

テレ会議システムの音質比較方法について

関西大学 環境都市工学部 建築学科

建築環境工学第 I 研究室

建 18-0010 植村 亜美

指導教官 豊田 政弘 教授

目次

1	はじめに	2
1.1	研究背景	2
1.2	研究目的	2
2	研究方法	3
2.1	研究の進め方	3
2.2	テレ会議システムについて	4
2.3	収録について	9
3	周波数特性による評価	12
3.1	評価手順	12
3.2	結果	12
4	「単語了解度」と「聴き取りにくさ」による評価	18
4.1	「単語了解度」について	18
4.2	「聴き取りにくさ」について	19
4.3	評価手順	21
4.4	結果	21
5	考察	23
6	おわりに	24
	参考文献	25

1 はじめに

1.1 研究背景

新型コロナウイルス (covid-19) が世界中で広がり、その前後で生活は大きく変化した。中でも、直接的な人々の接触を避けるために、教育機関では授業、行政機関や企業などでは会議がオンラインで行われるなど、コミュニケーションのオンライン化が急速に進んだ。この変化に伴い、多くのテレ会議システムが提案され、様々な場面で利用されることとなった。今では授業や会議のためだけでなく、遠隔で合奏を楽しむためのシステムも提案されている。

もちろん、どのシステムを用いるのが好ましいかは、利用目的や条件によって変わるが、どのような利用状況においても音質の良し悪しはコミュニケーションの質に大きな影響を与えるものと思われる。

1.2 研究目的

本研究では、特に音声の伝達に着目し、円滑なテレコミュニケーションを実現するために必要な音質、具体的には、周波数特性の傾向を明らかにすることを試みる。これにより、今後提案されるであろう新しいシステムの事前評価が可能となるとともに、音質に関する設計指針を与えることにもなるであろう。

2 研究方法

2.1 研究の進め方

本研究の大まかな流れとしては、以下の通りである。

周波数特性により評価する。

↓

「単語了解度」と「聴き取りにくさ」により評価する。

↓

周波数特性と「単語了解度」と「聴き取りにくさ」の関係性を考察する。

2.2 テレ会議システムについて

本研究では以下の5つのテレ会議システムを評価の対象とする。可能な限り、それぞれのシステムが提供する本質的な音質を比較するために、各システムの設定では自動で音量を調節する機能をオフにして、収録する。

・Zoom

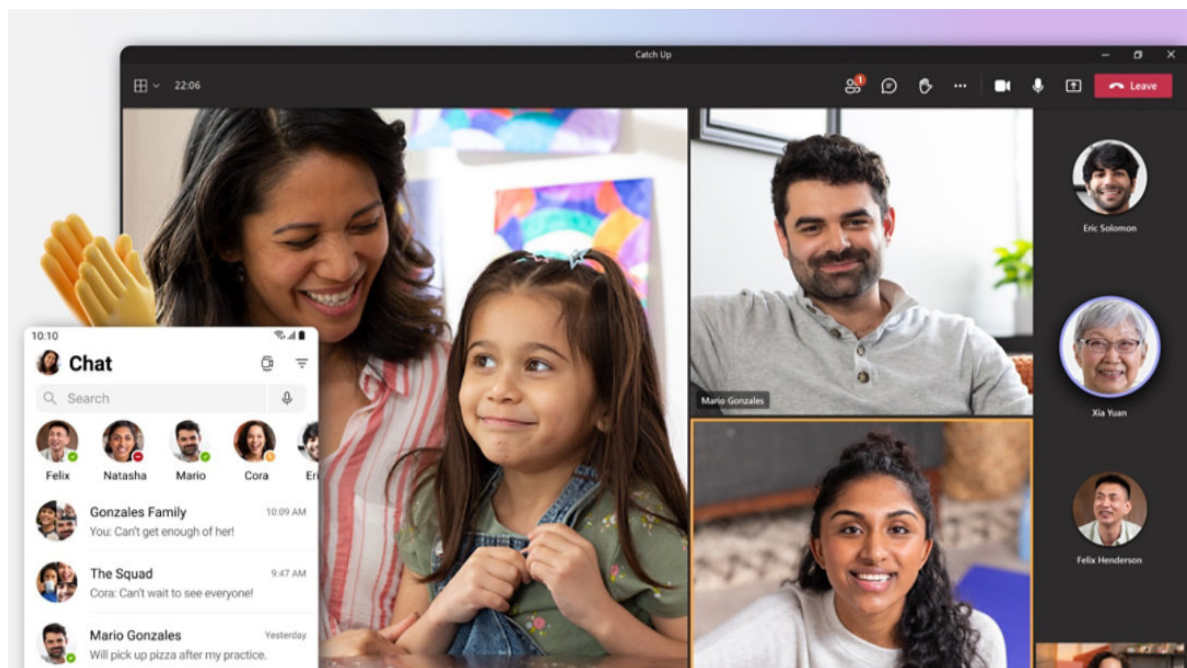
Zoom Video Communications が 2013 年 1 月から提供する、簡単で高品質な法人向けのビデオミーティングシステム。豊富なツールで対面と同等、もしくはそれ以上に充実したオンラインミーティングを可能にしている。



出典：<https://explore.zoom.us/ja/accessibility/>

- Teams

Microsoft が 2017 年 3 月から提供する、法人・教育機関・家庭のすべての方に向けたコラボレーションツール。会議だけでなく、チャット、通話、共同作業をすべて 1 カ所で完結することを可能にしている。



