

メディアを高品質で提示する 映像高画質化技術および音声合成技術

Video Improvement and Speech Synthesis Technologies for High Quality Media Output

影山 昌広 中嶋 満雄 谷田部 祐介
Kageyama Masahiro Nakajima Mitsuo Yatabe Yusuke
永松 健司 孫 慶華
Nagamatsu Kenji Sun Qinghua

日立グループは、映像メディアがもともと持っているポテンシャルを最大限に引き出すための代表的な高画質化技術として、超解像技術、映像圧縮ノイズ除去技術、局所コントラスト補正技術を開発している。また、音声合成技術では、機械学習技術によって最適なデータを自動学習することが可能となり、従来の人手による調整では困難であった日本語の細かな韻律特徴の考慮や話者の話し方の癖なども再現が可能である。

1. はじめに

ネットワークやストレージなどに代表されるデジタル技術の進展に伴い、身の回りには大量のメディアがあふれるようになった。これらの情報を機械から取り出して最終的に処理するのは人間であり、人間にストレスを与えない高品質なメディア提示技術の実現が望まれている。

ここでは、メディアを高品質で「見せる・示す」技術として、映像や画像を高画質化する技術と、音声を合成して出力する技術にスポットを当て、日立グループが開発している最先端技術について述べる。

2. 映像高画質化技術

近年、家庭やオフィスなどで用いられる映像（画像）の表示装置の進展が目覚ましく、大画面化、高精細化、高輝度・高コントラスト化、高い色再現能力などにより、さまざまな映像メディアをきめ細かく鮮やかに表示できるようになった。その一方で、もともと品質の低い映像はあらが目立ちやすく、ぼやけ感やノイズなどによって、平板な印象を受けることも多い。そこで、映像メディアがもともと持っているポテンシャルを最大限に引き出すために、信号処理によって映像を高画質化する技術がますます重要になってきている。

日立グループは、代表的な高画質化技術として、超解像

技術、映像圧縮ノイズ除去技術、局所コントラスト補正技術を開発している。

2.1 超解像技術

解像とは、像が分かれて見えることを指し、解像度は、解像できる度合い（どこまで精細に像が分かれて見えるか）を表す指標である。ここで述べる超解像技術は、撮影から蓄積・伝送・表示に至る全体の過程で定まってしまう解像の限界（ボトルネック）を超えて、より精細な映像を生成する信号処理技術である。

映像をデジタル化（離散化）する際に、1画面を構成する個々の画素よりも細かいテクスチャ（模様）は、「折り返し歪（ひずみ）」と呼ばれる干渉縞（じま）状のノイズになることが知られている。このノイズの発生を防ぐためには、カメラの撮像素子に光学ローパスフィルタを取り付けたり、拡大表示のような画素数変換の際にローパスフィルタ処理を行ったりして、折り返し歪の原因となる高精細成分を離散化の前に減衰しておくことが一般的である。しかし、高精細成分の減衰は解像度の低下につながるため、折り返し歪がわずかに発生してしまっていることが多い。

一般に、いったん発生してしまった折り返し歪は、映像にもともと含まれている成分と見分けがつかず、これらを分離することは困難とされてきた。

一方、日立グループは、映像中のテクスチャが一定の条件を満たせば折り返し歪を分離して逆変換できることを見だし、映像信号を実数と虚数のデータの組に変換して高度な処理を行う「複素信号処理」¹⁾などを利用して、高精細なテクスチャを再現する超解像技術を開発した（図1参照）。

日立グループでは、これまでに超解像技術をデジタルテレビ^{2), 3)}などに応用してきた。開発当時は、リアルタイ



図1 | 超解像技術

映像信号の中に含まれる「折り返し歪（ひずみ）」を分離して逆変換することにより、高精細なテクスチャを再現する。

ムで超解像処理を実現するために、専用のハードウェアが必須であった。その後、アルゴリズムの改良などによって、処理に必要な演算量を約1/10に削減し、現在では、一般的なPC (Personal Computer) 上で実行するソフトウェアにより、VGAサイズ（640×480画素）の映像を1秒間に30フレーム以上の速度で余裕を持って処理できるまでに軽量化している。このような取り組みを足がかりとして、映像監視や医療分野をはじめとする幅広い応用展開をめざしている。

2.2 映像圧縮ノイズ除去技術

映像データをコンパクトに蓄積したり伝送したりするためには、符号化による圧縮（データ量の削減）が不可欠である。また、映像を高画質に再生するには、圧縮の過程で生じるノイズ（モスキートノイズとブロックノイズ）の除去が重要な課題となっている（図2参照）。

