Whisphone: ささやき声で入力できるイヤホン

福本 雅朗*

概要. Whisphone は，ささやき声での音声入力が可能なイヤホンである． 先端部にマイクを搭載したカナル(耳穴挿入)型イヤホンによる外耳道閉鎖効果によって，骨導経由で外耳道内部に放射されたささやき声を効率的に収音できる．耳穴を塞ぐことによる外部騒音の遮断に加え，イヤホンの ANC (Active Noise Canceling) 機能の併用によって，80dB(A)の騒音下においても微小なささやき声の収録と音声認識が可能である．小型の機器は目立ちにくく常時装用が容易であり，小さなささやき声は入力に際して周囲の迷惑になりにくい．本装置を用いることで，オフィス・家庭・街頭等，日常生活の多くの場面においてハンズフリーで AI アシスタントとの対話を行うことができる．

1 はじめに

LLM (Large Language Model) と生成 AI は，コンピュータを「操作するもの」から,「対話するもの」に変えつつある．中でも音声による会話はその中心になっていくだろう．しかしながら，音声入力には「音漏れ(周囲への迷惑・情報漏洩)」という大きな問題が残っている．狭いオフィスや家庭，あるいは混雑する街頭で皆が大声で喋りまくる未来はぞっとしない．静かで心地よい世界を実現するには,「音漏れしない」音声入力手法の開発が必須と言える．

1.1 サイレント音声入力

サイレントな音声入力には，全く音を出さないもの(以下完全サイレント)と，周囲の迷惑にならない程度の微弱の音を出すものの 2 種類がある．

完全サイレントな手法の多くは，発話に伴う口腔周辺のパーツの動きを各種センサで捉えるものである．Sottovose[13]は，下顎に当てた超音波センサを用いて顎・唇・舌などの動きを検出し，音声に変換する.頬や下顎に設置した加速度センサ[18]や EMG (Electromyogram，筋電) 電極[28]を用いるものもある．このほか，マスク内部に設置した加速度センサ[6]や歪センサ[9]を用いるものや,イヤホンから外耳道内に放射した音波の反射を用いるもの[3]もある．また，頭部に設置した複数の電極で発話をイメージした際の ECG (Electro-encephalogram，脳波) を検出するもの[4]に至っては，もはや口すら動かす必要が無い．

いずれの手法でも，得られたセンサの信号を最終的な文字や音声に変換するには，専用に設計された認識機構が必要であり，多くの場合は話者個々人に対する追加学習を必要とする．また，認識できる音素や単語数が限られているものが多く，リアルタイム処理が難しいものもある．

1.2 ささやき声による音声入力

これに対し，ささやき声などの微弱な音を検出する手法は，得られる信号が本来の発話音声に似ていることから，認識機構の設計が前者に比べて容易であり，認識率や認識速度等の性能も上げやすい．

この分野の先駆的な研究としては NAM (Non-Audible Murmur, 非可聴つぶやき) [12]がある．耳介後下方部に接触させた「肉伝導マイク」で微小なつぶやき音声を収録することで，ほぼ音漏れの無い音声入力を可能としている．SilentVoice[11]は，従来の発話とは逆に「息を吸いながら話す」ことで，ポップノイズ無しにマイクを極限まで近づけられ，40dB(A)未満の超微弱音声の検出が可能である．WhisperMask[5]は，マスク内側に設置した面状の ECM (Electret Condenser Microphone) の持つ優れた close talk 性能により，高騒音環境においても小さい発話音声をクリアに検出することができる．多くは取得した ささやき声を文字として認識するが，WESPER[17]のように，リアルタイムに通常音声に変換するものもあり，コンピュータとの対話だけで無く，人間同士の会話にも使用可能である．

一部の音声認識機構 (Google Voice Input, OpenAI Whisper, Amazon Alexa の "Whisper mode" 等) は ささやき声の認識に対応しており,周囲に迷惑をかけずに入力を行なうことができる．また，通常音声と ささやき声を識別することで，ささやき声をコマンドとして用いることもできる[16].

