

SpeechJammer:聴覚遅延フィードバックを利用した発話阻害の応用システム

SpeechJammer:A System Utilizing Artificial Speech Disturbance with Delayed Auditory Feedback

栗原 一貴 塚田 浩二*

Summary. 本研究では、言葉を喋っている人に作用させて強制的に発話を阻害するシステム、「SpeechJammer」を開発する。一般に発話に対し、数百ミリ秒程度の遅延を加えて話者の聴覚に音声をフィードバックすると、話者は正常な発話が阻害されることが知られている。この現象は、肉体的苦痛を伴うことなく発話を阻害することができ、また発話をやめればただちにその認知的な影響が消失し、また話者のみに作用するためそれ以外の周囲の人たちには無害であるといった優れた特性を持っている。我々は指向性マイクと指向性スピーカーを組み合わせることで、外部の離れた場所から特定の話者の発話を阻害するシステムを試作した。これを応用し、会話のマナーとルールの制御、プレゼンテーショントレーニングなどに活用することを検討する。

1 はじめに

音声および文字を用いた他者との対話は、言語によるコミュニケーションの基本的な手段である。このうち、音声による対話は、身体のみで言語を表現でき、1対多の情報伝達が容易であるなどの優れた特徴があるため、人類が文字を発明して以降も絶えることなく用いられて今日に至っている。

しかし対話による紛争の平和的解決が重要視されている現代社会において、音声の持つ負の特徴によってその精神が脅かされたり、また健全な社会生活に害悪を成しているケースが後を絶たない。

我々はそのような音声の負の特徴のうち、不可避、占有、重畳無効化、調整困難の4つを以下のように定義し、これに注目する。

- 不可避：音声が一方向的に話し手から開始・継続でき、聞き手が制御できない。
- 占有：音声対話が通常、同時にただ一人だけの発話を許容する。
- 重畳無効化：いつときに二人以上の発話があった場合、全ての発話が無効化し理解困難になる。
- 調整困難：話し手が意図的に話速度などの音声プロパティを聞き手にふさわしく調整するのにスキルを要する。

これらの特徴により、代表的に以下の3種類の問題が指摘される。

(1) 議論における「言った者勝ち」問題

Copyright is held by the author(s).

* Kazutaka Kurihara, 産業技術総合研究所, Koji Tsukada, お茶の水女子大学 お茶大アカデミックプロダクション/科学技術振興機構 さきがけ



図 1. SpeechJammer システム概要

対話による紛争の解決には、公正な議論が不可欠である。音声による対話は不可避、占有、重畳無効化の特徴があるため、適切に発言権の交代ルールを規定し、それを遵守しなければならない。しかし、議論の内容よりも自身の存在を（マスコミ等を通じて）アピールすることを目的とする勢力が存在する場合などに、いたずらに自身の発言を長引かせたり、他者の発話の合間に割り込んだりといった行為が現れるようになる。また、他者の発話中に野次を飛ばすなどして、発話を無効化するなどの戦略が取られる。これは不可避、占有、および重畳無効化の濫用であり、結果的に「言った者勝ち」という風潮がまかり通っている。

(2) 会話が不適切である場における会話を止めるに
くい問題

公的な場の一部では過度の会話が不適切であると規定されていることがある。例えば図書館や電車内

での騒がしい会話はすべきではないとされている場合が多い。これは音声のもつ不可避な特徴から、聞くことの拒否が難しい以上、他者が不快に思うレベルの騒音は発生させないようにすべきだという社会的総意から生まれた規定であろう。

しかしこのような場において、既に発声している会話を止めることは難しい。なぜならば会話が不適切であることを会話者に説得し止めさせるためには、説得者も会話に参加せざるを得ない可能性が非常に高いからである。これは説得者自身も不適切行為に身を染めることであり、心理的負担が大きい。また、意を決して説得に乗り出したとしても、(1)のように公正な議論を行なう同意が得られなければ占有、重畳無効化の濫用により対話が無秩序化し、会話を止めるという目的からますます遠ざかる可能性を秘めている。