従来の映像圧縮ノイズ除去技術では、画面全体に一律の平滑化フィルタ処理を行っていたため、再生映像がぼやけてしまうことが課題であった。そこで、被写体のテクスチャが持つ特徴量や圧縮ノイズ特有の周波数成分を解析することにより、ノイズが目立ちやすい領域を高精度で予測し、その領域だけに絞ってノイズ除去することにした。こ

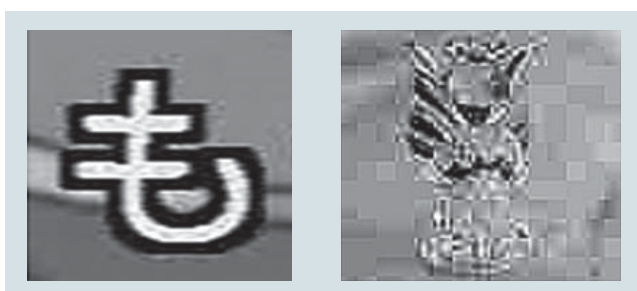


図2 | モスキートノイズ（左）とブロックノイズ（右）

モスキートノイズは、輪郭周辺に発生するジリジリとしたノイズであり、ブロックノイズは、矩形（くけい）のブロック状のノイズである。

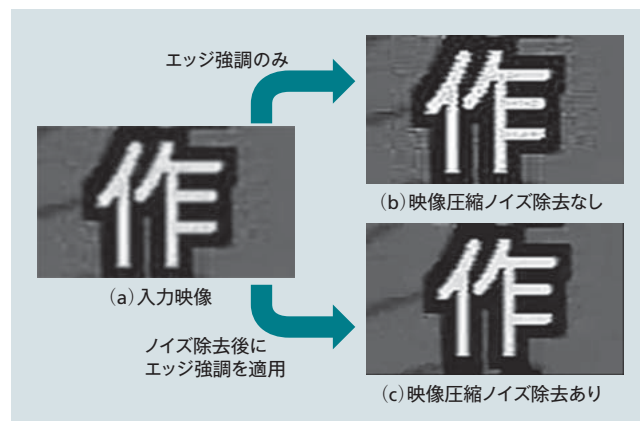


図3 | 映像圧縮ノイズ除去技術

映像圧縮ノイズ除去技術とエッジ強調技術を併用することで、ノイズの強調を抑えた良好な再生画像が得られている。

れにより、従来技術では擬似輪郭を発生させてしまうような微小なノイズも効果的に除去できるようになり、従来は困難であった精細感保持とノイズ除去の両立が可能となった⁴⁾。この技術とエッジ強調技術を併用することで、ノイズの強調を抑えた良好な再生画像が得られている（図3参照）。

これまでに、この技術をハードウェアとして実装し、デジタルテレビに搭載した²⁾。その際、フレームメモリが不要な面内演算処理とするとともに、リアルタイムで入力される画素データを格納するラインメモリの削減技術⁵⁾も適用し、低コスト化を実現した。

2.3 局所コントラスト補正技術

従来、映像表示装置のダイナミックレンジ不足を補う技術として、例えば、画面全体の輝度分布に基づいて表示階調を制御するコントラスト補正技術があった。しかしこの技術は、入力する映像によっては部分的に階調劣化などの副作用が生じてしまう課題があり、大きな補正効果を得ることが困難であった。

そこで日立グループは、人間の視覚特性を利用したコントラスト補正の開発を進めている。この技術は、「人間の目は照明光に依存せず、物体の表面反射率に基づいて物体を知覚することができる」というRetinex理論^{6), 7)}を応用したもので、局所的な映像補正を行うことによって、被写体の質感やコントラスト感を高める。従来技術と比べて補正が局所的であるため、階調劣化などの副作用も生じにくく、比較的大きな補正効果が得られる。

局所コントラスト補正技術では、まず、Retinex理論に基づいて映像の反射光成分を推定する。次に、推定した反射光成分に対して強調と再合成処理を施すことで局所的コントラスト補正映像を得る。その際、映像の特徴に応じて強調と再合成を制御することで、質感やコントラスト感を

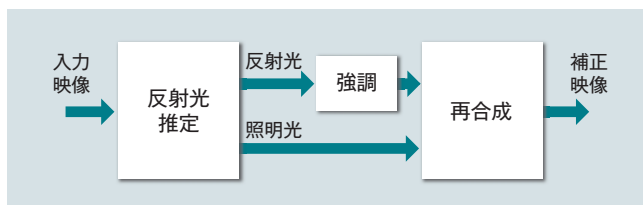


図4 | 局所コントラスト補正技術の処理の流れ

Retinex理論を応用した反射光推定と人間の視覚特性に基づく反射光成分の強調・再合成により、視認性のよい映像を得る。



図5 | 局所コントラスト補正技術

ガラスの反射光成分を強調することで、本来のガラスらしい光沢を持った映像を再現できる。

向上し、良好な視認性を得る（図4参照）。

この技術によって補正した映像では、例えばガラスの反射光成分を強調することで、本来のガラスらしい光沢を放った映像を再現することができる（図5参照）。

現在、上述した局所コントラスト補正技術のプロジェクトへの適用を検討している。プロジェクトは使用場所の明るさによって視認性が大きく左右され、特に明るい場所での視認性向上が効果的と考えられる。今後、ほかの映像機器への適用も検討するとともに、さらに効果的な視認性向上技術の研究を進めていく。

