3. 人にやさしい放送

3-2. 実用化されている情報バリアフリー技術

正会員 今井 亨[†], 正会員 都木 徹[†], 秋田 正二^{††}

キーワード ● 字幕放送, 音声認識, 話速変換, 高齢者, 老人性難聴, 背景音

1. ま え が き

デジタル放送などの放送の発展により, 視聴者がテレビから受け取る情報は, 年々増えつつある。一方, 高齢者や障害者など, 誰もが情報を等しく得られるようにするため, 情報バリアフリー技術が広く研究され, いくつかの技術がすでに実用化されている。本稿では, 音声認識を利用したリアルタイムの字幕放送, 早口の音声をやっくりにして聞きやすくする話速変換, 番組の背景音を小さくして音声聞きやすくする高齢者向け副音声サービスを紹介する。

2. 音声認識による字幕放送

字幕放送は, 聴覚障害者や高齢者にテレビ番組の音声で文字で伝えるサービスである。NHKのいくつかの生放送では, 音声認識技術を利用して, リアルタイムの字幕付与が実用化されている。

2.1 ニュース番組字幕化のための音声認識

NHKでは現在, 「おはよう日本」「正午のニュース」「ニュース7」「ニュース9」で, 独自に開発した音声認識システムによる字幕放送を毎日実施している¹⁾。これは, スタジオで原稿を読み上げるアナウンサの声を自動的に認識し(認識率95%以上), 結果に誤りがあれば人手で即座に修正して字幕を制作するものである(図1)。認識可能な語彙や言い回しは, 記者原稿から毎日自動的に学習し¹⁾, 新しい固有名詞や話題に対応できるようになっている。NHKの多くのアナウンサの音声で音響モデル(母音・子音の周波数特性)を学習してあり, 不特定多数話者を認識可能である。

図1 ニュース番組字幕化のための音声認識

[†]NHK 放送技術研究所

^{††}NHK 編成局

"Operated Information Barrier-free Technologies" by Toru Imai and Toru Takagi (NHK Science and Technical Research Laboratories, Tokyo) and Shoji Akita (NHK Programming Dept., Tokyo)

字幕表示の遅れを極力小さくするため, 認識エンジンが音声認識結果を早期確定²⁾し, 誤りの発見者と修正者がペアになってリアルタイムで認識結果を修正している。

音声認識は, 確率統計的手法が一般的であり, 事前に学習した記者原稿やアナウンサの声のモデルと照合することで, 入力音声に最も近い単語列が決定される。まず, 音響分析によって, 入力音声の特徴量が順次抽出される。次に, 登録されているすべての単語の発音の音響モデルと入力音声の特徴量が照合され, 可能性の高い単語だけが候補として保持される。単語から単語に遷移する時には, 言語モデル(単語の連続出現頻度)を参照して, やはり可能性の高い単語だけを候補に残す。以上の手順を発話終了までリアルタイムで処理し, 音響的かつ言語的に最も可能性の高い単語列を認識結果として出力する。ただし, 発話途中でも充分信頼できる単語候補は, 早期に確定している。

一方, 現場リポートや対談部分は, 現在は高速入力キーボードを利用して字幕を付与しており, 音声認識における背景雑音や発話スタイルの影響が研究課題となっている。

2.2 リスピーク方式音声認識による生字幕

ニュース以外の生放送番組では, 番組音声そのまま認識するのではなく, 字幕キャスタと呼ぶ別話者が言い直した音声を認識する「リスピーク方式³⁾」によって字幕を制作している(図2)。リスピーク方式によれば, 番組音声に含まれる雑音, 複数話者の声の重なり, 話し言葉のくだけた言い回しなど, 音声認識を困難にする音響的・言語的問題を軽減することができる。また, 字幕キャスタが内容を要約したり補完することで, 読んでわかりやすい字幕になるよう工夫したり, 拍手や歓声など場内の様子を補足することも可能である。言語モデルは, スポーツの種目や音楽, 生活情報などの番組ジャンルごとに学習する必要があるた

図2 リスピーク方式音声認識による生字幕

め、より効率的な学習法が研究課題になっている。

聴覚障害者の方々からは、字幕放送によって得られる情報が増し、テレビを見る楽しみが広がったといった声が寄せられている。

3. 話速変換

お年寄りにとって、放送の音声及早口で聞き取りにくいと感じられる場合がある。高齢者の聴力低下を補う手段として補聴器があるが、「早口が聞き取りにくい」ということを補うことはできない。そこで、聴取する音声の速度そのものを機械的にゆっくりにする話速変換が考えられる。

話速変換技術の開発は20数年前から内外においてさまざまな目的で行われており、難聴者の補聴用⁴⁾や家庭用ビデオの早見の補助機能⁵⁾としての製品化も図られてきた。しかし、放送番組の視聴において話速を低下させて音声全体の時間を延ばすことは、放送の即時性を損なうことにつながるため課題があった。ここでは、これを解決し、テレビやラジオ受信機に組込んで、リアルタイムの放送視聴が可能な話速変換技術について述べる。

