

テレビの音響・音声技術

平野 仁* 久保めぐみ**
古田 訓*
伍井啓恭*

Acoustic and Speech Processing Technologies for Television Receiver

Jin Hirano, Satoru Furuta, Hiroyasu Itsui, Megumi Kubo

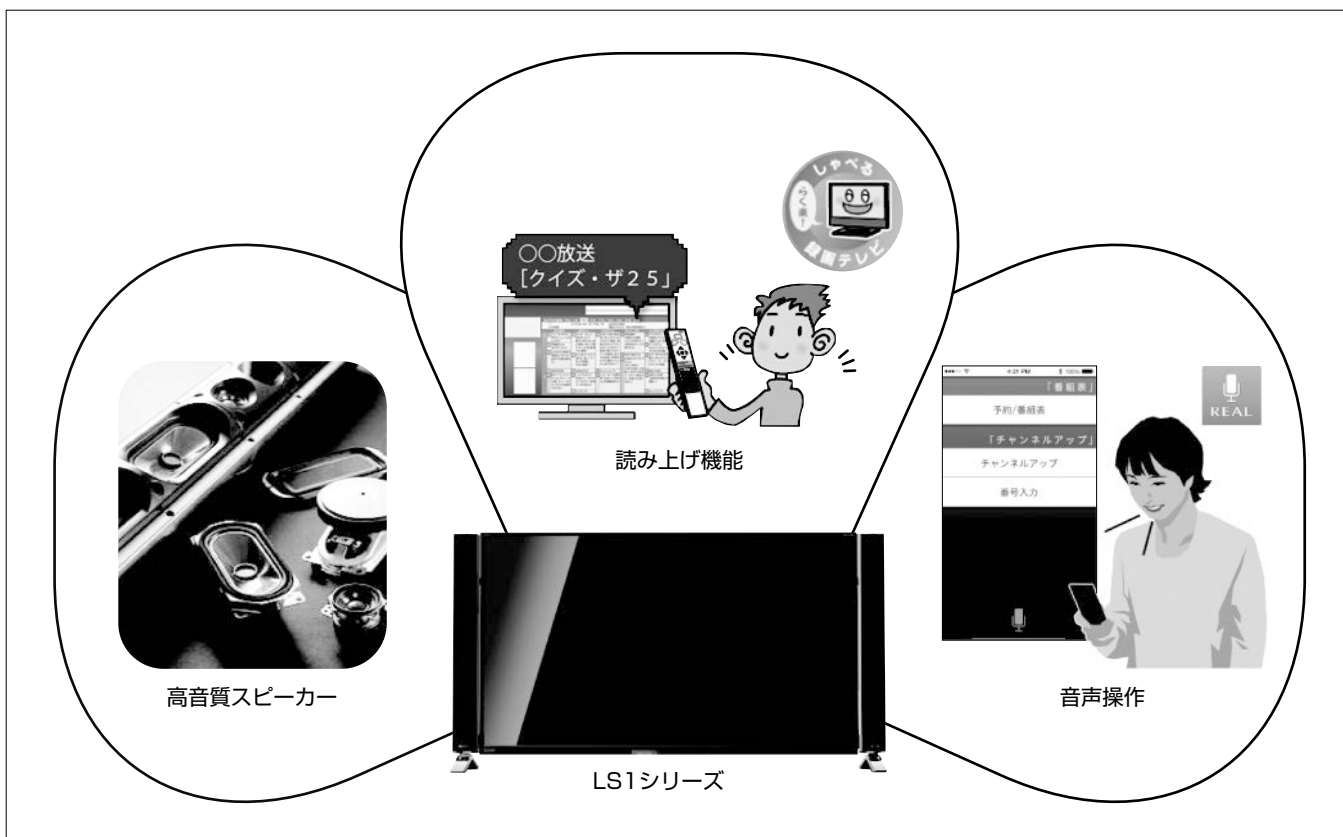
要 旨

4K実用放送が開始され、フルハイビジョンから4Kへとテレビの高画質化が進んでいる。画質の進化に伴い、映像にふさわしい高品位な音が求められるようになってきている。また、新しいサービスや機能の追加によって高機能化が進む反面、操作は複雑になり、誰でも簡単に使いこなせる使いやすさへの配慮が必要となっている。三菱電機が2014年に発売した4K対応レーザー液晶テレビ“LS1シリーズ”では、音響・音声技術を駆使して、これらの課題改善に取り組んだ。

高品位な音を実現するため、高音質スピーカーシステムを搭載した。高伝搬速度と高内部損失を両立させた当社開発素材を振動板に採用したスピーカーと信号処理の複合設計によって、高音質化を実現している。

