

受話音を聞きやすくする 音声強調技術／話速変換技術

Speech Enhancement Technology, Speech Rate Control Technology

外川太郎 大谷 猛 鈴木香緒里



携帯電話は最も身近なコミュニケーションツールとして生活に定着しており、様々な環境で様々な人が利用するようになった。それに伴い、周囲が騒がしい環境で通話する状況や、早口の相手と通話する状況が増えたが、そのような場合、相手の声が聞きづらくなる傾向があった。これに対して筆者らは、携帯電話の受話音を聞きやすくするために、聞き手の周囲の騒音に応じて相手の声を強調する音声強調技術、相手の声質のまま話速を遅くする話速変換技術を開発した。本稿では、音声強調技術及び話速変換技術について、技術の特徴と携帯電話へ実装する際の工夫を解説する。
キーワード：音声強調、話速変換、音質補正、聴覚補償

1. はじめに

携帯電話・スマートフォンはいつでもどこでも手軽に使えるコミュニケーションツールであり、外出先の騒がしい環境など日常生活のあらゆる場面で利用されている。周囲が騒がしい環境で通話した場合、通話相手に送られる音声（送話音）に騒音が混入して聞きづらくなるほか、通話相手から送られてくる音声（受話音）も騒音に埋もれて聞きづらくなる傾向があった。これに対して、これまで携帯電話には、音声品質向上を狙った様々な音声信号処理機能が搭載されてきた。例えば、騒音が混入した送話音に対して騒音のみを抑制し音声品質を向上させる騒音抑圧機能^{(1)~(3)}などが挙げられる。

富士通では、携帯電話の受話音を聞きやすくする音声信号処理技術として、駅ホームなどの騒がしい環境で、周囲騒音の大きさや種類に応じて通話相手の声を強調す

る音声強調技術「はっきりボイス」や、通話相手の早口の音声に対して、通話相手の声質のまま話速を遅くする話速変換技術「ゆっくりボイス」を開発した。

本稿では、まず2. で音声強調技術の、3. で話速変換技術の技術的な課題と技術の特徴を解説し、4. でそれらの技術を携帯電話へ実装する際の工夫を解説する。

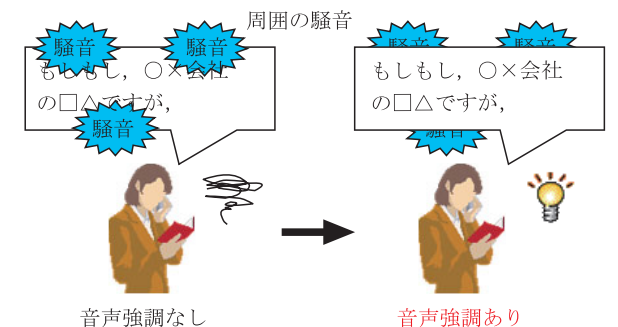


図1 音声強調技術の概要 周囲の騒音に応じて受話音を強調して際立たせることにより、聞きやすさを向上させる。

外川太郎 正員 (株)富士通研究所メディア処理システム研究所
E-mail togawa.taro@jp.fujitsu.com
大谷 猛 正員 (株)富士通研究所メディア処理システム研究所
E-mail otani.takeshi@jp.fujitsu.com
鈴木香緒里 富士通株式会社ユビキタスビジネス戦略本部
E-mail kaori@jp.fujitsu.com
Taro TOGAWA and Takeshi OTANI, Members (Media Processing System Laboratories, Fujitsu Laboratories Ltd., Kawasaki-shi, 211-8588 Japan), and Kaori SUZUKI, Nonmember (Ubiquitous Business Strategy Unit, Fujitsu Ltd., Kawasaki-shi, 211-8588 Japan).
電子情報通信学会誌 Vol.96 No.11 pp.874-881 2013年11月
©電子情報通信学会 2013

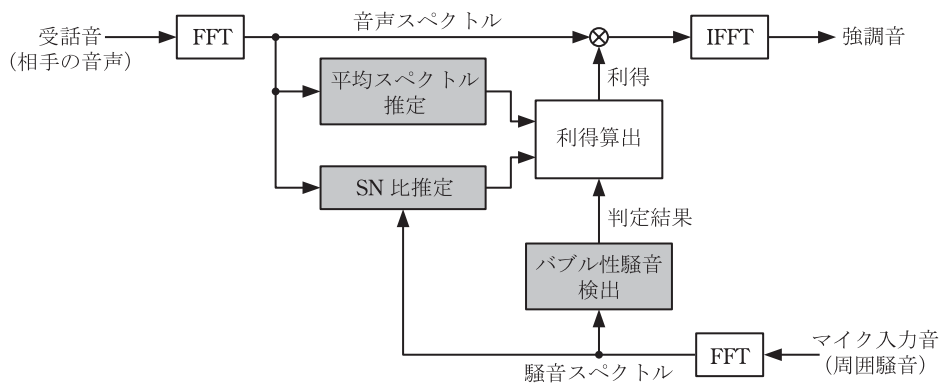


図2 音声強調技術の構成 入力音の周波数ごとのパワー（音声スペクトル）と、周囲騒音の周波数ごとのパワー（騒音スペクトル）に応じて、入力音に対して周波数領域で利得を適用することで、聞きやすさを向上させる。

