



ビジネステレフォニにおける明瞭度の BRAIN モデル

ホワイトペーパー

Jeff Rodman

特別研究員 兼 最高技術責任者

2003 年 1 月 16 日

jrodman@polycom.com

© 2003 POLYCOM, INC. ALL RIGHTS RESERVED.

本資料は、Polycom Inc. が発行したホワイトペーパーの翻訳です。

Polycom、Polycom のロゴは、Polycom Inc. の米国およびその他の国における商標または登録商標です。その他の社名および製品名は、各社の商標または登録商標です。本資料に記載した内容は、予告無く変更する場合があります。

ビジネステレフォニにおける明瞭度の BRAIN モデル

"サスキアナの刑務所にいるんだよ。「サスキアナ」って言ったんだ。
スペルは、S-u-s-q-
ちがう、Q だよ！キュー、ほら、ビリヤードに使うやつさ。
ビリヤードだってば。B-i-l-l-i-
L は Larynx の L だよ！
L-a-r-y-n-
ちがう！M じゃなくて N だって！
N は Neighbor の N！スペルは、N-e-i..."

--- 「Shall We Dance」 (1937 年) で、刑務所にいる Eric Blore が Fred Astaire に電話するシーンより

はじめに

日常生活では、受話器を通して理解し合う以外に方法がないこともあります。しかし、電話での正しい音声伝達に限界があることは、1937 年よりかなり前に認識されていました。変則的で限られた帯域、ノイズ、エンドツーエンドでの音量の変動、側音、歪みなどはすべて会話音声を劣化させる原因として知られていました。1910 年に Campbell が行った実験の結果、会話音声は、戸外では 96 % の精度で伝達されますが、電話では 59 % の精度しかないことがわかりました。⁽¹⁾ 「会話音声を正確に伝える有効な道具」としての電話の能力は、他に代わるものがないために許容されてきましたが、そこには一種のあきらめに似た感情が存在することも事実です。

デジタル中継技術や交換技術が発達した現代においても、これは日常的な問題です。アナログループの長さ、構内配線、回線の種類ごとに異なるイコライゼーション特性、受話器やスピーカーフォンの設計不良、異種ネットワークの混在、会議室内のノイズ、紙をめくる音、ペンでテーブルをたたく音、ファンノイズなど、さまざまな問題がデジタル化時代においても相変わらず存在するからです。

このホワイトペーパーでは、これらの問題について考察し、ビジネステレフォニにおける重要な要素の "BRAIN" モデルを紹介します。多様な要素の相関関係を利用して、電話やオーディオシステムのパフォーマンスを改善する方法を探るとともに、最新のシステムが明瞭な通信の実現にどのように役立つかを示します。

ビジネステレフォニの特徴

ビジネスにおいて、電話が果たす役割は多くの理由によって重要視されています。次のようなビジネステレフォニの特徴について考察してみましょう。

- 私用電話に比べると時間的な制約が厳しいことが多いため、誤解があった場合や「もう一度おっしゃっていただけますか」と言われた場合、より大きなストレスを感じます。会議の長さは決まっていることが多く、失った時間は取り戻せません。
- 発言内容が正確に伝わるのが必須です。何らかの理由で音声の歪みや精度の劣化が生じると、ビジネス上の損失につながりかねません。
- まったく面識のない相手と会話することもよくあります。重要なビジネス関係を構築する上で、1 本の電話によって双方の第一印象が決まります。
- 母国語の異なる人や方言のある人と電話で話す場合、テレフォニの抱えるさまざまな問題点に加えて、なまりや話し方も大きな課題になります。
- 複数のグループが会議に参加すると参加人数が多くなり、スケジュール調整が困難になるとともに、コストが上がり、会議の潜在的な価値は高くなります。しかし、グループで会議に参加するにはスピーカーフォンが必要になるので、音質の劣化は避けられ

