Vol.2021-HCI-191 No.12
2021/1/29

発声由来のウイルス拡散抑制を目指すスロートマイク音声処理

音質向上のための学習用データセットの構築に向けた予備実験

飯野 健広[†1]　木本 雅彦[†2]　中垣 淳[†1]　早川 吉彦[†1]　奥村 貴史[†1,a)]

概要：2020 年に世界的な大流行を引き起こした新型コロナウイルスにおいては，感染経路として発声により生じる口腔からの飛沫が重視されている．そのため，密集状態における発語を禁ずることにより，当該ウイルスの伝播経路を効率的に遮断しうる可能性がある．しかしながら，社会生活において発語を禁じることは現実的でない．そこで我々は，僅かな発声をも採音しうるスロートマイク (咽喉マイク) を活用することにより，感染性ウイルスの環境への拡散を極小化しうる可能性に着目した．
本研究では，スロートマイクの音質向上のため，高音質なコンデンサマイクとの同時録音を通じた対照データ生成の予備実験を行った．これら対照データを用いたスペクトル等の定量的定性的比較の結果，「首の運動に伴うノイズ」，「嚥下音」に加えて，「両唇音」などが，スロートマイクを介した会話の障害となることを同定した．今回得られた知見を元にフィルタ学習用のデータセットを構築することにより，スロートマイクを用いた低音量での各種会話支援技術の実現が期待される．

Throat Microphone Speech Processing for infection control of vocal origin viruses

Preliminary Experiments toward development of training data to improve the sound quality

1. はじめに

2020 年，新型コロナウイルスの流行によって，社会生活は大きな変化を余儀なくされた．この新型コロナウイルスは呼吸器感染症であることから，感染症の伝播においても口腔からの飛沫感染とエアロゾル感染が重視されている [1]．そのため，マスクの着用が感染拡大の抑制に効果的であることは，パンデミックの初期より示されており [2]，大きな発声による飛沫の拡散を防ぐため湿った布で口を覆うことが効果的であることなども報告されてきた [3]．

しかし，マスクを着用したとしても全ての飛沫を抑えることはできない．マスクでトラップしうるよりも微細な飛沫が発声により発生し，感染を成立させる可能性は以前から指摘されてきた [3]．逆に，首都圏の通勤電車のように過密で換気が悪い環境においても，患者の爆発的な感染が広がらなかったことは，過密であること (密集・密接)，換気が悪いこと (密閉)，すなわち政府が目安としてきた「3 密」[4] 以上に，発声の有無が感染成立に関与していることを強く示唆している．

そのため，発語を禁ずることにより，当該ウイルスの口腔外への飛沫拡散量の減少を通じて，伝播経路を効率的に遮断しうる可能性がある．しかしながら，社会生活において発語を禁じることは現実的でない．そこで我々は，咽喉部に装着し声帯振動を直接拾うことにより僅かな発声をも採音するスロートマイク (図 2) を活用することにより，感染性ウイルスの環境への拡散を極小化しうる可能性に着目した．ただし，スロートマイク (咽喉マイク) は，声帯振動を直接拾うことから音声的な特性の偏りが大きくことなることに加えて，唾液を飲み込む音等のノイズを避けることができない．そのため，スロートマイクを用いた会話を可能とするためには，音質の向上が欠かせない．

そこで本稿では，ウイルス拡散の極小化に向けて発声を抑制していくうえで，スロートマイクの有する各種の特性の補正に向けた予備実験を試みた．まず，次章において，スロートマイクを紹介しその可能な応用を示すと共に，関連研究として今までに試みられてきた微細な音声や無声による音声認識技術について整理する．次に，3 章において，スロートマイク音声と高音質なコンデンサマイク音声の比較により，スロートマイクの音質的特性を解析する．さらに，4 章において，音質改善のためのフィルタ学習用のデータセット構築について考察し，5 章に結語を記す．

[†1] 現在，北見工業大学
Presently with Kitami Institute of Technology
[†2] 現在，WIDE Project
[a)] tokumura@mail.kitami-it.ac.jp