更に、5. で受話音の聞こえを支援するそのほかの技術を紹介し、最後に6. で開発技術の今後の応用分野について述べる。

2. 音声強調技術

2.1 音声強調技術の概要

携帯電話は、駅ホームやショッピングセンターなど、様々な場所で利用されているが、周囲が騒がしい場所では、受話音が周囲騒音に埋もれて聞きづらくなる傾向がある。そのため、従来は受話音を聞きやすくするために、利用者がボタン操作等により再生音量を調節せざるを得ず、ユーザビリティが低い問題があった。

そこで、富士通では周囲の騒音に応じて受話音を自動的に強調して聞きやすくする技術を開発し（図1）、携帯電話・スマートフォン製品に活用している^{(4)~(8)}。以下、携帯電話向け音声強調技術の課題と特徴について解説する。

2.2 音声強調技術の課題

音声強調技術を携帯電話に適用する場合の課題として、以下の3点が挙げられる。

1点目の課題は、様々な音量の騒音に対する受話音の聞きやすさの実現である。携帯電話は、屋内だけでなく屋外でも使用されるため、利用者の周囲で発生する騒音の種類は固定電話と比べて多様である。そのため、携帯電話の利用環境によっては、受話音が周囲騒音より小さ

くなる場合があり、聞こえづらくなる傾向があった。

2点目の課題は、音声の自然性の維持である。通話相手によっては音声の高域のパワーが大きい場合があるため、騒音に応じて高域を過度に増幅するとキンキンと聞こえ耳障りな声質になりやすかった。

3点目の課題は、騒音特性（周波数ごとのパワーの大きさや時間変化）によらず聞きやすさを高めることである。周囲騒音の中でも、人込みのガヤガヤした騒音（パブル性騒音）は、騒音源が人（音声）であり、周波数ごとのパワーの大きさやパワーの時間変化の特性が受話音（通話相手の音声）と似ているため、パブル性騒音と受話音とを聴感上区別しにくい性質がある。そのため、周囲騒音がパブル性騒音の場合は特に、受話音が聞きづらくなる傾向があった。

2.3 音声強調技術の特徴

開発した音声強調技術の構成を図2に示す。開発技術は、受話音の周波数ごとのパワー（音声スペクトル）と、携帯電話の送話口のマイクを利用して得られる周囲騒音の周波数ごとのパワー（騒音スペクトル）に応じて、受話音に対して周波数領域で利得を適用することで、聞きやすさを向上させるものである。

開発技術は、(1)様々な音量の騒音に対して受話音を聞きやすくする点、(2)通話相手の様々な声質に対して音声の自然性を維持する点、(3)騒音特性（パブル性騒音）に応じて受話音の聞きやすさを高める点に特徴を有する。

以下、開発技術の特徴について述べる。

(1) 周囲騒音と受話音の音量比に応じた利得適応制御

周囲騒音に対して受話音を聞きやすくするためには周囲騒音の音量よりも受話音の音量を一定以上大きくする必要があり、そこで、周囲騒音と受話音の周波数ごとの

用語解説

モーラ 一定の時間長を持つ音の分節単位である。日本語の場合、仮名一つの単位が1モーラに当たる。（ただし、「ゃ」「ゅ」「ょ」などの小書きの仮名は1モーラとならず、例えば「ちゃ」で1モーラと数える。）話速は一般的に、単位時間当りのモーラ数で表すことが多い。

パワーを分析し、周囲騒音と受話音のパワー比（SN比）が所定の値以上になるように周波数ごとの利得を適応的に制御することにより、様々な音量の周囲騒音に対する受話音の聞きやすさを実現した。