ビジネステレフォニにおける明瞭度の BRAIN モデル

ません。室内残響、ファンノイズ、クリッピング、ハウリングに加えて、妨害や中断が発生したり、各参加者とマイクとの距離が一定でないという問題もあります。

- 発言者が複数いる場合、発言者を正確かつ迅速に特定することは重要ですが困難でもあります。
- 会議が長引いても、注意力を維持することが必要です。スピーカーフォンの性能が少し違うだけでも、時間が長くなると疲れや注意力に大きく影響するので、音質がさらに重要になります。

音声の明瞭度と BRAIN モデル

このような理由から、ビジネステレフォニでは、明瞭度（音声の聞き取りやすさ）が特に重要になります。グループでテレフォニを利用することが一般的であるにもかかわらず、明瞭度はグループの場合に特に大きな課題となります。それでは、音声の明瞭度にはどのような要素があるのか見てみましょう。

私たちが会話音声を理解するメカニズムは 3 段階に分かれています。ここでは、それぞれ物理的段階、認識段階、分析段階と呼ぶことにします。物理的段階では、音声話し手の口から聞き手の耳に伝達されます。このときの精度が非常に重要です。認識段階では、聞き取った中から不明瞭な点があれば、聞き手は文法上の規則やなまりの特性などを文脈から推測し、解決しようとします。最後の分析段階では、最初の 2 段階で解決できなかった単語が詳しく吟味されます。聞き手は問題の単語をより広範な文脈に照らし合わせることで、その意味を推察しようとします。

しかし、この第 2 段階および第 3 段階に必要な注意力は、徐々に散漫になっていきます。また、話し手が外国人である場合は、前述のような文脈による解釈を適用できないこともあります。母国語や方言の異なる話し手は、普通考えられないような単語や発音を使うことがあるからです。

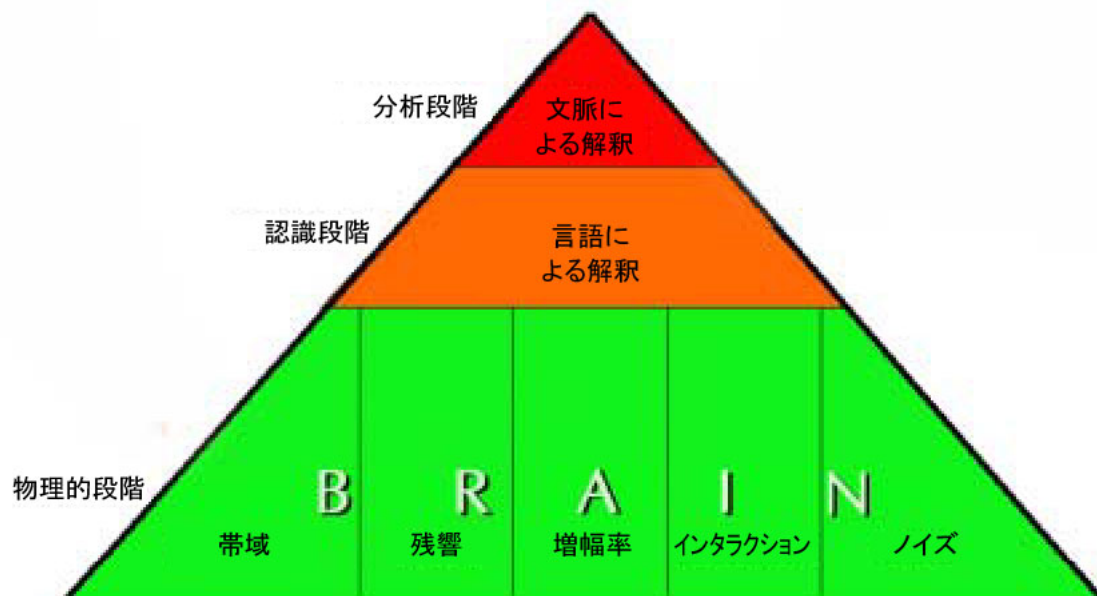


図 1. 物理的な音声通信の BRAIN モデル

ビジネステレフォニにおける明瞭度の BRAIN モデル

このような理由から、最も重要な防衛線は会話音声の物理的要素であるといえます。そもそも会話音声聞き手にはっきりと伝達されれば、不明確な要素が少なくなり、会議はスムーズに進行するのです。