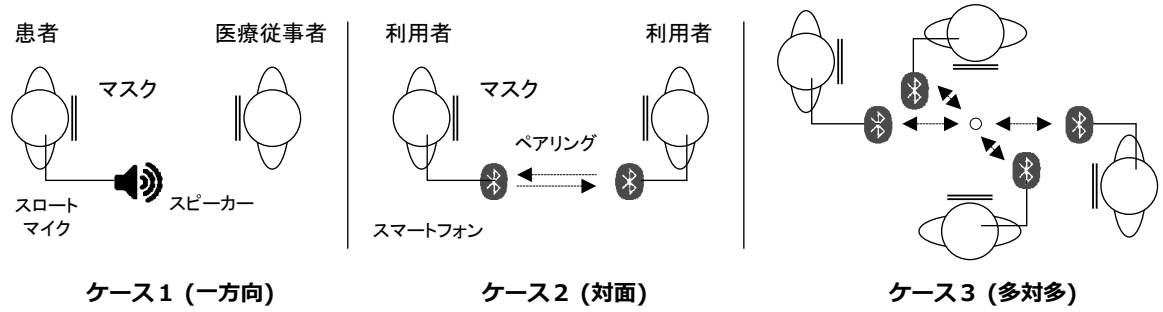

図 1 スロートマイクを用いた会話のユースケース

2. プロジェクト概要と関連研究

2.1 スロートマイクとは

スロートマイクは，音声信号を拾う一般的なマイクと異なり，咽喉周辺の骨伝導により音声を拾うマイクである．そのため，喉元での微細な発声である「ささやき声 (密語声)」が拾えることに加えて，周囲の雑音による影響を受けにくいという特徴がある．そこで，静音性が求められるある種のスポーツや，戦車内等の騒音環境におけるコミュニケーション等，特殊な用途に利用されてきた．

図 2 スロートマイク

2.2 想定するユースケース

我々は，スロートマイクを用いた発声の抑制を通じた感染制御に向けて，以下に示すような状況に活用しうる技術の研究開発を進めている．本稿で扱うスロートマイクの音質向上技術は，それぞれの状況に応じた会話支援システムに共通する要素技術に当たる．

まず，医療現場においては，医療従事者が感染した患者に直接接触することになる．たとえば，各種の診察に加えて，点滴や採血，体位交換等を患者に接触せずに行うことはできない．また，クラスタが多数発生している各種の介護福祉施設においても，食事や移動，排泄介助等において，患者との接触が生じる．これらに際して，作業者は防護衣を用いるが，患者より環境中に排出されるウイルス量の抑制は，感染リスクの軽減に直接の効果を有することが期待される．そのため，図 1・ケース 1 に示すように，患者にスロートマイクを装着し，「ささやき声」を用いてもらうことでウイルス排出を抑制する手法が考えられる．

次に，対話における発声の抑制が考えられる (図 1・ケース 2)．このケースにおいては，会話をする双方がそれぞれのスマートフォンにスロートマイクを接続し，短距離無線通信規格である Bluetooth 等を用いて両者のスマートフォンを一時的にペアリングする．それによって，お互いの会話をイヤホンで聞きあうことで，スマートフォンを補聴器のように用いることを実現する．その際，会話の音量調整が必要になるが，スマートフォン間の相対的な位置情報を用いることで距離感に応じた音量調節を可能とする等，効率的な会話を可能とするための工夫が考えられる．

最後に，複数名による会食的な環境での発声の抑制が考えられる (図 1・ケース 3)．新型コロナウイルスによるパンデミックにおいては，不特定多数が集まり近距離で会話を交わす会食が感染拡大に大きく関わることが明らかとなっている [4]．そのため，会食の場となる飲食業を中心に自粛が求められ，経済的に大きな打撃が生じた [5]．スロートマイクを短距離無線通信を用いた多対多会話に活用することで，この問題の解決がもたらされる可能性がある．

2.3 関連研究

近年，HCI (Human–Computer Interaction) 分野において，サイレント音声認識と呼ばれる技術が研究されている [6]．この技術は，主として発話障害者の支援や，静穏性が求められる環境，ないし，周辺雑音が大きな環境における会話の実現を目的として発展してきた．たとえば，「吸気」による発話を用いる入力デバイス「SilentVoice」[7] や，超音波画像を用いた無発声音声システム「SottoVoce」[8] が提案されており，発話音量の削減が試みられている．

しかし，これらの手法は，入力に訓練が必要であったり高額な機器を利用する点で不利がある．実際，サイレント音声認識技術を用いたウイルス拡散抑制法は知られていない．一方，本研究での提案手法は，高い普及率を誇るスマートフォンに，廉価なものは 1000 円を下回る価格で購入可能で安価なスロートマイクを接続するだけで実現しうるもので，価格面，普及面で優位性を有すると期待される．

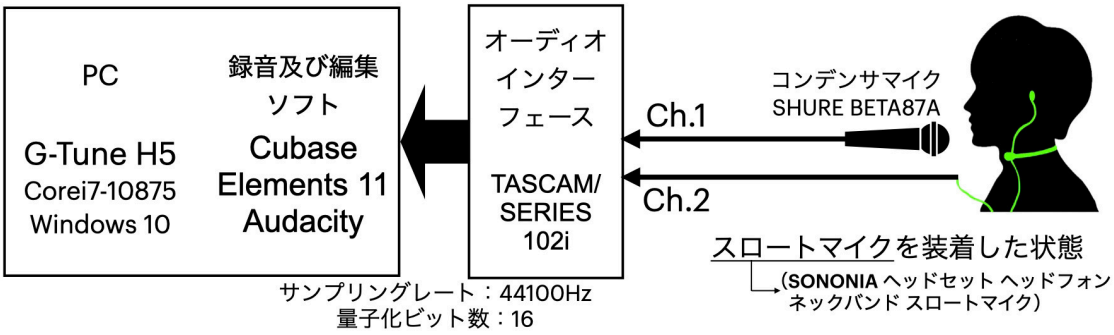

図 3 実験機材及び環境構図

3. 予備実験

3.1 実験概要

本章では，スロートマイクの特性を評価し，音質向上への糸口を探る．そのために，スロートマイク音声と高音質なコンデンサマイク音声の比較により，スロートマイクの音質的特性の定量的，定性的な解析を試みた．