しかし，本来小音量である ささやき声を良好に収録するには，静かな場所で使用する，スマートホンやスマートスピーカ等の機器に口を近づける，ある

* Microsoft Corporation

(English Title)
Whisphone: Whispering Input Earbuds

いはヘッドセット等を用いてマイクを口元に近づける必要がある．また，イヤホン型ヘッドセットの多くはビームフォーミング等を用いて口元の音声を増幅している．しかし，たとえこれらの機器を使用したとしても，騒々しいオフィスや街頭での使用は難しく，ヘッドセットマイクを装着したまま日常生活を送ることは容易では無い(食事等の際に邪魔になるほか，「見た目」の問題から一般大衆への浸透は簡単では無い)．

1.3 骨導マイクロホン

上述の ささやき声入力機構の多くは，発話に伴って周囲大気中に放射された音(気導音)を検出している．これに対し，発話に伴う振動が頭蓋や軟組織を経てきたものを，皮膚表面に設置したセンサで検出するもの(骨導式, 上述の NAM もこの一種)がある．骨導式の収音機構としては，古くから使われている咽頭マイク[22](喉部に接触させる)のほか, 外耳道壁の振動を検出するもの[21]，小型の ECM マイクで外耳道内放射音を収録するもの[20]，イヤホンに設置した MEMS 振動センサによって耳穴周辺の振動を検出するもの[19]等がある．また，市販されているノイズキャンセルイヤホンの中にも，筐体内部に設置した加速度センサを用いて発話時の振動を検出し，音声の明瞭度向上を図っているもの[7]がある．骨導式は外部騒音に強い反面高周波成分の減衰が強く，声が籠って聞こえてしまうという問題があり，特殊状況下での使用に留まっている．また NAM を除き，通常発話の収録を目的としており，微小な ささやき声を十分なレベルで検出できるものは少ない．

1.4 骨導による ささやき声収録

一般的に知られている骨導聴取の経路は，頭蓋や軟組織を経て伝わった音響振動が，直接耳小骨や蝸牛に到達して音として知覚されるというものである．一方でこれらの音響振動が，外耳道壁を震わせて外耳道内部に一旦「音として」放射され(外耳道内放射)，それが(通常の気導音と同様に)鼓膜を伝わって聴取されるという経路もあり，特に健聴者においては，こちらの寄与も大きいとされている[25]．また，指や耳栓などで外耳道開口部を塞いだ際に，骨導聴取において 1kHz 以下の低周波領域が 5-20dB 強調されて聞こえる現象(外耳道閉鎖効果)が知られており，これは後者の経路によって外耳道内に放射された音響エネルギーが，外耳道開口部から逃げなくなる為だと考えられている[8]．

この効果は自身が発した音声(聴覚フィードバック)に対しても有効なので, 外耳道内部にマイクロホンを設置し，併せて外耳道開口部を塞ぐことで，ささやき声などの微弱な発話を増幅された状態で収録できると考えられる．また，これは「耳を塞いだ」状態に相当するので，マイクロホンへの外部騒音の侵入も同時に阻止でき都合が良い．

しかしながら，ささやき声の発話音量は 40dB(A) 程度であり，60dB(A)である通常発話に比べてかなり低い．仮に外耳道閉鎖による骨導成分増幅や外部騒音低減があったとしても，高騒音下で良好な S/N 比を得るのは簡単では無い．

一方，近年の TWS (True Wireless Stereo) イヤホンの普及に伴い，ANC (Active Noise Canceling) 技術の発展が著しく, 特にカナル型のイヤホンでは，公称値で 45dB を超えるノイズ抑圧性能を謳うものも多くある[2]．カナル型イヤホンを装着した状態が「外耳道閉鎖」と同じであることを考えると，ANC 機能を併用することで，同様のノイズキャンセル効果が外耳道内部に設置したマイクロホンに対しても得られることになる．

本稿では，常時装用型 ささやき音声入力デバイスである Whisphone の提案を行う．ANC を動作させたカナル型イヤホンの耳栓先端部分に設置したマイクロホンで外耳道壁から放射される骨導音声の収録を行うことで，たとえ騒音下であっても，周囲の迷惑にならないような微小な ささやき声を安定して取得できる．以下の章では，外耳道閉鎖による骨導音声の増幅と，耳栓と ANC の併用による外部騒音低減効果について実験結果と共に説明する. 次いで，汎用の音声認識機構と組み合わせることで，追加学習や個人適応を行うこと無く，騒音下であっても ささやき声の音声認識が可能であることを示す．いくつかの実装例を紹介すると共に現状の課題や解決策を示し，将来への展望を述べて結言とする．

2 Whisphone

2.1 構成

Whisphone の構成を図 1 に示す．ANC 機能を持つカナル型イヤホンの耳栓先端部に設置したマイクロホンによって，ささやき声発話時に外耳道内に放射される骨導音声を収録する．カナル型イヤホンによって外耳道開口部が密閉されるので，外部騒音の低減に加え，外耳道閉鎖効果による音声信号の増幅が期待できる. ANC 機能による更なる外部騒音低減

