声を体内の骨や肉を経由して取得するものである．これによって，騒音環境など外部の音を拾うことなく利用することができる．しかし，これらのマイクは体に直接取り付ける必要があるため，歩行など利用者の体の動きによるノイズが生じる．また，体の中を通る音であるため，囁き声のような呼気によって生じる音を取得することが難しい．そこで導電性フィルムを用いたコンデンサマイクを提案する．このコンデンサマイクは近くの音のみを拾うことができマスク型に実装することで，自身の声を中心に取得することができる．実験の結果，ピンマイクと比べて騒音環境下でも高い SNR を得ることができた．

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

## 2 関連研究

接触型のマイクとして，咽喉マイク [6, 3] や NAM マイク [5, 4, 7, 2] が提案されている．咽喉マイクは首元に接触させた振動板を経由して，体の表面を伝わる音声を取得する手法である．NAM マイク [5] はシリコンを振動板とするマイクを耳の後ろの部分に取り付けることで，体の肉や骨を伝わる音を取得することができる．これらは外部の音を取得しないため，自身の音のみを取得できる．特に，このような特殊なマイクに対しての音声認識がおこなわれている．またマイクに入る音を物理的に工夫，ピンマイクのような単一指向性マイクが提案されている．他にも，多数のマイクを用いることで音源方向を特定して音を抽出する手法も提案されている．

| | Contact Noise | Ambient Noise |
|---|---|---|
| NAM mic<br>Throat mic | △ | ○ |
| Pin mic<br>Headset | ○ | △ |
| **WhisperMask** | ○ | ○ |

**図 1. 本研究の位置付け**

## 3 提案手法

提案するマイクロホンは，帯電したフィルムと導電性繊維からなるエレクトレットコンデンサで構成されている（図 2）．フィルムの厚さは 12.5 μ m で帯電しやすく，常に静電気を帯びているエレクトレットとして振る舞います．フィルムの外周に粘着テープを貼り，その両面に厚さ 200 μ m のメッキされた布を固固定して誘電体として用いている．コ

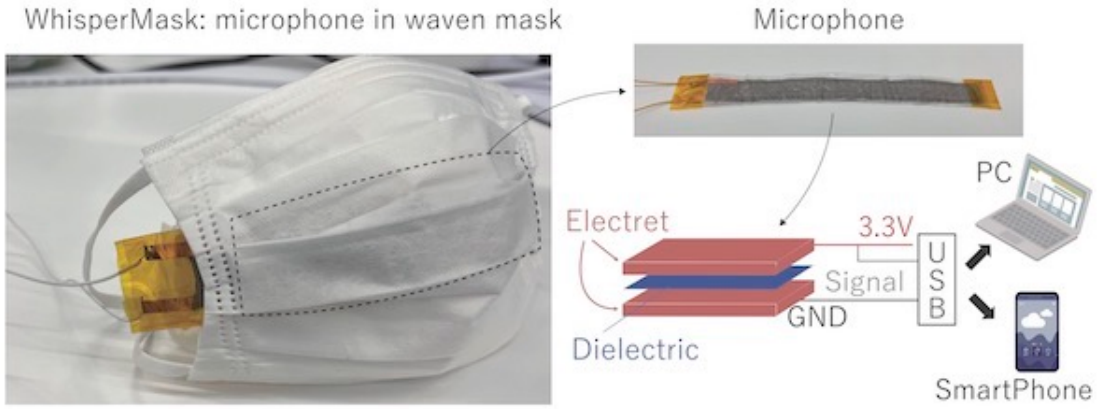


**図 2. 提案するハードウェア**

ンデンサーの大きさは，図 1 のように 18cm × 7cm に設計した．

マイクロホンとしての動作原理は，一般的なエレクトレットコンデンサマイクロホンと同じである．エレクトレットに音声振動を与えると，振動によってエレクトレット膜と電極の距離が変化し，電極上の電位が変化する．これにより，音声を電圧の振幅として取り出すことができる．

実験に使用するマイクロホンは，市販のコンデンサーマイクロホンの回路にエレクトレットコンデンサーマイクを取り付けて使用した．

マイクは口元に近い形で入力を行う必要があるため，マスク型に設計することでウェアラブルな実装にした．

## 4 実験

信号対雑音比率 (SNR) を測定するため，10 名のユーザーに対して TIMIT コーパス [1] から選択した 100 の英文を読み上げを行った．読み上げの方法は通常の発声と囁き声である．得られた音声データに対して音素アラインメントを行って無声区間と有声区間を識別し，SNR の計算を行った．得られた結果は図 3 の様になった．Natural Speech において，t 検定を行ったところ，$p = 2.38e - 51 (< 0.01)$ となり有意な差が得られた．

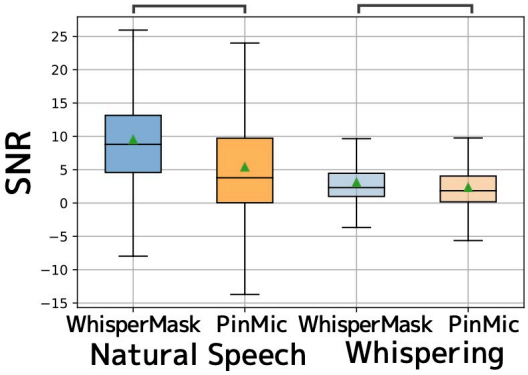


**図 3. Natural Speech, Whisper での SNR**

## 5 今後の課題

今後は，デバイスの面積や素材を変化させた場合の SNR の変化について評価を行っていく．また，日常的に利用可能なデバイスとして長期的な利用を行なった際のデバイスの安定性や，マイクに特化した音声認識の必要性について評価を行っていく．

## 参考文献

[1] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue. TIMIT Acoustic-phonetic Continuous Speech Corpus. 11 1992.

[2] T. Hirahara, M. Otani, S. Shimizu, T. Toda, K. Nakamura, Y. Nakajima, and K. Shikano. Silent-speech enhancement using body-conducted vocal-tract resonance signals. Vol. 52, pp. 301–313, 2010. Silent Speech Interfaces.

[3] J. Kawaguchi and M. Matsumoto. Noise Reduction Combining a General Microphone and a Throat Microphone. Vol. 22, 2022.

[4] Y. Nakajima, H. Kashioka, K. Shikano, and N. Campbell. Non-audible murmur recognition input interface using stethoscopic microphone attached to the skin. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, Vol. 5, pp. V–708, 2003.

[5] S. Shimizu, M. Otani, and T. Hirahara. Frequency characteristics of several non-audible murmur (NAM) microphones. Vol. 30, pp. 139–142, 2009.

[6] A. Vijayan, B. M. Mathai, K. Valsalan, R. R. Johnson, L. R. Mathew, and K. Gopakumar. Throat microphone speech recognition using

mfcc. In *2017 International Conference on Networks Advances in Computational Technologies (NetACT)*, pp. 392–395, 2017.

[7] N. Yoshitaka, K. Hideki, C. Nick, and S. Kiyohiro. Non-Audible Murmur (NAM) Recognition. Vol. E89-D, 2005.