(2) 受話音の自然性を維持する帯域間利得制御

騒音環境下での音声の自然性と音声の各帯域の利得の関係性を調査するために、騒音環境（街頭騒音）において、低域（0～2 kHz）と高域（2～4 kHz）の利得を変化させた音声の自然性を比較する実験を行った。実験の結果、高域のパワーが大きすぎる場合にキンキンと聞こえ耳障りとなり、低域のパワーが大きすぎる場合にこもって聞こえ不明瞭となることが分かった。そこで、増幅後の受話音の高域と低域のパワー差が所定範囲になるように周波数ごとの利得を調整することで、様々な通話相手の声質（音声の高域パワーの大きさ）に対して受話音の自然性を維持できるようにした。

(3) 騒音特性に応じた高域利得制御

人の声は、声帯振動に起因するピッチ周波数とその整

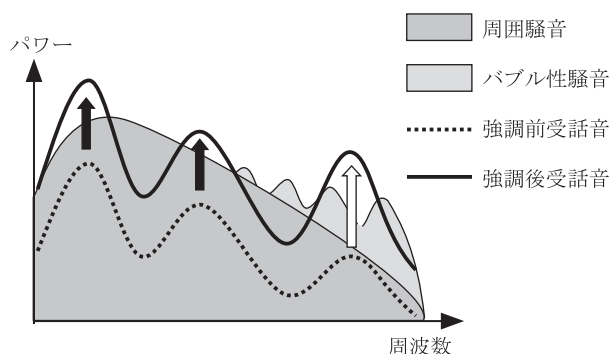


図3 騒音の種類に応じた高域利得制御の原理 周囲にバブル性騒音（人混みの騒音）がある場合、受話音に対する高域の増幅量を大きくすることで聞きやすさを高める。

数倍である倍音のパワーが大きく、声帯の振動数が時間変動することから、ピッチ周波数や倍音の周波数が時間的に変化する性質がある。そのため、人の声が混ざった音であるバブル性騒音は、周波数成分の強弱の時間変化が激しい特徴がある。そこで、周囲騒音の周波数成分の時間変化量に基づいてバブル性騒音の有無を推定し、周囲騒音がバブル性騒音の場合は、聴覚的に感度の高い高域（2～3.5 kHz）において受話音を増幅することで、騒音特性によらず受話音を聞きやすくした（図3）。

2.4 音質評価

開発した音声強調技術の音質向上効果を評価するため、騒音環境で強調前の受話音（原音）と開発技術による強調音を聞き比べて、聞きやすさに関する相対評価を行った。

音質評価法として、VoIP等の音質評価で一般的に用いられている比較範ちゅう尺度法（CCR: Comparison Category Rating）⁹⁾を使用し、評価尺度は5段階（-2：原音より聞きづらい，-1：原音よりやや聞きづらい，0：原音と同等の聞きやすさ，+1：原音よりやや聞きやすい，+2：原音より聞きやすい）とした。被験者は20～80歳代の一般被験者48名である。

評価結果を図4に示す。評価の結果、開発技術による聞きやすさの評価値は、全ての騒音条件で原音の聞きやすさ（0）を上回り、かつ原音の聞きやすさに対して有意な差があったため、開発技術は騒音環境によらず音声の聞きやすさを高める効果があることを確認できた。