結果的に現状では積極的に会話禁止に乗り出すのではなく、ヘッドフォンを着用するなどして音声の不可避を断ち切り、会話と共存を図るなどの消極的解決手段に頼らざるを得ない様子がしばしば観察される。

(3) プレゼンテーションにスキルが必要な問題

現代社会においてプレゼンテーションは広く普及が進んでいる1対多のコミュニケーション手段である。ここにおいても音声は主要な情報伝達手段として用いられているが、その運用は普段行っている1対1もしくは1対少人数の音声コミュニケーションとは異なる。例えば通常の1対少人数の音声コミュニケーションでは、しばしば聞き手の理解が難しかった点をその場その場で問いただすことでインタラクティブに解消できるが、通常のプレゼンテーションではそのような機会は極めて少ない。従って話者自身が内容、および話速度などの音声プロパティについてある程度「最適なパフォーマンス」を想定、実施し、それを維持することが必要である。プレゼンテーションの初心者が行なうのは困難であり(調整困難)、習得には訓練が必要である。

栗原ら [2] はそのような様式の自習のためのシステムを開発したが、音声の問題を視覚で指摘するなど、指摘対象と指摘のモダリティが異なっており、強制力が弱かった。

我々は、外部からの働きかけにより話者の発話を制御する研究を行っている。その第一段階として本論文では、人間の聴覚特性の応用である聴覚遅延フィードバックという手法を用いて遠隔地から強制的に発話を阻害するシステムを構築する(図1)。本手法は肉体的苦痛を伴うことなく発話を阻害することができ、また発話をやめればただちにその認知的な影響が消失し、また話者のみに作用するためそれ以外の周囲の人たちには無害であるという優れた性質を持っている。発話者の外部から強制的な発話の阻害が可能になることで、例示した問題の根源であ

る音声の負の特徴が改善される可能性がある。すなわち、不可避性は緩和され、その結果適切に話者交替を行なうことで占有性を制御でき、かつ発話の重畳がなくなることで重畳無効化性も問題とならなくなる。またプレゼンテーショントレーニングシステムと組み合わせることで、調整困難性を改善できる。

本論文は、まず関連研究を列挙する。その後実装したシステムについて詳細を示す。そして今後の評価実験計画について検討する。

2 関連研究

2.1 聴覚遅延フィードバック

我々は発声時、単に発声という運動出力を行なうだけでなく、実際に耳から聞こえた発声音(聴覚フィードバック)を脳内で活用していると考えられている[9]。ここで、人工的に自己の発声音を聞かせるタイミングを遅延させると、脳内の処理系に何らかの作用が働き、以降の発声に支障をきたす。これが聴覚遅延フィードバック(DAF)による発話阻害である。DAFは吃音症(どもり)と関係が深い。健常者ではDAFにより人工的などもり(発話阻害)が誘発される形となるが[9]、逆に吃音症患者はDAFにより吃音が矯正されることが知られており[1]、医療用のDAF機器も市販されている[5]。

我々はDAFを活用し、話者の意思とは関係なく外部から話者の発話を阻害するデバイスを開発する。システム構成によっては通常のDAF用システムとは異なり、マイクとスピーカーが話者から離れた場所に設置される点が特徴的である。

2.2 議論支援

本研究はデバイスによる音声コミュニケーションの制御を扱ったものである。これは広義には議論の支援研究とも考えられる。議論の支援に関する関連研究として、議論内容を記録し再利用性を研究した長尾らによる[4]や、議論記録から有意義なコミュニケーションパターンをボトムアップに構築することを目指した中田らによる[7]などが挙げられる。

またテキストチャットにおいてコミュニケーションにあるルール・制約を適用したり、議論に関する何らかの情報を可視化することで議論を活性化させたり、集団意思決定などのある共通の目的を達成することを支援する研究も盛んである[3][6][8]。

本研究で開発するシステムは、音声による議論に「強制的な発話の妨害措置」という新たな強い制約を課すことが可能である。この制約を活用することにより、議論のコミュニケーションパターンに有意義な変化をもたらすことができ、これら議論に関する研究を支援・拡張することが展望される。

