

能動的音響センシングを用いた黙声認識の検証

永山 晃誠^{†1,a)} 高田 峻介^{†2}

概要: マスクを用いて口元を隠し、黙声認識技術を用いることで、周囲に悟られることなく電子機器の操作が可能になる。本研究では、従来の黙声認識技術よりも手軽に使用できる手法として、能動的音響センシングを用いる。本手法では、左右どちらかの耳の周辺に取り付けた圧電ブザーと圧電マイクを用いて、口の形状を認識する。まず、圧電ブザーを肌に密着させ、20kHz~48kHzの掃引信号で振動させる。次に、同様に密着させた圧電マイクによって、伝搬してきた信号を測定し、信号の周波数伝達特性を得る。最後に、SVMを用いて周波数伝達特性を識別することで、使用者の口の形状を認識する。本研究では、4人の実験協力者を対象とした実験によって、5種類の母音の口の形状の認識精度を調査した。

1. はじめに

音声入力は、他の作業で手がふさがっているときや、感染症の懸念から、公共の機器に直接接触することがためられるような場合において有用である。しかし、音声認識を用いた音声入力は、プライバシーとセキュリティの観点から、ユーザが積極的に使用しづらいという問題がある [1][2][3][4]。さらに、音声認識の精度は、周囲の環境音の影響を大きく受けてしまう。

そこで、これらの問題を解決するために黙声認識の研究が行われている。黙声認識とは、声帯の振動を伴わない発話動作をセンシングし、発話の推定を行うというものである。黙声認識は、音声認識に代わる有望な代替手段として示されており [5][6]、多くのユーザが従来の音声認識による入力よりも、黙声認識による入力を好むという研究結果も報告されている [7]。また、声帯振動での発話が困難な者への補助手段としても用いることができる [8][9]。

黙声認識を行う手法は、様々なものが提案されている。口周辺の動きに注目してセンシングを行うものは、RGB画像を用いて黙声認識を行うものや [10]、皮膚の運動計測を行うもの [11] ならびに静電センサアレイを用いるもの [12] などがある。また、他に超音波イメージングを用いるものや [13]、口周辺の筋電を用いてジェスチャ認識を行うものがある [14]。しかし、これらの手法は周囲の光の影響を受けるもの、デバイスが目立って不自然なもの、食事や会話などの日常生活に影響を与えるものなど各々問題がある。また、特に画像を用いる黙声認識手法では、感染症対策の

ためにマスクを装着した際に、口元が隠れてしまうオクルージョンの問題も抱える。

そこで、本研究では、能動的音響センシングを用いた黙声認識を提案する。本手法では、肌に密着させた圧電ブザーと圧電マイクを用いて黙声認識を行うことを目標とする。本研究は圧電ブザーと圧電マイクを耳の周りに装着することで、黙声を認識できる。そのため、ヘッドホンのような耳に装着デバイスに搭載することができ、目立ちにくい。また、本手法では人体の信号伝達特性を用いて黙声認識を行うため、画像を用いる手法とは違ってオクルージョンや照明条件の問題を受けない。

本研究では、まず能動的音響センシングによって口の形状を認識可能かどうかを実験によって調査した。

2. 関連研究

黙声認識はカメラを用いるものや各種センサを用いるものなど様々な手法が提案されているため、本研究との差異を述べる。また、本研究では黙声認識に能動的音響センシングを用いているため、同手法を用いた研究を述べる。

2.1 黙声認識手法

様々な黙声認識を行う手法が提案されている。RGB画像を用いて黙声認識を行うものとして、カメラにて話者の口唇あるいは顔全体を撮影し、口唇の周りの皮膚の変形と動きを読み取るリップリーディングによって発話内容を推定できる [6][15]。また、口周辺のモノクロ画像から濃度ヒストグラムを取得し、口部ジェスチャを認識できる [16]。しかし、これらの手法では、カメラを話者の顔の前に配置する必要があり、環境光の影響を受けることに加え、マス

