

人の声と背景音のボリュームバランスを調整する 音源分離応用技術のTVへの適用

Technology to Control Volume Balance between Voices and Background Sounds
for TV Products Based on Sound Source Separation Technology

天田 皇 竹内 広和

■ AMADA Tadashi ■ TAKEUCHI Hirokazu

デジタルTV(テレビ)の発展に伴い、TVの音声信号に対して複雑なデジタル信号処理を行い、ユーザーの好みや視聴シーンに合わせた音質補正を実現する音響処理技術が用いられるようになった。

東芝は、新たに音源分離技術を応用した音響処理技術を開発した。この技術は、人の声と背景音を分離しそれぞれのボリューム(音量)を独立に調整してバランスを再構成できるようにするものであり、ユーザーは好みのボリュームバランスで番組を視聴できる。新興国向けに商品化したTV L4300にこの技術を適用した。

The progress made in the field of digital TVs in recent years has led to advances in complex digital signal processing technologies for audio signals. One such technology is sound compensation technology to provide sound quality matching the preferences of viewers and the scenes of TV programs.

Toshiba has developed a new sound source separation technology that allows viewers to watch a variety of TV programs with their preferred volume balance, which is achieved by separating human voices and background sounds and independently adjusting the volume balance for each. We have also launched TV products equipped with this new technology.

1 まえがき

デジタルTVの発展に伴い、TVの音声信号に対して複雑なデジタル信号処理を行い、様々な補正や効果を生み出す音響処理技術が用いられるようになった。例えば、スピーカやアンプの性能を最大限に引き出すためのイコライジング技術⁽¹⁾やユーザーが好みに応じて音質を変える音響効果技術などである。後者には、サラウンド、低音強調、ボリューム正規化など、様々な音響効果を生む技術があり広く普及しているが、大衆化しすぎていて、特徴ある商品としてアピールしづらいという課題がある。

東芝は、新たな音響効果として音源分離を応用した音響処理技術を開発した⁽²⁾。この技術は、人の声と背景音を分離しそれぞれのボリュームを独立に調整してバランスを再構成する。これにより、ユーザーは好みのボリュームバランスで番組を視聴できる。当社は、この技術をパソコンやタブレットに既に適用しているが、現在TVへの適用を始めている。

ここでは、音響分離を応用した音響処理技術とTVへの実装にあたっての工夫について述べる。

2 音源分離を応用したTV向け音響処理技術

TVの音響効果技術として、フィルタ処理などにより音質を補正する技術が広く用いられている。例えば、周波数特性をユーザーの好みに応じて変える方式や、ステレオ信号の差分