出典：<https://www.microsoft.com/ja-jp/microsoft-teams/group-chat-software>

- Skype

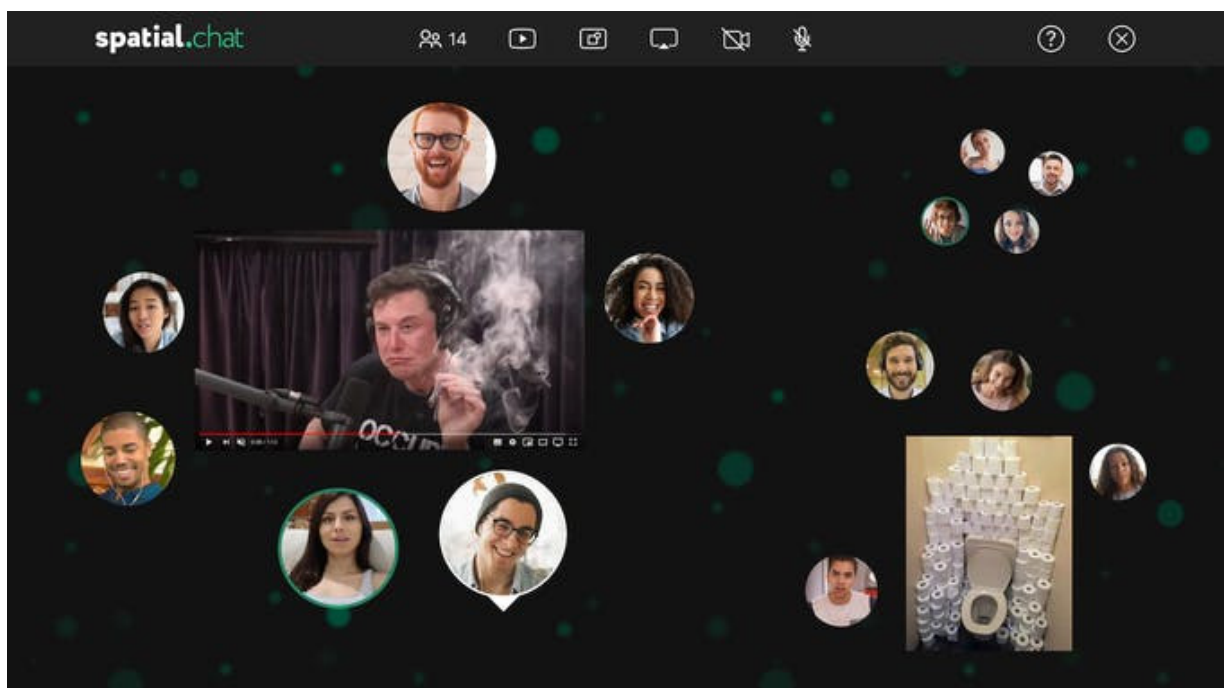
Microsoft が 2004 年 7 月から提供する、ビデオ通話サービス。他のシステムよりも長い歴史を持っており、テレ会議システムのパイオニア的存在であろう。お互いに ID を持っていれば、世界中どこにいても無料で何時間でも音声通話やビデオ通話、チャットをすることが可能である。



出典：<https://www.skype.com/ja/features/calling-and-instant-messaging/#calls>

- SpatialChat

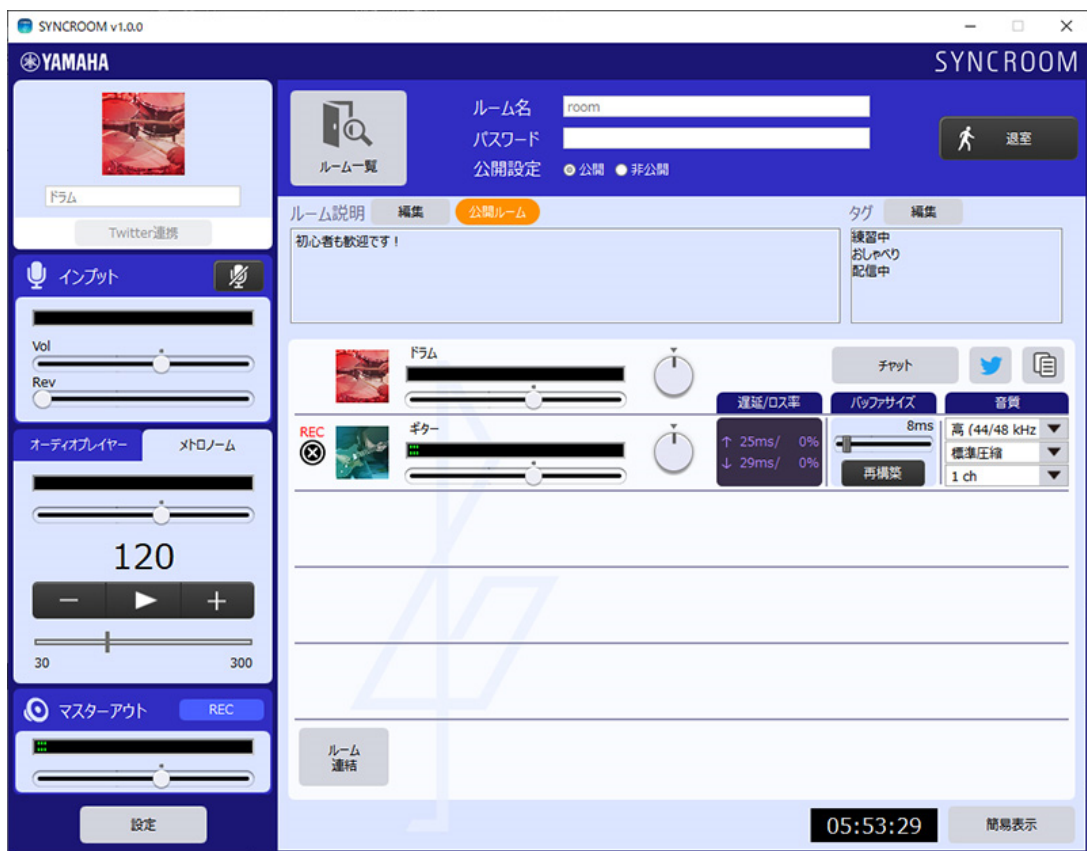
FunCorp Lab が 2020 年 4 月から提供する、リアルな距離感でのコミュニケーションを実現しているオンラインビデオチャットサービス。「カクテルパーティーをオンラインで再現する」という目的で開発に至っているため、実際のパーティー会場のように自由に動き回ることができ、話している人に近づくと声が聞こえてくる仕組みになっている。



