

オンラインビデオ会議の状況に応じた 背景音の自動生成と共有

神場知成^{†1}

概要：オンラインビデオ会議の利用が急増しているが、実際の利用方法はさまざまである。形式的な会議もあればプライベートな飲み会もある。映像をオンにするルールのものであれば、任意で一部の人だけがオンになっている場合もある。音声も、全員がオンの場合もあれば、話をする人が発言のときだけオンにする場合もある。発言者がしゃべり続けても、他の人の反応がまったくわからない場合も多い。このようなオンライン会議では、「話をする際に緊張する」「会話が途切れると沈黙が耐えられない」などという意見も聞くようになった。一方、従来から、対面の会議室やカフェ等での背景音（ノイズやBGM）が、会議参加者の心理面や会話に与える影響が検討されている。本研究ではオンライン会議の背景音に注目し、一般的に使われる会議システム内に流れる音量の変化をモニタリングしながらリアルタイムで適切な背景音を自動挿入することが可能なシステムの実装を行った。背景音の選択、再生の開始と終了、フェードイン・アウト、ボリューム等を自動制御することが可能なため、今後は本システムを用いて、会議の効率や心理的な側面にポジティブな影響を与える手法の検討を進める予定である。

1. はじめに

オンラインビデオ会議は以前から利用されていたが、コロナ禍で急増した。しかもここでは「会議」と呼んでいるが、単なる雑談や「オンライン飲み会」と言われるような娯楽のための利用も増えた。利用者はシステム操作に不慣れだったり利用環境の設備が整っていなかったりするため、さまざまな問題も発生した。技術的には問題なくオンラインビデオ会議ができて、オンラインコミュニケーションは対面の会議と大きく異なるため、感覚的に違和感を持つ人もいる。

だがオンラインビデオ会議には移動のための時間や費用を大きく削減するという明確な利点があり、すでにかかなりの数の人が利用に慣れたことを考慮すれば、今後広がっていくことは確実だろう。その際に、画像や音声の低遅延や高画質化、セキュリティ強化、画面共有の柔軟性などの機能強化や性能向上も重要だが、参加する人が落ちついて参加でき快適なコミュニケーションができるように、心理的な面を含むユーザエクスペリエンスの向上は非常に重要だろう。

この問題に対し本稿では、音響面の施策によるアプローチを試みる。具体的には、会議内の音響をモニタリングし、自動的に背景音を挿入することで利用者の心理的な快適性を向上するシステムの開発を目指す。なお、以下では、特にことわらずに「オンライン会議」と記載した場合は、Zoom、Google Meet、Teams、WebEx 等の汎用的なシステムを用いたオンラインビデオ会議を指し、利用法としては、娯楽的なミーティングまで含めるものとする。

2. オンライン会議の利点と欠点

オンライン会議の一般的な利点は明らかで、地理的に離れた場所にいる人同士が、開催場所を確保したり移動する時間をかけたりすることなくすぐに開催できるので、会議開催の自由度は大きく上がる。これによる費用の削減も、特に遠隔地同士であれば非常に大きい。

欠点について考えるためには、もう少し実際の運用方法を見る必要があるが、特に最近のコロナ禍での運用には、次のような特徴があるだろう。

- 従来は「対面で集まっている会議に、地理的な理由などで参加できない一部の人がビデオや音声で参加する」というハイブリッド型の運用が多かったのに対し、「参加者全員が一人ひとり、自分の PC からオンラインで参加する」という形態の完全なオンライン会議が増えた。
- そもそも対面で参加するという選択肢が提供されず、PC 等のシステムに不慣れな人も一人で会議に参加することになり、1つの会議あたりの平均的なオンライン参加者の数も増えた。
- さまざまな種類の会議で利用されるようになったので、たとえば議事進行役がいて発言者や発言時間を制御するような形式的な会議だけでなく、参加者が思いついたときに自由に話し、終了時間も厳密に決まっていないブレインストーミングなども増えた。さらに、業務上の会議だけでなく、友人同士のオンライン交流会なども行われるようになった。