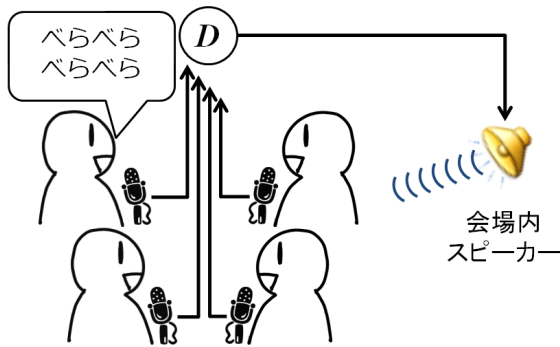


図 2. 会議室における発言権の制御

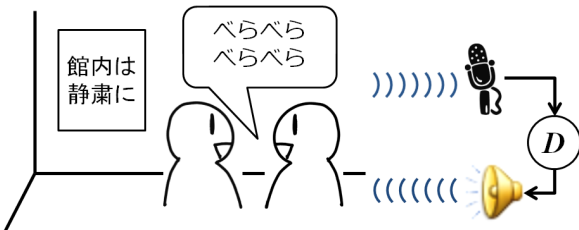


図 3. 携帯発話阻害銃

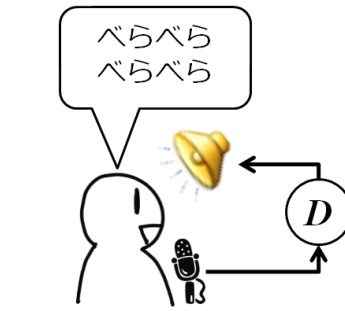


図 4. 発話者自身による発表練習支援

マイクとスピーカーを発話者近傍に設置できるヘッドセットを活用することができる。この場合、空気中の音波の伝達はほぼ無視できる(図4)。このセッティングのもとでプレゼン先生システム [2] と組み合わせることにより、話速度が閾値を超えて早口になっているとシステムが判断したとき、発話阻害機能を発動させる。

表 1. 発話阻害の使用が想定される局面のマイク及びスピーカー設置状況による分類

	話者近傍 スピーカー	遠隔 スピーカー
話者近傍 マイク	発話者自身による 発表練習支援	会議室における 発言権の制御
遠隔 マイク	(なし)	携帯発話阻害銃

3 SpeechJammer システムの実装

3.1 設計

我々は話者の発話を阻害するシステム「SpeechJammer」を設計する上で、その適用局面を検討する必要がある。なぜならば、発話の場にマイクとスピーカーがどのように置かれているかによって、ハードウェア設計に違いが生じ、また音波の伝達時間に違いにより、DAFのパラメータが変化するためである。我々の想定する本システムの適用局面は、

1. 会議室における発言権の制御
2. 携帯発話阻害銃
3. 発話者自身による発表練習支援

の3つであり、これらは1で挙げた3つの問題点にそれぞれ対応している。図2,3,4に、それぞれの局面におけるシステムの構成を示す。図中、実線は電気信号の伝達を、波紋は空気中の音波の伝達を表す。

表1は、これらをマイクとスピーカーの設置状況をもとに分類したものである。会議室における発言権の制御には、それぞれの発話者の近傍に設置されたマイクと、会場に備え付けられたスピーカーをインフラとして活用できる(図2)。この場合、空気中の音波の伝達はスピーカー・聴衆間の一方向のみである。携帯発話阻害銃は、聴衆の手許にマイクとスピーカーを備えたデバイスを用意し、生活上で遭遇する不適切な発話に対し発話阻害を行なう(図3)。この場合、空気中の音波の伝達は話者・聴衆間の往復分である。発話者自身による発表練習支援では、

本論文では、これらのうち最も複雑な構造をもつ携帯発話阻害銃タイプのSpeechJammerのプロトタイプを開発し、検証する。このデバイスさえあれば、他の2つの用途(発表練習支援、会議の発言権制御)の検証にも活用することができる汎用性を持っているからである。