使いやすさを向上させる音声技術として、読み上げ機能と音声操作を搭載した。読み上げ機能では、自然なイントネーションと高域の明瞭性を改善した合成音声によって、操作メニューや電子番組表などを聴き取りやすい音声でガイドすることができる。音声操作では、音声認識によってテレビを操作する携帯端末用の当社専用アプリケーションを開発した。特殊なコマンドを覚える必要はなく、ユーザーの日常的な発話や言い回しでテレビを操作できるようにした。

今後も更なる高画質化が期待されるが、当社ならではの音響・音声技術によって、画質だけではなく“美しい音”と“使いやすさ”にもこだわった開発を進める。



三菱液晶テレビ“LS1シリーズ”を支える音響・音声技術

当社の4K対応レーザー液晶テレビLS1シリーズに搭載している音響・音声技術を示す。当社製テレビの高音質化と使いやすさは、様々な音響・音声技術によって支えられている。

1. ま え が き

当社製テレビは、高音質、高画質、使いやすさを特長としている。本稿では、2014年に発売した当社初の4K対応液晶テレビLS1シリーズに盛り込んだ高音質化技術として高音質スピーカーシステムを、使いやすさを向上させる技術として読み上げ機能と音声操作を述べる。

2. 高音質スピーカーシステム

4Kテレビ市場では、高画質と同時に高音質を訴求した製品が発売され、音質が競争軸の1つとして再認識されている。当社製テレビでは、一貫して音質を重視した設計をしている。LS1シリーズのスピーカーシステム設計では、当社開発の素材“NCV(Nano Carbonized high Velocity)”のスピーカー振動板への採用、スピーカーと信号処理の複合設計、低音信号処理によって高音質化を実現している。

2.1 NCV振動板

スピーカーの振動板として理想的な素材は、音の伝搬速度が速く、内部損失が大きいものとされているが、その2つの物性を両立させることは難しい。当社開発の素材であるNCVは高伝搬速度、高内部損失の両立を実現しており、スピーカー振動板の材料として理想的な物性となっている(図1)。

伝搬速度とは、材料内を振動が伝わる速さを示し、材料の剛性が高いほど速く、振動に対する追従性がよい。内部損失は、材料内部での振動エネルギーの損失を示し、材料を叩(たた)いた際に鳴る音の響きが短いほど内部損失が高く、余計な響きを持たない振動板となる。

高音用・低音用スピーカーは異なる振動板材料を使うのが一般的であり、金属・セラミック等の伝搬速度の高い材料は高音用スピーカーに、紙や樹脂などの内部損失の高い材料は低音用スピーカーに用いられることが多い。

NCVは、高伝搬速度と高内部損失を兼ね備えているため、低音から高音までを同じ材料の振動板でカバーすることができ、全帯域で統一感のある音を再生することが可能となる。

2.2 スピーカーと信号処理の複合設計

LS1シリーズでは、大画面(58・65インチ)4K画質に相

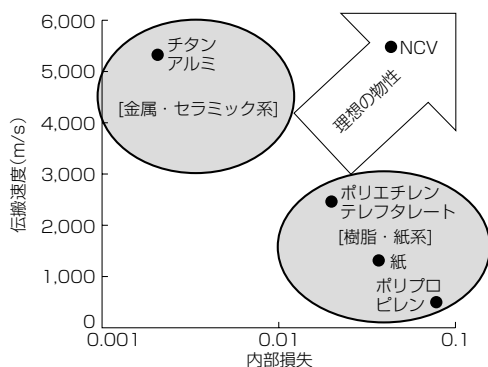


図1. 材料物性

応(ふさわ)しい低音再生を目指して設計を行った。スピーカーボックスの容積と低音再生能力は比例するため、限られたサイズで最大限に低音再生能力を引き出すパッシブラジエータ方式を採用し、さらに、音響信号処理との複合的な設計を行った。

同じ容積のスピーカーで比較した場合、パッシブラジエータ方式は一般的に用いられる低音増強方式バスレフ型スピーカーよりも更に低音再生能力が高い反面、設計・コスト面に難しさがある。

この開発では、パッシブラジエータ型スピーカーの集中定数系モデル(図2)を見直し、低音再生のボトルネックを特定してスピーカー設計にフィードバックを行った。その結果、当初使用していたスピーカーユニットでは駆動力が不足し、パッシブラジエータを十分に鳴らし切れていないことが明らかとなった。信号処理による低音増強も織り込んで磁気回路の再設計を行い、磁束密度を6,000G(ガウス)から10,000Gまで強化することで、サイズの限界までスピーカーを鳴らし切る設計を実現している(図3)。