これらの利用状況を考えたときに、たとえば次のように、ユーザエクスペリエンスの観点から、利点とも欠点とも判断しづらいいくつかの特徴が見えてくる。

^{†1} 東洋大学 情報連携学部

- PCに搭載された、あるいは手元に設置された Web カメラから本人の顔だけを比較的大きく映して参加することが多いので、身振りなどのボディランゲージは伝わりにくい。オンライン会議によっては、一部あるいはすべての人がビデオ映像をオフにして音声だけで参加することもあり、この傾向はさらに顕著になる。
- ノート PC から会議に参加することが多く、PC 内蔵のマイクを使うと周囲環境のノイズを拾うことが多いため、専用のマイクを使う人も増えた。結果として、本人の声はよく聞こえ、環境音は入りにくくなった。
- さらに、上記のように環境音が入りにくくなったとは言っても多地点で多人数がオンにしていると全体としてはノイズがある程度増えてしまうため、「発言していないときは音声をオフにする」という習慣も起きつつある。これにより、発言者の発言以外の音はまったく入ってこない、という状況も増えている。
- 映像だけでなく、サウンドも発言者以外はオフにしている状態では、ちょっとしたあいずち(「そうですね」「うん、うん」など)も聞こえなくなるため、発言している人は「なんのフィードバックもない沈黙の中で自分の発言を続ける」という状況になりやすい。
- しかも、議事進行にもとづく会議のようなものと異なり、参加者が自由にアイデア出しをするミーティングであったり、プライベートなオンライン飲み会などであれば、少しの間、沈黙が続くような状況も自然に発生し、その間は完全な無音の状態となる。

このような運用の結果として、参加者からは「相手の反応が見えづらいので話しづらい」「皆が黙ってしまったときに沈黙が耐えられない」などの声も聞かれるようになった。

従来、オンライン会議において、単なるコンピュータ画面の共有だけでは伝わらなかったり、通常の映像ではずれてしまう視線の動き・対話者の目線の一致やボディランゲージを伝える研究は、石井の ClearBoard [1]をはじめとして多々の研究がおこなわれているが、これらはある程度専用の環境や専用の会議システムを必要とするため、現在の「一般のノート PC と Web カメラだけで、さまざまな汎用のオンライン会議ソフトを用いる」と運用形態を保ったままでも、できる限り会議参加者の心理的な軽減を行う方法が必要となってきた。

この課題に対する第一段階として本研究では、オンライン会議内に流す背景音(環境音や BGM)だけで対処することを試みる。本稿ではその実験の準備段階として、汎用のオンライン会議システム内のサウンド状況をモニタリングして会議状況に応じて自動的に背景音を流すことが可能なシステムを構築したので、その実装方法について報告する。実際に流す背景音については、再生開始の条件(会議内の音のボリューム変化、周波数変化等)や再生方法(サウンド選択、ボリューム調整、自動的な楽曲生成など)をプロ

グラムにより柔軟に制御可能である。実際に、会議の効率や快適性を向上するための適切な運用方法については、このシステムを用いた実験により調査および検証をしていく予定である。

3. 背景音制御システムの概要

3.1 システム構成と動作の概要

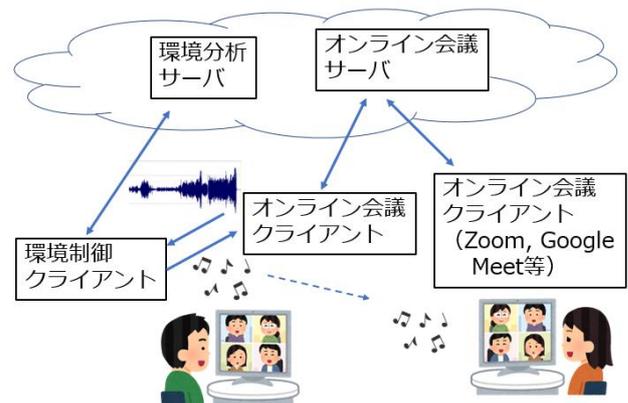


図 1 システムの概要

提案するシステムの概要を図1に示す。利用場面として想定するのは、汎用のオンライン会議クライアント (Zoom, Google Meet, Teams, WebEx 等) を用いて複数メンバーがオンライン会議をしている状況である。ここでは「会議」と呼ぶが、複数メンバーがそれぞれ話をするという状況を指しており、業務的な会議であっても私的な雑談であってもかまわない。