^{†1} 神戸市立工業高等専門学校専攻科 電気電子工学専攻

^{†2} 神戸市立工業高等専門学校 電子工学科

^{a)} r117229@g.kobe-kosen.ac.jp

ク装着時や手で口元を覆った場合に、オクルージョン問題が発生する。

超音波イメージングを用いて、口腔内の状態から発話内容の推定を行った研究もある [13][17]。超音波イメージングとは、超音波プローブを顎の下の皮膚に取り付け、身体に放射した超音波の反射時間を測定することで身体の内部状態を認識するというものである。しかし、超音波イメージングのためのプローブは非常に高価であり、正確に体内画像を取得するためにゲルを皮膚に塗布し、プローブを皮膚に密着させる必要がある。

マイクを用いた黙声認識 [18][19] では、非可聴つぶやき (Non-audible murmur, NAM) という、声帯の振動を伴わない呼吸音 (ささやき声) を、利用者の喉に取り付けたマイクで認識する。しかし、システムが正確に黙声認識するためには、利用者がある一定以上の音量でつぶやき発話を行う必要がある、このつぶやき声を近隣の人に聞き取られてしまう可能性がある。

筋電図 (EMG) を用いた黙声認識 [14][20][21] では、口腔付近の筋肉の状態を認識して黙声認識を行う。口腔付近の筋肉は複雑な構造をしており、それぞれの筋肉の伸縮状態を計測するために、複数のセンサが必要になる。また、顔表面に装着した電極が目立ってしまう。

口腔内、あるいは周辺にセンサを配置する方式も提案されている。Derma[11] は、顎下の皮膚に装着した加速度センサおよび角速度センサで黙声認識を行う。高い精度で発声コマンドを認識できるが、顎下のセンサは筋電図を用いる手法同様に目立ってしまう。SilentMask[22] も同様のセンサによって、口の形状変化に伴うマスクの動きをセンシングし、ジェスチャ認識を行う。マスクを装着しているため、口の動きが見えず情報秘匿性が高い。しかし、マスクと皮膚の密着状態が常に維持されないため、認識精度が落ちる問題がある。TongueBoard[12] は、口に配置した静電センサアレイで音声と舌のジェスチャを認識する。しかし、センサを口腔内に配置するため、日常会話や食事の際の妨げになる。

一方、提案手法は振動の伝達特性を用いるため、環境光の影響やオクルージョンの問題が発声せず、耳の周囲にセンサを貼り付ければよいいため、ヘッドホンなどの耳に装着するデバイスへの搭載が容易である。

2.2 能動的音響センシング

能動的音響センシングとは、物体の形状、材質、境界条件によって決まる固有の共振特性、周波数伝達特性を用いるセンシング手法である。

Touch&Activate[23] は、物体の境界条件の変化による、共振スペクトルの変化を用いて物体がどのように触られているか認識できる。

また、AudioTouch[24] は、指の姿勢を変えた際に、手の

骨や筋肉が動くことによる、手の周波数伝達特性の変化を用いて手のジェスチャ認識を行う。AudioTouch で用いられるセンサはピエゾ素子 2 枚のみで、これらを手の甲に貼り付けて使用する。手のひらや指先にセンサを取り付けることがないため、侵襲性が低い。The Sound of Touch[25] は、腕に取り付けた圧電ブザーと圧電マイクで、指で腕をなぞる動作や腕をつかむジェスチャの認識およびタッチ圧力を推定している。Bitey[26] では、骨伝導マイクを用いて歯を噛み合わせた際の音を認識し、クリック入力として使用可能であることを示している。また、異なる歯同士を噛み合わせた場合の音の違いを認識し、入力語彙を増やすこともできる。

これらの研究は、能動的音響センシングを用いて人の振る舞いが認識できることが示されており、本研究は本手法を表情変化の認識へ用いる点で異なる。

3. 能動的音響センシングを用いた黙声認識

3.1 本システムの概要

発話動作を行う際に、口の周辺の筋肉や皮膚ならびに骨が動くため、顔に流す音響の周波数伝達特性が変化すると考えられる。この周波数伝達特性の変化を利用することで、能動的音響センシングによる黙声認識を実現できる可能性がある。