2.3 低音信号処理

音響処理“音ハッキリ”には、スピーカー再生能力の限界まで鳴らし切る技術VBL(Variable Bottom Limiter)が搭載されている。

スピーカーは再生周波数が低いほど大きく振幅するが、スピーカーの構造上、振幅できる範囲には限界がある。限界を超えるような過大な入力信号が入った場合、再生音の歪(ひず)みや音割れを生じる。通常、歪みや音割れの回避策として、ハイパスフィルタなどを用いて低音再生能力を

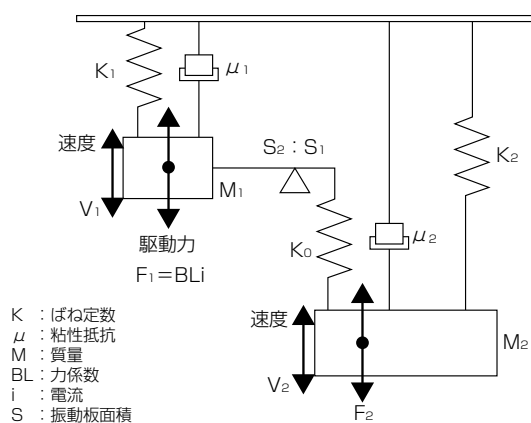


図2. パッシブラジエータ型スピーカーの集中定数系モデル

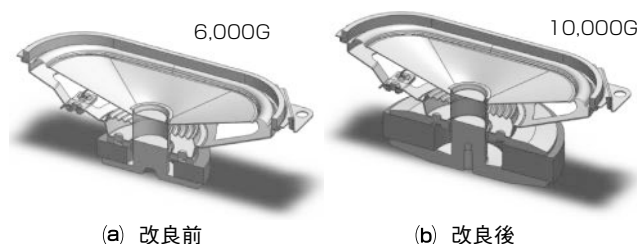


図3. スピーカー磁気回路

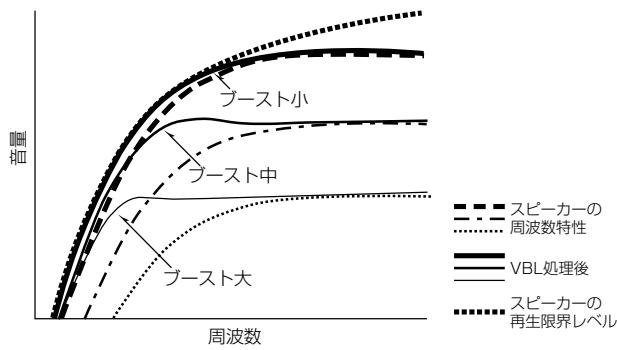


図4. VBL動作

抑制して振幅を抑えるが、この方法では本来再生できる音量に対しても同様の抑制がかかりスピーカー本来の低音再生能力を十分に発揮できない。

VBLは、豊かな低音再生と音割れ防止を両立させる。スピーカー振幅をリアルタイムに推定し、それに応じてハイパスフィルタのカットオフ周波数を適応的に変化させることで、スピーカーへの過大な入力による音割れを抑制しつつ、スピーカーの振幅限界を最大限に使い切っている(図4)。

3. 聴き取りやすい読み上げ機能

LS1シリーズには、視覚障がい者やシニアを含めた全てのユーザーにとってテレビの多様な機能が更に使いやすくなるための“しゃべるテレビ”機能を搭載している。しゃべるテレビとは、当社が推進している“らく楽アシスト”の取組みの一環として、操作メニュー項目、電子番組表(Electronic Program Guide: EPG)の番組タイトルや内容、録画予約状況や購入設置直後の初期設定を合成音声でガイドする機能である⁽¹⁾。

3.1 テキスト音声合成

図5にLS1シリーズに搭載しているテキスト音声合成システムを示す。合成対象であるテキスト(漢字かな交じり文)が入力されると、読み・アクセント解析部では言語辞書を適用してテキストの構文・読み解析を行い、テキストに対応した音素名列やアクセント位置などが決定され、中間言語と呼ばれる機械も人も内容が理解できるようにした表現として出力する。イントネーション・リズム制御部では、音の長さである各音素の継続時間長、音声の高低・イントネーションを表すピッチ周期系列を生成する。それらの情報を基に、音片選択・接続部で、音響辞書中の音片データを読み出しながら順次変形・接続して音声波形を生成し、合成音声として出力する。