出典：<https://www.producthunt.com/posts/spatialchat>

・ SYNCROOM

YAMAHA が 2020 年 6 月から提供する、離れていても音楽セッションをすることを目的とした演奏アプリケーション。YAMAHA が独自に開発した遠隔合奏技術を採用することで遅延の少ない音声データのやりとりができ、快適に遠隔合奏を楽しむことが可能である。



出典：<https://syncroom.yamaha.com/topics/2020/01.html>

2.3 収録について

収録は理想的な通信環境・オーディオ環境のもとで、2 台のパソコンとオーディオインターフェイスを用いて行う（図1）。使用した機器は以下の通りである。

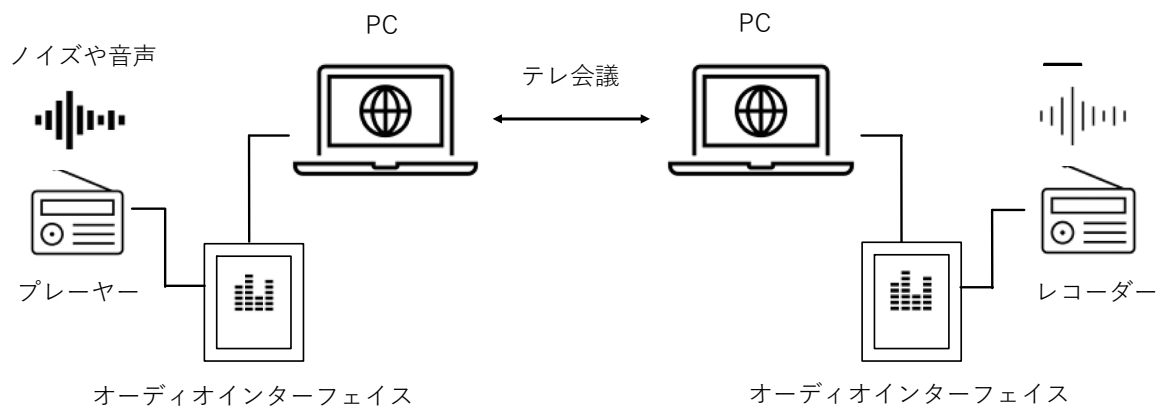


図1 収録方法

・ Apple MacBook Pro (型番：MR9U2J/A)



2台使用。ノートブック型パーソナルコンピュータ。シンプルなデザインで簡単に使いこなせるだけでなく、圧倒的にパワフルで作業効率を向上させることが可能である。エントリーモデルのMacBook Airよりも性能が高く、プロユーザー向けとして販売されている。

(出典：https://support.apple.com/kb/SP775?locale=ja_JP)

・ RME Babyface Pro (型番：BABYFACEPRO)



優れた音質の提供を可能にするオーディオインターフェイス。2011年に登場したBabyfaceは、プロフェッショナルの現場での要求に応えられる初のモバイル・オ

ーディオ・インターフェイスとして制作・リスニング環境に革命をもたらした。その初代Babyfaceの正統な後継機種として新たに登場したのが、Babyface Proである。オーディオ回路や、ドライバ開発だけでなく、メカニクすべてに高いクラフトマンシップを反映し、Babyfaceに比べ大幅にオーディオ・クォリティを向上させることに成功した。Babyface Proは、新たにデザインされたアナログ回路とデジタル回路により、全高調波歪み率、ダイナミックレンジ、S/N比等、すべてのスペックにRMEの一切の妥協がない技術が体现しており、録音だけでなく、オーディオ鑑賞、さらにはヘッドフォン・アンプとしても非常に高いパフォーマンスを提供している。

(出典：<https://synthax.jp/babyface-pro.html>)

- ・ SONY リニア PCM レコーダー (型番：PCM-M10)



レコーダーとして使用。小型・軽量コンパクトボディで、音楽録音のために機能性と操作性を徹底的に追求したポータブル。DAT (48 kHz 16 ビット) や音楽 CD (44.1 kHz 16 ビット) を超える 96 kHz 24 ビットのリニア PCM 録音で、原音を忠実に記録する。

