

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

おおむろ ひとし おかもと まなぶ
大室 伸 / 岡本 学
さいとう しょういちろう さかうち すみたか
齊藤 翔一郎 / 阪内 澄宇
えむら さとる
江村 暁

NTTサイバースペース研究所

従来の電話は、人の声に含まれる周波数成分の一部のみを送っているため、個性や明瞭性が不足することがあります。NGNで提供される高音質電話では、肉声に近い音質での通話が可能になるほか、ステレオ音声による「場」の共有も期待されています。本稿ではそれらに必要なコーデック、音響設計、信号処理技術について紹介します。

リアルな声とは

「フレッツ 光ネクスト」の「ひかり電話」は、単にVoIP（Voice over IP）技術を利用するだけの電話サービスではなく、従来の電話では送ることのできなかったさまざまな情報を、電話と同じ使い方（かけ方）で送ることができるようになりました。その1つが広帯域音声です。従来の電話では、人間の声に含まれる周波数成分のうち300 Hz～3.4 kHzのみを送っています。電話の声は個性が不足しているとか、似た言葉と聞き違いをするといわれる原因が、この一部の周波数成分しか送っていないところにあります。一方、広帯域音声は、従来の2倍に相当する50 Hz～7 kHzの周波数成分を送ることができるため、肉声に近いリアルな音質で通話することができます。

音声符号化技術

音声信号をデジタル化（パケット化）する機能を、音声コーデックと呼びます。従来のアナログ電話（加入電話）では、音声コーデックを電話局の加入者交換機に持ち、符号化方式にはITU-T G.711が用いられてきました。一方、ひかり電話では、音声コーデッ

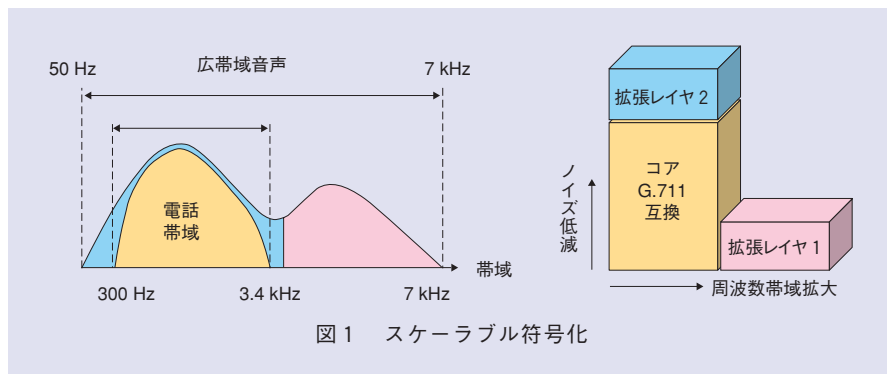
クは端末（IP電話機またはホームゲートウェイ）に実装します。ひかり電話でも従来同様の電話として利用する場合には、符号化方式としてG.711を利用します。

広帯域音声で通話をするためには、広帯域音声符号化方式を利用します。G.711を利用するか広帯域音声符号化方式を利用するかは、発側と着側の端末間のネゴシエーションで決定されます。古くから利用されている広帯域音声符号化にはITU-T G.722があり、テレビ会議システム等で利用されています。最近はIP電話機でも利用されるようになってきました。

電話の音を広帯域音声にする観点で重要なことは、広帯域音声で通話する場合でも、従来の電話と同様のサービスを可能とすることです。例えば、保

留・転送や多地点通話などが従来どおり利用できなければいけません。その際に問題となるのがコーデックの切り替え制御です。従来はG.711を使うことが原則でしたので、コーデックの違いや切り替えを意識する必要がありませんでした。しかし、広帯域音声符号化が混在する環境では、多様な通話形態におけるコーデックの切り替え制御が課題となっています。

これに対して、NTTサイバースペース研究所では、G.711をコアとするスケーラブル符号化を開発し、ITU-T G.711.1として標準化をしました。スケーラブル符号化（図1）とは、従来と互換性のあるG.711の符号に、音声品質を向上するための「拡張レイヤ」の符号を組み合わせるというもので、積み木に例えることができます。



G.711とG.711.1は「積み木の一部を抜き出す」「足りない積み木を積む」だけで容易に変換可能なため、G.711とG.711.1が混在する環境でも、保留・転送や多地点通話などの電話機能を容易に広帯域音声対応にできると期待されます⁽¹⁾。