3. 話速変換技術

3.1 話速変換技術の概要

携帯電話の普及により、様々な利用者が様々な相手と通話する機会が増加した。通話相手の中には早口で話す人がいるほか、移動中などの忙しい状況では早口で話す

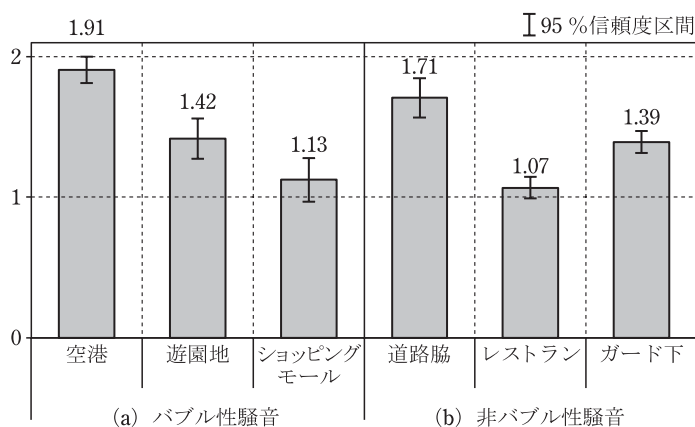


図4 音声強調技術の評価結果

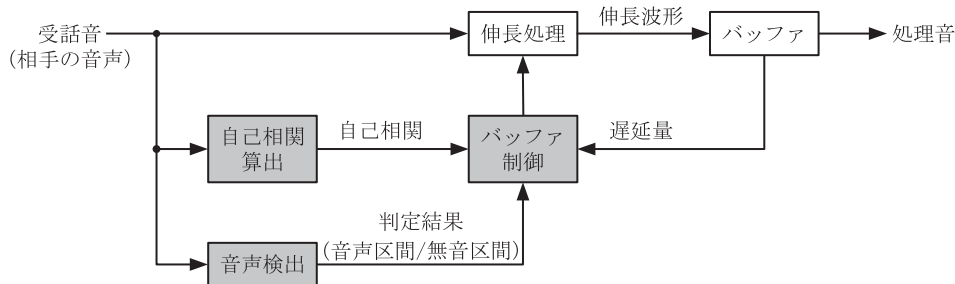


図5 話速変換技術の構成 受話音に含まれる音声区間と無音区間を判定し、音声区間を伸長しながら無音区間を短縮することで、受話音の音声の話速を遅くする。

場合があり、この場合、聞き手である利用者が聞きづらさを感じるがあった。特に中高齢者は、加齢に伴い聴力が低下するほか、音の認知能力も衰えることが知られており、中高齢者にとっては、早口の音声の聞きづらさは深刻な問題であった。

それに対して富士通では、音質を低下させることなく相手音声の話速を遅くする話速変換技術を開発し、携帯電話・スマートフォン製品に搭載してきた^{(4)~(8), (10)}。開発当時、テレビやラジオの放送音声の話速を遅くする技術⁽¹¹⁾が実用化されていたが、携帯電話のようなリアルタイムでの双方向通話向けの話速変換技術については実用化されていなかった。

以下、携帯電話向け話速変換技術の課題と特徴について解説する。

3.2 話速変換技術の課題

話速変換技術の課題として、以下の2点が挙げられる。

1点目の課題は、通話相手の声質のまま話速を遅くすることである。話速を遅くするためには、音声を時間的に長くする必要がある。単純に音声波形を時間方向に引き伸ばせば話速を遅くできるが、声の高さが低くなり、まるで別人のような声になるため音質面で問題があった。

2点目の課題は、受話側で話速を遅くしても会話において不自然さを感じないようにすることである。話速を遅くすることで音声の再生時間が元の長さより長くなると、受話側で音声を聞き終わる時間が遅延する。その結果、双方向の通話では会話が間延びしてリアルタイム性が失われるため、通話相手が不自然さを感じたり、会話のずれが大きすぎる場合には通話に支障を来す問題があった。

3.3 話速変換技術の特徴

開発した話速変換技術の構成を図5に示す。開発技術は、音声検出部で受話音に含まれる音声区間と無音区間（音声が含まれない区間）を判定し、音声区間を伸長し

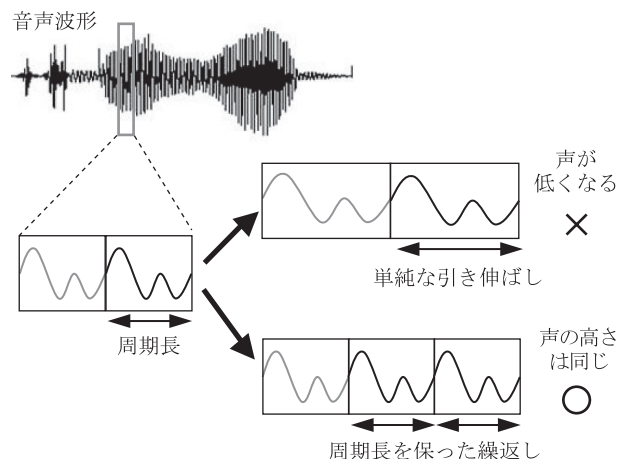


図6 話速変換の基本原則

ながら無音区間を短縮することで、受話音の音声の話速を遅くするものである。

開発技術は、(1)声質を変えずに音声を伸長する点、(2)話速を遅くしても会話に支障を感じさせない点に特徴を有する。

以下、開発技術の特徴について述べる。

(1) 通話相手の声質を変えない音声伸長

人の音声には、声帯の振動に起因する周期的な成分が含まれている。単純に、音声波形を時間方向に引き伸ばすと周期長が変化するため、声の高さが低くなる。そこで、自己相関分析により相関の高い（周期性の高い）区間を抽出し、抽出した区間の音声波形を繰り返すことにより、声の高さ（周期長）を変えずに音声を伸長する技術を開発した（図6）。

(2) 話速を遅くしても会話に支障を感じさせない遅延制御

双方向通話において遅延があった場合に、聞き手が会話に支障があると感じるかを調査したところ、遅延が1秒より短ければ会話に支障を感じにくいことが分かった。そこで、無音区間を短縮して遅延を削減するととも