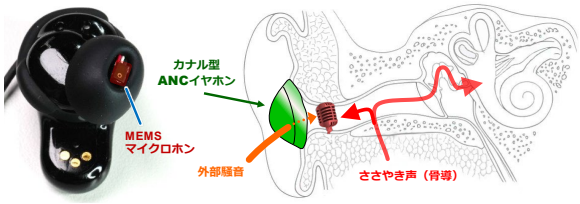
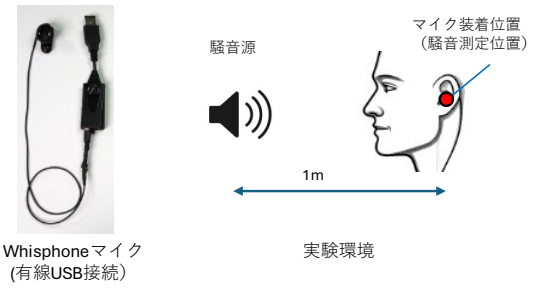


図 1. カナル型 ANC イヤホンの耳栓先端部分にマイクを設置し，ささやき声の外耳道内放射音を収録する.



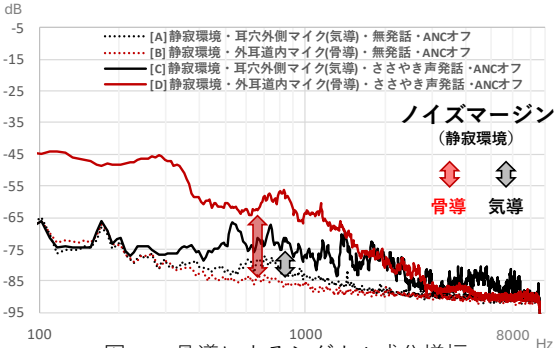


図2(a): 骨導によるシグナル成分増幅

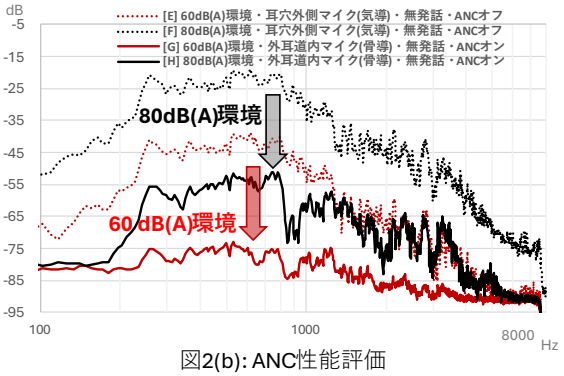


図2(b): ANC性能評価

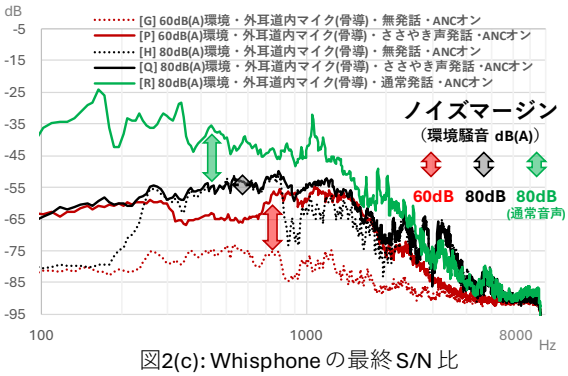


図2(c): Whisphone の最終 S/N 比

図 2. Whisphone の収音性能 (a): 外耳道閉鎖効果によって，ささやき声を 10dB 程度増幅して収音できる．(b): ANC によって外部騒音を 30dB 程度低減できる．(c): 総合的な S/N 比改善効果は 40dB 程度になり，騒音下でも微弱な ささやき声の収録が可能になる．

により，騒音下においても ささやき声による音声収録が可能である．

2.2 収音性能

初めに，Whisphone の収音性能の確認を行った．ANC 機能を持つカナル型イヤホン[2]の耳栓先端部に小型の MEMS (Micro Electro Mechanical Systems) マイク[10]を設置し，USB オーディオインタフェース[1](ローカットフィルタを 235Hz に変更)を介して PC に接続した．イヤホンを外耳道に装着した状態で ささやき声による入力を行い，音声波形を記録した．ささやき声の発話音量は，1m 距離で標準的な 40dB(A)となるように調整した．騒音源には，電車のホームで収録した音源を用い，耳元で 60dB(A)もしくは 80dB(A)になるように調整している．なお，実験環境の暗騒音は 33dB(A)であった(以下，静寂環境と表示)．ANC 機能は同イヤホンの “Quiet” モード(最大強度)を用いている．

実験は，被験者一名(成人男性)に対して行った．発話音声としては『こんにちは，私の声が聞こえますか？』を用いた．外部騒音レベル(静寂環境/60dB(A)/80dB(A))・マイク設置位置(外耳道内/耳穴外側)・発話/無発話・ANC 機能の On/Off を変更しつつ収録を行った．