高音質ハンドセット設計技術

ハンドセットによる通話は電話機の基本です。電話機の広帯域音声対応は音声コーデックをソフトウェアで実装するだけでは実現できません。従来のハンドセットは、最低限300 Hz～3.4 kHzの音声を送受話すればよいと規定されており、そのまま用いても4 kHz以上の音声を十分に送受話できない設計になっています。したがって広帯域音声用のハンドセットを新たに設計する必要があるため、そのためには基準となる特性が必要となります。

情報通信ネットワーク産業協会(CIAJ)では、2007年に150 Hz～7 kHzの音声帯域で通話するハンドセット電話機の品質規格として「電話機通話品質標準規格CES-Q004」を規定しました(図2)。それによると送話は125 Hz～6.3 kHz、受話は160 Hz～6.3 kHzの音声の送受話が可能で、かつ100 Hz～8 kHzの間で周波数特性が一定の範囲に入っていることが品質条件として求められています。

3.4 kHz以上の音響実装は、拡声の場合、オーディオや放送機器のノウハウが流用できますが、ハンドセットの場合新たな設計のノウハウが必要となります。その際には、音響素子(マ

イク・スピーカ)の選定、実装の工夫、回路による雑音対策の見直し、信号処理による特性補正等がポイントとなります。

2007年のNGNトライアル用に開発した最初の高音質電話機NP-1では、送受話用のマイクとスピーカは電話機用のものではなく、オーディオ用のものを用い、さらに実装方法に工夫をしています。送話器はマイク素子前面に空気室を設け、容量や形状を調整しました。受話器は、スピーカ前室の容積、開口穴の位置等を調整しました。これらにより送受話器共に7 kHzまでの平坦な特性を実現しています(図3)。

一般に送話器より受話器の方が、平坦な特性の実現が困難です。また使用できる音響素子により、適切な実装方法は異なるため、受話器の設計時には空気孔や空気室を設ける余地を残しておくことが重要となります。

そのほか、従来の電話の回路では300 Hz以下や3.4 kHz以上の雑音対策がされてない場合もあり、広帯域化で雑音が目立つようになる場合があるため、回路の見直しが必要となります。さらに送受話器の実装のみで規格の特性を得られない場合は、電話機本体に補正機能を実装し、信号処理により規格の特性に合わせる手段も有効です。

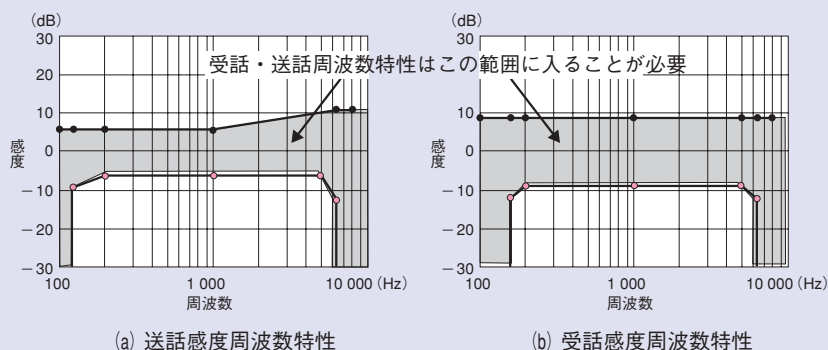


図2 CIAJによる広帯域IP電話機の周波数特性規定(CES-Q004)



図3 トライアルに用いた高音質電話機(NP-1)

2009年に開発した無線LAN用高音質コードレス電話機の試作機（図4）は、携帯サイズの小型筐体のため、実装の工夫は主に開口部の面積を大きく取ることにとどめ、音響素子の選定や、電話機内部の周波数補正処理により高音質の特性を実現しています。

エコーキャンセラ技術

ハンドセットによる通話は会話内容が周りに漏れないため一対一のコミュニケーションに適しているのに対して、ハンズフリー通話は多人数で円滑に情報を共有したい場合に適しています。特に近年、小型で高機能な端末が多く市場に出回るようになり、出先で手軽



図4 高音質コードレス電話機

に電話会議を行ったり、家庭で手軽にテレビ電話を行ったりすることが可能になりつつあります。そのため、今後音声コミュニケーションの1つのあり方として、小型端末によるハンズフリー通話が浸透していくものと考えられます。