3.2 遅延の計算

システムを実現するためには、話者とデバイス間の距離に応じた音声の伝達時間を考慮し音声遅延を発生させる必要がある。ここで図5のように話者とデバイスとの距離を $x[m]$ とし、気温を $t[^\circ C]$ 、DAFで発話を阻害するために必要な遅延を $D_{daf}[sec]$ とすると、空気中(1[atm])の音の伝達速度を加味した上でシステムに生じさせる遅延 $D[sec]$ は

$$D = D_{daf} - 2x/v \quad (1)$$

で表せる。ただし音速 $v[m/sec]$ は

$$v = 331.5 + 0.61t \quad (2)$$

である。

上式によると、たとえば $D_{daf} = 0.2[sec]$ のように話者に与える遅延を固定した場合、気温 $20[^\circ C]$

において $D \geq 0$ となるのは $x \leq 34.37[m]$ である。これは気温 $20[^\circ\text{C}]$ において本手法による最大射程が約 $34[m]$ であることを示しており、一般的な室内環境であれば実用上十分であると言える。

この最大射程内の任意の標的に対し常に発話障害を引き起こすためには、標的とデバイスの間の距離 x が固定され事前に知られている場合は固定値 D を計算すればよく、 x が未知、もしくは変動する場合は距離センサにより逐次的に x を測定し D を求めればよい。

一方、システムを簡素化し固定の D により発話障害を発生させる場合の考え方は以下の通りである。DAF による発話障害が $0.1 \leq D_{daf}$ の音声遅延で発生するとすると、(1) 式より $D = 0.1[\text{sec}]$ とすれば、 $0 \leq x$ となり、固定の D によって任意の距離での発話障害が可能となる。ただし、 D_{daf} による発話障害の程度の違いや D_{daf} の有効な上限値は我々の調査した範囲では明らかになっていないため、実験により与える必要がある。また実際はマイクとスピーカーの性能により x の上限（最大射程）は規定される。

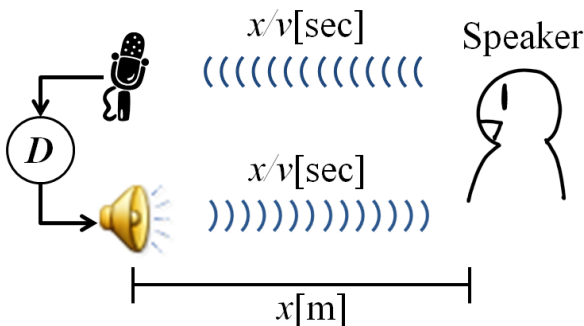


図 5. 必要な遅れ D の計算

3.3 プロトタイプ

我々は、SpeechJammer の 2 種類のプロトタイプを作成した。

プロトタイプ 1 は SpeechJammer の基礎的な機能を備えた最初のプロトタイプである (図 6)。指向性マイク (Sony ECM-Z60) と指向性スピーカー (トライステートパラメトリック・スピーカーキット) を一体化し、アクリル製のケースに取り付けた。ケースには三脚用のネジを切っており、一般的な三脚に固定して角度などを調整できる。指向性マイクの出力は宿主 PC のマイク入力に、指向性スピーカーの外部入力に接続される。宿主 PC 上では、Pure Data (以下、PD) で記述された音響処理プログラムが動作しており、指向性マイクの出力に任意の遅延をかけた上で、指向性スピーカーに出力することができる。なお、AC

ラインに起因するノイズを防ぐため、電源として単三電池 $\times 8$ 本を組み込んでいる。

上記プロトタイプ 1 での基礎的な動作を確認した上で、スタンドアロン型のプロトタイプ 2 を実装した。指向性マイク (Sony ECM-CZ10) と指向性スピーカー (トライステートパラメトリック・スピーカーキット) を中心に、レーザーポインタ、距離センサ、各種スイッチ、及びメイン基板を一体化して、独自に製作したアクリル製ケースに組み込んだ (図 7, 8, 9)。レーザーポインタは、SpeechJammer の大まかな照準として利用し、背面のスイッチで任意にオン/オフできる。距離センサは、ターゲットとの距離を計測し、遅延時間を調整するために利用する。