(出典：<https://www.sony.jp/ic-recorder/products/PCM-M10/>)

- ・ SONY リニア PCM レコーダー (型番：PCM-D1)



プレーヤーとして使用。リアルな音の記録のために一切の妥協を排除した、4 GB フラッシュメモリー内蔵のポータブル。上記 PCM-M10 と同じく、96 kHz 24 ビットのリニア PCM 録音が可能である。

(出典：<https://www.sony.jp/ic-recorder/products/PCM-D1/>)

3 周波数特性による評価

3.1 評価手順

周波数特性の評価の手順は以下の通りである。

- ① 5つのシステムを用い、プレーヤーからホワイトノイズを入力し、システムを経由した後のレコーダーでの出力を収録する。(図1参照)
- ② 入力と出力をそれぞれ周波数分析する。
- ③ 各システムの周波数特性を比較する。

3.2 結果

分析した結果、周波数特性は以下のようになった。(図2～図6)収録前のホワイトノイズを青線、出力後のノイズをオレンジ線で示す。

まず、高音域の音圧レベルを比較する。完全にカット(30 dB以上)されている周波数帯があったのは Teams と Skype で、Teams・Skype 共に 8000 Hz 以上が完全にカットされている。少しカット(5 dB程度)されている周波数帯があったのは Zoom、Skype、SpatialChat で、Zoom は 8000 Hz 以上、Skype と SpatialChat は 2000 Hz から 8000 Hz が少しカットされている。

低音域で完全にカットされている周波数帯があったのは Teams だけで、125 Hz 以下をやや緩やかではあるが完全にカットしている。少しカットされている周波数帯があったのは Zoom と SpatialChat で、Zoom は 125 Hz 以下、SpatialChat は 62.5 Hz 以下を緩やかにカットしている。Skype と SYNCROOM は 62.5 Hz 以下の音圧レベルが少し上げられている。

