

音声をよりクリアに、音楽をより豊かに ——残響制御が切り拓く「音」の世界

少しでも残響の残る部屋で音声をマイク収録すると、収録音には目的の音声だけでなく残響も含まれてしまいます。残響は収録音声の明瞭性を下げ、また自動音声認識の性能を低下させる要因になります。このように、コミュニケーションの手段としての音声にとっては、残響はしばしば邪魔なものになります。一方、音楽の鑑賞や演奏にとって、残響は必要不可欠な要素です。例えば、音楽ホールのステージ上で音楽が演奏されれば、その音はホール独自の豊かな響きである残響を伴って、厚みのある音として聴衆の耳に届きます。本稿では、残響のこのような功罪両面の効果を自在に操る革新的音響信号処理技術「残響除去・制御技術」を紹介します。

きのした けいすけ
木下 慶介

NTTコミュニケーション科学基礎研究所

革新的音響信号処理技術 「残響除去・制御技術」

■研究の歴史

残響により響いた音声をクリアで聞き取りやすい音声に戻したり、いったん収録された音楽をコンサートホールで聴いているかのように豊かなものにしたたりするには、収録音に含まれている残響を制御する「残響除去・制御技術（残響制御技術）」が必要です（図1）。

残響制御技術は、雑音除去技術と並び、長い研究の歴史があり、世界的にみると1980年代から盛んに研究が行われてきました。しかし、雑音除去と違い、実環境における残響制御は近年まで実現されず、困難な未解決問題と

して残されてきました。これに対し、我々は線形予測という数学的手法を發展させることで、世界で初めて実環境で高精度に動作する残響制御を実現しました⁽¹⁾。

■残響制御技術の核：残響推定、分離の原理

高精度な残響制御のために重要となる処理は、目的の音響信号に、どのような残響がどの程度含まれているかを自動で推定する処理（残響推定処理）です。

図2に残響の特徴をまとめました。図2(a)は、残響のない環境で人が「あ、い」と発音した場合の録音信号の特徴を表しています。図からも分かる通り、残響のない環境で録音された信号

中の時間的に離れた場所の間には、一般に特定の関係性（相関）はありません。これは、音響信号が時々刻々と変化するためであり、多くの音響信号がこのような特徴を持ちます。

一方で、図2(b)に示すように、残響のある環境で同じ「あ、い」が発音された場合、信号中の時間的に離れた場所は相関を持つようになります。これは、過去の信号が残響により引き延ばされ、その後の音に重畳してしまうためです。この時間的に離れた信号間での相関を、残響に起因する相関とみなし、過去の信号と相関のある成分は残響、相関のない成分は直接音としてそれぞれの成分を推定するのが、我々が提案した残響推定処理です。

さまざまな応用分野への波及

残響制御技術はさまざまな分野に波及し、効果が示されています。

■音声をクリアにする技術

映画、TV番組制作のためのポストプロダクション*業界では、作品制作のためにさまざまな場所で音声を収録

* ポストプロダクション：放送やパッケージメディアなどの映像作品、映画の製作における撮影後の作業の総称。

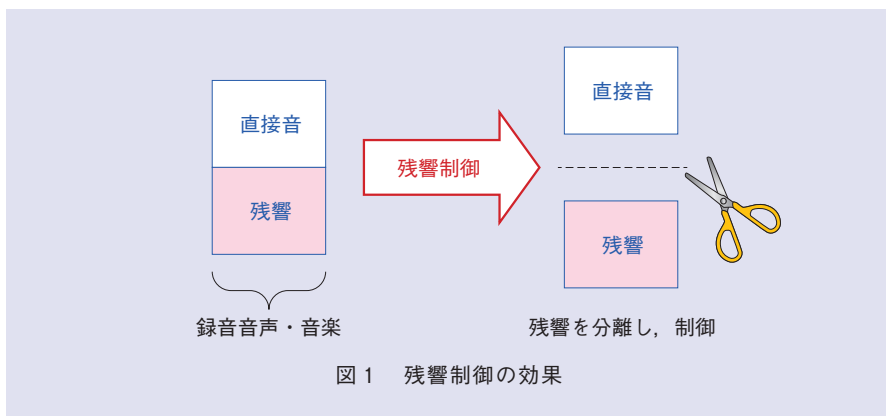


図1 残響制御の効果

します。その際に、収録音声には作品に不要な残響が含まれてしまうことがあります。このような状況に対処するため、残響制御技術は長らく登場を待ち望まれていた技術でした。現在、我々の技術はポストプロダクション業界向けのプロ用残響除去ソフトウェアとして、世界の多くの制作スタジオなどに導入され、日々作品制作の役に立っています。