本システムは、Web サイト上の環境分析サーバと、ブラウザ上で動作する環境制御クライアントから成る。利用にあたっては、オンライン会議に参加している 1 台の PC 上のブラウザから、環境分析サーバが動作するサイトにアクセスする。サーバは、クラウド上に立ち上げてもローカル PC 内に立ち上げてもかまわない。そのサーバ上から読み込まれたページに記述されたスクリプトを通じて PC から取得した会議音がサーバに送信され、サーバ側からは、会議音の分析結果に基づいて、ブラウザ上の環境制御クライアントに対し背景音再生の指示が送られる。サーバ側とクライアント側の機能分担は柔軟に設定してあり、たとえば、何の背景音を再生するかをサーバ側で判断したのちに再生指示をクライアントに送ることも可能であるし、サーバ側から送付した条件(その時点の会議内がどの程度活性化しているか等のパラメータ)に基づいて、クライアント側で背景音を選択することも可能である。

環境制御クライアントが動作するブラウザ上の表示例を図2に示す(画面の一部のみ)。画面下部左側のグラフは、環境分析サーバから受信した結果に基づいて、クライ

アント側で生成・再生中のサウンドのボリュームであり、右側のグラフは、会議環境から取得中のサウンドの高速フーリエ変換 (FFT) 結果のグラフである。これらのグラフの下に、会議環境から取得しているサウンドのボリュームグラフなども表示されるが、ここでは省略している。

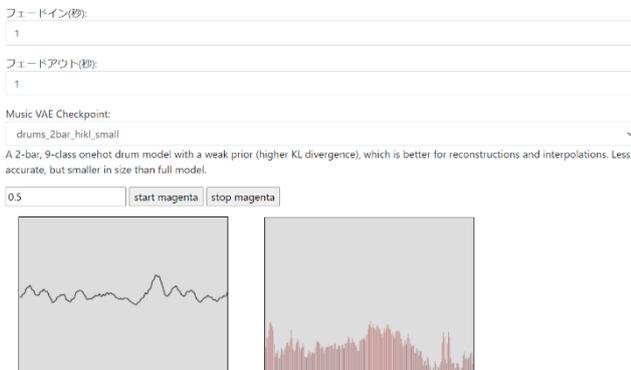


図 2 ブラウザ表示画面例

画面上部では、再生するサウンドを選択したり、サウンド再生の際のフェードイン、フェードアウトにかかる経過時間 (秒) 等のパラメータを指定することができ、それらの数値をサーバ側からコントロールすることもできる。

上記の、環境制御クライアントとして動作するブラウザは、会議参加者のうちの一人がオンライン会議参加用に利用している PC 上で立ち上げて、それとは別の PC で立ち上げてかまわない。同じ PC の場合は、環境制御クライアントから発生する背景音は、会議参加者が利用している会議ソフトに対し、その PC にいる参加者の声に重畳して送り返される。別の PC で立ち上げた場合には、その PC 上で会議ソフト (つまり Zoom, Google Meet 等のアプリ) を動作させうえて、仮想的な一人の会議参加者として会議にアクセスする。環境制御クライアントからの背景音を、その仮想的な会議参加者のサウンド入力に入れる。この場合、オンライン会議システムの側から見るとその参加者は、会議音をモニタして適宜、背景音を生成して流すだけの、いわゆる「ボット」として動作することになる。

会議参加者が会議に参加している PC 上でこの環境制御クライアントを動かす場合には、その PC からは会議参加者自身の音声と、環境制御クライアントが生成するサウンドの両方を重畳して会議クライアントに送り込むことになる。この際、「会議参加者本人には、会議クライアント内の音 (つまり他の参加者の声) と、環境制御クライアントが送り込むサウンドの両方が聞こえるが、本人の声はエコーバックしない」「他の会議参加者全員には、この本人の声と、環境制御クライアントが生成したサウンドの両方が聞こえる」ということを実現するために、Macintosh 上のビジュアル・ソフトウェアミキサーである LOOPBACK を用いて制御している[2]。LOOPBACK の画面例を図 3 に示す。Windows 上では類似機能を実現するために Voicemeeter