メイン基板上には、マイコン (Microchip PIC18F-452)、デジタル・ディレイ IC (Rohm BU9262AFS)、プリアンプ、メインアンプ、及び周辺回路が配置されている。

指向性マイクの入力は、プリアンプを通して、メイン基板上のデジタル・ディレイ IC に接続される。デジタル・ディレイ IC は、シリアルインタフェースを介して、遅延時間を $9.2\text{msec} \sim 192\text{msec}$ まで変更できる。遅延時間は、背面のロータリースイッチを用いて 8 段階で選択する¹。デジタル・ディレイ IC の出力は、メインアンプを通して、指向性スピーカーに接続される。プリアンプ/メインアンプは通常はミュートされており、トリガスイッチが押された場合のみにオーディオ入力/出力が行われる仕様としている。また、入力/出力音量はケース上面/背面のつまみで個別に制御できる。マイコンは、デジタル・ディレイ IC、トリガスイッチ、ロータリースイッチ、距離センサなどを制御するために利用する。

このように、ユーザは SpeechJammer をターゲットに向け、拳銃のようにトリガスイッチを引くだけで、容易に発話障害機能を発動することができる。

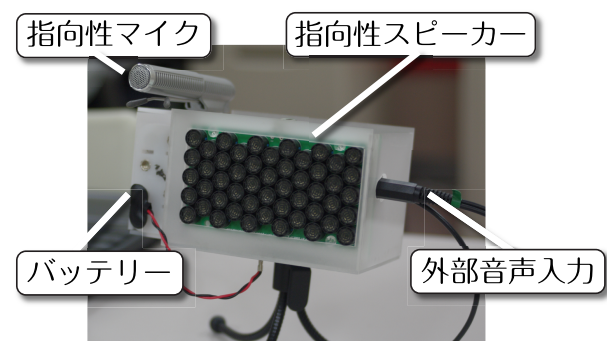


図 6. SpeechJammer プロトタイプ 1 の概要

¹ ロータリースイッチの設定次第で、距離センサを用いてターゲットとの距離を計測し、遅延時間を自動的に調整したり、任意の範囲で周期的に遅延時間を変更する仕様も準備している。

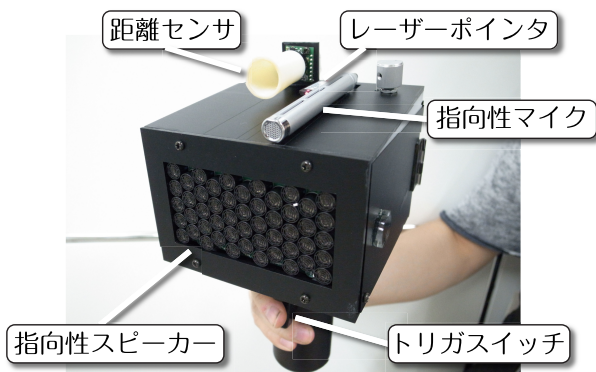


図 7. SpeechJammer プロトタイプ 2 の概要, 前面

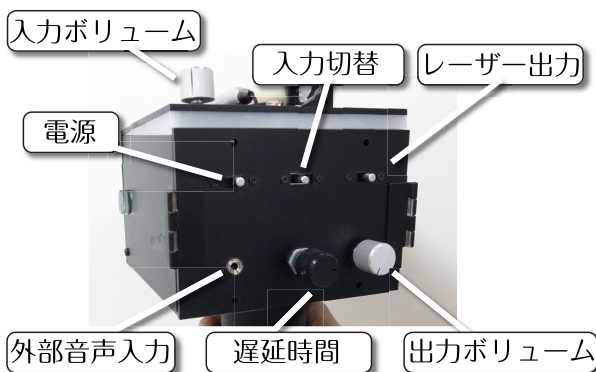


図 8. SpeechJammer プロトタイプ 2 の概要, 背面

4 評価実験計画

我々は実装した SpeechJammer システムのプロトタイプ 1 により, ホスト PC 上で各種パラメータを変化させながら発話阻害効果について調べる小規模なインフォーマル実験 (被験者数 5 人) を行った. この結果を踏まえて, 今後フォーマルな評価実験を行なう際の計画および評価の観点を以下に示す.