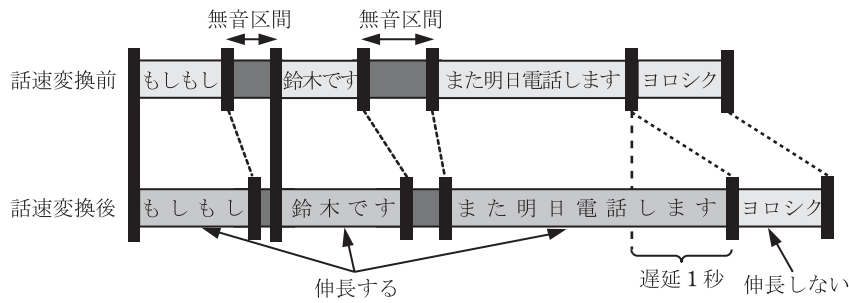


図7 話速変換技術の動作例 音声区間を伸長する一方で、無音区間を短縮して遅延を削減する。遅延が1秒を超える場合は、音声区間を伸長しない（元の話速に戻す）。

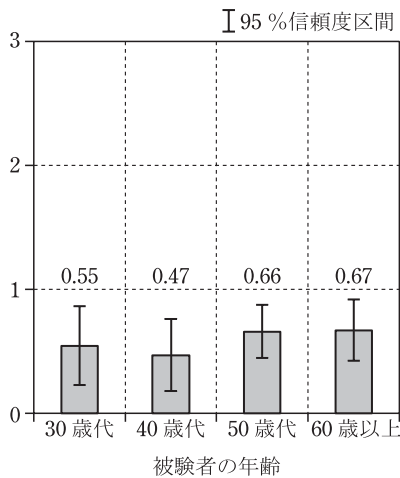


図8 話速変換技術の評価結果

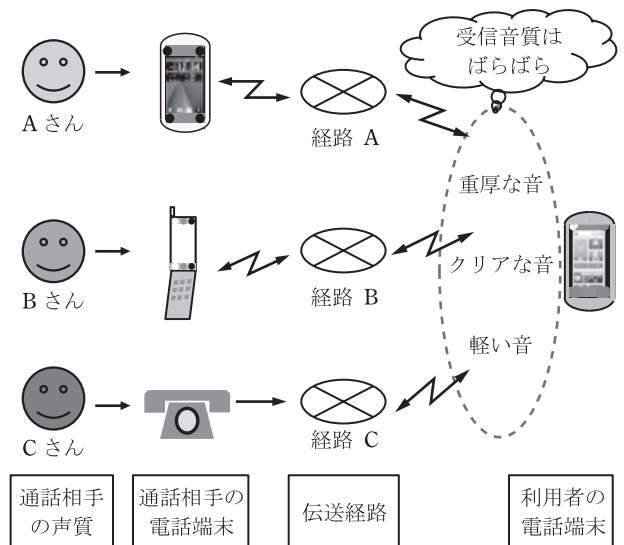


図9 受話音質を決める要因

に、遅延が1秒以上になる場合は、話速を徐々に元に戻すように制御した（図7）。これにより、音声伸長に伴う遅延を1秒以内に抑えられるため、通話相手が会話に支障を感じにくくなり、通話のリアルタイム性の維持が可能とした。

3.4 音質評価

開発した話速変換技術の音質向上効果を評価するため、元の話速の受話音（原音）と開発技術により話速を遅くした処理音を聞き比べて、聞きやすさに関する相対評価を行った。

評価尺度は7段階（-3：原音より非常に聞きづらい，-2：原音より聞きづらい，-1：原音よりやや聞きづらい，0：原音と同等の聞きやすさ，+1：原音よりやや聞きやすい，+2：原音より聞きやすい，+3：原音より非常に聞きやすい）とした。被験者は30～80歳代の一般被験者28名である。原音の話速は、中高齢者が早口と感じられる8～10モーラ^(用語)/秒とした。

評価結果を図8に示す。評価の結果、開発技術による聞きやすさの評価値は、全ての年齢の被験者について原音の聞きやすさ（0）を上回り、かつ原音の聞きやすさ

に対して有意な差があった。そのため、開発した話速変換技術は、受話音の話速が速い場合に、聞き手の年齢によらず聞きやすさを高める効果があることが確認できた。

4. 装置実装上の工夫

音声強調技術及び話速変換技術が、実環境（実際に利用者が使用する条件や環境）で性能を発揮し、利用者が効果を実感できるようにするための装置実装上の工夫を紹介する。

4.1 様々な受話音質条件で音質改善効果を引き出すための工夫

受話音質は、通話相手の声質、通話相手の電話端末の特性、通話相手と利用者の電話端末の間の伝送経路、利用者の電話端末の特性によって様々な音質となる（図9）。そのまま音声強調技術及び話速変換技術を適用すると、通話ごとに音質が異なるだけでなく、条件によって