明瞭度の物理的要素

物理的な音声通信の主要素には、帯域 (bandwidth)、残響 (reverberation)、増幅率 (amplitude)、インタラクション (interaction)、ノイズ (noise) の 5 つがあります。これらの頭文字を合わせて BRAIN モデルと呼びます。ビジネスにおける音声通信では、これら 5 つのパラメータを元にして、会話内容を理解したり、話し手を特定します。

"帯域" とは、聞き手に伝達される音声の帯域を意味します。電話は 300 Hz~3.3 kHz の帯域に限定されているので、人の声に含まれる周波数の 20 % しか伝達できません。これに対し、ビデオ電話システム、次世代電話、携帯電話など、現在商品化されているビジネス用オーディオシステムには、7 kHz または 14 kHz までの周波数を伝達できるものもあります。

"残響" とは、話し手とマイクの間で発生する室内反響音の程度を意味します。残響があると、会話の内容はさらに理解しにくくなります。どの程度の残響が発生するかは、部屋の実験 (壁、床、天井が硬く反射しやすいかどうか) や大きさ、スピーカーとマイクの対面角度に左右されます。マイクが話し手の方を向いていない場合や離れている場合は、直接的な会話音声ではなく残響がマイクに集音される度合いが高くなり、明瞭度が低下します。

"増幅率" とは、話し手の声聞き手にどの程度大きく聞こえるかを意味します。すべての条件が同じであれば、話し手の声が小さいほど内容が理解しにくく、聞き手がラウドスピーカーから離れるほど聞こえにくくなります。電話回線によるネットワークの場合は、接続状況に応じてゲインに最大 20 dB (100 倍) もの差があるため、この状況がさらに悪化する可能性があります。

"インタラクション" とは、2 人以上の参加者が電話会議で自然に対話できるかどうかを示します。必要に応じて、会話の流れを止めることなく相手の発言を遮って発言できなければ、対話が堅くなったり不自然になります。性能の悪いスピーカーフォンを使用した場合などは、「何ですか?」、「お先にどうぞ」、「いえ、そちらからどうぞ」といったやりとりで時間を取られ、会話の流れが途切れがちになります。

"ノイズ" とは、話し手の声と一緒に周囲の騒音が集音される度合いを意味します。エアコンやプロジェクタからのファンノイズなど室内騒音は集音されやすく、音声の明瞭度が低下する大きな原因になります。このことは、会議室でオーディオシステムのマイク付近に LCD プロジェクタが置かれるようになって以来、さらに深刻な問題になっています。LCD プロジェクタでは、小さなファンから大きな音がすることが多いためです。

これら 5 つの BRAIN 要素で興味深いのは、組み合わせによって音声の明瞭度が高くなったり、話し手を特定しやすくなる点です。これまでの研究で、5 つの要素が驚くほどのレベルで互いの欠陥をカバーし合うことがわかっています。たとえば、誤って集音されたノイズが話し手の声に乗せられた場合、ノイズの量が多いと会話の内容は聞き取りにくくなります。ただし、信号全体の帯域を広げると補正することができます。室内残響時間が長いと、信号が濁って聞き取りにくくなりますが、信号の帯域または音量、あるいはその両方を引き上げることでカバーできます。話し手がマイクから遠すぎるためによる音量の不足は、ノイズの除去、帯域の拡張、残響時間の短縮などによって改善を図ることができます。

BRAIN モデルを最適化する方法

BRAIN に含まれるパラメータは操作しやすいので、このような相互作用があることは幸運といえるべきでしょう。直接的に解決できない問題（たとえば室内反響を少なくするには、相当の費用を要する場合があります）は、パラメータの改善で補完することができます。