収録結果を図 2 に示す．図 2(a) は，外耳道内マイク及び外耳道閉鎖による音量増幅の確認である．静寂環境におけるバックグラウンドノイズ [A] に対し，通常のイヤホンで使われるような耳穴付近に設置したマイクでの気導音の収音 [C] ではノイズマージンは 10dB 程度であり，微弱な ささやき声を十分な音量で取得することは難しい．一方で，マイクを外耳道内に設置し，外耳道開口部を塞いだ状態 [D] では，1.5kHz 以下の領域で，同じ ささやき声を 20dB 程度のノイズマージンで取得することができている．これは，耳栓による外部ノイズレベルの低下 [B] に加え，外耳道閉鎖効果による外耳道内放射成分の増強によるものと考えられる．一方，骨導収音においては 2kHz 以上の成分は急激に減少し，4.5kHz 以上ではほぼ収録されていない．なお，同図の [D] において，500Hz 以下の成分が急激に増加している．これは主に発話に伴う外耳道の変形によるものであるが，現在は積極的に使用していない(詳細後述)．

図 2(b) は，ANC 機能による外部騒音低減の確認である．60dB(A)及び 80dB(A)の外部騒音に対しては，耳栓と ANC 機能を併用することで，30dB 程度のノイズ低減効果が得られている ([E]vs[G], [F]vs[H])．

図 2(c) は，Whisphone の収音性能の確認である．外耳道マイクと ANC 機能の併用によって，60dB(A)

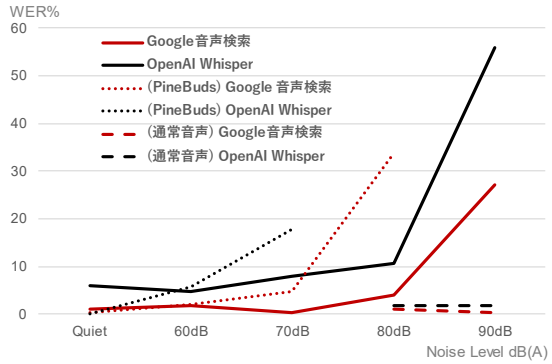


図 3. Whisphone を用いた音声認識の性能評価

の外部騒音下においては，ささやき声による発話を最大 20dB 程度のノイズマージンを保って取得することができている([P]vs[G]). ささやき声の音量が 40dB(A)であることを考えると，Whisphone による総合的な S/N 比改善効果は，通常の気導マイクに比べて最大 40dB 程度(外耳道閉鎖効果によるシグナル成分増強が+10dB, ANC によるノイズ成分低減が -30dB)であることがわかる. また, 外部騒音 80dB(A)でも同様に 40dB 程度の S/N 比改善効果を得られているが，ノイズマージンはほぼ無くなってしまう([Q]vs[H])為，このあたりが使用限界と考えられる.

外部騒音レベルがこれ以上になると，ささやき声による音声認識は困難になってしまうが，通常発話(60dB(A)程度の音量)を用いた場合は更に 20dB 程度のノイズマージンがある為，90dB(A)を超える外部騒音下であっても音声の取得が可能([R]vs[H])である(外部騒音が 80dB(A)以上の場合，60dB(A)程度の通常発話は周囲の人からは ほぼ聞こえなくなる).

2.3 音声認識性能

次いで簡易的な音声認識性能の評価を行った．使用した機材やノイズは前節で用いたものと同一である．音声認識機構には，Google 検索ページの「音声検索」(以下 Google 音声検索)及び OpenAI の Whisper (large-v3)を用いており，追加学習やエンロールなどの個人適応は一切行っていない．テストには，音声アシスタントや生成 AI を用いた操作で良く使われるような語句(『明日の予定は何がありますか』・『東京駅はどう行けば良いですか？』・『新しいアイデアを出してください』等)を 15 個選んでいる．ささやき声による発話は各 3 回行い，WER (Word Error Rate, 単語エラー率) を算出した．実験環境における暗騒音は 33dB(A)であった．なお被験者は成人男性 1 名である.

図 3 に結果を示す．ノイズレベルが 80dB(A)程度(地上を走る電車の車内等)までは WER が概ね 10%以下であり，小さな ささやき声でもほぼ正確な入力が行えている．一方，90dB(A)(地下鉄の車内等)になると WER が大きく上昇すると共に，VAD (Voice Activity Detection，発話区間検出) の失敗も頻発する為実用的とは言えない.

このことから，Whisphone を用いた音声認識は，外部騒音が 80dB(A)までは特に追加学習や個人適応処理を行うことなく使用可能であると考えられる．なお, 外部騒音が 80dB(A)を超える場合であっても，通常発話を用いた場合には引き続き低 WER での音声認識が可能である(図 3 の破線部参照).