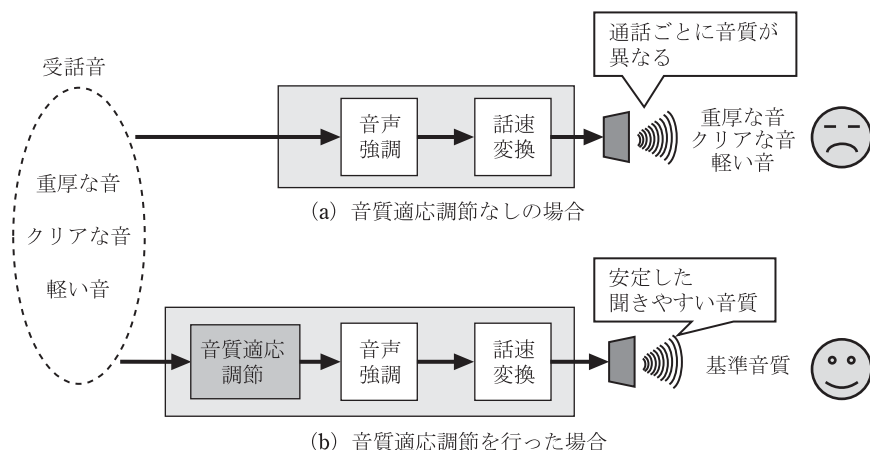


図 10 音質適応調節による基準音質の実現

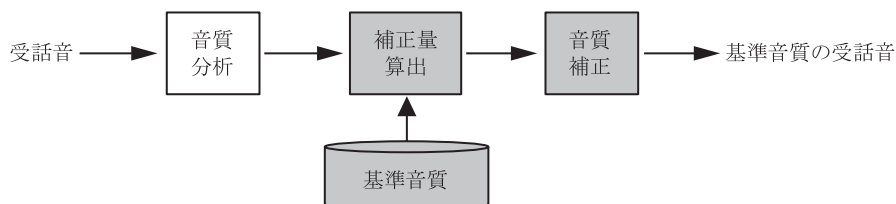


図 11 音質適応調節技術の構成

は聞きやすさの改善効果が損なわれる可能性がある（図 10(a)）。そこで受話音質を基準の音質に調節する音質適応調節技術により、通話条件による音質特性差を吸収することで、音声強調技術及び話速変換技術の性能を發揮させるとともに、通話条件によらずに安定した聞きやすい音質とした（図 10(b)）。

これを実現するための音質適応調節技術の構成を図 11 に示す。受話音の音質分析を行い、受話音の音質を基準音質に合わせるための補正量を算出し、受話音に適用することで、通話ごとに異なる受話音質を基準音質に調節する。例えば、全帯域成分に対する低域成分の比率が大きい場合にこもった音と感じられやすく、逆に高域成分の比率が大きいとキンキンした音とを感じる傾向がある。富士通では、基準音質として「低域に伸びのある、厚みのある、落ち着いた音」を基準音質と定め、音質適応調節技術により各周波数ごとの成分の比率を変えて音質を制御することで、全ての機種で基準音質を実現している。

4.2 騒音環境で性能を發揮させる工夫

携帯電話は、駅ホーム、車通りの激しい街頭などの高騒音環境から、自宅などの静かな環境まで、様々な環境で最適な性能を出すことが求められる。特に高騒音環境では、音声騒音に埋もれ（SN 比の低下）、聞きやすさ向上技術の性能を發揮することが難しくなる。特に話

速変換技術では、音声はゆっくり聞こえるよう音声区間を伸長し、会話の間である無音区間を削除して音声伸長に伴う遅延を解消することで、相手との通話を妨げずにゆっくりと聞きやすい効果を持続させる。このため、音声検出の性能（無音区間の検出性能）が話速変換技術の効果を左右する。

音声検出は、音声と周囲騒音の性質の違いを利用して音声区間を検出するが、周囲騒音の大きさを推定して、音声を検出する判定しきい値を適応制御することと、周波数分析により SN 比が高く音声の特徴が損なわれにくい周波数の情報を抽出して算出した特徴量で音声らしさを判定することにより、高騒音条件でも良好に音声を検出できる技術を開発した（図 12）。これにより騒がしい環境でも無音区間のうち 60% 以上を検出でき、遅延を解消して効果を持続させることで、様々な場所でも話速変換技術の効果を發揮できるように実装している。