ハンズフリー通話（図5）では、スピーカから出た音がマイクで再び收音されて相手に戻る「エコー」が発生するため、デジタル信号処理によってエコーを消去するエコーキャンセラを用います。小型端末にエコーキャンセラを実装するうえで技術的に問題となるのは、使用環境の雑音と筐体設計です。

雑音環境で問題となるのは、エコーキャンセラがエコー経路（スピーカからマイクへの音の空間伝播特性）を正しく推定できないことです。一般にエコーキャンセラは、マイクで收音された音が端末側の話者の音声のみである場合は何もせず、エコーのみである場合は、そのエコーをもっとも小さくするように音の伝達経路を推定しつつエコーを消去します。そうすることで、両方の話者が話した場合でも、エコー成分だけを取り除き、話者音声のみを相手に伝えることができます。しかし、雑

音環境では、エコーキャンセラがエコーだと判断した音に雑音が含まれてしまうため、エコー経路の推定にも誤りが生じてしまいます。そのため、エコーを適切に消去することができず、相手の話者は消し残ったエコーにより通話がしにくくなってしまいます。

雑音高耐性エコーキャンセラ技術（図6）では、マイク入力に含まれる雑音とエコーの比率を推定することでエコー経路の推定速度を制御し、雑音への耐性を向上させています。具体的には、スピーカへ入力される信号と、エコーキャンセラから出力される誤差（送話）信号の比の統計的分布を用い、現在の信号がどれほどエコーに近い、もしくは雑音に近いかを推定します。そして、雑音が誤差信号に多く含まれるほどエコー経路の更新を小さくすることで、エコー経路の推定値が真の値から大きく外れないような工夫をしています。この技術を導入することで、導入前に比べてエコーの低減量が2～3倍に改善され、一般的な使用環境では聞き取れない程度にまでエコーを低減することができました。