3. 音声合成技術

音声合成技術とは、入力テキストを音声データに変換し、コンピュータがテキストを読み上げる技術である。

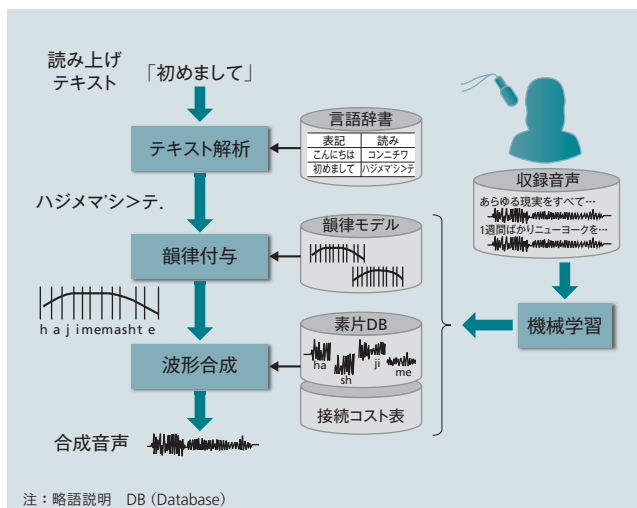
近年、カーナビゲーションシステムやスマートフォンでの音声技術の利用が進み、それまで福祉分野などでの応用が多かった音声合成技術が広く社会で利用されるようになってきた。

3.1 音声合成技術の概要

音声合成技術は、大きく分けて、テキスト解析、韻律付与、波形合成の三つの処理から構成される（図6参照）。

テキスト解析とは、入力テキストから音節列への変換である。アクセントや無声音化などの発音情報もここで決定される。

韻律とは、声の強弱、リズムや抑揚を指す。韻律付与では、韻律モデルを参照し、音節列をどのようなリズム・抑



注：略語説明 DB (Database)

図6 | 音声合成技術の処理の流れ

音声合成技術は、大きく分けて、テキスト解析、韻律付与、波形合成の三つの処理から構成される。

揚で読み上げるかを決定する。

波形合成では、各音節に対応する音声部品（素片）を素片DB (Database) から取得し、音のつながりのよさ（接続コスト）を考慮して最適な組み合わせを選択し、指定された韻律になるように変形・結合することで合成音声を生成する。

これらの処理で用いるデータ（韻律モデル、素片DB、接続コスト）は、従来、収録音声を基に人手での特徴量選択やチューニング作業を経て作成していた。しかし、機械学習技術によって最適なデータを自動学習することが可能となり、高品質な音声合成を実現できるようになってきた。

3.2 統計的機械学習を用いた韻律モデリング技術

機械学習の一つである「決定木手法」では、さまざまな特徴量が付与された大量の数値を学習データとし、ある特徴量リストが与えられたときに最も精度よく数値を予測できる構造（決定木）を自動構築する（図7参照）。

韻律付与では、大量の収録音声から学習された決定木（韻律モデル）に、入力テキストに対応する特徴量リストを入力して、音素の継続長や基本周波数を予測する。

この機械学習により、従来の人手による発見的な特徴量選択では困難であった日本語の細かな韻律特徴の考慮や話者の話し方の癖なども再現できるようになった。

また、機械学習の活用には、朗読調や会話調など、発話スタイルを変えて音声データを収録するだけで、新たな発話スタイルへの対応が可能になる利点もある。日立グループでは、現在、カーナビゲーションシステムやスマートフォンでの音声対話技術をターゲットに、会話調の韻律モデル構築を進めている。

音節列：ハジメマシ>テに対応する特徴量のリスト

音素	...	m	a	sh	t	e
前の音素	...	e	m	a	sh	t
後の音素	...	a	sh	t	e	—
アクセント句の音節数	...	5	5	5	5	5
アクセント高低	...	H	H	L	L	L
...	...					