図 3 に，実験環境を示す．まず，高品質なオーディオインターフェースのチャンネル 1 と 2 にコンデンサマイク (超単一指向性) とスロートマイクを接続し，2 種類の音声の同時録音が可能な環境を用意した．録音には，PC 上に用意した Cubase Elements 11 を行い，Audacity 上で周波数スペクトル等の解析を行った．実験は，環境音量約 50dB の研究室にておこなった．防音設備はないものの，コンデンサマイク側の録音データより，周囲環境音はほぼ排除されていることを確認できている．

被験者としては，20 代男性の被験者 1 名 (著者 T.I) を対象とした．発音に際しては，大きく・ゆっくり・はっきりとした発音を「明確音声」，音量を下げ，声帯の振動を意識的に控えて発話したものを「ささやき声（密語声）」とし，複数の発声方法を区別し録音した．

以下では，まず，スロートマイクの「周波数特性の解析」を行った．次に，スロートマイクを介した会話に際して支障となり得る「ノイズの特定」を行った．そのために，録音音源に対して発話部分と非発話部分，ノイズ部分を対象とした分類を試みた．最後に，発話における「音声学的解析」を試みた．それぞれ，実験構成と結果，考察を示す．

3.2 周波数特性解析

本研究で用いるスロートマイクは，将来的な普及過程を見据えて，安価なモデル (1000 円前後) を利用している．そのため，利用説明書等において周波数特性は記載されていない．そこでまず，利用するスロートマイクの周波数特性の解析を行った．

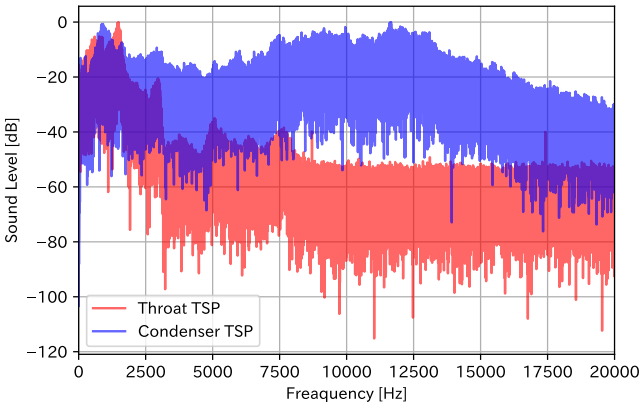

図 4 スロートマイクとコンデンサマイクの周波数特性

周波数特性は，インパルスをシステムに入力，その出力インパルス応答をフーリエ変換することで理論上得られるが，出力が十分でないなどの問題があり一般的には困難である．そこでインパルスを使うことなくインパルス応答を算出することのできる TSP 信号を用いた解析を行った．得られたスロートマイクとコンデンサマイクの周波数スペクトルを図 4 に示す．

スロートマイク (図 4・赤) の周波数特性として，低周波にピークがあることが分かる．一般的な電話音声は，人間の声に含まれる周波数成分のうち 300Hz〜3.4kHz を送っているとされ [9]，スロートマイクはこの範囲の一部をピックアップしていることになる．しかしながら、電話音声が実際の会話よりも不明瞭であるのは，まさに伝送している周波数がこの範囲に制限されているためであり，広帯域音声技術はその倍の範囲に相当する 50 Hz〜7kHz の周波数成分を伝送することで音質の向上を実現している．この点は，業務用コンデンサマイク (図 4・青) が，倍音成分を含むより高音域に至るまでフラットにピックアップしうる点との比較によりより明瞭となる．スロートマイクの音質向上に際しては，この極端に狭い周波数特性をいかに補完するかが重要となると考えられる．