言語辞書は、EPGの番組内容や有名人・俳優などの固有名詞を含む大量のテキストデータから出現頻度が高い単語や文章などを抽出して作成され、文字・読み・品詞・アクセント位置等のデータを蓄えている。韻律辞書は、ナレータが発声した大量の音声信号波形を統計的に分析し、その

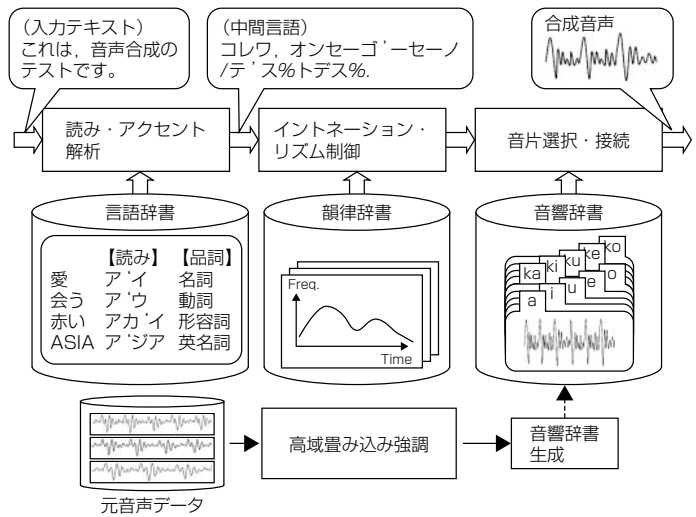


図5. テキスト音声合成システム

学習した韻律パターンを複数保持したものである。音響辞書は、ナレータが発声した音声信号を大量に収録しておき、その音声信号から所定の学習方法に基づいて幾つかの音素波形を切り出して選択して保持したものである。EPGデータ及びナレータ発声から学習した言語辞書や韻律辞書を用いることで、読み誤りの少ない自然なイントネーションの合成音声を生成することが可能となる。

3.2 高齢者にも聞き取りやすい音声強調

一般に、高齢者は加齢による聴覚器官の衰えによって聴力が低下し、特に高域が聞き取り難(にく)くなる傾向がある。高域の音声の明瞭性を改善する方法として、イコライザで高域パワーを持ち上げることで高域明瞭性を高めることが考えられるが、再生周波数が制限されたシステムではイコライザによる強調では大きな効果は望めない。そこで、再生帯域より高域の信号をあらかじめ音響辞書の再生周波数帯域内に重畳して、音声合成のためのメモリ量や処理量に負担をかけず合成音声の高域の明瞭性を改善する音声強調方式(高域畳み込み強調)⁽²⁾を導入した。図5に示すように、周波数帯域制限前の元音声データにこの高域畳み込み強調を施すことで、合成音声に再生周波数帯域以上の疑似的な広帯域感を与えるとともに、子音の明瞭性を大きく向上させることが可能となる。

この高域畳み込み強調の効果を主観評価試験で確認した。評価手法は、音声強調あり・なしの合成音声を成人6名の被験者に聴取させ、“どちらの合成音が自然で明瞭に聞こえるか”の対比較試験を実施した。また従来のイコライザによる音声強調の場合も比較評価した。図6に主観評価結果(プレファレンススコア)を示す。図6から、高域畳み込み強調あり・なしの場合の強調ありの選択率は74.2%であり、また、従来のイコライザによる音声強調と比較した場合は62.4%である。高域畳み込み強調による処理を行った方が、合成音声の明瞭度がより高くなることが分かる。

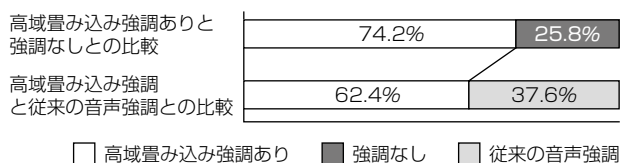


図 6. 高域量み込み強調の効果 (主観評価結果)

4. 音声操作

3章で述べた“らく楽アシスト”の取組みの一環として，“REAL Remote 音声操作”（以下“音声操作アプリ”という。）を搭載した。音声操作アプリは、携帯端末（iOS^(注1)スマートフォン／タブレット）にインストールすることで携帯端末をテレビリモコン化する当社アプリケーションである。これは、ユーザーの発話を音声認識してテレビを操作する音声認識システムと、先に述べた端末側で音声ガイダンスを実現するテキスト音声合成システムで構成している。