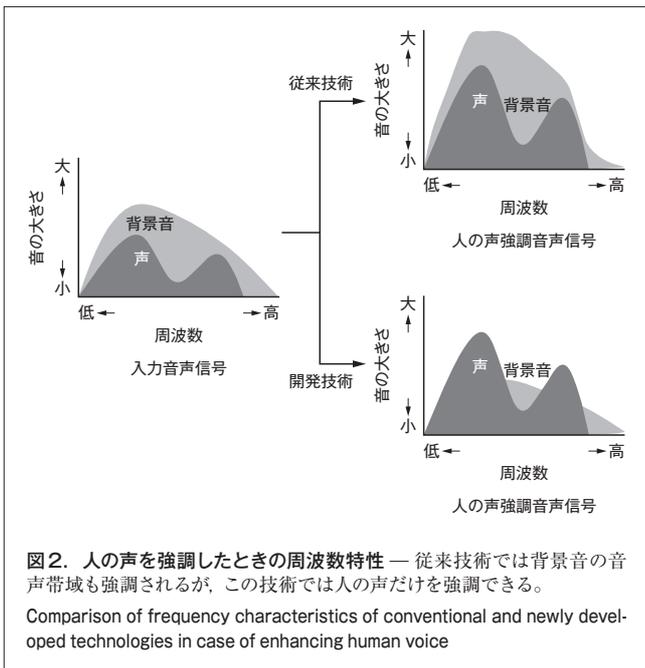
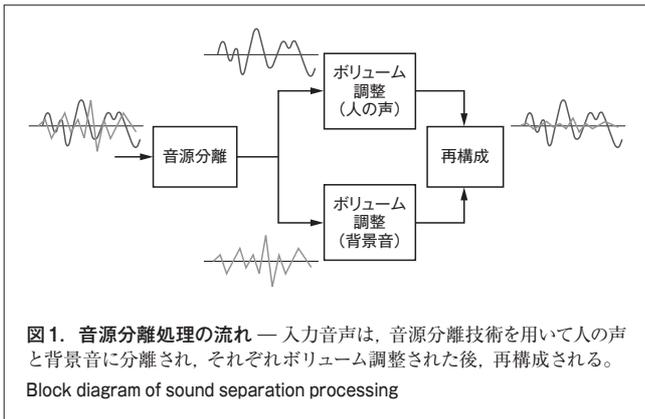
成分を強調しステレオ感を増加させる方式、反対に同相成分を強調し中心付近に定位するせりふを強調する方法などが挙げられる。これらは比較的少ない演算リソースで実現可能であり、TVのように限られた演算リソースでリアルタイム処理を行うシステムに向いている。

一方、特定の信号だけに効果を出すなど複雑な処理には向いていない。例えば、人の声を強調する場合、背景音と独立に制御することは難しく、人の声を強調すると背景音も含めて全体の音質が変わって聞こえる場合がある。

今回開発した技術は、音源分離技術をTV向けの音響効果として応用したものである。音源分離技術は近年の音響技術の一分野として研究が盛んであり、雑音除去や話者分離など、主に音によるコミュニケーションを支援する技術として注目されている⁽³⁾。音源を人の声と背景音に分離することで、人の声と背景音を独立に制御できるようになる。

入力信号を分離し再構成するまでの処理手順を図1に示す。入力信号は音源分離により人の声と背景音に一度分離され、それぞれボリュームを変えて再構成される。例えば、スポーツ番組などで実況の声が気になり内容に集中できない場合に、実況の声を控えめにして再構成することで、臨場感を高めることができる。反対に映画を見ているときに背景音が気になりせりふが聞き取りづらい場合、背景音のボリュームを下げることでせりふの聞き取りを助ける効果も期待できる。

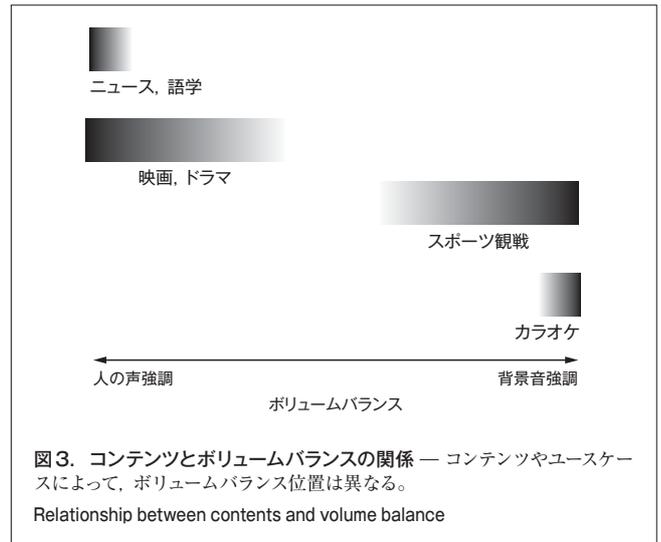
従来のイコライザを用いて人の声を強調する場合と、この技術を用いた場合の違いを、周波数特性の模式図を用いて図2