3.1 放送の音声品質を保つ話速変換

従来、音声の再生速度を変えると、声の高さや声質が変化し、かえって聞き取りにくくなるが多かった。放送の音声品質を極力維持する話速変換方式として、デジタル音声信号処理による以下のような方式を採用した。

- (1) 入力した音を、声のある区間と声のない区間（話の“ま”や背景音だけの部分）とに分割。
- (2) 声のある区間はさらに、波形に声帯振動に基づく周期性がある区間（“a, i, u, e, o”の母音など）と周期性がない区間（“s, p, t, k”等の子音など）に判別。そして、①周期性がある区間では波形の周期を抽出、②周期性がない区間と、③声のない区間では、波形に明瞭な周期性は認められないが、①と同様に擬似的な周期を抽出。
- (3) 指定した話速変換倍率に応じて、①～③のそれぞれの区間で求めた周期を単位として、遅くしたい場合は波形を挿入。速くしたい場合は波形を間引く。
- (4) 区間ごとに伸び縮みした波形をつなぎ合わせて出力。

声帯の振動周期は2～20ms程度の範囲で変化するが、通常一つの母音は80ms前後の長さを持つため、声帯の振動周期単位での挿入や間引きはことばの内容に影響しない。周期性を利用することによって、もとの声の高さや特徴を保ったまま、話速のみを変更することができ、話し手があたかも“ゆっくり”話したかのように変換できる。

3.2 番組の時間枠を超えない話速変換

音声のどの部分も一様に話速をゆっくりにすると、放送の時間枠を越えてしまう。テレビでは、映像から音声がどんどん遅れることになる。そこで、音声の話し始めをゆっくりとし、次の息つきまでに次第に元のはやさに戻すとともに、途中、声の高さが高くなった部分は、一時的に話速

図3 ゆっくりにしても番組全体の時間が延びない原理

をゆっくりにする。そして、“ま”は違和感のない程度に短くして実時間に追いつく（図3）。この操作を“ま”で区切られる文単位で繰り返すことで、声のある部分は一切欠落せず、かつ全体として長さが延びていないにもかかわらず、原音声より“ゆっくり”に聞こえるようになる。この手法を用いると、音声側の処理だけであるにもかかわらず、テレビの場合でも映像と大きくズレることがない。

しかし、“ま”が短い場合など、発声の仕方によっては、実時間からの遅れが蓄積してしまうので、その場合は直後の話し始めのゆっくりの度合を減らしたり、早めに元の速さに戻すといった操作を加えた。この適応的な話速変換のアルゴリズムは、“自然さ”や“ゆっくり感”を指標にした聴取実験により自動調整のパラメータを決定した。

60～80歳代の被験者約280名による評価では、原音声との比較において“原音声よりゆっくりに聞こえる”と判定され、また、ブラインドの絶対評価でも変換音声は“聞きやすい”、“落ち着いている”などの回答を多く得た⁶⁾。

3.3 実用化への取組み

この話速変換技術は、お年寄りのもとより、世代によらず外国語の放送をゆっくり聞いたり、逆に外国出身の方に日本語放送をゆっくり聞いていただくためにも有効である。すでに一部のメーカーからこの話速変換機能を内蔵したラジオ⁷⁾やテレビが発売され、好評を得ている。

一方、地上デジタル音声放送においては、一部の番組で主チャンネルの音声を“ゆっくり”に変換したものが同時にサブチャンネルで提供されている。

さらに、インターネットのNHKのホームページにおいては、ラジオニュースの2次利用として、一様に20%伸張した“ゆっくり”音声や、3.2の適応的な話速変換を応用することで約6割の時間でも内容が聞き取りやすい“早口”音声のサービスを行っている。お年寄りだけでなく、目の不自由な方や忙しい方の利用も期待されており、さまざまなサービスが展開され始めている。

4. 高齢者向け音声サービス

放送番組の音声について、「音楽や効果音などの背景音が大きくてナレーションなどの声が聞き取りづらい」という視聴者からの意見が寄せられている。このような意見を分類すると、60歳以上の高齢者からのものが多数を占めている。加齢による番組音声の聞き取りづらさを解消するため、背景音を通常よりも小さくし、聞き取りやすい音声による放送サービスの実現に向けた検証を進めている。

実用化されている情報バリアフリー技術

4.1 老人性難聴と放送におけるサービス手法の検討

加齢によって聞き取りが悪くなる要因として老人性難聴がある。老人性難聴の特徴は以下のようなものである。

- (1) 耳の感度が悪くなる
- (2) 高い周波数の音が聞こえにくくなり、言葉の明瞭度が下がる
- (3) 小さな音は聞こえにくく、大きな音はうるさく感じ、音の大きさに対する許容の幅が狭くなる
- (4) 脳の弁別能力が低下し、早口や複数の音が混じっていると目的の音が聞き取りにくい