電話会議においても、残響による音声の明瞭性低下が問題となることがあります。我々の技術を基に開発されたリアルタイム残響除去処理技術は、よ

りクリアな音声通信を可能とする次世代電話会議端末へも導入されようとしています。

また、残響除去処理技術は自動音声認識の性能向上にも寄与しています。2014年度に開催された残響下音声認識に関する国際評価企画では、約30の研究機関が参加する中、我々は残響除去処理技術を音声認識の前処理として用いることで、多くのタスクで最高性能を達成しました⁽²⁾。

■音楽をより豊かにする技術

残響制御技術は、録音されたステレオ音源から残響を取り出し、過去の名

演を元音源に忠実なサラウンド音楽としてよみがえらせることにも活用されています⁽³⁾。図3のように、コンサートホールの観客席で音楽を聴いている場合、我々の耳に到来する音は、音源から直接届く音成分（直接音）と、四方八方あらゆる角度から到来する残響の2つに大別されます。CDなどには、一般的に直接音と残響の両方が混ざって記録されています。このようなCDを従来のステレオスピーカを用いて再生すると、直接音と残響が両方とも前方から再生されてしまい、元の音響環境とは異なる音環境が作り出されてしまいます。しかし、我々の技術を用いて音楽信号を直接音成分と残響成分に分離し、図4のようなサラウンド環境にて、前方スピーカから直接音成分、後方スピーカから残響成分を再生すれば、元の録音環境に近い音環境を再構成することができます。

これまで、我々の開発した残響制御技術は、海外アーティストの過去のライブ録音をサラウンド音楽に変換して映画館などで配信する「再現ライブ」などで使われ、好評を博しています。また、このサラウンド化機能を搭載したホームオーディオ製品も他社から発売されています。

今後の展望

残響は、目的音源の位置とマイクの位置が離れている場合に、より顕著に観測される現象です。残響制御技術とは、どんなに離れたマイクで目的音を収録しても、収録後にその距離感を自在に操り、必要に応じて無効にしてしまう面白い技術です。残響制御技術の精度がさらに向上すれば、さまざまな可能性が広がります。例えば、電話や自動音声認識など、音によるコミュニケーションにおいて、現在のようにマ