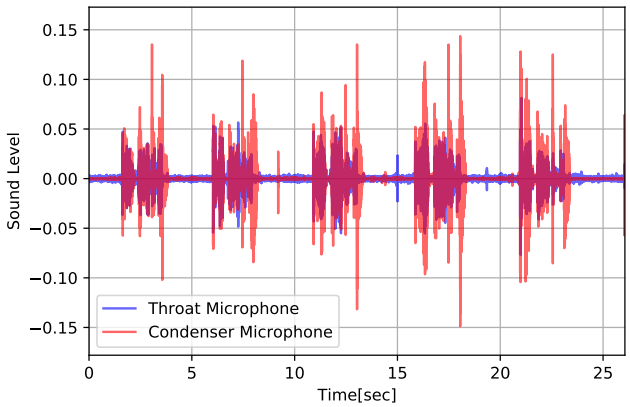

図 5 「ささやき声」に含まれる嚥下音成分

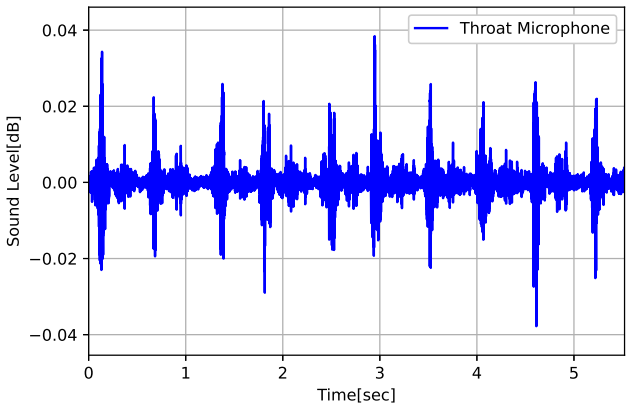

図 7 スロートマイクで収音した首元で生じる摩擦音波形

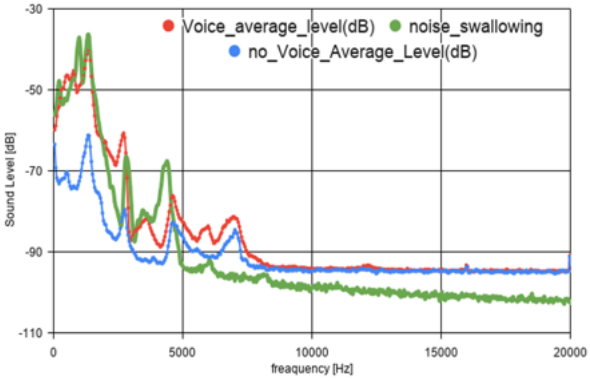

図 6 スロートマイク音の発話、非発話、嚥下音のスペクトル

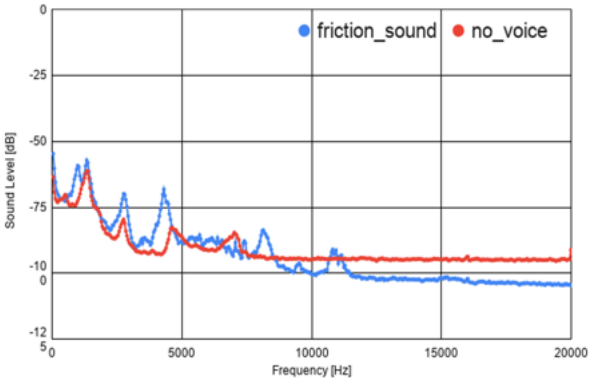

図 8 スロートマイクの非発話部分と装着部摩擦音のスペクトル

3.3 ノイズの特定

次に，スロートマイクに含まれるノイズについての検討を行った．そのために，「これはデモ音源です」という文章を 5 回発話したものをスロートマイクで録音し，分析した．このフレーズは，佐藤らによる報告 [10] にて用いられた「犬の名前はモモである」というフレーズを元に選択した．このフレーズには，「音声認識装置による認識率低下が認められた両唇音を含む」，「音声振動を伴う母音または有声子音で構成されている」(基本周波数パターンの抽出から得られた知見による) という特徴が含まれる．

この発話パターンを対象に，オーディオエディタである「Audacity」を用いて，発話部分 5 回と非発話部分 6 回，スロートマイクにおいて混入が避けられない「嚥下音」部分に対してそれぞれ周波数スペクトル (窓関数はハミング窓) を算出し，それぞれの特徴を分析した．

なお，先行研究 [11] では，雑音を「嚥下」「咳」「呼吸」「咀嚼」「発話」の 5 種類の行動に帰している．ノイズ源となるこれらの行動のうち，咀嚼や咳は，音声会話中に起こり得るが頻度は少なく，マイクの遮断により避けうるノイズと考えられる．また，発話や呼吸は，音声会話上に不可避な事象といえる．そこで，本研究においては，ノイズとして特に頻度が高い「嚥下」に着目した．

スロートマイクの波形 (図 5) では，±0.02 以上の部分は発話部分であると判断できる (縦軸ついて波形データを-1 から 1 までに正規化した)．それ以外の部分については，非発話部分として環境音などによるランダムノイズと判断できる．この波形のうち、スロートマイク (青) の非発話部分に混入している突発的なピークが「嚥下音」であり，図 5 に示した波形以外のデータにおいても，発話部分と並ぶかもしくは超える音量であった．

なお，コンデンサマイク (赤) において生じている微小なピークは，発話を開始する前に生じた雑音である．そこで，嚥下音のみを録音したデータを作成し，非発話部分と発話部分のスペクトルを比較した (図 6)．得られたグラフから，嚥下音はささやき声の発話と同等レベルであるが，5kHz 直下にあるピーク成分 (緑) が，発話部分と非発話部分よりも突出している．ここが嚥下音の特徴ということが明らかになった．

次に，身体へと接触しているスロートマイクにおいて同じく混入が避けられない，「摩擦によるノイズ」について分析した．そのために，静止状態での非発話部分と首の筋肉の動きによる摩擦音の抽出を行い，周波数及びスペクトラム等の定量的な非発話部分との比較による評価を試みた．