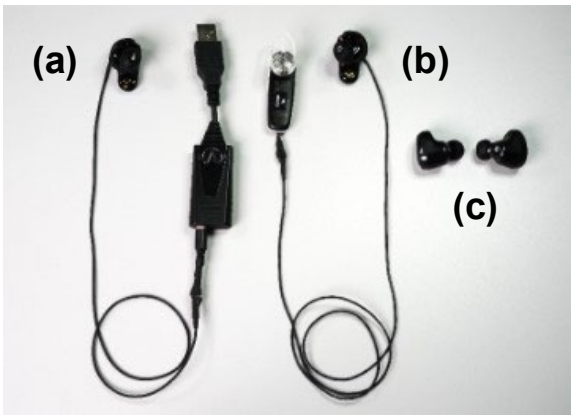


図 4. 実装例 (a): 有線 USB マイクロホンタイプ．(b): Bluetooth ヘッドセットタイプ(モノラル)．(c): Bluetooth TWS イヤホンタイプ.

3 実装

次に Whisphone の実装例を示す．第二章で示したように, Whisphone の騒音抑圧性能を上げる鍵は ANC にある. 幸いなことに, 近年では公称値で 45dB を超えるノイズ抑圧性能を持つ TWS イヤホンが入手できるようになっており，小型マイクをこれらと組み合わせるだけで簡易的な実装が可能である．図 4(a)は小型の MEMS マイク[10]を市販の TWS イヤホン[2](左耳側)と組み合わせた例である．マイクカプセルの信号線は，外耳道の密閉性を損なわないように極細ケーブルを用いて外側に取り出し, USB オーディオ I/F[1]に接続している(一般的な有線マイクロホンとして動作)．使用時には，TWS イヤホンのノイズキャンセル機能のみをオフラインで動作させることで，ノイズ抑圧された骨導音を取得することができる.

図 4(b)は，市販の Bluetooth ヘッドセット[15]の ECM カプセルを取り外し，図 4(a)の小型マイクを接続したものである．使用時には，同じく左耳側はオフラインで ANC 機能のみ動作させ，右耳側のみペアリングして(モノラルの)Bluetooth ヘッドセットとして使用する.

図 4(c)は，一体型の Bluetooth TWS イヤホンと

しての実装例である．市販の ANC イヤホンの中には，小型のマイクカプセルが耳栓先端部に設置されているものがある．これは ANC の FeedBack マイクとして動作させる為であるが，同じマイクを音声収録用として使えば Whisphone を実現することができる(一部の TWS イヤホンでは，通話時にこのマイクを用いて音声のピックアップを行っている[7]．但し，音声収録のメインはあくまでもイヤホンの外側に設置した複数マイクによるビームフォーミングであり，内部マイクや筐体内部の振動センサは，あくまで明瞭度を上げる為の補助的な使用に留まっている)．つまり，このタイプの ANC イヤホンでは，音声収音用マイクとして，イヤホン外側のマイクの代わりに FeedBack 用マイクを使用するようにファームウェアを書き換えるだけで Whisphone を実現できる．ここでは，ファームウェアがオープンソースで公開されている Pinebuds Pro[14]を用いて実装を行った．ANC の性能が劣る(実測値で 10dB 程度)為に騒音下での WER は悪化するが，70dB(A)程度までは比較的良好な WER が実現できている(図 3 の点線部参照)．

4 議論

- エコーキャンセル

イヤホン一体型の機器として用いる場合，イヤホンから再生された音楽や音声が先端のマイクで収音され，ノイズとなってしまう．音声認識時の障害となるほか，リアルタイム通話の場合はエコーとなって相手方に帰ってしまう．イヤホンからの再生信号は既知なので，スピーカーホン等で使われているエコーキャンセル技術を用いて対処可能である．

簡易的な手法としては，左右双方に装着している場合に，音声認識や通話時には片方をマイク，他方をイヤホンとして用いることでもノイズやエコーを防止できる(図 4(c)の TWS イヤホンタイプで実装)．

- ダブル NR

外部騒音に音声が含まれている場合，たとえ低減後であっても音声認識機構が反応してしまうことがある(多くの音声認識機構側に自動ゲイン調整機構がある為)．これを防ぐには，イヤホン外部に別のマイクを設置(ANC イヤホンの FeedForward マイクに相当)して内外のマイクの音量差を比較し，内部マイクの音が大きい成分のみ通過させれば良い(耳栓がある為，外部騒音の音量は外側のマイクが常に大きくなる)．

- 片耳/両耳

Whisphone は片耳/両耳のどちらでも使うことができる．片耳装用の場合，もう一方の外耳道の状態(解放/耳栓/ANC)を変化させても，収録された音声や騒音のレベルに違いは見られなかった．装着の負担が少なく外音聴取も可能であり，音楽鑑賞を行わないのであれば有効な選択肢と言える．