もう1つの問題は筐体設計です。小

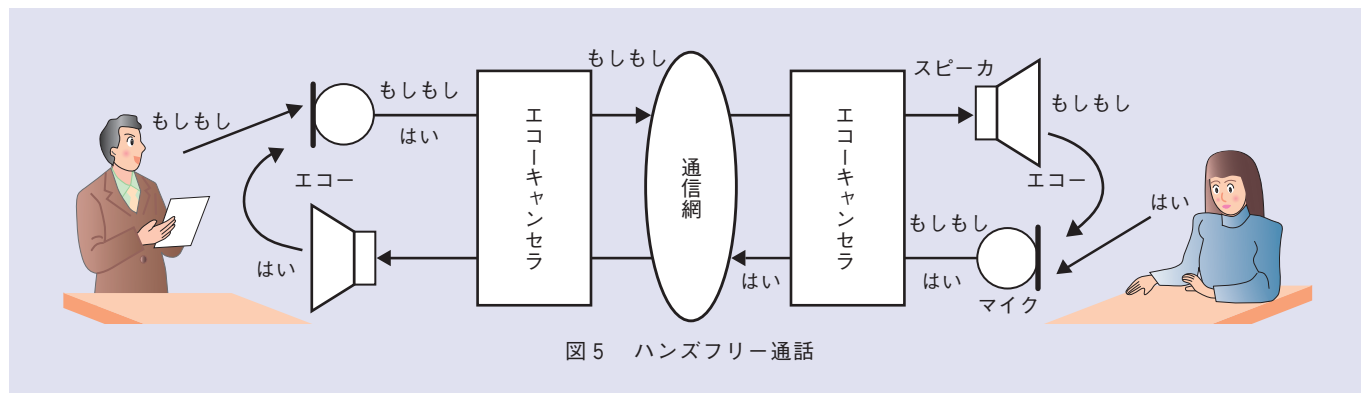


図5 ハンズフリー通話

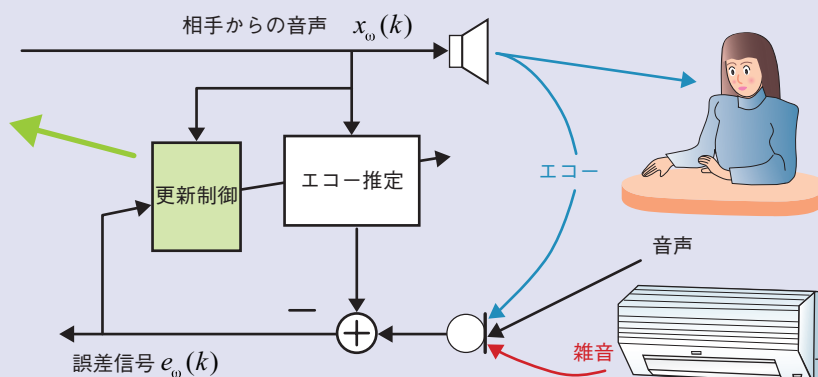
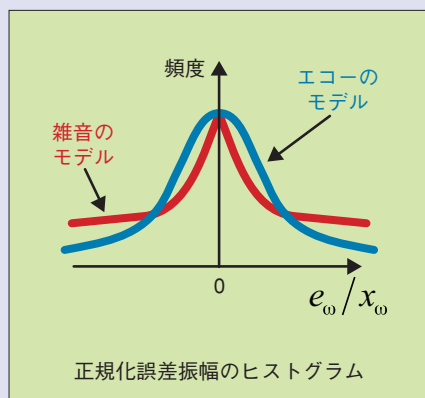


図6 雑音高耐性エコーキャンセラ技術

型端末では、専用のテレビ会議システム等に比べてスピーカとマイクの距離が非常に近く、エコーの音量が非常に大きくなり消去が困難になります。この問題に対処するには、まずはマイクロホンに指向性をつけ、端末の前にいる話者の声を大きく、スピーカの音なるべく小さく收音するよう設定します。それに加えて、筐体内部を伝わってマイクに入る音を低減するため、リブを設けて筐体の強度を上げ振動を低減したり、プラスチックの板やスポンジなどでスピーカとマイクの空間を何重にも隔てたり、マイクが直接筐体に触れないようゴムなどで本体に固定する、などの対策を行います。これらは現状ではある程度の予測を持って試行錯誤をするしかありませんが、将来的には計算機シミュレーションを用いて行うことも考えています。