提案手法では、AudioTouch と同様に、片方の圧電ブザーと圧電マイクを取り付け、素子間に流した超音波の周波数伝達特性をもとに黙声認識を行う。

提案手法の検証のために、能動的音響センシングを用いた母音発音時の口の形状の認識精度の調査ならびに、ブザーの最適な固定位置を調査した。

3.2 本システムの構成

本システムの構成をハードウェアとソフトウェアに分けて説明する。

ハードウェア

本システムでは図 1 に示す圧電ブザー (図 1 左 ムラタ 7BB-20-6L0,20mm) および圧電マイク (図 1 右 Thrive K2512BP1,25 × 12 × 0.23mm) を用いて、耳の周辺の皮膚の振動、および人体を経由してきた振動を計測する。

これらの素子を医療用両面テープ (3M, 2477P) を用いて耳の周辺の肌に直接貼り付けた。

素子をオーディオインタフェース (Roland Octa-Capture ua1010) を介してコンピュータ (HP ENVY Laptop15-ep, CPU: Intel Core i9) と接続し、コンピュータを用いて音の再生および録音を行った。

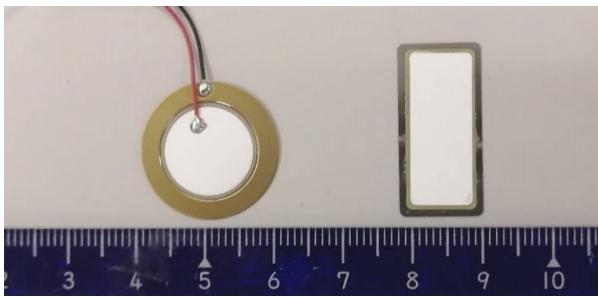


図 1 使用する圧電ブザーおよび圧電マイク

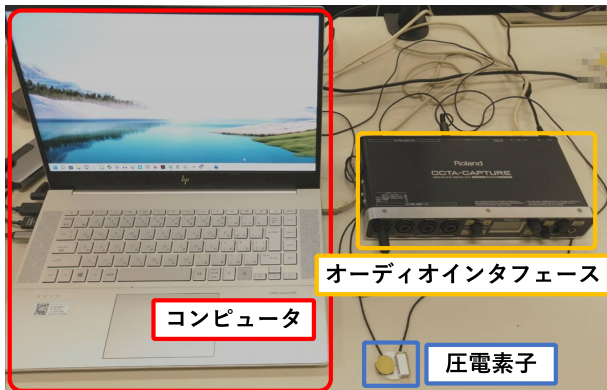


図 2 ハードウェアの構成

ソフトウェア

まず、コンピュータ上にて信号を生成し、Python の SoundDevice ライブラリを用いて信号の再生および録音を行う。本研究で用いた信号は 20kHz~48kHz の超音波を 4096 点 (約 43 ミリ秒分) で直線的に遷移する掃引信号で、これをサンプリング周波数 96kHz で生成する。この周波数は先行研究 [23] を参考にして決定した。次に、録音した信号に対して高速フーリエ変換 (FFT) を行い、特徴量を抽出する。信号はサンプリング周波数 96kHz で 16,384 点 (約 171 ミリ秒) ずつ抜き出し、抜き出したデータに対して高速フーリエ変換 (FFT) を行い、スペクトルを得る。使用した掃引信号は 4,096 点 (約 43 ミリ秒) で 1 周期分であるため、16,384 点で 4 周期分の掃引信号を含むことになる。スペクトルからデータを 40 点おきに抜き出すダウンサンプリングを行い、抜き出したデータを平滑化フィルタに通す。最終的なデータ長は 205 点となり、これを特徴量として SVM 分類器を作成する。

5つの母音を発音する際の口の形状の識別に、RBF カーネルを用いた SVM 分類器を用いる。SVM を用いた理由は AudioTouch と同じで、少ないデータで高精度の分類が可能であるためである。また、各ユーザごとに分類器の再学習が可能であるという面もある。SVM 分類器を製作する際に、Optuna ライブラリを使用してハイパーパラメータの探索も行う。SVM モデルとして用いたのは Scikit-Learn ライブラリのモジュールで、テストデータによる予測を行