両耳装用の場合，高騒音下では困難になる自身のささやき声の発話状態を(骨伝導による聴覚フィードバックを介して)確認でき，発話音量や滑舌の制御が容易になる．なお，両側に設置したマイクの出力を合成すれば更なる S/N 比改善の可能性もある(理論上最大 1.4 倍)．但し，タイミング調整が難しい(特に TWS の場合)ことに加え，左右音声成分の相関低下(外耳道や頭蓋の形状は完全な左右対称では無い)，騒音成分の相関上昇(骨導によるミックス)がある為，効果は限定的だと考えられる．

- 認識率向上

Whisphone で収録された骨導 ささやき声は，2kHz 以上の高周波成分が著しく減少している．特に高周波成分の多い音素(子音の /s/ や /h/，母音の /i/)の聞こえ方が通常の ささやき声とは異なり，認識率低下の一因と考えられる．骨導 ささやき声による認識機構の再トレーニングが最も有効ではあるが，大量の音声データを集めるのは容易ではない．既に通常 ささやき声のデータがある場合，フィルタによって 2kHz 以上の成分をカットしたもので追加トレーニングを行うことで，新たなデータ収集を行わずとも認識率向上が可能だと考えられる．なお，ささやき声では有声音が無声音化する(例えば /g/→/k/)が，多くの音声認識機構では ささやき声の大量学習や単語辞書を用いて認識率低下を防いでいる．

- リアルタイム音声変換

本稿では主に骨導 ささやき声の認識に焦点を当ててきた．一方，収録された骨導 ささやき声を，通常 ささやき声や通常音声にリアルタイム変換すれば，音声通話等にも利用することができる．前者には欠落している高周波成分の復元が必要であり，骨導音声から気導音声への変換技術[26]が利用できる．後者では更に声帯振動成分やピッチの復元が必要であり，ささやき音声から通常音声への変換技術[17]が利用できるだろう．

- 発話による外耳道変形と入力への応用

図 5(a)上側は，静寂環境における ささやき声発話時(『こんにちは，私の声が聞こえますか？』)の波形例，同下側は，同じ文章のサイレントスピーチ時(声を出さずに同じ口の動きをする)の波形例である．両者で同様のスパイク状の低周波信号が確認できる．これは，発話時の口や顎の動作に伴う外耳道の変形によって生じた圧力変化だと考えられる．これらの成分は，音声波形に比べて過大であり，信号クリップによる歪の原因となる為，現在はフィルタで取り除いており，積極的な利用は行っていない．但し，ここには発話情報が含まれていると考えられ，これ

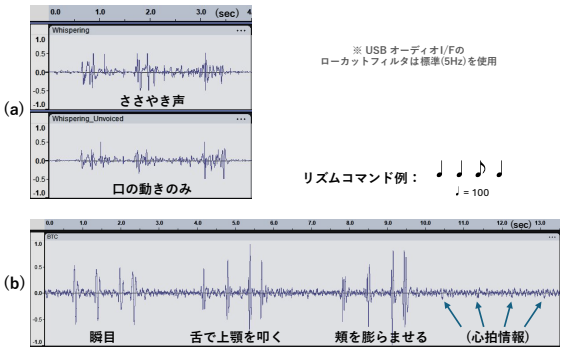


図 5. 音声信号以外の利用可能性 (a): 低周波成分には発話時の口や顎の運動情報が含まれている. (b): 瞬目リズム等をコマンドとして使用できる.

らの成分を含めて学習させることで，認識性能を高められる可能性がある．更に，この成分「のみ」を用いれば，サイレントスピーチの認識が可能となるかもしれない．

また，図 5(b)は，瞬目等を ♩ ♩ ♪ ♩ のリズムで行った場合の波形であり，外耳道内の圧力変化に動作情報が含まれていることがわかる．これらを利用すればハンズフリーでの操作が可能になる(特に舌の動きを用いたものは秘匿性が高い)．簡単なリズムを用いたコマンド[27]であれば，コンパレータやタイマで実現可能であり，特に低消費電力での常時動作が要求される wakeup 操作に適している上，画像や音声の認識に比べてプライバシー問題への対処がしやすい(図 4(c)の TWS イヤホンタイプで AI アシスタントの呼び出しコマンドとして実装)．なお，微弱ながら心拍情報も含まれているので，簡易的な心拍センサとしても利用できるだろう．