に示す。従来技術では人の声とともに背景音も強調されるが、この技術では人の声だけを分離できるため独立に強調できる。

このように、入力信号を人の声と背景音に分離した後にそれぞれのボリュームを変えて再構成することで、従来技術では実現が難しい音源別の強調効果を実現できる。このため、ユーザーは好みや視聴シーンに応じて自由にボリュームバランスを変えられることができる。

コンテンツの種類とボリュームバランスの例を図3に示す。中心よりも右側ほど人の声が抑圧され、左側ほど背景音が抑圧されることを示している。音楽コンテンツに対しては、背景音に対して人の声のボリュームを下げることでカラオケのような効果を得ることができる。スポーツの実況中継では、解説者の声を抑え臨場感を高める効果が得られる。逆に、背景音を下げる使い方では、映画やドラマのせりふと聞き取り支援のほか、ニュースや語学番組など声を明瞭にする使い方も考えられる。更に、深夜など、ボリュームが上げにくい状況でも、聞き



たい音だけ強調することで全体のボリュームを上げる必要がなくなり、周囲への迷惑を軽減できる。

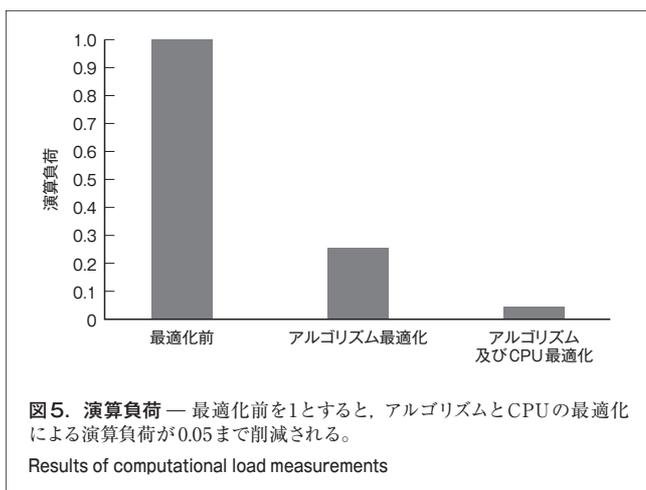
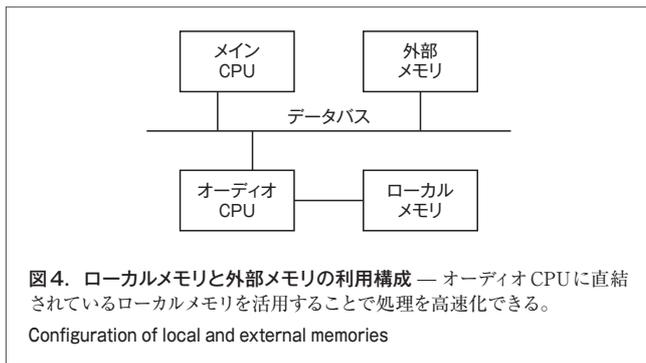
3 TVへの実装

3.1 オーディオCPU向け最適化

TVにはメインCPUのほかに複数のプロセッサが搭載されている。メインCPUの処理能力は高まってきており音源分離技術の実装も難しくはない。一方で、音響信号処理を専門に行っているオーディオCPUは、メインCPUほどの処理能力はなくメインCPUに比べて実装が難しい。オーディオCPUの利点の一つとして、映像と音声の同期を維持し、様々な入力ソースに対して遅延なく処理できるように設計されている点が挙げられる。今回は、リアルタイム動作を実現するため、オーディオCPUに実装した。

実装にあたっては、アルゴリズムの最適化と、CPU固有の機能を用いた最適化を行い、演算負荷を削減した。アルゴリズムの最適化では、メモリアクセスや分岐の少ない計算方法への変更、複雑なデータ構造の排除、スタックを消費しない構造への変更などを行った。CPU固有の最適化では外部メモリとは別にオーディオCPU内部に存在する高速なローカルメモリの活用や、積和命令などの拡張命令を利用し少ない命令数で計算結果を得る工夫を行った。ローカルメモリは作業用のメモリとして使い、演算途中の結果はここに一時的に保持する。演算終了後に結果を外部メモリに書き戻すことで、低速な外部メモリへのアクセス回数が減り、演算負荷が下がる。