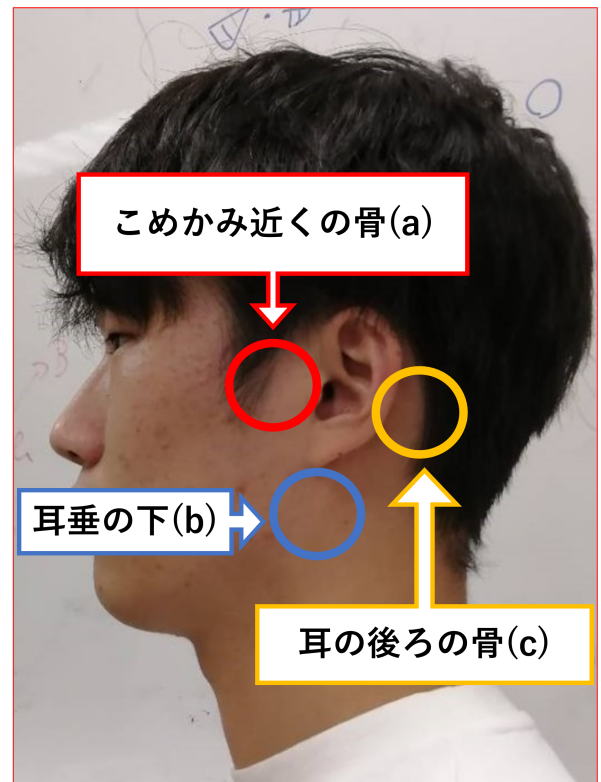


図 3 圧電ブザーおよび圧電マイクの固定位置

う際もこのライブラリを用いてテストデータと学習用データの分割を行う。テストデータと学習用データの分割は録音したファイルごとに行い、分類モデルの精度の評価は、K-分割交差検証で行う。最後に、検証の結果を母音ごとの認識の正答率として出力する。

4. 能動的音響センシングを用いた黙声認識の検証

提案手法の検証のために、各ジェスチャの認識精度の調査、および送信用の圧電素子の最適な固定位置を調査した。

4.1 学習用データおよびテスト用データの収集

ジェスチャ認識 SVM 分類器を作成するために、著者を含む 4 人の実験協力者に対して学習用データおよびテスト用データの収集を行った。送信用の圧電素子の固定位置は、こめかみ近くの骨、耳垂の下、耳の後ろの骨とし、それぞれの圧電素子の位置を入れ替えながら 6 パターンの固定位置にて録音をした。

録音中は、実験協力者には実際に発音はせずに、口の形 (母音ジェスチャ) を維持するよう指示した。1 母音ジェスチャあたりの録音時間は約 3 秒程度で、これを 1 パターンあたり 50 回繰り返し、これを 1 ラウンドとした。また、圧電素子の密着具合やわずかな位置ずれの影響を考慮し、10 回ごとに受信用の圧電素子を貼り付けなおした。順序効果を打ち消すために、10 回の中で、5 つの母音を計 2 回

表 1 著者による事前検証

	あ	い	う	え	お	平均
a-b	77.1%	20.0%	59.7%	45.9%	62.9%	53.1%
a-c	30.3%	15.5%	18.1%	14.8%	9.3%	17.6%
b-a	52.2%	45.2%	20.0%	21.0%	31.2%	33.9%
b-c	48.6%	16.6%	32.4%	16.4%	19.3%	26.7%
c-a	45.7%	15.5%	26.9%	47.6%	31.6%	33.4%
c-b	19.7%	14.0%	41.0%	31.4%	44.1%	30.0%

表 2 分類精度

	あ	い	う	え	お	平均
a-b	31.0%	21.1%	27.1%	29.7%	31.7%	28.1%
a-c	28.7%	18.3%	9.1%	12.1%	20.3%	17.7%
b-a	27.7%	15.4%	19.1%	19.1%	24.0%	21.1%
b-c	33.5%	15.8%	14.8%	21.1%	25.9%	22.2%
c-a	22.1%	32.3%	14.5%	29.0%	28.1%	25.2%
c-b	18.0%	19.9%	22.3%	30.0%	25.4%	23.1%

ずつ、ランダムな順番で行ってもらった。また、本節にて述べた 6 種類の圧電素子の貼り付けパターンごとに 5 ラウンドずつ実施した。一人あたり収集するデータの数の合計は、10 回×5 ラウンド×6 パターンの 300 回であった。