表 1 日本語発音（子音の分類）

	両唇音	唇歯音	歯音	歯茎音	後部歯茎音	そり舌音	硬口蓋音	軟口蓋音	口蓋垂音	咽頭音	声門音
破裂音	p, b	-	-	t, d	-	ʈ, ɖ	c, ɟ	k, g	q, ɢ	-	ʔ, -
鼻音	-, m	-, ɱ	-	-, n	-	-, ɳ	-, ɲ	-, ŋ	-, ɴ	-	-
ふるえ音	-, ʙ	-	-	-, r	-	-	-	-	-, ʀ	-	-
はじき音	-	-	-	-, ɾ	-	-, ɽ	-	-	-	-	-
摩擦音	ɸ, β	f, v	θ, ð	s, z	ʃ, ʒ	ʂ, ʐ	ç, ʝ	x, ɣ	χ, ʁ	ħ, ʕ	h, ɦ
側面摩擦音	-	-	-	ɬ, ɮ	-	-	-	-	-	-	-
接近音	-	-, ʋ	-	-, ɹ	-	-, ɻ	-, j	-, ɰ	-	-	-
側面接近音	-	-	-	-, l	-	-, ɭ	-, ʎ	-, ʟ	-	-	-

表 2 日本語発音における調音位置

呼称	上の調音器官	下の調音器官	例
両唇音	上唇	下唇	[p], [m]
唇歯音	上歯	下唇	[f], [v]
歯音	上歯	舌端	[θ], [t]
歯茎音	歯茎	舌端	[s], [t]
後部歯茎音	後部歯茎	舌端	[ʃ], [ʒ]
そり舌音	後部歯茎	舌尖	[ɻ], [ʂ]
硬口蓋音	硬口蓋	前舌	[j], [ʎ]
軟口蓋音	軟口蓋	後舌	[k], [ŋ]
口蓋垂音	軟口蓋後部と口蓋垂	口舌後部	[N], [R]
咽頭音	咽頭壁	舌根	[ħ]
声門音	左右の声帯	左右の整体	[h], [ʔ]

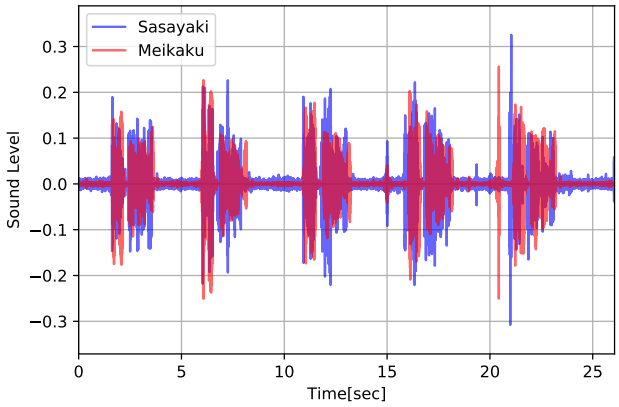

図 9 明確音声とささやき声による「これはデモ音源です」を 5 回繰り返した波形の組み合わせ

実験では，首の回転と首筋肉運動を 5 回ずつ連続して行い，摩擦を生じさせた (図 7)．このスペクトルを確認すると (図 8)，5kHz，8kHz，11kHz 近傍にピーク (青) が認められ，摩擦によって生じるノイズに特徴的な周波数であると考えられる．

以上より，スロートマイク使用時に特徴的なノイズである「嚥下音」と「首元で生じる摩擦音」について，実測するとともに，それらノイズの周波数的特徴を示すことができた．これらは特徴的な周波数成分を含むことから，適切なフィルタ処理により除外しうるものと期待される．

3.4 音声学的解析

スロートマイクは，口腔の外においたマイクと異なり，喉元 (咽喉) 周辺に接触している．そのため，口腔や唇，舌といった発音に利用される構成要素を通過する「前」の音声をピックアップすることになる．明確な音声においては振幅の大きさにより収音できていた音も，同じ語をささやき声で発話した場合には，振幅が少ないために効率的に収音できない可能性もある．そこで最後に，スロートマイクを用いた発語における音声学的解析を試みた．

日本語の発音は標準語母音で構成されており，子音は国際音声記号によって，声の有無 (「有声音」,「無声音」に分類)，調音位置，調音様式という三つの基準を基に類別することができる．子音は声道の特定の位置で狭窄が形成され，それが呼気の流れの障害となることによって生じる音である．その中でも最も狭まっている位置のことを調音位置と呼ぶ．狭窄にもさまざまな形態があり，それらを調音様式と呼ぶ．子音は，IPA 国際音声学会 [12] の子音分類により，両唇音をはじめとした 11 項目 (表 1 の 1 行目) と破裂音をはじめとした 8 項目 (表 1 の 1 列目) に分類される (表 1)．また調音位置については，調音器官などに基づいた 11 項目に分類される (表 2)．