ローカルメモリと外部メモリの利用構成を図4に示す。外部メモリは共有のデータバスに接続されているため他のプロセッサとアクセスが競合するが、ローカルメモリはオーディオCPU専用のため競合することはない。ローカルメモリの利用は、データバス競合による負荷増加というシステム動作時の問題を



軽減する効果もある。

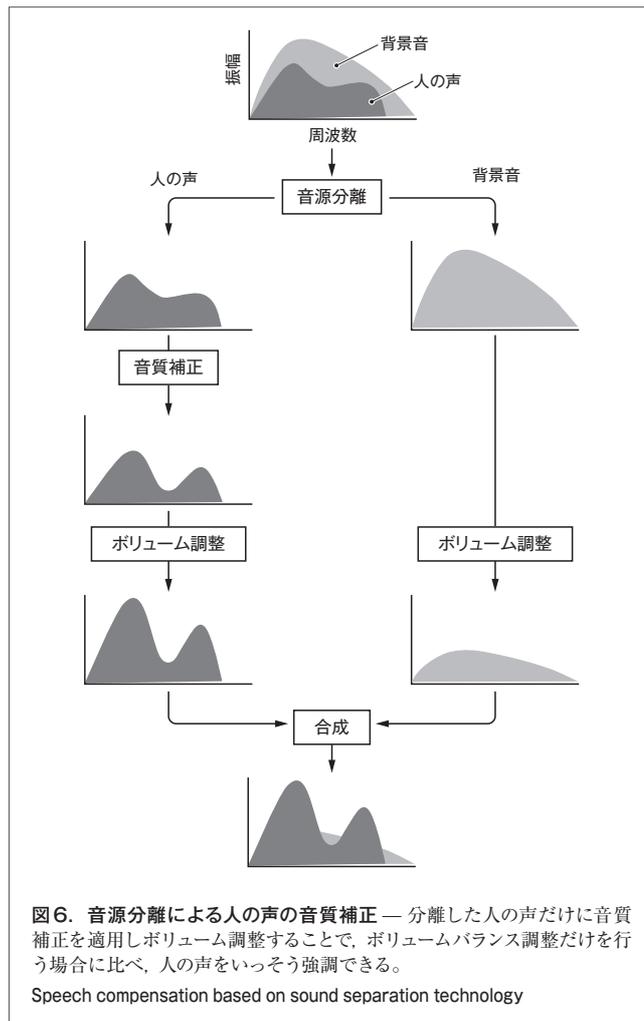
最適化による演算負荷の軽減効果を図5に示す。最適化前を1とすると、アルゴリズム最適化を行った場合は0.25、これに加えてCPU最適化を行った場合0.05まで演算負荷が軽減され、処理速度が20倍に高速化された。

3.2 声の聞きやすさの向上

音源分離による効果を更に高めるため、音源ごとのボリューム調整だけでなく、音源ごとの音質補正を行うことも可能である。音質補正を用いて人の声を強調する場合の処理手順を図6に示す。音源分離を行うことで、人の声だけを補正することが可能であり、ボリュームバランス調整だけの場合に比べ、更に人の声を際立たせることができる。従来の音響効果処理では人の声だけに選択的にフィルタをかけることは難しいが、この技術では音源分離技術を用いることで背景音に影響を与えずに人の声を聞きやすくなる。