4.2 SVM 分類器の作成

収集したデータ (wav 形式) を用いて、SVM 分類器を作成し、その分類器の精度を調査した。各圧電素子の貼り付けパターンごとに音声データを分類し、各データで分類器の精度評価の交差検証を行うためにテストデータと学習用データの分割を行った。分割は音声データごとに行い、1 ジェスチャあたり 10 個のファイルのうち、9 個を学習用、1 個をテスト用ファイルとして分割した。

4.3 著者による事前検証

実験協力者による録音を行う前に、著者のみのデータを用いて SVM 分類器を作成、ならびに認識精度の調査を行った。以下の表 1 に、各貼り付けパターンごとの認識の正答率を示す。なお、表における圧電素子の取り付け位置はそれぞれ a (こめかみ近くの骨)、b (耳垂の下)、c (耳の後ろの骨) を (圧電ブザーの位置) - (圧電マイクの位置) のように表記している。

結果を見ると、圧電マイクをこめかみ近くの骨、圧電ブザーを耳垂の下に取り付けた場合に、約 53 パーセントの正答率で認識できたことがわかる。特に、口の形を「あ」、「お」とした際の精度はそれぞれ 77.1 パーセント、62.9 パーセントを記録した。

4.4 分類精度の評価

表 2 に、今回の検証における圧電素子の貼り付けパターンごとの認識の正答率を示す。結果からは、圧電素子の固定位置が a-b、すなわち圧電ブザーをこめかみ近くの骨、圧電マイクを耳垂の下に取り付けたときの認識精度が、他の圧電素子の貼り付けパターンよりも約 10 パーセント高く、約 30 パーセントとなっていることがわかる。30 パーセントという認識精度はかなり低い値であるが、著者による事前検証では同じ固定位置で約 53.1 パーセントの正答率を記録している。このことから、今回の検証では実験を行う際に、実験協力者に実験を行う上でブザーの固定方法や正しい口の形状などをうまく伝えられていなかった可能性が考えられる。

5. 議論・今後の課題

5.1 ユーザごとのキャリブレーション

本検証では、ユーザごとのキャリブレーションレスな分類器を作成することを目的に、4 人の実験協力者のデータをまとめて学習させた。しかし、今回の検証で、ユーザごとに SVM 分類器を学習させた方が、複数ユーザのデータによって学習させた場合よりも正答率が下がることがわかった。そのため、本手法を実際に用いる場合は、ユーザごとにキャリブレーションを行う必要があると考えられる。

そのため、作成した SVM 分類器の他者への適用が困難であると考えられる。

本手法にて実際に黙声入力を行う際は、ユーザごとに SVM 分類器の再学習、あるいは 1 から学習をさせなければならぬと考えられる。

5.2 信号処理手順の見直し

本研究では、先行研究を参考にして再生する信号を 0~48kHz の掃引信号にした。しかし、40kHz 周辺の信号が経皮伝播特性に最も優れるため [27]、今後はこの周波数帯の信号を用いることによって、認識の正答率が向上するのかがどうかを検証する。

5.3 圧電素子の固定方法

今回の検証で得られた特徴量は、個人差が大きいことに加えて、圧電素子と皮膚の接触状況に依存してノイズが多量に含まれていることがわかった。従って、圧電素子の固定方法を検討しなおす必要がある。AudioTouch[24] では、デバイスの侵襲性を低減するために、柔軟性のあるピエゾ素子を使用することを提案している。本手法でも、柔軟性を持つピエゾ素子を用いれば、より素子を皮膚に密着させ、効率的に振動させられる可能性がある。