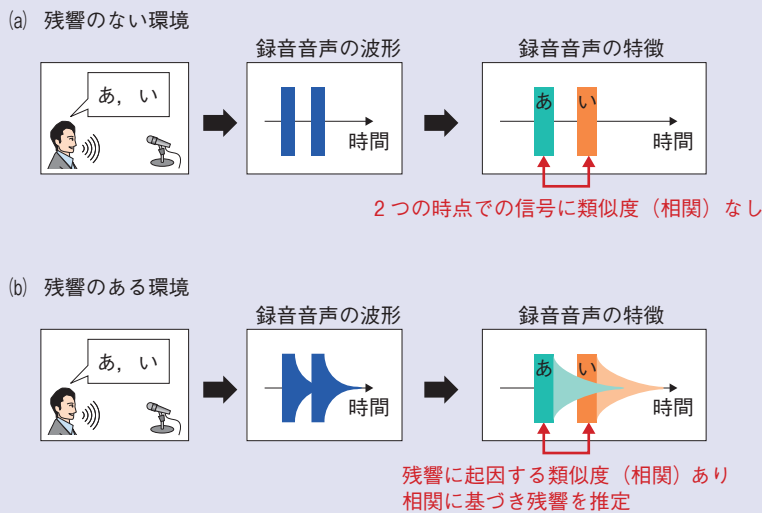


図2 残響を含まない音響信号と残響を含む音響信号の時間構造の違い

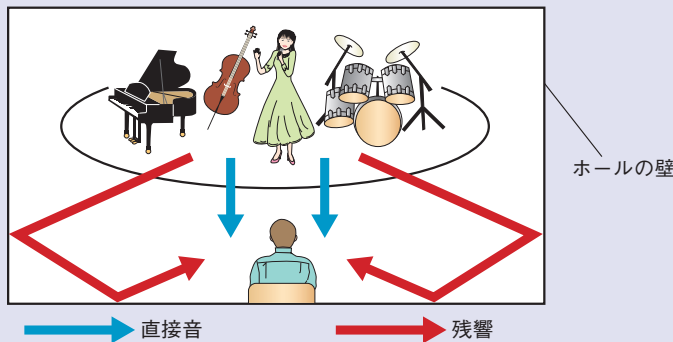


図3 ホールの客席に届く音

コラム

残響とエコー、残響除去とエコーキャンセルの違い

エコーとは、音声通信の中で生じる音響現象です。例えば、電話会議の状況を想像してみましょう。電話会議は、我々が自身の会議室内（近端会議室）で声を発し、それが近端会議室のマイクで収録されることから始まります。近端会議室で収録された音は、通信路を經由し、遠方の会議室（遠端会議室）にあるスピーカから拡声されることで、会議相手にメッセージとして伝わります。これが電話会議の基本的なメカニズムです。しかしこのとき、遠端会議室のスピーカから拡声された音は、会議相手の耳に届くだけでなく、遠端会議室にあるマイクにも届いてしまいます。結果、我々の声は遠端会議室、通信路を介して、近端会議室のスピーカに戻ってきて拡声されることとなります。この通信路と遠端会議室を經由して戻ってきた我々の声のエコーです。残響とエコーは、目的音源が生み出した音を基にした反射音である、という点では似た性質を持っています。

しかし、エコーキャンセルと残響除去には大きな違いがあります。エコーは、いったん近端会議室のマイクで収録されている音を基とした反射音です。そのため、近端マイクで以前に収録した音を参照すれば、比較的容易にエコーを推定・除去することができるようになります。このように、近端マイクで収録しておいた音を手掛かりとして反射音を除去するのが、エコーキャンセルです。

一方、残響除去では、残響の基となる音（目的音源が発した音）は一般的に知ることができません。つまり、残響を除去するために我々が用いることのできる信号は、残響をすでに含んでしまっている音のみであり、我々はその中から残響成分を正確に推定し、抑圧しなければなりません。響き成分の基となる音を、参照して響きを打ち消すことができるのがエコーキャンセル、参照せずに響きを打ち消すのが残響除去です。

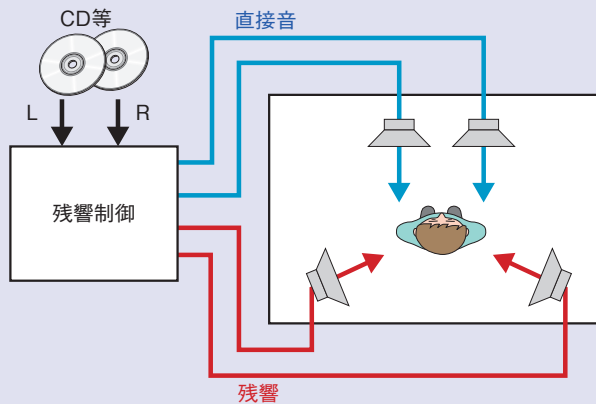


図4 残響制御を用いたサラウンド化再生

イクの位置を気にする必要はなくなり、好きなときに好きな場所で自由に楽しめるようになるでしょう。

また、手元にある音楽ソースをサラウンド化して再生すれば、家にいながらにして、その音楽を収録したライブ会場や音楽ホールで音楽を聴くのと全く同じ音環境を体験することもできるようになります。これらの応用を含むさまざまな場面で、残響制御技術は音テクノロジーの根幹を支える技術となり、これから先の音の未来を切り拓く一助となっていくことでしょう。

■参考文献

- (1) K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi : "Suppression of Late Reverberation Effect on Speech Signal Using Long-Term Multiple-step Linear prediction," IEEE Trans. Audio, Speech, Lang. Proc., Vol. 17, No. 4, pp.534-545, 2009.
- (2) M. Delcroix, T. Yoshioka, A. Ogawa, Y. Kubo, M. Fujimoto, N. Ito, K. Kinoshita, M. Espi, T. Hori, T. Nakatani, and A. Nakamura : "Linear prediction-based dereverberation with advanced speech enhancement and recognition technologies for the REVERB challenge," Proc. REVERB challenge workshop, May 2014.
- (3) K. Kinoshita, T. Nakatani, and M. Miyoshi : "Blind upmix of stereo music signal using multi-step linear prediction based reverberation extraction," In Proc. of ICASSP2010, pp.49-52, March 2010.



木下 慶介

音声・音楽信号処理技術は日々進歩しており、数年前には不可能であったことも、今や実用レベルに到達しているものも多くあります。残響制御技術やその周辺技術の最先端動向にご興味のある方は、お問合せください。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
メディア情報研究部
信号処理研究グループ
TEL 0774-93-5322
FAX 0774-93-5158
E-mail kinoshita.k@lab.ntt.co.jp