本実験では，「これはデモ音源です」の音源及び CosCom「ひらがな カタカナ 50 音表」[13] に掲載されている日本語発音に対して音声品質評価を行った．調音位置がスロートマイクの収音可能範囲に存在しない場合には，聞き取り可能な音声の収音が困難である可能性が懸念される．これは子音について顕著であることが推測された，そこで「これはデモ音源です」という発語を，明確音声，ささやき声の 2 つの発声方法で行い，その音声をスロートマイクとコンデンサマイクで同時録音し，各音声を比較評価した (図 9)．発話部分での明確音声の音量最大値は 0.24，ささやき声の最大値は 0.06 であるため，ささやき声の値を 4 倍することで，明確音声と同等レベルへと補正してある．非発話部分については，ランダムノイズの除去を行っていないため，本考察では考慮しない．

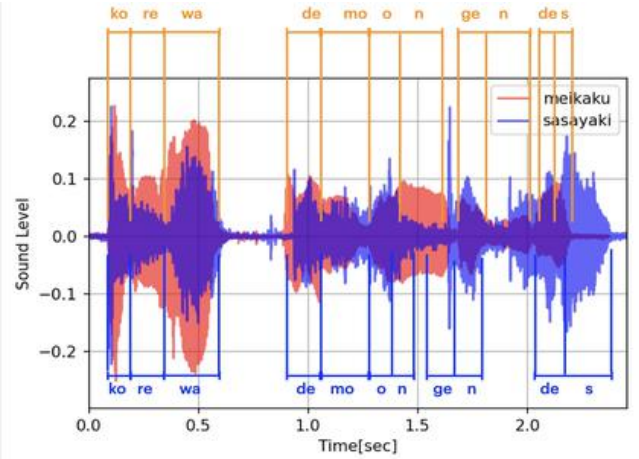

図 10 明確音声とささやき声による発話音声の図からの 6s - 8.5s を抽出した波形（赤：明確音声の発話ラベル，青：ささやき声の発話ラベル）

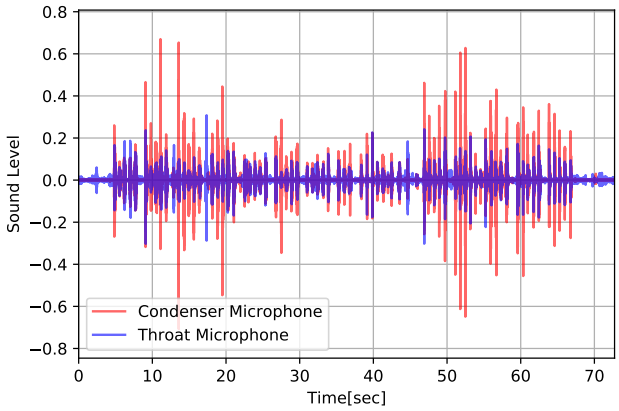

図 11 ささやき声による日本語 50 音の発話のコンデンサマイクとスロートマイクの波形）

このうち、2 種類の発声方法の開始タイミングと発話時間がほぼ同じであった 2 回目 (6s - 8.5s) に着目し，抽出した波形を図 10 に示す．この図において，音また音節ごとの 2 種類の発声方法の差違に着目する．明確音声では，子音を伴う音または音節に対する波形に，コントラストや音節同士の区切りがはっきりみて取れるのに対して，ささやき声ではそれらが明確には見られなかった．とりわけ，[m] 音や [s] 音，[r] 音を伴う波形が他の子音よりも小さい．これらの子音は，調音位置が喉元でない両唇音，そり舌音，歯茎音に該当する．

さいごに，各発音の子音に対する分析を試みた．母音 ([a]，[i]，[u]，[e]，[o])，清音 ([s]，[t]，[n]，[h]，[m]，[y]，[r]，[w])，濁音 ([g]，[z]，[d]，[b]，[p]) を順に発語し録音した (清音と濁音は母音を伴う 5 音)．コンデンサマイクとスロートマイクの波形 (図 11) を比較した結果，[s]，[r]，[h]，[m] を伴う音では，母音のみの音のスペクトルと比較して著しく音量が下回っていることが分かった (図 12)．それぞれ「両唇音」「そり舌音」「歯茎音」「後部歯茎音」に該当する．このことから，これらの子音を伴う音では，狭窄の形成による 咽喉周辺の振動が，母音単体の発話時よりも少なくなることがわかった．いわゆる「スロートマイクを経由して拾いにくい音」であるとみなせる．

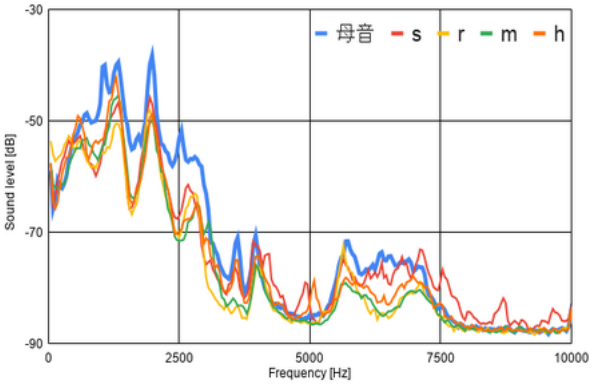

図 12 日本語発話における母音と子音 [s]，[r]，[m]，[h] を伴う音のスペクトル

4. 考察 – フィルタ学習用データセットの構築に向けて

4.1 フレーズの作成及び選択

本研究の目的は，スロートマイクの音声を聞き取りやすくするというものであり，そのために同時録音したコンデンサマイクの音声を併用して機械学習によるフィルタ生成を行うことを想定している．ここで，学習用のデータセットを作成する際には，今回見出したスロートマイク特有のノイズデータとして，「嚥下音」と「首元における摩擦音」のみのデータをラベル付きで用意することが望まれる．そのうえで，スロートマイクの弱点を考慮した発音を含む文章を選択する必要がある．