4.1 遅れ時間 D と発話阻害効果の関係

文献 [9] によると, 吃音症を DAF で改善する研究では, 遅れ D として $0.004[\text{sec}]$ から $0.195[\text{sec}]$ までを扱った事例があり, その区間では D が大きければ大きいほど吃音が改善される傾向があった. 我々が行ったインフォーマルな実験においても, $1[\text{sec}]$ を超える D で発話阻害効果が確認された事例が得られている. これらを参考に, SpeechJammer システムを用いて遅れ D を変化させながら発話阻害効果を調べ, D はどこに最適値があるのかを見出すことが必要である. D と発話阻害効果に有意な相関が見いだせれば, 単純に発話阻害を ON/OFF するだけでなく, 「どの程度」阻害するかをシステムが制御できるようになり, 我々の研究目標である「外部からの働きかけによる話者の発話の制御」の達成に大き

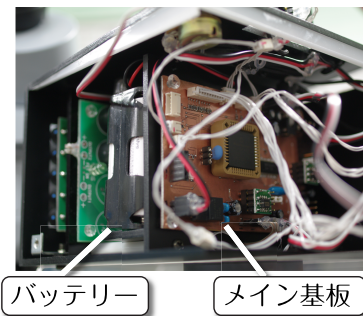


図 9. SpeechJammer プロトタイプ 2 の概要, 内部

く近づく. これを応用すれば, プレゼン先生システム [2] と組み合わせることで, 設定した話速度で正確に発表をする支援なども実現できるかもしれない.

4.2 フィードバックの種類と発話阻害効果の関係

我々が行ったインフォーマルな実験では, DAF の遅れ D を固定する設定に加えて, T を 0 から開始する時刻 $[\text{sec}]$ として $D = 0.15 + 0.05\sin(2\pi T)$ のように遅れ D を正弦波関数で変動させる設定も試行した. その結果, 固定の D に比べて D が変動する設定の方が発話阻害効果が大きいケースが観察された. 通常の固定 D によるフィードバックに加えて, このように様々な D を変動させるフィードバックに対する発話阻害効果を調べ, 最適なものを見出すことが必要である.

4.3 距離 x 及び音響系ゲインと発話阻害効果の関係

我々が行ったインフォーマルな実験では, 発話阻害の効果は話者の声の音量やマイクおよびスピーカのゲインによって変化する傾向にあった. これらのパラメータは, 話者が自分の声の自分の耳で聞く時の音声レベルと, DAF により聞こえる音声レベルの比に影響を与え, それが発話阻害効果に関わっているものと考えられる. 話者とシステムとの距離 x や各種音響系のゲインを変化させ, どのようなパラメータが発話阻害に最適かを調べる必要がある.

4.4 発話の種類と発話阻害効果の関係

我々が行ったインフォーマルな実験において, 発話の種類として「ニュース読み上げ」と「原稿のない自由会話」を扱ったところ, ニュース原稿を読み上げるタスクの方が, 原稿のない自由会話タスクよりも発話阻害効果が高い傾向が見出された. また, 自明なことだが「あー」などの言語的に意味の薄い単音からなる発話の場合には発話阻害効果が得られない. このような発話の種類によって発話阻害効果がどの程度影響を受けるのかを調べる必要がある.