最も収録前の周波数特性との差がなかったのは SYNCROOM で、62.5 Hz 以下に少し変化が見られたのみである。

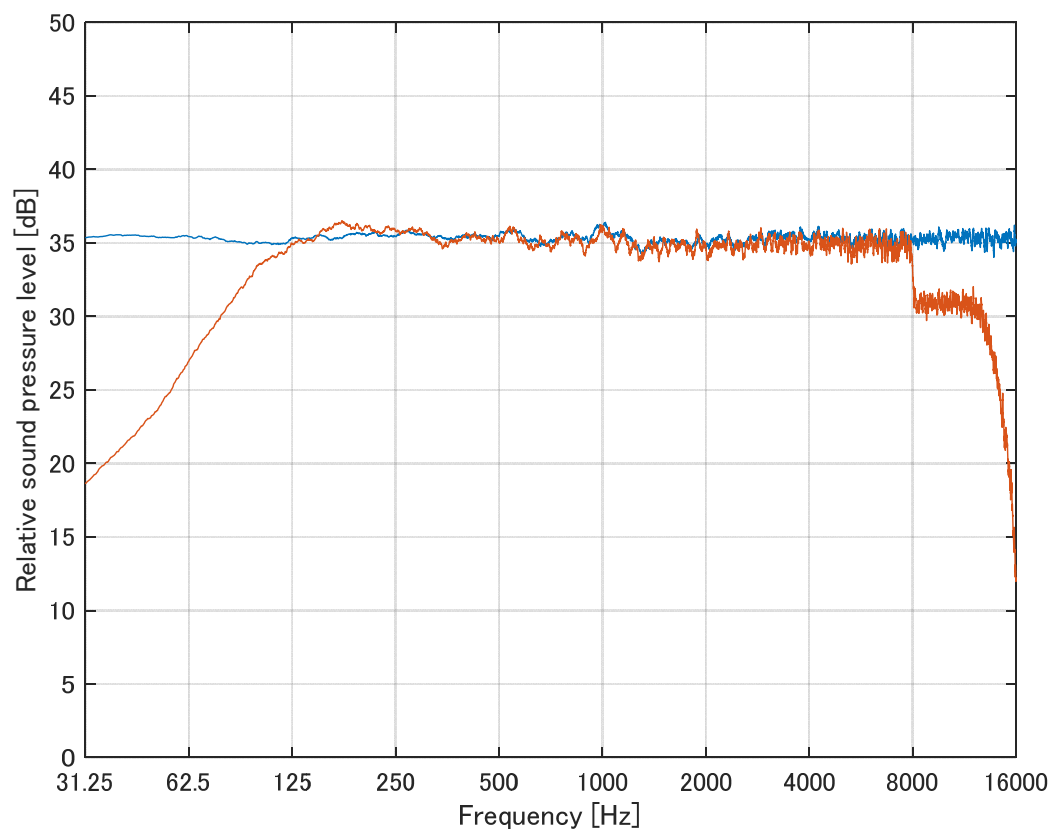


図2 Zoomの周波数特性

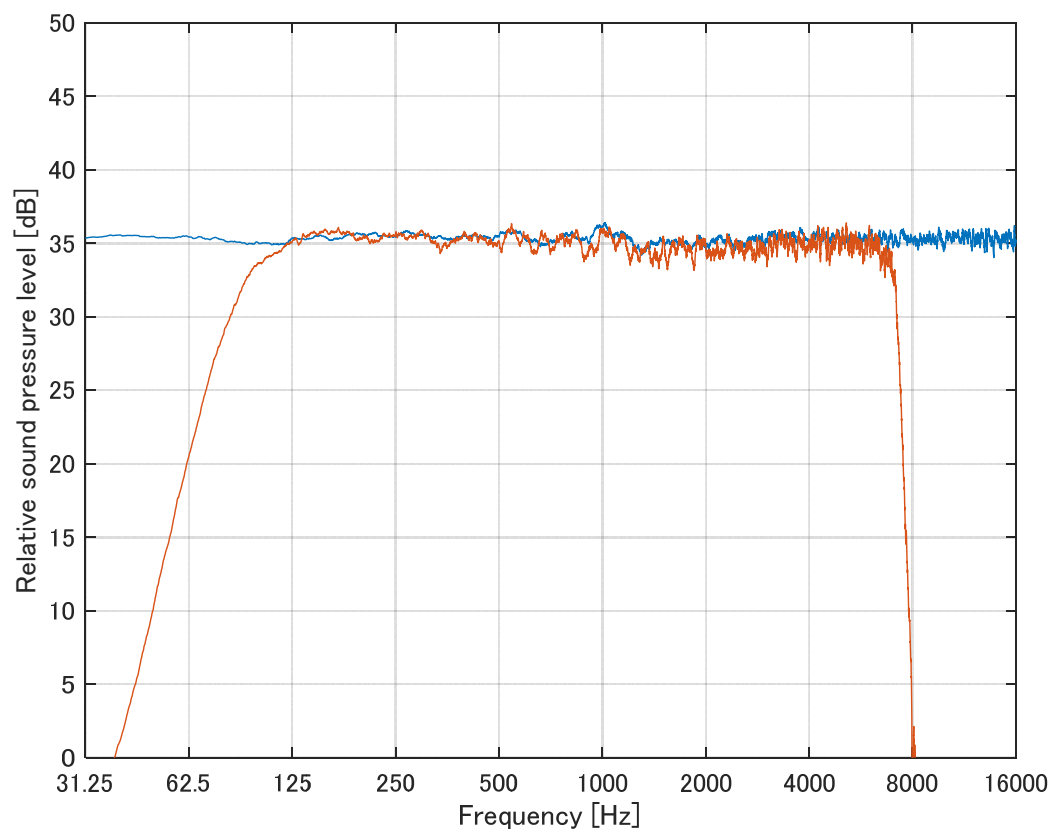


図3 Teamsの周波数特性

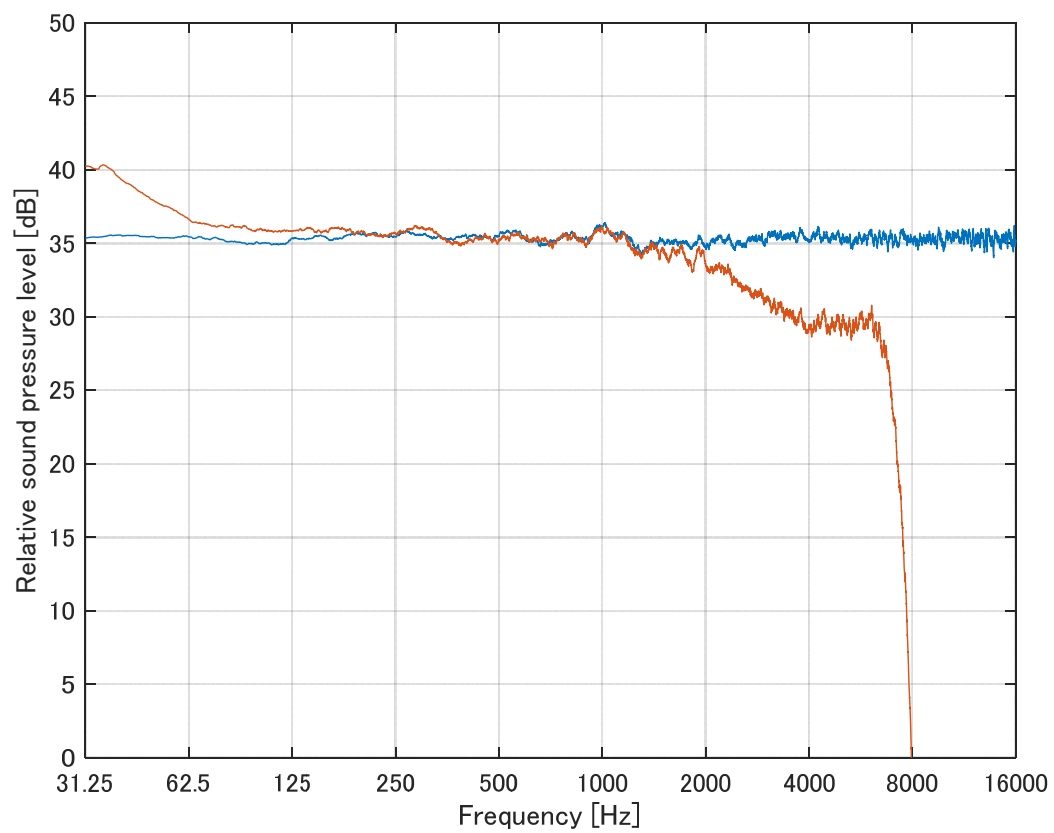


図4 Skypeの周波数特性

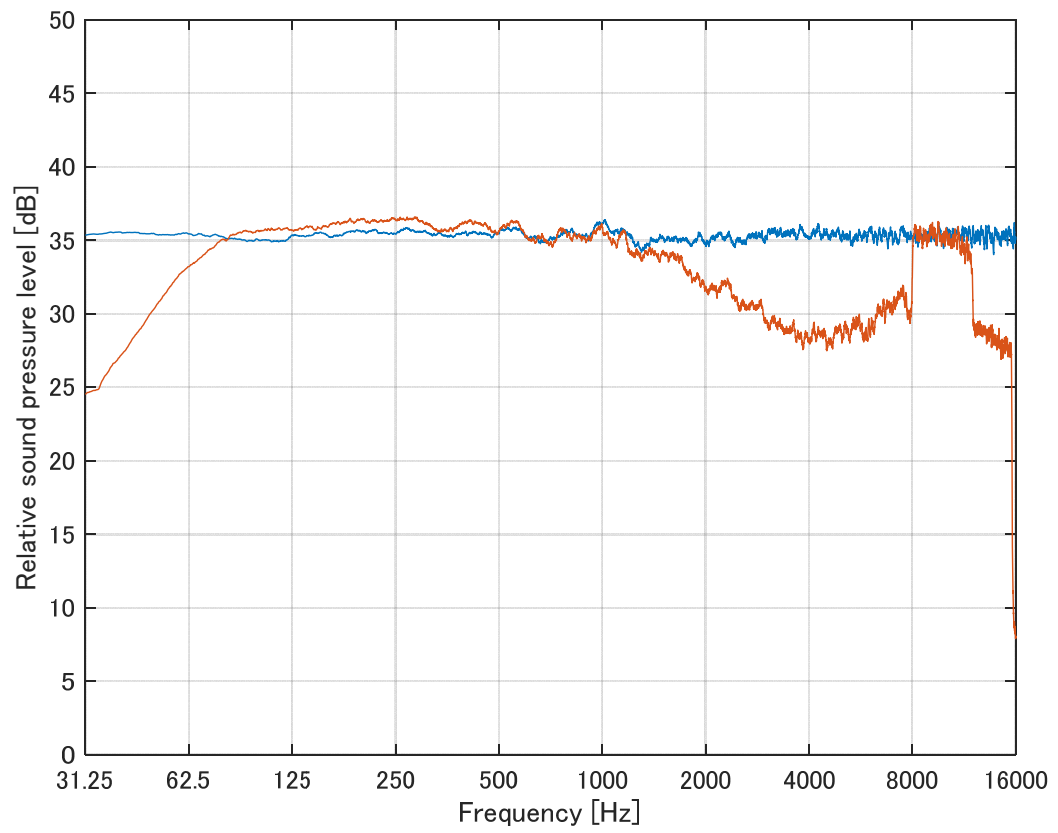


図5 SpatialChatの周波数特性

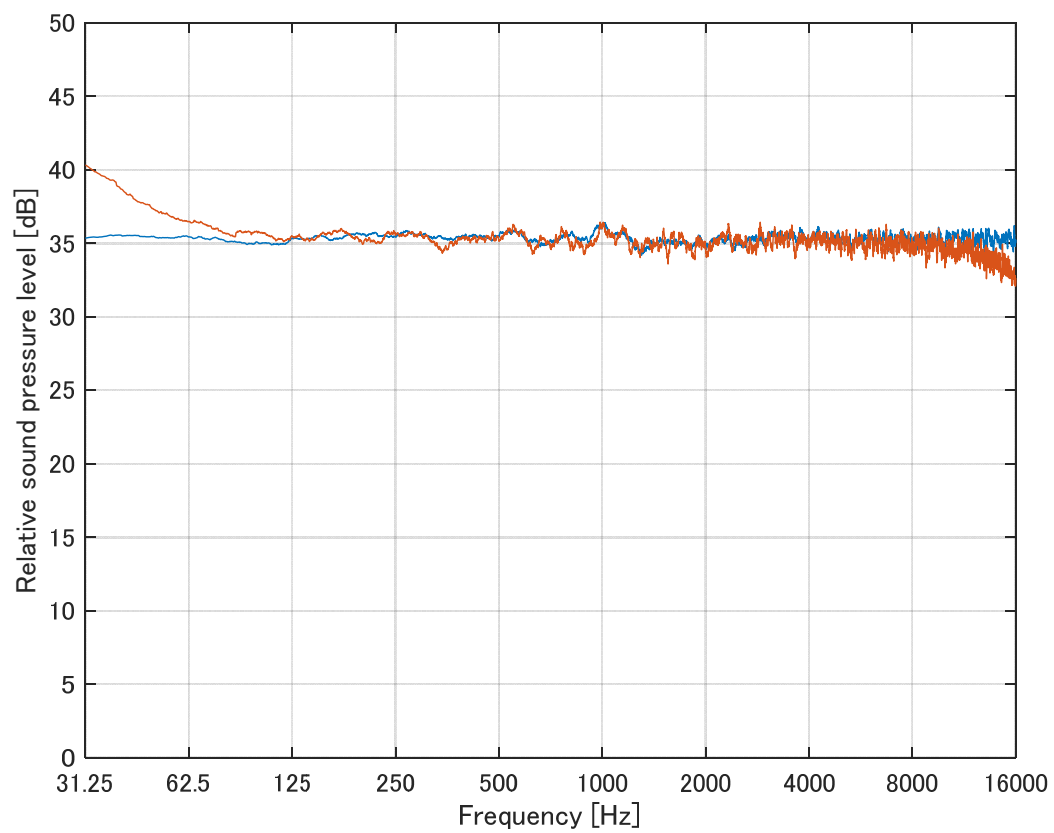


図6 SYNCROOMの周波数特性

4 「単語了解度」と「聴き取りにくさ」による評価

4.1 「単語了解度」について

「単語了解度」は伝送された単語が正しく聴取される割合である。実験には意味を持つ試験用音声を用いるため、実際に即した絶対的な音声伝送性能の評価が期待できるが、欠損した音響情報の類推が可能になる。ただ、類推による音声情報の補完は、実際の音声コミュニケーションにおいても行われていることであり、実際に即した評価を行うためには、この過程が評価に含まれることが望ましい。しかし、この類推のしやすさは、単語の意味に依存するため、音声伝送性能そのものの評価を行うためには、それらを統制する必要がある。そこで、坂本ら[2]は、単語の「なじみ」の程度を示す単語親密度[3]をもとに、類推のしやすさを統制した単語了解度試験用音表を提案した。単語親密度は、数字が大きいほど「なじみ」があることを示し、1から7の範囲の値をとる。この音表は、単語親密度と音韻バランスが統制された50単語からなり、単語親密度によって4段階に分類されている。坂本らは、この音表を用いて単語了解度試験を行い、単語親密度が高いほど「単語了解度」が高くなることを示した。単語了解度試験用音表は、欠損した音響情報の類推を不可能にすることなく、それが「単語了解度」に及ぼす影響を統制することに成功しており、現段階では、最も実際に即した音声伝送性能の評価が可能な試験用音表と考えられる。