この文章選択においては，「ATR technicalreport 研究用 ATR 日本語音声データベースの作成（別冊 1 連続音声テキスト)」の A-D ラベル 200 文を用いた．この 200 文を，まず，Python 上で pykakasi モジュールを用いてローマ字表記へと変換した．そのうえで，「つ」の発音表記を [t][u] などの変換条件を入れた後に，[h]，[m]，[s]，[k] 音の総数を算出し，19 個以上の 4 子音を含む 10 文を抜き出した．結果を表 3 へと示す．

スロートマイクの音質向上のための機械学習用データセットの作成のためには、表 3 に記載した 10 文を被験者に読み上げてもらった音声を記録することにより，スロートマイクの弱点である周波数を網羅したデータセットが作成できることが期待される．

表 3 研究用 ATR 日本語音声データベースの作成（別冊 1 連続音声テキスト）

ラベル番号	文
A28	「母は脳血栓の後遺症で老人性痴呆症になり一年前から入院中です」
A38	「自動車や 精密機械などで 技術系の採用を抑える ところが目立ち 売り手市場の 技術系にもかげりが見え始めた」
A42	「首相 自ら 国民 一人一人 百ドル 舶来品を 買うように すすめた」
B29	「彼の 数学の 授業は 抜群に面白く試験前には月給外補習授業をする程 熱心である」
B36	「これまで 少年野球 ママさんバレーなど 地域スポーツを 支え 市民に密着してきたのは 無数の ボランティアだった」
B41	「事故の 直接原因となった 圧力隔壁の ずさんな 修理 そのずさんさを見落としたチェック システムなどがそうだ」
C01	「やがて 証拠の 書類や 物品が 押収され 諸君は 取調べのため，国税局へ連行される」
C03	「文芸編集者としては 作家たちに最も信頼されている名編集者だ」
C36	「新聞 週刊誌 雑誌にかぎらずほとんどすべての 取材記者は自分の予定原稿を 持って やってくるのだ」
D40	「おれは コツコツと シングルを 狙うよと言うがもちろん本音ではなく 持ち前の パワーを 発揮し始めてきた」

4.2 データ数量・手法

最後に，データセットの設計において必要となるデータ量と取得手法について検討する．本提案手法で目的とする音質向上のためのフィルタは，スロートマイクで録音を行えていない周波数成分の復元操作となり，近年研究が進んでいる機械学習を用いた声質変換の処理と本質的に等価と考えられる．

声質変換は，元話者の入力音声を目的話者の声質へと変換する操作である．先行研究として，多対一の声質変換に機械学習を用いた事例 [14] や，環境雑音下における入力音声に対して高精度の声質変換を行った研究 [15] などがある．これらの研究においては，教師データとして、日本声優統計学会が公開をしている音素バランス文 100 文の 5 セット分 [16] や、8 種類のアメリカ英語方言の話者 630 人のブロードバンド録音収録を行った TIMIT コーパスを音素認識モデルの作成に利用している．このうち，日本声優統計学会で公開している音声ファイルは，総長約 2 時間である．日本声優統計学会のコーパスは，3 名の女性声優が、それぞれ 3 パターンの声質で録音した音声を収録しており，女声のみが含まれている．

以上を基準に，我々の研究で必要となるデータセットの規模を推定する．表 3 の 10 文を読み上げる時間を 90 秒とし，加えて，ノイズである「嚥下音」と「首元における摩擦音」を含めて，ひとりあたり 120 秒程度の録音を行う．また，我々のデータセットでは，性別，年齢のバリエーションに加えて，スロートマイクの機種による特性の差違も含める必要がある．年代ごと (20 - 30 代，40 - 50 代，60 - 70 代) に男女 5 名ずつ，合計 30 名の被験者の発声を，3 機種のスロートマイクを使って録音すると，約 3 時間分のデータセットとなる．

これは，日本声優学会のデータセットと同等規模となる [16]．また，上記のデータセットは，約 900 文と概算しうる．ATR 日本語音声データベースセット B を用いた男性から女性の声質へと変換する先行研究 [17] では，503 文を 50 - 450 文を学習用，残り 53 文を評価に使っている．これらより，フィルタ学習用のデータセットとして十分な量を備えるものと期待される．