5 まとめ

我々は音声によるコミュニケーションを円滑にするために、不可避、占有、重畳無効化、調整困難といった音声の負の特徴を緩和するシステム、SpeechJammerのプロトタイプを実装した。SpeechJammerは聴覚遅延フィードバックの仕組みを用いて、肉体的な苦痛を伴うことなく話者の発話を阻害することができる。今後は様々なパラメータを変化させ、パラメータと発話阻害効果の関係を調査し、応用方法を検討していく。

参考文献

- [1] Andrew Stuart et al. Effect of Monaural and Binaural Altered Auditory Feedback. *J. Acoust. Soc. Am.*, 101(6), June 1997.
- [2] K. Kurihara, M. Goto, J. Ogata, Y. Matsusaka, and T. Igarashi. Presentation Sense: A Presentation Training System using Speech and Image Processing. In *Proc. of ICMI'07*, pp. 358–365, 2007.
- [3] G. Leshed, D. Cosley, J. T. Hancock, and G. Gay. Visualizing language use in team conversations: designing through theory, experiments, and iterations. In *CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pp. 4567–4582, New York, NY, USA, 2010. ACM.
- [4] Nagao, K., Kaji, K., Yamamoto, D., and Tomobe, H. Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities. In *Proc. of PCM'04 Part 1*, pp. 522–531, 2004.
- [5] SmallTalk. <http://www.casafuturetech.com/Catalog/smalltalk.shtml>.
- [6] 西田 健志, 山崎 敬一. 集団意思決定への参加感を高めるトーナメント型議論. グループウェアとネットワークサービスワークショップ2009論文集, 情報処理学会シンポジウムシリーズ, 2009(8):123–128, 2009.
- [7] 中田 篤志, 角 康之, 西田 豊明. 非言語情報の出現パターンによる会話状況の特徴抽出. インタラクシオン'10 予稿集, 2010.
- [8] 村田 和義, 川口 修, 倉本 到, 渋谷 雄, 辻野 嘉宏. 遅延時間を用いたチャット参加者の発言制御. ヒューマンインタフェース学会論文誌, 6(4):411–422, 2004-11-25.
- [9] 庄司 壽一, 伊福部 達. 吃音軽減のための音声フィードバック装置について. <http://mito.cool.ne.jp/stutstudent/ronbun.html>.

未来ビジョン

我々は、本研究を通じて世界平和の実現を希求している。世界平和に最も必要なのは、人と人との対話である。対話から諸問題の解決方法が生み出され、またあらゆる相互理解が構築される。しかし現状では、対話はしばしば万人に対等ではなく、「声の大きいものが勝つ」といった先時代的な戦略が政治や外交のような世界のリーダーたちの対話においても未だに用いられていることは悲しむべきことである。

そこで、本論文で論じた SpeechJammer が拳銃と同程度に小型化・高機能化され、全ての人がこれを所持する近未来を想像するとどうだろう。発話を強制的に阻害するというのがある種の暴力であるから、これは紛れもなく一種の兵器である。しかし従来の兵器とは異なり、「だれでも使用可能（注1）で、対話の可能性を常に留保する（注2）」兵器である点は特記すべきである。対峙する侍が互いに刀に手をかけた時、または対峙するガンマンが互いにホルスターに手をかけた時、片方でもその力を行使してしまえば、その後には対話の余地のない殺戮が待っているのみである。しかし

SpeechJammer は互いに引き金を引いたとしても、対話の再開に向けて頭を冷やすために十分な沈黙が我々を暖かく待ってくれる。

兵器を互いに所持することで抑止力を働かせ平和を構築する時代は過去のものである。多様化と対話の時代にあつて我々人類は「一方や双方が力を行使しても平和を導ける能力」を兵器に求めていくべきなのではないだろうか。

（注1）非殺傷かつ対話の余地を残す兵器という意味で、ジャーナリズム、特にビデオカメラによる映像の放送は近代社会の偉大な発明のひとつである。最近では Youtube, Ustream やニコニコ動画/生放送といった、個人が映像発信を簡単に行える仕組みも整いつつある。しかしカメラによる個人放送のもつ抑止力のみで、「今日の前で起こっている声の暴力」を防ぐことは難しい。やはり我々は、直接的に、そしてなるべく無害な方法でこれを止めなければならない。

（注2）もちろん限界もある。講演など的一方通行会話は一方的に邪魔できてしまう。