「単語了解度」は実際に即した音声伝送性能の評価が可能であるが、比較的音声伝送性能が悪い場合でも、正答率が100%に達してしまうことが多い。正答率が100%であることは、「音声が正しく伝わる」という絶対的な評価として意味を持つが、音声伝送性能の差を相対的に評価できないという欠点がある。

4.2 「聴き取りにくさ」について

上記の問題点を解決する音声伝送性能の評価方法として、Morimoto ら[4]は「聴き取りにくさ」を提案した。「聴き取りにくさ」は、坂本ら[2]による単語親密度で統制された単語了解度試験用音表のうち、最も単語親密度の高い音表を聴取した際に感じた聴き取りにくさを、図7に示す4段階のカテゴリで評価させる。中島ら[5]の「聴き取りやすさ」とは異なり、「聴き取りにくくない」から「非常に聴き取りにくい」の単極の選択肢を用いる。回答の集計は、通常のカテゴリ評価と異なり、「聴き取りにくい」と回答した割合を算出する。具体的には、2から4のカテゴリのいずれか、すなわち程度に違いはあるが、「聴き取りにくい」という回答がなされた数の合計が、全回答数に占める割合を算出する。従って、最高の音声伝送性能は「聴き取りにくさ」が0%である状態と明確に定義されている。また、日常生活において音声伝送性能について意識することがあるとすれば、それは音声を聴き取りやすいと感じた場合よりも、むしろ聴き取りにくいと感じた場合であることが多い。従って、「聴き取りやすさ」よりも「聴き取りにくさ」のほうが、より実際の音声伝送性能に対する評価態度と一致しており、直感的にわかりやすい指標であると考えられる。