- 発声手法

通常発話の主たる音源が声帯振動であるのに対し，ささやき声のそれは，声道中の狭窄部位を空気が通過する際の乱流音である．狭窄部位の生成手法は個人や状況によっても異なるが，声帯を狭める(「ハスキーボイス」の音源としても使われる)・上顎に舌を近づける・唇を狭める(いわゆる「ひそひそ声」)等があり，声質にも違いがある．声帯を狭める手法は大きな乱流を発生させることができ，発話可能な音素も多い(ほぼ全ての無声音)．一方で高音域が弱く明瞭度が下がるのに加え，声帯への負担も大きい．これに対し，口先で喋る手法は音量を絞りやすく声帯の負担にはなりくいが，発声可能な音素が限られる(特に鼻音 /n/ が難しい)．上顎を使う手法は両者の中間である．認識率の更なる向上には，それぞれの発声手法による追加学習が必要だと考えられる．

なお，ささやき声は SilentVoice[11]で述べられている吸気発話(Ingressive Speech) を用いても発声可能である．音量はほぼ同じである為秘匿性向上には繋がらないが，呼気発話との弁別が可能になれば，人間向け(呼気発話)とネット向け(吸気発話)の切り替えに利用できるかもしれない．

- オートベント

日常生活で密閉度の高いカナル型イヤホンを使用する場合，周囲とのコミュニケーションや危険把握の為に適切な外音の取り込みが必要となる．現在は主に「外音取り込みモード」(外のマイクで取得した音の一部をイヤホンから再生する)が用いられているが，消費電力増大が課題である．また，密閉度を上げると自身が発話した音声の聴覚フィードバックが過大となる・心拍音が聞こえる等の問題も出てくる．通常はベント穴を設けることでこれらを軽減しているが，一方でノイズ抑圧性能の低下を招いてしまう．Whisphone を常時装用する場合には，これらの問題への対応が求められる．例えば，圧電素子を用いた(Normally Open タイプの)アクティブベント機構[23]を用いれば，電源オフ時にはベント穴を開放して(電気的な再構成では無い)外音をそのまま通過させ，動作時には密閉してノイズを遮断することができるだろう．

- 他言語への対応

今回用いた Google 音声検索や OpenAI Whisper は多言語での ささやき声認識に対応している．2 章で行った実験と同様のセンテンスに対して，英語における WER の計測を行った結果を示す．静寂時での通常マイクによる ささやき声認識の WER がそれぞれ 2.4%・0.0%なのに対し，Whisphone 環境では 55.0%・56.2%と極端に悪くなった．英語は子音(1kHz 以上の高周波成分が支配的)が多い言語であり，骨伝導による高周波成分減少との相性が悪い．このほか中国語(普通話)は声調(tone)ベースの言語であり，tone 成分がほぼ存在しない ささやき声との相性が悪い．その意味では Whisphone は，母音の寄与率が高く tone の変化が少ない日本語(及び韓国語)での使用に適していると考えられる．

5 おわりに

本稿では，常時装用型 ささやき音声入力デバイスである Whisphone の提案を行った．まずは社会的適合性(周囲への迷惑を気にする)と言語的メリット(骨導 ささやき声入力に向いている)が多い日本での普及を足掛かりに，いずれ他の地域・言語にも広げて行きたい．本稿では主に音声入力を対象としたが，ささやき声を用いたリアルタイム会話にも，周囲への迷惑低減や情報漏洩防止等のメリットがある為，将来 ささやき声での会話が日常的に使われるようになるかもしれない．

静かな未来がやってくるのだ．

参考文献

[1] Andrea Communications USB-MA (2024/10/27 確認) https://andreacommunications.com/products/usb-ma-premium-external-usb-microphone-adapter/

[2] BOSE QuietComfort Ultra Earbuds (2024/10/27 確認) https://www.bose.com/p/earbuds/bose-quiet comfort-ultra-earbuds/QCUE-HEADPHONEIN.html

[3] DONG, Xuefu, et al. ReHEarSSE: Recognizing Hidden-in-the-Ear Silently Spelled Expressions. *In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2024.* pp. 1-16.(2024)

[4] FITRIAH, Nilam; ZAKARIA, Hasballah; RAJAB, Tati Latifah Erawati. EEG-Based Silent Speech Interface and its Challenges: A Survey. *International Journal of Advanced Computer Science and Applications, 2022,* 13.11. (2022)

[5] HIRAKI, Hirotaka, et al. WhisperMask: a noise suppressive mask-type microphone for whisper speech. *In: Proceedings of the Augmented Humans International Conference 2024*. pp.1-14. (2024)

[6] HIRAKI, Hirotaka; REKIMOTO, Jun. SilentMask: mask-type silent speech interface with measurement of mouth movement. *In: Proceedings of the Augmented Humans International Conference 2021*. pp. 86-90. (2021)

[7] HUAWEI FreeBuds Pro 3 (2024/10/27 確認) https://consumer.huawei.com/jp/audio/freebuds-pro-3/