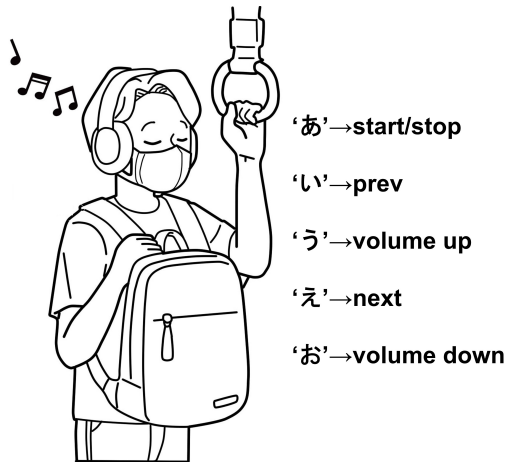


図 4 ヘッドホンでの使用例

5.4 アプリケーション例

本手法を用いたアプリケーション例を述べる。まず、本手法によって図 4 に示すように音楽プレイヤーなどの操作が可能である。外出時、両手が塞がっている場合に、マスクをしていてもデバイスへのアクセスが可能のため、情報秘匿性、利便性の面において優れる。また、SkullConduct[28]のように、生体認証への応用も考えられる。この場合、本手法で使用しているセンサは、SkullConduct で使用されている骨伝導スピーカーおよび骨伝導マイクよりも安価で小型であるため、より手軽に実装できる。

6. 結論

本研究では、能動的音響センシングを用いて、5 種類の母音発音時の口の形状の認識精度を調査した。その結果、約 30 パーセントの精度で認識を行うことができ、圧電素子をこめかみ近くの骨と耳垂の下に貼り付けた場合に、最も高い認識精度が得られることが分かった。また、周波数伝達特性の個人差は大きく、実際に使用する場合はユーザーごとにキャリブレーションを行う必要がある。

今後は、信号処理方法および圧電素子の固定方法の見直しを行い、認識精度の向上に努める予定である。

参考文献

- [1] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction*, Vol. 31, No. 4, pp. 307–335, 2015.
- [2] Christos Efthymiou and Martin Halvey. Evaluating the social acceptability of voice based smartwatch search. In *Asia Information Retrieval Symposium*, pp. 267–278. Springer, 2016.
- [3] Salil Prabhakar, Sharath Pankanti, and Anil.K. Jain. Biometric Recognition: Security and Privacy Concerns. *IEEE Security & Privacy*, Vol. 1, No. 2, pp. 33–42, 2003.
- [4] Laxmi Pandey, Khalad Hasan, and Ahmed Sabbir Arif. Acceptability of Speech and Silent Speech Input Methods in Private and Public. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, May 2021.

- [5] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. In *proceedings of ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 1, No. 4, Jan. 2018.
- [6] Ke Sun, Chun Yu, Weinan Shi, Lan Liu, and Yuanchun Shi. Lip-Interact: Improving Mobile Device Interaction with Silent Speech Commands. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, UIST '18, pp. 581–593, New York, NY, USA, 2018. Association for Computing Machinery.
- [7] Laxmi Pandey, Khalad Hasan, and Ahmed Sabbir Arif. Acceptability of Speech and Silent Speech Input Methods in Private and Public. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, New York, NY, USA, 2021. Association for Computing Machinery.
- [8] Charlotte.M. Reed. The implications of the tadoma method of speechreading for spoken language processing. In *Proceeding of Fourth International Conference on Spoken Language Processing*. ICSLP '96, Vol. 3, pp. 1489–1492 vol.3, 1996.
- [9] Jose A. Gonzalez, Lam A. Cheah, Angel M. Gomez, Phil D. Green, James M. Gilbert, Stephen R. Ell, Roger K. Moore, and Ed Holdsworth. Direct Speech Reconstruction From Articulatory Sensor Data by Machine Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, No. 12, pp. 2362–2374, 2017.
- [10] Yannis M Assael, Brendan Shillingford, Shimon Whiteson, and Nando De Freitas. Lipnet: End-to-end sentence-level lipreading. *arXiv preprint arXiv:1611.01599*, p. 13, 2016.
- [11] 暦本純一, 西村悠. Derma: 皮膚運動計測によるサイレントスピーチインタラクション. 情報処理学会インタラクション 2020, pp. 11–20, 3 2020.
- [12] Richard Li, Jason Wu, and Thad Starner. TongueBoard: An Oral Interface for Subtle Input. In *Proceedings of the 10th Augmented Human International Conference 2019*, AH2019, New York, NY, USA, 2019. Association for Computing Machinery.
- [13] Tamás Gábor Csapó, Tamás Grósz, Gábor Gosztolya, László Tóth, and Alexandra Markó. DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. In *INTERSPEECH*, 2017.
- [14] Arnav Kapur, Shreyas Kapur, and Pattie Maes. AlterEgo: A Personalized Wearable Silent Speech Interface. In *23rd International Conference on Intelligent User Interfaces*, IUI '18, pp. 43–53, New York, NY, USA, 2018. Association for Computing Machinery.
- [15] Michael Wand, Jan Koutník, and Jürgen Schmidhuber. Lipreading with Long Short-Term Memory. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6111–6119. IEEE Press, 2016.
- [16] 渡邊睦, 西奈津子. 口部パターン認識を用いた日常会話伝達システムの研究. 電気学会論文誌C (電子・情報・システム部門誌), Vol. 124, No. 3, pp. 680–688, 2004.
- [17] Naoki Kimura, Michinari Kono, and Jun Rekimoto. Sot-toVoce: An Ultrasound Imaging-Based Silent Speech Interaction Using Deep Neural Networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–11, New York, NY, USA, 2019. Association for Computing Machinery.