ただし、これらの聴覚特性が現れる程度については個人差が大きい。

老人性難聴の特徴に対して、放送の送り手側として、どのようなサービスが可能なのか調査・検討を行った。番組音声の周波数特性の補正やレベル圧縮を行うことは、番組全体の音質に大きな影響を与えてしまう。放送という一律のサービスにおいて、音質を大きく変えてしまうことは望ましくない。一方、声以外の背景音（BGMや効果音）のレベルを下げることは、音質への影響が少なく、声を聞き取りやすくするという観点に立てば多くの高齢者の方に効果が期待でき、放送サービスの手法として好ましい。

背景音レベルを下げた音声の実現方法として、次の三つの方式を検討した。

(1) レベル差方式

声以外の背景音レベルを通常より低くバランスした音声を作成し、副音声チャンネルで放送する。視聴者は好みに応じた音声チャンネルを選択して聴取する。

(2) 独立方式

声と背景音を独立した音声チャンネルで放送する。受信機側では好みに応じたバランスに調整して聴取する。

(3) 識別音声付加方式

声の音声成分のデータを送り、受信機側でこのデータを基に通常音声の成分を好みに応じて強調して再生する。

この中で、受信機側に負担をかけることなく、現在のシステムで制作・送出を低コストで実現できることから、「レベル差方式」での具体化を目指した。

4.2 最適な背景音レベル差とデジタル放送での試行

通常音声に比較して背景音をどの程度下げれば高齢者にとって効果的な音声サービスとなるかを確認するために、番組サンプルを使用した高齢者による評価実験を行った。通常音声の背景音レベルを基準0dBとして、-3dB、-6dB、-10dBに調整した音声との比較によって、「声の聞きやすさ」と「番組音声としての好み」について評価した。その結果、背景音を通常より6dB下げた場合が、声が聞きやすく番組音声としても好ましいということが確認された⁸⁾。

以上のような研究を踏まえて、2004年2月にBSデジタル放送と地上デジタル放送での試行を実施した(図4)。今回の試行は、紀行番組(BSデジタル)、自然番組(地上・BSデジタル)、情報バラエティ番組(地上デジタル)で行った。

図4 高齢者向け音声サービス試行時のイメージ

同時に行ったアンケート(サンプル母数の問題から上記紀行番組を対象)から、高齢者に対して効果が期待されることが確認された。一方、サービス効果がより高い番組ジャンルの検証を進めることが今後の課題である。

5. む す び

デジタルデバインド(情報格差)を解消していくことは、世界共通の課題となっている。多くの人が放送を楽しむため、情報バリアフリー技術の開発を今後さらに進めていく。

(2004年6月30日受付)

〔文 献〕

- 1) 安藤彰男ほか：“音声認識を利用した放送用ニュース字幕制作システム”，信学論，J84-D-II，6，pp.877-887（2001）
- 2) 今井亨ほか：“最優秀単語列逐次比較による音声認識結果の早期確定”，信学論，J84-D-II，9，pp.1942-1949（2001）
- 3) 松井淳ほか：“言い換えを利用したスピーク方式によるスポーツ中継のリアルタイム字幕制作”，信学論，J87-D-II，2，pp.427-435（2004）
- 4) Y. Nejime, et. al.：“A Portable Digital Speech-Rate Converter for Hearing Impairment”，IEEE Trans. Rehabilitation Eng. 4，2，pp.73-83（1996）
- 5) 田中浩司ほか：“時間差適応型話速変換技術の開発”，1995信学総大会，D-695（1995）
- 6) 今井篤ほか：“ニュース音声を対象にした時間遅れを蓄積しない適応型話速変換方式”，信学論A，J83-A，8，pp.935-945（2000）
- 7) http://www.jvc-victor.co.jp/audio_w/product/radio/index.html
- 8) 沢口真生ほか：“高齢者にやさしい放送サービス—放送番組音声の背景音レベルと聞きやすさについて—”，音響論集，1-5-5，pp.435（2003）

今井 亨 1987年、早稲田大学理工学部電気工学科卒業。同年、NHK入局。1990年より、同放送技術研究所に勤務。音声認識の研究に従事。1996年、米BBN滞在研究員。現在、同研究所（人間情報科学）副部長、博士（情報科学）。正会員。

都木 徹 1981年、電気通信大学大学院電気通信研究科修了。同年、NHK入局。長野放送局を経て、1984年より、同放送技術研究所勤務。現在、同研究所（音響情報）主任研究員。声質変換、音声知覚、話速変換等の研究に従事。正会員。

秋田 正二 1985年、大阪大学工学部通信工学科卒業。同年、NHK入局。1991年より、同制作技術センターに勤務。主に、クラシック音楽番組のミキシングを担当。現在、同編成局チーフエンジニア。