各パラメータは必ずしも問題に直接影響するわけではありませんが、状況の改善に役立ちます。言い換えると、帯域を拡張しても室内反響は減少しませんが、会話の内容は聞き取りやすくなります。

会議の効率と快適性を高めるためには、どのような手段があるのでしょうか。ここでも BRAIN モデルに沿って考察し、問題の原因と解決方法を探ります。

帯域。 ナローバンドの POTS (Plain Old Telephone System) や IP 接続で電話を使用する場合は、帯域が不足することがよくあります。このような帯域の不足は、100 年来のインフラによって制約されているために、解決不能な問題であると長い間見なされてきました。ところが、新しいテクノロジーの台頭がこの問題を取り除こうとしています。ポリコム製の VTX テクノロジーを使ったアナログテレフォニで使用できる帯域は、従来の 300 Hz~3.3 kHz の 2 倍の 80 Hz~7 kHz になります。この周波数帯は最高性能のビデオシステムに匹敵するものですが、それを従来の電話回線で得ることができます。デスクトップでは、IP テレフォニを使った 7 kHz のデスクトップ電話の規格として、TIA (米国電気通信工業会) 920-200 が制定されました。この規格は近い将来、製品に採り入れられる予定です。

残響。 残響は直接操作することも、BRAIN モデルの他の要素を介して操作することもできます。残響の問題を直接解決する方法の 1 つは、話し手のマイクの位置を調整することです。たとえば、各発言者にとって最も集音のよいマイクを効率的に選択するマルチマイクロフォンシステム (Vortex[®] 独立型音声システム、SoundStation VTX 1000™、SoundStation[®] Premier 卓上システムなど) を使用することができます。また、メインの発言者に最も近い位置にマイクを設置したり、個別のマイク (Polycom のワイヤレスマイクなど) を使用する方法もあります。もちろん、残響の問題を根本的に解決するには、部屋そのものを工夫する方法もあります。音響反射板や吸音材を壁、天井、四隅などに取り付ければ、室内残響が大きく改善されます。

これらの方法を試みても満足のいく結果が得られない場合は、BRAIN モデルの別の要素を使って、過剰な残響をカットすることもできます。たとえば帯域を広げることにより、残響感のある音声の明瞭度を大幅に改善できます。あるテストでは、残響時間の長い室内で帯域を 4 kHz から 8 kHz に引き上げると、明瞭度が 52 % から 80 % に改善したと報告されています。⁽²⁾ 過剰なノイズが発生する環境では、増幅率を上げることもノイズ除去に役立ちます。

増幅率。 増幅率 (ラウドネス) が不足すると、話し手の声が聞こえにくくなります。チャンネルゲインが 20 dB 下がると、明瞭度の測定基準の 1 つである明瞭度指数は 80 % から 30 % まで低下します。⁽³⁾ 増幅率が低くなる原因としては、話し手がマイクから遠すぎる、聞き手がスピーカーから遠すぎる、回線の不可逆が激しい、などが考えられます。話し手と聞き手が物理的に移動することは明らかな解決方法ですが、現実的に難しいこともあります。このような場合、マイクのゲインを自動的に調整できるシステムがあれば非常に便利です。テレフォニ オーディオ システムで、マイクの AGC (オート ゲイン コントロール) が適切に設計されていれば、大幅に改善するためのツールになります。もちろん、マイクの AGC で、電話回線のゲイン変動を補うことはできません。しかし VTX テクノロジーには、電話回線のゲインそのものを正規化するというあまり知られていない機能があります。この機能を使用すると、多様なアナログネットワークで有効な変動値をプラスマイナス 1 dB 程度に抑えることができます。

ビジネステレフォニにおける明瞭度の BRAIN モデル

BRAIN の他の要素を適用して、増幅率が低い状況を解決することもできます。帯域の拡張 (従来のスピーカーフォンを 2 台の VTX ユニットに取り替えるなど)、残響時間の短縮 (最適なマイクを選択するなど)、あるいは周囲の騒音除去などを行うと、増幅率の問題をかなり改善できます。