こういった工夫を施すことで、通常のハンズフリー装置と同等の性能を小型化して実現することに成功しました。本技術を搭載したIPテレビ電話機が、

「ひかりフレッツフォン VP3000」としてNTT東日本・NTT西日本から販売予定です。

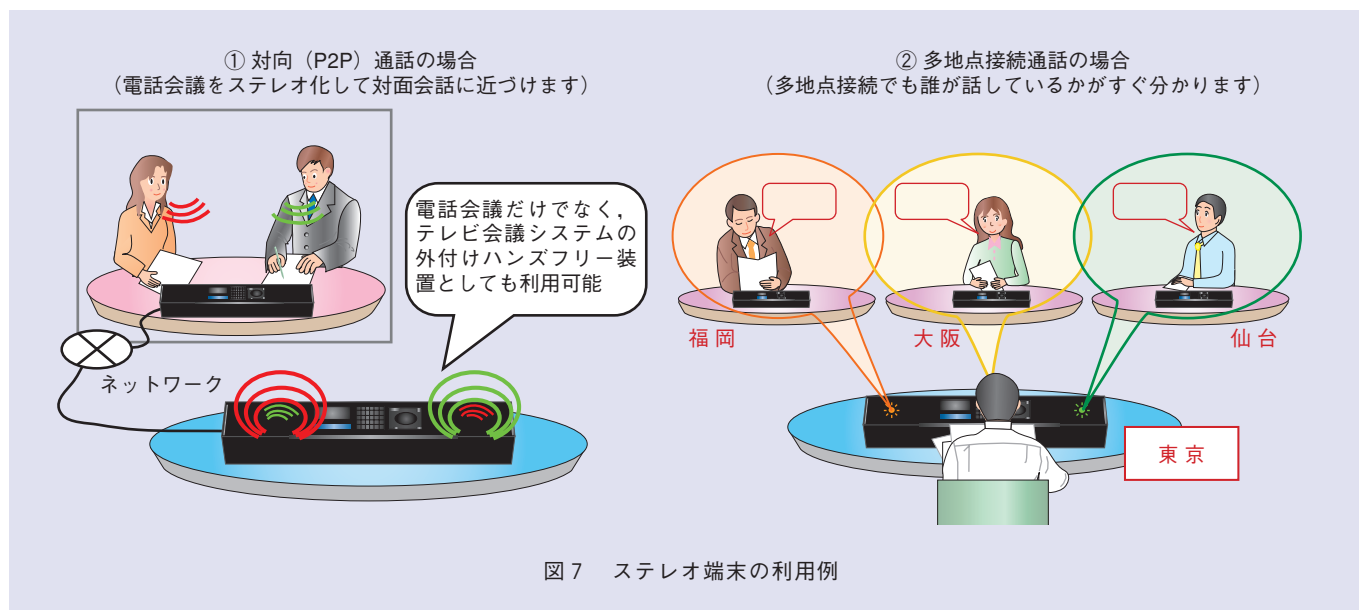
ステレオ端末技術

重要な会議や打合せには、例えば遠方であっても直接出向いて参加するといった場面がまだ多く見かけられます。出張（移動）の削減は、コスト抑制のみならず環境への負荷軽減にもつながるため、昨今、電話会議をはじめとする遠隔通信のさらなる利用が強く求められるようになりました。電話会議は、端末の設置が容易で操作も手軽なため、テレビ会議に比べると頻繁に利用されていますが、音声だけによる通信のために通信先の誰が話しているかわかりにくい、相手側の雰囲気がかめれないといった課題もあります。そこで、このような課題を解決し、よりリアルで豊かなコミュニケーションを実現する電話会議向けステレオ端末の開発を進めています。この端末は、7 kHz よりも広いFM放送並み（14 kHz）の

ステレオ音声を用いることにより、話者の識別性を向上し、相手側の雰囲気を与え、対面での会議に近い電話会議を実現します。

開発中のステレオ端末は、ステレオ音声に対応したエコーキャンセラや、端末から遠い人の声は大きくし近い人の声は抑える方向別AGC（Auto Gain Control：自動音量調整）の機能を具備しています。さらに、発話者の識別性を向上させるための機能を新たに開発して実装しました。この機能は、はじめに、複数のマイクロホンで收音した信号の音量差と時間差から相手側の発話者位置を検出します。次に、自分側端末のスピーカ近傍に配置した複数のLEDを、相手側の発話者位置それぞれの音声レベルに連動させて独立に点灯させるという機能です。音声の再生とともに各LEDが点灯するため、視覚と聴覚のマルチモーダル*な

* マルチモーダル：外界からの情報を、五感や体性感覚（平衡感覚など）といった複数の感覚を組み合わせる知覚する仕組み。



認知によって発話者の識別性が飛躍的に向上します。なお、14 kHz帯域のステレオ音声を相手側端末と送受するためには、新たな音声コーデックが必要となります。そのために、G.711.1をさらに14 kHz帯域・ステレオに拡張した符号化方式の標準化も進めています。

このようなステレオ端末を用いた遠隔通信サービスの利用例について、図7を用いて説明します。図7①は、対向 (P2P) 通話を表しています。FM放送レベルの音声品質によって、声だけではなく相手側の環境音までも伝えることができるため、臨場感が飛躍的に高まります。さらに、ステレオ再生とLED点灯によって相手側話者の識別も容易になり、会議のしやすさが格段に向上します。図7②は、多地点接続通話を表しています。ステレオ再生によって、例えば相手側3地点の音像 (音声) をそれぞれ左・中央・右に定

位させることができるため、モノラル再生では不可能な各地点を分離した識別性の高い電話会議が可能となります。なお、PCを用いたテレビ会議の場合には、ステレオ端末をPCに接続して外付けのハンズフリー装置として利用することも可能です。手軽に持ち運べる一体型装置の利点を活かせるため、場所を選ぶことなく高精細な映像コミュニケーションの実現も可能となります。

今後の取り組み

今後は事業会社と連携し、高音質電話端末の普及を目指します。また、単なる電話の延長ではなく、広帯域音声・ステレオならではの新たなコミュニケーションスタイルの提案を行っていきたいと考えています。

参考文献

- (1) 佐々木・日和崎・森・大室・片岡：“広帯域音声符号化の国際標準ITU-T G.711.1 (G.711 wideband extension),” NTT技術ジャーナル, Vol.20, No.5, pp.34-37, 2008.



(後列左から) 大室 伸/ 齊藤 翔一郎/
江村 暁
(前列左から) 岡本 学/ 阪内 澄宇

今後は電話もエンドエンドデジタル化を機に、リアルな声が標準の時代になるよう努力したいと思います。

◆問い合わせ先

NTTサイバースペース研究所
第一推進プロジェクト
TEL 0422-59-4548
FAX 0422-60-7811
E-mail hq-telephone@lab.ntt.co.jp