5. そのほかの聞こえ支援技術

これまでに述べた音声強調技術や話速変換技術に加え、利用者の動作に応じて受話音を聞きやすくする機能「ぴったりボイス」と、利用者の年齢に応じて聞こえを改善する機能「あわせるボイス」について、技術の概要を紹介する。

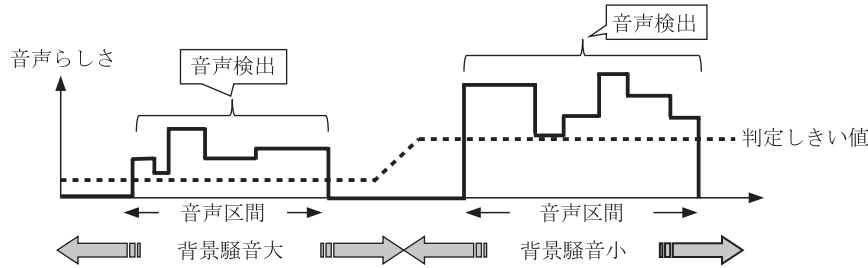


図12 音声検出の性能向上の工夫

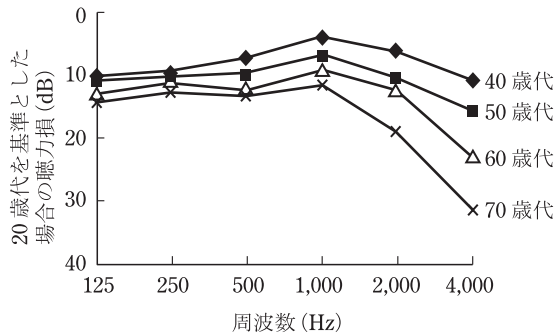


図13 加齢による聴力低下 40～70歳代の被験者136名(各年代20～48名)に対する聴力測定の結果である。年齢が高くなるほど、特に高域の聴力低下が大きい傾向がある。

5.1 ぴったりボイス

通勤・通学や買い物などの移動中に、歩きながら、または走りながら携帯電話で通話することがある。そのような状況では、歩行に伴い利用者の頭部や腕が動くため携帯電話の受話口が利用者の耳穴から離れやすく、受話音が聞きづらくなる場合がある。それに対して、利用者の動作(歩行/走行)に合わせて、受話音の周波数ごとのパワーを増幅する機能を開発した。開発技術では、携帯電話に搭載されているモーションセンサを利用してきょう体の揺れ(動き)の大きさを検知し、利用者の動作が歩行であるか、または走行であるかを判定する。そして判定した動作に合わせて、受話音の周波数ごとのパワーを増幅することで、聞きやすさを高めた。

5.2 あわせるボイス

一般に年齢が高くなるにつれて聴力が低下して音声聞きづらくなることが知られている(図13)。そこで、利用者の聴力に応じて受話音を増幅することにより、聞こえを改善する機能を開発した。

開発技術の特徴を以下に示す。

(1) 年齢に基づく音質補正技術

聴力と年齢に相関がある性質を利用し、利用者があらかじめ入力した年齢から聴力低下量を推定し、聴力低下量に基づいて受話音を増幅することにより、簡単な操作

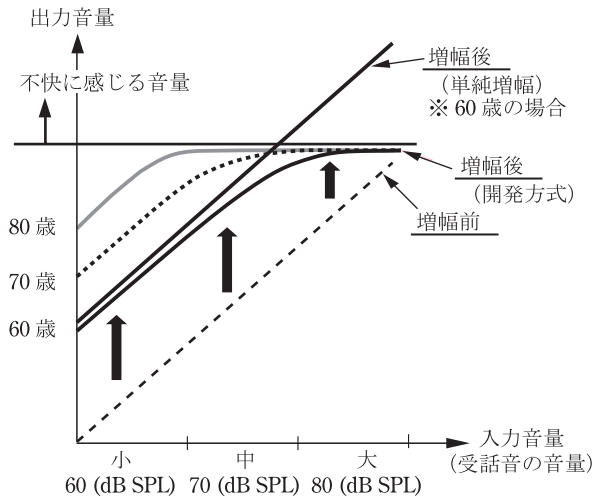


図14 入力音量に応じた増幅量制御 入力音量(受話音の音量)が小さい場合は聴力低下量に基づいて増幅し、入力音量が大きい場合は不快と感じる音量を超えないように増幅する。なお、dB SPLは音圧レベル(Sound Pressure Level)の単位である。