5. おわりに

本研究ではスロートマイクの音質向上に向けた学習用データセットの構築のための予備実験により，コンデンサマイクとの比較から日常生活においてスロートマイクを活用する上での課題について検討した．その結果，スロートマイクの音声として「両唇音」「そり舌音」「歯茎音」「後部歯茎音」などの特定の音が，「スロートマイクを経由して拾いにくい音」であり，スロートマイクの録音音声には含まれない周波数要素を復元する必要があることが明らかとなった．また「嚥下音」や「摩擦により生じるノイズ」について，適切なフィルタ処理が望まれる．

今後，入力をスロートマイク，出力をコンデンサマイクとしてフィルタを機械学習させることにより，効果的な音質向上が期待されるが，コンデンサマイクでの録音音声そのままでは，不明瞭なささやき声が復元されることになる．そのため，今回見出した音声学的な知見に基づいて，ノイズゲート，ハイパスフィルター，イコライザーなどを用いた音声処理を行うことにより，音質向上による「聞きやすさ」の向上とその評価を図りたい．

我々の公開するスロートマイクと高音質な処理済みコンデンサマイクの比較データセットの活用により，新型コロナウイルスの感染拡大阻止に寄与しうる各種の会話支援技術の発展を期待している．また，将来的に，発話の中での嚥下音が発せられる息継ぎなどのタイミングの予測や解析，それらを用いた研究開発を検討したい．

参考文献

[1] R. Zhang, Y. Li, A. L. Zhang, Y. Wang, and M. J. Molina. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences*, Vol. 117, No. 26, pp. 14857–14863, June 2020.

[2] D. Chu, E. Akl, S. Duda, K. Solo, S. Yaacoub, H. Schünemann, A. El-Harakeh, A. Bognanni,

T. Lotfi, M. Loeb, A. Hajizadeh, A. Bak, A. Izcovich, C. Cuello-Garcia, C. Chen, D. Harris, E. Borowiack, F. Chamseddine, F. Schünemann, and M. Reinap. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*, Vol. 395, pp. 1973–1987, June 2020.

[3] P. Anfinrud, V. Stadnytskyi, C. Bax, and A. Bax. Visualizing speech-generated oral fluid droplets with laser light scattering. *New England Journal of Medicine*, Vol. 382, No. 21, pp. 2061–2063, April 2020.

[4] 厚生労働省. 健康や医療相談の情報. https://www.mhlw.go.jp/stf/covid-19/kenkou-iryousoudan.html (参照 2021-01-02).

[5] 谷口優. "データに見る新型コロナウイルスが飲食店に及ぼす影響と支援について". https://www.tablecheck.com/ja/blog/covid-19-impact-to-japan-restaurants-blog1/ (参照 2021-01-02).

[6] 齊藤剛史. サイレント音声認識（silent speech recognition）. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, Vol. 30, No. 2, p. 110, 2018.

[7] M. Fukumoto. Silentvoice: Unnoticeable voice input by ingressive speech. In *UIST 2018 Conference Proceedings*, pp. 237–246. Association for Computing Machinery, October 2018.

[8] N. Kimura, M. Kono, and J. Rekimoto. Sottovoce: An ultrasound imaging-based silent speech interaction using deep neural networks. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–11. Association for Computing Machinery, May 2019.

[9] 大室仲, 岡本学, 齊藤翔一郎, 阪内澄宇, 江村暁. リアルな声による豊かなコミュニケーションを実現する音声端末技術. NTT 技術ジャーナル, Vol. 22, No. 12, pp. 15–19, December 2010.

[10] 佐藤成美, 山内さつき, 高林範子, 石井裕. 音声分析によるマスク着用時のコミュニケーション方法についての検討. 岡山県立大学保健福祉学部紀要, Vol. 21, pp. 45–55, August 2014.

[11] 小林悠一, 山田侑太朗, 西村雅史, 峰野博史, 飯田一朗. 嚥下音を用いた水分摂取量推定手法の研究. 情報処理学会論文誌, Vol. 57, No. 2, pp. 532–542, February 2016.

[12] International Phonetic Association. The international phonetic alphabet (revised to 2015), 2015. http://www.internationalphoneticassociation.org/content/ipa-chart (参照 2020-12-05).

[13] CosCom Language Service, Inc. ひらがな カタカナ 50音表. https://www.coscom.co.jp/hiragana-katakana/kanatable-j.html (参照 2020-11-07).

[14] 吉田天哉, 田村仁. 機械学習を用いた声質変換手法. 第 82 回全国大会講演論文集, 第 2020 巻, pp. 187–188, February 2020.

[15] 佐藤邦彦, 暦本純一. 多様な雑音に対して耐性のある声質変換システム. インタラクション 2018 論文集, pp. 115–124, February 2018.

[16] y_benjo and MagnesiumRibbon. Voice-actress corpus. http://voice-statistics.github.io/ (参照 2021-1-04).

[17] 廣芝和之, 能勢隆, 宮本颯, 伊藤彰則, 小田桐優理. 畳込みニューラルネットワークを用いた音響特徴量変換とスペクトログラム高精細化による声質変換. 情報処理学会 音声言語情報処理研究会 研究報告, No. 27, pp. 1–4, June 2018.