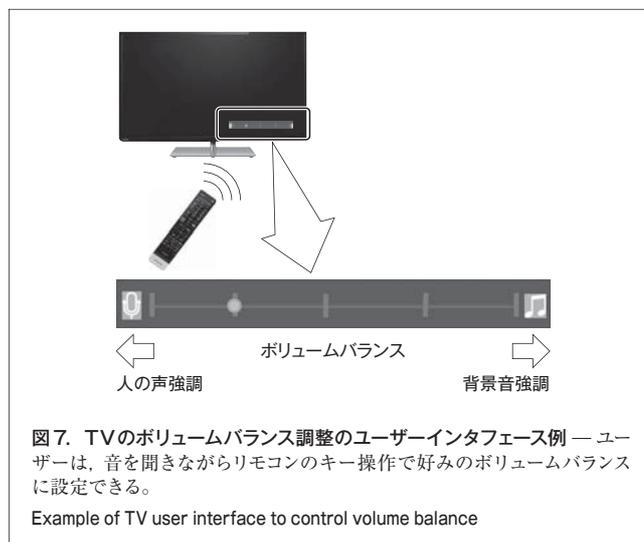
3.3 背景音の音質補正

背景音も人の声とは独立に音質補正が可能である。人の声とは異なり、背景音の種類は様々であるが、音源分離により低域のボリューム感が低下する場合があります、これを回復する補正などが例として挙げられる。人の声の強調と同様に、分離信号の背景音成分だけに背景音補正フィルタを適用することで人の声には影響を及ぼさずに背景音を補正できる。



3.4 ユーザーインターフェース

ボリュームバランスを調整するためのユーザーインターフェースの例を図7に示す。TVのリモコンキーの操作でバランス値を左右に動かし、人の声と背景音のバランスを簡単に調整するこ



とができる。ユーザーが指定したバランス値はオーディオCPUにパラメータとして瞬時に送られ、処理に反映されるので、ユーザーは音を聞きながらバランスを調整できる。

4 製品への適用例

この技術を適用した新興国向けTV L4300の外観を図8に示す。L4300はAndroidTM(注1)を搭載したTVである。このTVでは、通常のリモコンに加えてポインティングリモコンが利用可能である。図7に示したボリュームバランス調整用インタフェースを操作する場合、通常リモコンではキーを連続して押すことでバランスの変更を行うが、ポインティングリモコンではマウスを操作する感覚で好みのバランスをワンクリックで指定できる。

この技術は、TV放送だけでなく、メディアプレーヤや動画配信アプリケーションなどに対しても有効である。



(注1) Androidは、Google Inc.の商標。

5 あとがき

人の声と背景音のボリュームバランスを調整する音源分離を応用した音響処理技術のTVへの適用について述べた。人の声を強調することでせりふを聞きやすくしたり、背景音を強調することでスポーツ番組の臨場感を盛り上げたりするといった使い方を提案した。この技術は音源分離技術をTVのユースケースに合わせて応用した技術であるが、ユーザーインタフェースも含め他のユースケースあるいは他製品での適用可能性がある。今後、更なる改良を行い他製品への適用も検討していく。

文献

- (1) 山本敏文. 液晶テレビ用スピーカの音質補正技術. 東芝レビュー. 68, 5, 2013, p.56 - 57.
- (2) 広畑 誠 他. 人の声と背景音のボリュームバランスを調整する音源分離技術. 東芝レビュー. 68, 9, 2013, p.26 - 29.
- (3) 牧野昭二 他. ブラインドな処理が可能な音源分離技術. NTTジャーナル. 15, 12, 2003, p.8 - 12.



天田 皇 AMADA Tadashi

パーソナル&クライアントソリューション社 ライフスタイルソリューション開発センター エンベデッドソフトウェア技術開発部主務。音響信号処理技術の開発に従事。日本音響学会、IEEE会員。Lifestyle Solutions Development Center



竹内 広和 TAKEUCHI Hirokazu

パーソナル&クライアントソリューション社 ライフスタイルソリューション開発センター エンベデッドソフトウェア技術開発部参事。音響信号処理技術の開発に従事。Lifestyle Solutions Development Center