Banana が有効なようだが、現時点では動作未確認である[3]。

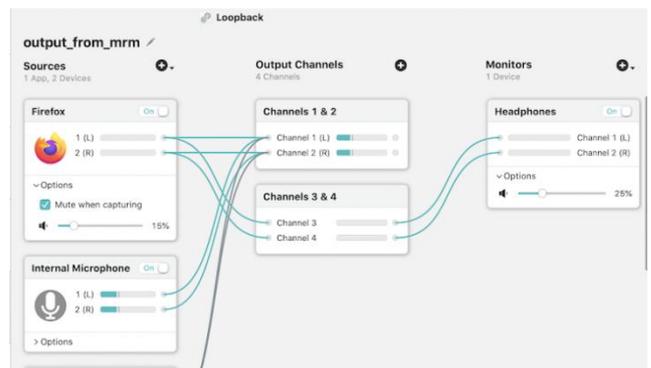


図 3 オンライン会議システムへのサウンド送出設定

3.2 動作の詳細

具体的には、次のような動作をする。図 4 に、クライアントとサーバ間の通信のシーケンスを示す。

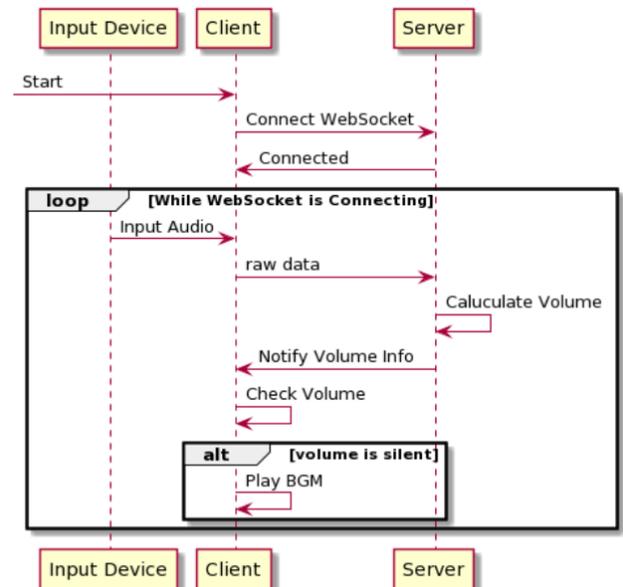


図 4 シーケンス図

- Web ブラウザから、クラウドまたはローカル上に立ち上げた Web サイトにアクセスすると (サイト A とする)、ブラウザにサイト A が読み込まれ、そのページに Web Audio API を用いて実装された JavaScript プログラムが記述されている。
- 上記の JavaScript プログラムは環境分析サーバに対して WebSocket の通信を開き、接続中は常に PC 上のサウンドデータを取得し、特定間隔 (自動調整されている) ごとにサーバに送信する。それを受信したサーバ上の Python プログラムは一定時間ごと (設定可能。デフォルト値は 5 秒) の平均音量を算出し、さらにその変化状況に応じていくつかまとめた上で、JSON

形式でクライアントに送る。ただしここで、サーバ側で何を計算してクライアントに送信するかは今後、要検討である。たとえば「音量から判断した会議の活性化状況を示す数値」「会議内で続いている沈黙時間」「会議内で流れている声の高さの平均値」などさまざまなものが考えられる。

- クライアント側では、上記データを受け取り、会議内のサウンド変化状況が、設定したルールに当てはまると背景音を再生する。ルールは Web ページ内の JavaScript で記述しているため柔軟に設定可能だが、現時点では、静音状態が何秒間続いたかだけを見て、生成した背景音を会議内に送出している。送出をし始めるまでの秒数は Web ページ上で設定可能である。
- 再生する背景音は、ローカル PC 内にあらかじめ蓄積したものを利用することも、ネット上のストリーミングサイト等を指定することも可能。現時点では、あらかじめ蓄積するサウンドは、フリーサンプルの環境音（オフィスのざわめき、波の音、山の鳥の声）を選択可能とした他、楽曲生成のライブラリである Magenta [4]による自動生成を指示することも可能である。
- Magenta は Google が中心となって開発を開始した、深層学習による音楽等の生成用ライブラリであるが、Python モジュールを用いて学習をさせることができ、JavaScript モジュールの Magenta.js を用いることで、事前に学習が済んだモデルに対し、適宜パラメータを指定しながら音楽生成をすることができる。ここでは、モデルとし Magenta にあらかじめ用意されている MusicVAE を用い、チェックポイントを利用者が設定可能とした。Magenta では、音のばらつき等のパラメータを指定して、その条件下で楽曲の自動生成をすることができるが、今後、会議状況のさまざまな視覚的、聴覚的パラメータをそのパラメータに反映させたいので楽曲生成をすることも想定している。