「聴き取りにくさ」では、提示された試験用音声を聞こえた通りに書き取った直後に、聴き取りにくさの評価を行うため、同時に単語了解度も求めることができる。Morimoto ら[4]は、単語了解度と「聴き取りにくさ」を比較し、音声伝送性能が比較的良い音場では「聴き取りにくさ」のほうが、音声伝送性能の違いに対して感度が高いことを示した。つまり、単語了解度では、値が最大に達してしまうため、相対的な差を評価できないような音声伝送性能の差でも、「聴き取りにくさ」を用いれば評価が可能である。

1	聴き取りにくくはない	(Not difficult)
2	やや 聴き取りにくい	(A little difficult)
3	かなり 聴き取りにくい	(Fairly difficult)
4	非常に 聴き取りにくい	(Extremely difficult)

図7 「聴き取りにくさ」評価のカテゴリー

4.3 評価手順

「単語了解度」と「聴き取りにくさ」の評価手順は以下の通りである。

- ① 5つのシステムを用い、単語了解度試験用音表から任意に選定したそれぞれ10単語ずつの無響室録音音声プレイヤーからシステムに入力し、システムを経由した後のレコーダーでの出力を収録する（収録方法は図1参照）。
- ② 収録した音声を用いて「単語了解度」と「聴き取りにくさ」についての聴感実験を行う。
- ③ 「単語了解度」と「聴き取りにくさ」を集計し、比較する。

4.4 結果

10代から40代の聴力が正常な男女10人を対象に「単語了解度」と「聴き取りにくさ」の実験を行い集計した結果、以下のようになった（図8）。

比較してみると、「単語了解度」と「聴き取りにくさ」ともにSkypeの評価が最も低いことがわかる。次いで「単語了解度」はSpatialChatの値が低く、その他3つは98%と優れた結果となっている。「聴き取りにくさ」については、Skypeに次いでTeamsの値が大きく、その他3つは僅差ではあるが、Zoom、SpatialChat、SYNCROOMの順に優れている。