インタラクション。地理的に離れた場所にいる複数のグループ間でインタラクティブに対話することは、さまざまな理由により困難を伴います。最も単純な例として、透過的にインタラクティブな対話を行える全二重システムがない場合が挙げられます。ここでも、最適なマイクへの交換、帯域の拡張、信号の強度確保などの方法を検討する価値があります。

インタラクションの質の低さは、音声信号のエンドツーエンドの遅延が長すぎることに起因する場合があります。遅延原因の 1 つとしては、通信チャネルそのものの故障があります。また、最近では少なくなりましたが、衛星による接続の場合も、地上ベースの接続に比べると遅延が大きいといえます。このことは、放送などの一方通信では問題になりませんが、インタラクティブなライブ会議を行う場合などには致命的な問題に発展しかねません。また、IP テレフォニシステムの使用も大きな遅延原因になります。IP テレフォニシステムでは、150 ms 秒もの遅延が発生する場合があります。POTS や IP の会議接続装置が原因で、処理中に大きな遅延が生じることもあります。室内残響も、話し手と聞き手の間に明らかな遅延を生む原因です。これらすべてのケースにおいて、まず設計と評価を入念に行うことが重要になります。

ノイズ。一般的なノイズの発生源は、会話音声の周波数帯と重なる部分が多いため、マスクでノイズを除去しようとする、と、会話が聞き取りにくくなります。まず第一に、ノイズの発生源を断つことが重要です。エアコンのダクト、オーバーヘッドプロジェクタ、コーヒーマーカーなどからマイクを遠ざけます。指向性マイクの場合は、マイクの向きをノイズの発生源からずらすと効果があります。ただし、発言者に向いていることが条件です。

このような直接的な方法で問題を解決できない場合は、帯域を広げると、明瞭度を大幅に高めることができます。発言者の声を伝える信号量が多くなれば、人の声とノイズを区別しやすくなるので、会話の内容がわかりやすくなります。また、音声帯域が拡張された VTX テクノロジには、電話回線からノイズを除去して、アナログテレフォニに伴うクリック音やアーティファクトを取り除けるという特徴もあります。

この問題のもう 1 つの解決方法として、最近のシステムに採用され始めているアクティブ ノイズ リダクションのアルゴリズムがあります。これは、会話音声への影響を最小限またはゼロに抑えながら、タイムドメイン理論および周波数ドメイン理論に基づいてマイク音声を分析し、6 dB または 9 dB のファンノイズを除去できる高度なフィルタを適用する機能です。たとえば、Polycom Vortex や SoundStation VTX 1000 はこの機能を搭載しています。一般的なノイズの問題を解決するには、ここで説明した多様なアプローチを組み合わせると有効です。

結論

BRAIN モデルの各分野におけるテクノロジーは著しい進歩を遂げています。また、BRAIN モデルには、互いにサポートしたり補完し合うという特性があります。これらを組み合わせて活用することで、優れた音声会議を容易に実装し、満足のいく結果を得ることができます。Eric Blore と Fred Astaire のサスキアナ刑務所にそのような設備が整うのはもう少し先のことになるでしょうが、ビジネスの世界にいる私たちは、今すぐにも、手ごろな価格で音質を大幅に改善することが可能なのです。

⁽¹⁾ George A. Campbell 著、「Telephonic Intelligibility」、Philosophical Magazine、19, ser. 6 (1910): 158

ビジネステレフォニにおける明瞭度の BRAIN モデル

⁽²⁾ P. W. Barnett 著、「Overview of Speech Intelligibility」、Proceedings of the Institute of Acoustics, Rayleigh House, Bush Hill Park, Vol. 21 Part 5 (1999)

⁽³⁾ Harvey Fletcher および Rogers H. Galt 共著、「The Perception of Speech and Its Relation to Telephony」、The Journal of the Acoustical Society of America, 22 No. 2 (March 1950): 138