4. 考察と今後の課題

4.1 会議等におけるサウンド挿入について

本稿では、オンライン会議へのサウンド挿入について技術的な方法を中心に述べているが、対面の会議等へのサウンド挿入の効果については、これまでさまざまな角度からの議論がある。

たとえば、藤井等は単純な精神活動（1桁の加算演算、講演の音声聴取）に雑音が及ぼす影響の実験を通じ、有意雑音が無意味雑音よりもうるさく感じられることなどを述べている[5]。また、堀中等は、一般には飲食店においてBGMを流すことは快適な環境を保つ効果があると思われるものの、学生食堂での実験ではむしろネガティブな影響が生じた実験について述べている[6]。これらはいずれ

も、雑音やBGMがマイナスの影響を持つ場合があり、流す音響や音量、その他には慎重な検討を要することを示す。

一方、辻村他は、対面会議で知識創造活動を行う状況を想定して、環境音として「会話の内容は理解できない複数の音声による会話雑音」を用いた実験を行い、「会議のしやすさ」への影響を調べ、「知識創造活動では必ずしも静かな環境が望ましいとは言えず、静かな印象が損なわれず（50dBより低い）、活気が感じられやすい（45dB以上）環境が「会議しやすい」という印象につながる事が示唆された」と述べている[7]。

八木は、会話が騒音であることを前提としたうえで、それを和らげるための背景音を入れるサウンドマスキング（実験では、42.5dB）に一定の効果があることを、物理的な仕切りの効果と合わせて論じている[8]。

カフェにおける実験についても、前述の堀中等によるものとは異なる視点のものもあり、小林等は「アップテンポのBGMによって、より視線を合わせ、会話量が増し、会話音量が大きくなるという傾向に合った」等の報告をしている[9]。

さらに、会議とは異なるが中川等は、演習形式の授業にBGM導入をすることで学習の持続効果が見られた等の報告をしている[10]。

また大野等は、「会議の背景音としてサウンドを流す」というものとは異なり、会話の際に個人が、自分自身のみが聴取可能な音楽を聴取することで各人の緊張が緩和され、会話が円滑に行われるようになるという手法の提案をしている[11]。

これらは基本的に対面の場や会議を想定したものであるが、その知見はオンライン会議においても参考になるであろう。

4.2 サウンド生成条件の判定

上記に示したような従来の研究では、背景音はノイズであれBGMであれ、人手で制御することを想定しているが、本稿での検討は自動制御を想定している。サウンドのオン・オフあるいはボリューム制御、またはサウンド選択（ノイズ、BGM、あるいは前述のような、その場での自動作曲）などをすることになるが、制御のためのインプットとして現時点で想定しているのは会議内に流れている音のボリュームであり、同時に周波数解析（FFT）も行っている。会議内に挿入するサウンドはサーバ側のPythonプログラム、またはクライアント側のJavaScriptで制御するので処理の自由度は高い。

現状では、会議参加者が沈黙した状態が10秒、20秒、30秒など経過した時点で背景音再生をする簡易な実験を行った段階で、沈黙の判定は、手動で設定した適当な数値を下回っているかどうかということだけで行っている。しかし実際には、オンライン会議における「沈黙」の自動判

定はさほど容易ではない。たとえば参加者が、自分が話していないときはずっとマイクをオフにしているような会議と、何人かがオンにしているような会議とでは定常的に流れている音量も異なる。沈黙の判定を自動化するためには、会議によって実際に沈黙が続いているときにどのような音が流れているかを測定し、その会議における沈黙の状態を判定する必要があるだろう。