[8] K.Ito; S.Nakagawa. Bone-conducted ultrasonic hearing assessed by tympanic membrane vibration in living human beings, *Acoust, Sci. & Tech*, Vol. 34(6), pp. 413-423, (2013)

[9] KUNIMI, Yusuke, et al. E-MASK: a mask-shaped interface for silent speech interaction with flexible strain sensors. *In: Proceedings of the Augmented Humans International Conference 2022*. pp. 26-34. (2022)

[10] LinkMems LMA2718T381-OY1 (2024/06/18 確認) https://www.lcsc.com/product-detail/MEMS-Microphones_LinkMems-LMA2718T381-OY1_C5455565.html

[11] Masaaki Fukumoto. SilentVoice: Unnoticeable voice input by ingressive speech. *In Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology, UIST '18*, pp. 237-246. (2018)

[12] NAKAJIMA, Yoshitaka, et al. Non-audible murmur (NAM) recognition. *IEICE TRANS. on Information and Systems*, 89.1: pp. 1-8. (2006)

[13] Naoki Kimura, Michinari Kono, and Jun Rekimoto. Sottovoce: An ultrasound imaging-based silent speech interaction using deep neural networks. *In Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI '19*, pp. 1-11. (2019)

[14] PINE64 PineBuds Pro (2024/10/27 確認) https://pine64.org/devices/pinebuds_pro/

[15] Plantronics M70 (2024/10/27 確認) https://web.archive.org/web/20200416222955/https://www.plantronics.com/us/en/product/m70

[16] REKIMOTO, Jun. DualVoice: A speech interaction method using whisper-voice as commands. *In Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts. 2022*. pp. 1-6. (2022)

[17] REKIMOTO, Jun. WESPER: Zero-shot and realtime whisper to normal voice conversion for whisper-based speech interactions. *In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2023*. pp. 1-12. (2023)

[18] REKIMOTO, Jun; NISHIMURA, Yu. Derma: silent speech interaction using transcutaneous motion sensing. *In: Proceedings of the Augmented Humans International Conference 2021*. pp. 91-100. (2021)

[19] SCHILK, Philipp, et al. In-Ear-Voice: Towards Milli-Watt Audio Enhancement With Bone-Conduction Microphones for In-Ear Sensing Platforms. *In: Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation. 2023*. pp. 1-12. (2023)

[20] SONY ECM-TL1 (2024/10/27 確認) https://web.archive.org/web/20110325062731/https://www.sony.jp/microphone/products/ECM-TL1

[21] TEMCO JAPAN EM20N-T3.5P (2024/10/27 確認) https://www.temco-j.co.jp/products/em20n-t3-5p/

[22] TEMCO JAPAN TM80N-T (2024/10/27 確認) https://www.temco-j.co.jp/products/tm80n-t/

[23] xMEMS SKYLINE (2024/10/27 確認) https://xmems.com/acousticvent/

[24] 安藤毅; 田野俊一; 市野順子; 橋山智訓. 痕跡器官の末梢系生体情報を用いた常時装着型入力インタフェース. HI シンポジウム 2008, pp.229-234. (2008)

[25] 伊藤一仁. 骨導聴覚の知覚機序に関する実験的考察. 純真学園大学雑誌. Vol.11, pp.67-72, (2021)

[26] 川本大貴; 中山仁史. 畳み込みニューラルネットワークを用いた光骨伝導音の音質改善検討.日本機械学会論文集, 2024, 23-00304, (2024)

[27] 福本雅朗; 外村佳伸. “指釦”:手首装着型コマンド入力機構. 情報処理学会論文誌, Vol.40, No.2, pp.389-398, (1999)

[28] 真鍋宏幸; 平岩明; 杉村利明. 無発声音声認識: 筋電信号を用いた声を伴わない日本語 5 母音の認識. 電子情報通信学会論文誌 D, 2005, 88.9: pp. 1909-1917. (2005)

未来ビジョン: 進化か先祖返りか?

我々人類は古くから主に音声によって人対人のコミュニケーションを行ってきた. この先, 人類が情報空間と共に生きていく為には, 対話先(人かネットか)の瞬時確実な切替手段が必要になるだろう(ウェイクワードやジェスチャーでは誤入力の可能性を排除しにくい). しかし我々が持つ唯一のアクチュエータである筋肉は, 既に日常生活中で用いられている為, それらと干渉しない新たなモードを探すのは簡単では無い. 例えば耳動筋等の痕跡器官を使う手法[23]は若干の訓練こそ必要ではあるが, 新世代人類の基本技能としては有望だと思われる(例 : 耳を持ち上げながらひそひそ声で話すと音声入力). 今から耳を動かす練習をしておいた方が良いかもしれない(朝練耳上げ百回!).