で高齢者の聞きやすさを向上した。

(2) 入力音量に応じた増幅量制御

高齢者は小さい音は聞きにくいものの、大きい音は十分に聞こえるため、単純に全ての音を増幅すると、音量が大きい場合にうるさすぎて不快を感じる課題があった。そこで、図14に示すように、入力音量(受話音の音量)が小さい場合は聴力低下量に基づいて増幅し、入力音量が大きい場合は不快と感じる音量を超えないように増幅することにより、高齢者の不快感を解消した。

6. おわりに

本稿では、携帯電話において受話音の聞きやすさを向上する音声強調技術及び話速変換技術の特徴と、装置実装上の工夫、そのほかの聞こえ支援技術の特徴について述べた。

開発技術は、人と人とのコミュニケーションの基本である会話を支援する技術であるため、携帯電話に限らず、放送・通信、教育、窓口のように、音声を用いたコ

コミュニケーションにより業務が成り立っている分野への応用が期待できる。更に、日本国内だけでなく世界各国で高齢化が進展していることを踏まえると、聞こえを支援する開発技術は、今後ますます世界中で必要とされることが予想される。

また、近年の携帯電話・スマートフォン製品の形態の多様化やデザインの進化により、様々な音質条件への対応が必要になると予想されるため、今後も「聞きやすさ」「話しやすさ」を極めるための取組みを継続していく。

文 献

- (1) 栗原秀明, 片山 浩, 大田恭士, 片江伸之, “携帯電話の音声信号処理,” 音響誌, vol. 58, no. 12, pp. 786-791, 2002.
- (2) 山崎 泰, 松澤 均, 佐々木 繁, 棚橋純一, “音声通信向け帯域別雑音抑圧機能の開発,” 音響論集, no. 2-P-20, pp. 321-322, March 2000.
- (3) 大谷 猛, 鈴木香緒里, 大田恭士, 佐々木 繁, 天野文雄, “音声スペクトル推定を適用したノイズサプレッサ,” 第21回信号処理シンポジウム, no. C8-3, Nov. 2006.
- (4) 天野文雄, “(招待講演)らくらくホンにおける聞こえ支援機能について,” 音響論集, pp. 1563-1564, March 2012.
- (5) 松田喜一, “AV 関連技術の動向と富士通の取組み,” 富士通技術報, no. 55, pp. 3-7, Oct. 2010.
- (6) 林田 健, 中条 薫, 渡邊儀一, “らくらくホンの開発コンセプトと機能,” 雑誌 FUJITSU, vol. 61, no. 2, pp. 184-191, March 2010.
- (7) 中条 薫, 大田恭士, 高木和幸, 椎谷秀一, “ヒューマンセントリックエンジンとサービス展開,” 雑誌 FUJITSU, vol. 63, no. 5, pp. 505-511, Sept. 2012.
- (8) 佐相秀幸, “らくらくホンについて,” 電気通信, vol. 72, no. 751,

pp. 42-48, 2009.

- (9) F. De Rango, M. Tropea, P. Fazio, and S. Marano, “Overview on VoIP: Subjective and objective measurement methods,” IJCSNS International Journal of Computer Science and Network Security, vol. 6, no. 1B, pp. 140-153, Jan. 2006.
- (10) 富士通研究所: やさしい技術講座,
<http://jp.fujitsu.com/group/labs/techinfo/techguide/list/voice.html>
- (11) 清山信正, 今井 篤, 三島 剛, 都木 徹, 宮坂栄一, “高品質リアルタイム話速変換システムの開発,” 信学論 (D-II), vol. J84-DII, no. 6, pp. 918-926, June 2001.

(平成 25 年 7 月 31 日受付 平成 25 年 8 月 9 日最終受付)



とがわ たろう
外川 太郎 (正員)

平 14 東工大・工・情報卒, 平 16 同大学院修士課程了。同年(株)富士通研究所入社。以来, 音声・音響信号処理方式の研究に従事。平 24 文部科学大臣表彰科学技術賞(開発部門)受賞。



おおたに たけし
大谷 猛 (正員)

平 10 北大・工・電子卒, 平 12 同大学院工学研究科修士課程了。同年(株)富士通研究所入社。以来, 音声信号処理方式, 実現技術, 応用の研究開発に従事。平 24 文部科学大臣表彰科学技術賞(開発部門)受賞。



すずき かおり
鈴木 香緒里

昭 59 北大・理・地球物理卒。昭 61 同大学院修士課程了。同年富士通株式会社入社。以来, 携帯電話端末向けの聞きやすさ向上技術の研究開発, 実装に従事。平 24 文部科学大臣表彰科学技術賞(開発部門)受賞。