4.3 サウンドと再生音量の選択

本システムで現在までに試したのは「オフィスのざわめき、波の音、山の鳥の声」の3つであり、再生するときの音量はあらかじめ適当な数値に設定している。前節で「沈黙が一定秒数以上続いたときに再生をする」と述べたが、実際には「話が続いているときは邪魔にならないように音量を下げ、沈黙が続いているときは音量を上げる」等の工夫が必要になるであろう。また、本来これらは「あたかもオフィスにいるように感じる」あるいは「海辺で皆が話をしている」というような感覚を得ることを目的としているが、現時点では、利用者は単に、そのような BGM が鳴っていると感じる程度である。たとえば「オフィスのざわめき」の背景音を利用して、参加者たちが、あたかも対面のオフィスで話をしているような感覚をつくりだすのであれば、単に適切なサウンドをあらかじめ用意するだけでなく、3次元音響など新たな技術的な工夫が必要になると考えている。

4.4 フェードイン・フェードアウト

沈黙（会議内のボリュームが一定値以下）の状態が一定秒数（10～30秒）続いたときにサウンドを再生する簡易な実験を行ったところ、単に沈黙の判定に用いる秒数だけでなく、サウンド再生の際に用いるフェードイン・フェードアウトにかける秒数が印象に大きく関わることも確認している。特に、フェードインやフェードアウトをまったくせずに背景音が始まったり終了したりすることは非常に不自然な印象を受ける。また、そこにかける秒数は、特にフェードインの場合、ごく短い秒数（1～3秒程度）ではなく、やや長め（7～10秒程度）の方が適切というのが簡易実験による印象である。会議参加者の側からすると、多少沈黙が続いたときに、いきなり背景音が入るのは違和感があり、いつの間にか自然にサウンドが鳴り始める程度が好ましい。これについては、沈黙と判定する時間長などとあわせ、今後実験によって評価をしていく予定である。

5. おわりに

オンライン会議において、沈黙が参加者に与える不安の解消や、会議内の雰囲気向上のための BGM の再現などを想定し、汎用的なオンライン会議システムにおいて、

会議内の音の状況に基づいて自動的に背景音を挿入することが可能なシステムについて述べた。本システムは、いわば「会議内の音量変化を聴きながら適切な背景音を挿入するボット」であり、さまざまな会議システムで利用可能な実装方法をとっている。今後は、本システムを用いてオンライン会議を運用することにより、実際に会議参加者がリラックスして会議に参加できたり、会話がスムーズに行われたりするような背景音挿入について検討していく予定である。

謝辞 本研究は、東洋大学井上円了記念研究助成により助成を受けたものです。また、本システムのプログラム開発においては、塩田健氏の協力を得ました。同氏に感謝いたします。

参考文献

- [1] H. Ishii and M. Kobayashi: ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact, CHI'92, pp. 525-522 (1992).
- [2] Rogue Amoeba: LOOPBACK (<https://rogueamoeba.com/loopback/>)
- [3] VB-AUDIO Software: Voicemeeter Banana (<https://vb-audio.com/Voicemeeter/banana.htm>)
- [4] Magenta (<https://magenta.tensorflow.org/>)
- [5] 藤井他：有意味・無意味外来雑音が単純精神作業者に及ぼす影響, 人間工学, Vol. 38, No. 1, pp.63-68, 2002.
- [6] 堀中他：飲食店内の BGM の効果, 日本心理学会大会発表論文集, 日本心理学会第 71 回大会, 2007.
- [7] 辻村他：複数人による知識創造活動を行う会議に及ぼす室内音環境の影響, 日本建築学会環境系論文集, Vol.80, No. 711, pp.397-405 (2015).
- [8] 八木：会話騒音のあるオフィスにおけるサウンドマスキングと仕切りの効果, 労働科学, Vol. 88, No. 1, pp.25-36, 2012.
- [9] 小林, 小口：光色と BGM の種類がカフェでの会話行動に与える影響, 日本建築学会環境系論文集, Vol.71, No. 599, pp.143-150, 2006.
- [10] 中川他：演習主体とする授業における BGM 導入による教育効果, 神戸高専研究紀要第 55 号, pp.1-6, 2017.
- [11] 大野他：自身のみ聴取可能な音楽を用いたコミュニケーション円滑化手法の提案, 情処学会研究報告 グループウェアとネットワークサービス (GN), Vol.2019-GN-106, No.22, pp. 1-7, 2019.