	単語理解度[%]	聴き取りにくさ[%]
Zoom	98	8
Teams	98	16
Skype	89	22
SpatialChat	93	11
SYNCROOM	98	12

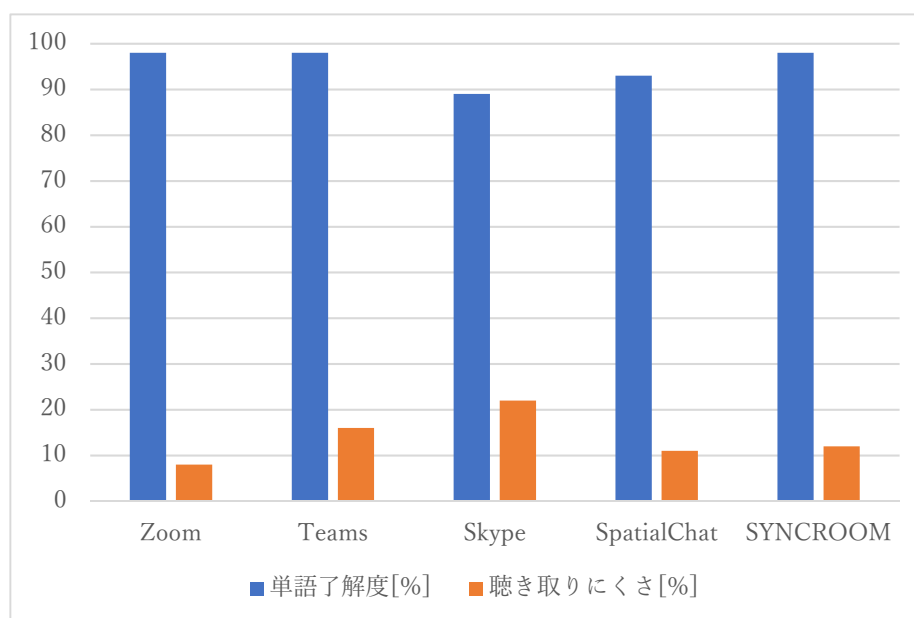


図8 単語理解度と「聴き取りにくさ」の集計結果

5 考察

図2～図6と図8を照らし合わせ、周波数特性と「単語理解度」、ならびに、「聴き取りにくさ」の関係性を考察する。

Skype や SpatialChat のように 2000~8000 Hz の周波数の音圧レベルを下げることで、「単語理解度」は低くなることが分かる。SpatialChat の場合、「聴き取りにくさ」は2番目に低いにも関わらず、「単語理解度」が低い。これは、日常会話に使われる周波数は 500~2000 Hz だが、SpatialChat は 1500 Hz あたりから音圧を下げてしまっていることが影響していると考えられる。

また、Teams や Skype のように 8000 Hz 以上の音圧を切り取ることで「聴き取りにくさ」が高くなる。この点に関しては佐藤[6]により、健聴者については、4000 Hz 以上をカットした場合に違和感が少し上昇することが明らかになっている。これは音声の個人性の手がかりが 2500 Hz 以上の周波数帯域にあることが原因だと考えられている。125 Hz 以下の音圧レベルに関しては、「聴き取りにくさ」に影響は与えないと考えられる。

一見、SYNCROOM の周波数特性が最も聴き取りやすいように見えるが、会話をする程度であれば、高音域・低音域を少し切り取っているだけの Zoom と差異はないのだろう。ただ、SYNCROOM は楽器音を伝送するためのシステムなだけあって、感覚的には一番音がきれいに感じた。

6 おわりに

本研究では物理的評価（周波数特性）と心理的評価（「単語理解度」、「聴き取りにくさ」）の両方を用いて、テレ会議システムの音質比較を行った。その結果、どの周波数の音圧レベルを下げているかを周波数特性から判断することで、「単語理解度」と「聴き取りにくさ」の数値の傾向がわかるようになった。これにより、物理的な測定結果からシステムの音質に関する心理的評価を見積もることが可能となり、今後の新しいシステムで目指すべき物理特性が明らかとなった。

ただ、本研究で用いた5つの周波数特性と全く違う周波数特性を有するシステムについては、心理的評価の傾向を予測することは困難である。よって、より多くのデータを集め、周波数特性と「単語理解度」、ならびに、「聴き取りにくさ」の関係性の精密度を上げることが今後の課題である。

オンライン化が急速に進み、新しいテレ会議システムが次々と提案されていく中、音質評価方法として本研究が役に立てば嬉しい限りである。

参考文献

- [1] 佐藤 逸人, “ユニバーサルデザインを考慮した音声伝送性能に関する研究”, 神戸大学博士論文, 2008
- [2] 坂本 修一, 鈴木 陽一, 天野 成照, 小澤 賢司, 近藤 公久, 曾根 敏夫, “親密度と音韻バランスを考慮した単語理解度試験用リストの構築”, 日本音響学会誌 54, 842-849 (1998)
- [3] 天野 成照, 近藤 公久, “日本語の語彙特性” (1999)
- [4] M. Morimoto, H. Sato, and M. Kobayashi, “Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces”, J. Acoust. Soc. Am. 116, 1607-1613 (2004)
- [5] 中島 立視、前田 節雄, “室内での音声の「聞き取り易さ」と STI“, 日本音響学会建築音響研究会資料 AA-84-30 (1984)
- [6] 佐藤 逸人, “高齢社会に対応した音声伝達性能とその評価法に関する研究”, 東北大学修士論文, 2001