- [18] Yoshitaka. Nakajima, Hideki. Kashioka, Kiyohiro. Shikano, and Nick. 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. 708–711, 2003.
- [19] Tatsuya Hirahara, Makoto Otani, Shota Shimizu, Tomoki Toda, Keigo Nakamura, Yoshitaka Nakajima, and Kiyohiro Shikano. Silent-Speech Enhancement Using Body-Conducted Vocal-Tract Resonance Signals. *Speech Commun.*, Vol. 52, No. 4, pp. 301–313, apr 2010.
- [20] Geoffrey S Meltzner, James T Heaton, Yunbin Deng, Gianluca De Luca, Serge H Roy, and Joshua C Kline. Development of sEMG Sensors and Algorithms for Silent Speech Recognition. *Journal of Neural Engineering*, Vol. 15, No. 4, p. 046031, Jun. 2018.
- [21] Michael Wand, Tanja Schultz, and Jürgen Schmidhuber. Domain-Adversarial Training for Session Independent EMG-based Speech Recognition. In *Interspeech*, pp. 3167–3171, 2018.
- [22] Hirotaka Hiraki and Jun Rekimoto. SilentMask: Mask-Type Silent Speech Interface with Measurement of Mouth Movement. In *Augmented Humans Conference 2021, AHs'21*, p. 86–90, New York, NY, USA, 2021. Association for Computing Machinery.
- [23] Makoto Ono, Buntarou Shizuki, and Jiro Tanaka. Touch & Activate: Adding Interactivity to Existing Objects Using Active Acoustic Sensing. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology, UIST '13*, pp. 31–40, New York, NY, USA, 2013. Association for Computing Machinery.
- [24] Yuki Kubo, Yuto Koguchi, Buntarou Shizuki, Shin Takahashi, and Otmar Hilliges. AudioTouch: Minimally Invasive Sensing of Micro-Gestures via Active Bio-Acoustic Sensing. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '19*, New York, NY, USA, 2019. Association for Computing Machinery.
- [25] David Merrill, Hayes Raffle, and Roberto Aimi. The Sound of Touch: Physical Manipulation of Digital Sound. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08*, pp. 739–742, New York, NY, USA, 2008. Association for Computing Machinery.
- [26] Daniel Ashbrook, Carlos Tejada, Dhwanit Mehta, Anthony Jiminez, Goudam Muralitharam, Sangeeta Gajendra, and Ross Tallents. Bitey: An Exploration of Tooth Click Gestures for Hands-Free User Interface Control. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '16*, pp. 158–169, New York, NY, USA, 2016. Association for Computing Machinery.
- [27] Valentina Suchkova, Farhan N. Siddiqi, Edwin L. Carstensen, Diane Dalecki, Sally Child, and Charles W. Francis. Enhancement of Fibrinolysis With 40-kHz Ultrasound. *Circulation*, Vol. 98, No. 10, pp. 1030–1035, 1998.
- [28] Stefan Schneegass, Youssef Oualil, and Andreas Bulling. SkullConduct: Biometric User Identification on Eyewear Computers Using Bone Conduction Through the Skull. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16*, pp. 1379–1384, New York, NY, USA, 2016. Association for Computing Machinery.