継続長
60 ms

継続長を予測する決定木

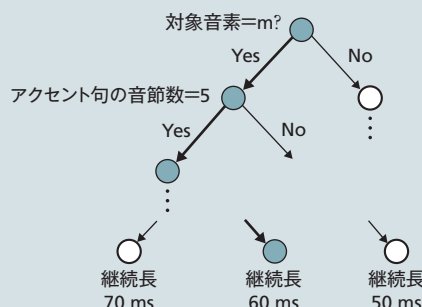


図7 | 決定木手法による韻律情報の予測（継続長予測の例）

さまざまな特徴量が付与された大量の数値を学習データとし、ある特徴量リストが与えられたときに最も精度よく数値を予測できる構造（決定木）を自動構築する。

3.3 セレクティブ重畳型波形合成技術

波形合成には、波形重畳と波形接続という二つの手法がある。波形重畳では、PSOLA法（Pitch Synchronous Overlap and Add）と呼ばれる信号処理を用いて素片の韻律を変形させる。従来のように素片DBが小さく、目標韻律に近い素片がない場合にこの手法が用いられる。一方、波形接続は、素片に韻律変形を施さず、スムージング処理によって結合するだけで合成音声とする手法である。

波形重畳は、目標韻律に一致した音声を合成できる半面、こもり感などの音質劣化が生じる。波形接続は、目標韻律に合致した素片が見つからない場合に不連続感などの音質劣化につながるという課題があった。

日立製作所は、この二つの手法を動的に切り替えて使い分けるセレクティブ重畳型波形合成技術を開発した。この技術は、目標韻律との差や隣接する素片間の韻律の差を基に、波形重畳で韻律をどの程度変形するか、あるいは変形なしの波形接続とするかを素片ごとに決定する。

セレクティブ重畳型波形合成技術の開発で、合成音声の肉声感が大幅に向上した。前述の韻律モデルの改良による自然性の向上とともに、日立音声合成の音質を大きく改善する要因となっている。

今後は、手法切り替えのしきい値や音のつながりのよさ

の判定などに機械学習手法を応用していく⁸⁾。これにより、アナウンサーの声質をさらに生かせる音声合成技術の実現をめざしている。

4. おわりに

ここでは、メディアを高品質で提示する映像高画質化技術および音声合成技術にスポットを当て、これらの最新技術に関する日立グループの取り組みについて述べた。

人にストレスを与えないメディア提示技術は、社会やビジネスのさまざまな分野に応用できると考えている。

参考文献

- 1) 大西、外：複素信号処理による周波数変換法の一検討、昭和63年電子情報通信学会秋季全国大会講演論文集A-42（1988.9）
- 2) 中嶋、外：Wooo高画質映像表示技術の進化、日立評論、92、10、736～741（2010.10）
- 3) 鈴木、外：デジタルテレビ「Wooo」の高画質、エコへの取り組みより豊かな生活に貢献する商品めざして、日立評論、93、10、683～687（2011.10）
- 4) 谷田部、外：1080/60p対応画像圧縮ノイズ除去技術、映像情報メディア学会年次大会講演予稿集 19-4（2010.8）
- 5) Y. Yatabe, et al. : A low cost noise reduction technology for compressed video, IEEE International Conference on Consumer Electronics 530 - 531（2012.1）
- 6) E. H. Land, et al. : Lightness and Retinex Theory, Journal of the Optical Society of America, Vol. 61, pp. 1-11（1971.1）
- 7) 坂庭、外：視覚特性を利用した映像の光沢部分抽出手法、2012年映像情報メディア学会年次大会予稿集 21-6（2012.8）
- 8) 孫、外：統計的モデルを用いた波形接続方式音声合成における分割学習によるモデル構築法、日本音響学会秋季研究発表会講演論文集 3-Q-24（2012.9）

執筆者紹介



影山 昌広

1987年日立製作所入社、中央研究所 情報システム研究センター 知能システム研究部 所属
現在、画像・映像処理応用システムの研究開発に従事
博士（工学）
映像情報メディア学会会員



中嶋 満雄

1981年日立製作所入社、横浜研究所 組込みシステム研究センター 組込みソリューション研究部 所属
現在、映像処理技術の研究開発に従事



谷田部 祐介

2001年日立製作所入社、横浜研究所 組込みシステム研究センター 組込みソリューション研究部 所属
現在、画像圧縮伸張・画像処理の研究開発に従事
映像情報メディア学会会員



永松 健司

1998年日立製作所入社、中央研究所 情報システム研究センター 知能システム研究部 所属
現在、音声合成システムの研究開発に従事
博士（工学）
電子情報通信学会会員、日本音響学会会員



孫 慶華

2008年日立製作所入社、中央研究所 情報システム研究センター 知能システム研究部 所属
現在、音声合成システムの研究開発に従事
博士（工学）
日本音響学会会員