(注1) iOSは、Cisco Technology, Inc. の登録商標である。

4.1 音声認識システムの概要

テレビ放送のデジタル化、インターネットとの連携の進展に伴い、テレビの高機能化が進み、複雑な操作が要求されている。高齢者の場合、ボタン操作よりも音声認識を用いた操作の方が使いやすい⁽³⁾。視覚障がい者は、音声認識に対する期待が高い⁽⁴⁾。しかし、音声認識による機器操作には次の課題がある。

- (1) コマンドの言い方が分からない、覚えられない。
- (2) 誤認識した場合に誤操作になってしまう。

これらの課題を解決するため、(1)については、“いい音が聴きたい”などのように日常的に発話する言い回しで操作できるよう方式に工夫を加えた。(2)については、音韻の変形の大きい話し言葉に対応することで誤認識を低減するとともに、音声認識結果を複数出力して、それに紐(ひも)づくコマンドを列挙してユーザーに選択確認してもらうことで誤操作を防止した。

4.2 日常発話に対応した音声認識システム

図7に、この音声認識システムを示す。発話音声が入力されると、音響尤度(ゆうど)計算部が、話し言葉音響モデルを用いて事前学習した音声パターンとの音響的類似度(音響尤度)を算出する。このシステムでは、話し言葉を多数の話者から収集し、話し言葉に特有の長音化や音韻の不明瞭化などの変形を精緻に表現した音響モデルを用いている。言語尤度計算部では、算出された音響尤度とテレビ固有単語辞書から各単語の尤度を計算し、その単語連鎖の尤度を次に述べる話し言葉クラス言語モデルで計算することで、もっともらしさ順に単語列を出力する。一般的に、言語モデルは文例から単語n個の連鎖確率を表すモデルとして生成する。このため、n個の単語連鎖の組合せをカバーする膨大な文例が必要となる。このシステムでは、文例の中で同じ概念

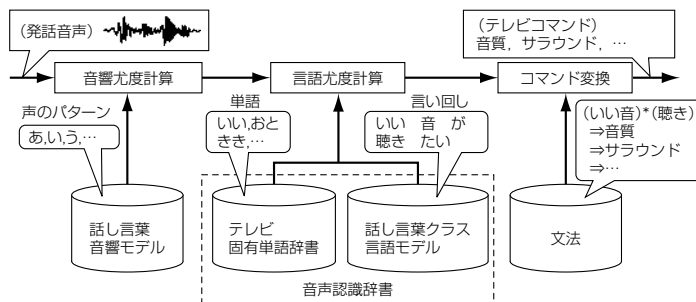


図 7. 音声認識システム

の並びとなる固有単語をクラスという単位で扱う方式を用いた⁽⁵⁾。これによって、クラスを指定したテレビ固有単語を登録するだけで認識が可能となった。また、言語モデルの記憶容量を削減して携帯端末でも動作可能にした⁽⁶⁾。認識結果単語列は、コマンド変換部で文法で受理判別し、受理時に文法に紐づくテレビコマンドに変換する。このように、日常発話でもテレビコマンドに変換することを可能とした。

5. むすび

当社製テレビに搭載されている音声技術として、高音質スピーカーシステム、読み上げ機能、音声操作について述べた。これらの技術は、高画質、高音質で使いやすいテレビの実現に大きく寄与している。

映像処理技術の向上に伴い、テレビには大画面、高画質に見合う高音質に対する要望が高まっており、今後更なる高音質化に向けた技術開発を進める。

参考文献

- (1) 三菱電機HP：しゃべるテレビ機能のご紹介
http://www.mitsubishielectric.co.jp/home/ctv/syaberu_tv/
- (2) Furuta, S., et al.: Intelligibility Improvement of Bandlimited Synthesized Speech by Superposing High Frequency Component of Input Signal on Baseband Signal, IEEE 3rd Global Conf. on Consumer Electronics, 223~224 (2014)
- (3) 小峯一晃, ほか: テレビ画面上のGUI操作環境における高齢者のリモコン操作性評価, 映像情報メディア学会誌, **55**, No.10, 1345~1352 (2001)
- (4) 吉田 諒, ほか: 音声とテンキーを用いた視覚障害者向けリモコンの試作と評価, ヒューマンインタフェース学会論文誌, **9**, No.2, 87~92 (2007)
- (5) 吉田裕美, ほか: 金融業向け音声認識ボイスロギングソリューション, 三菱電機技報, **83**, No.7, 445~448 (2009)
- (6) 花沢利行, ほか: 統計言語モデルを用いた連続音声認識における単語グループ間バイグラムの削減, 電子情報通信学会総合大会公演論文集, D-14-6 (2006)