

3 音声合成技術の将来について

人の内面を考慮した音声合成の実現に向けて

NTT 人間情報研究所（以下、人間研）では、音声合成技術の研究開発を進めている。音声ガイダンス、CM 作成やアバター広告などデジタルツインの DX が進む中、音声合成技術の研究開発やその適用について、井島特別研究員に現状の課題から将来の展開についてお聞きした。

これまでの人間研での音声合成の研究開発

人間研は、これまで音声合成の研究開発として、自然性の高い合成音声で低コストで作成することを中心に進めてきた。特に任意の話者の声色の合成音声を低コストで再現する技術開発の進展により、これまでのアナウンス用途以外にも使用されるようになってきている。例えば、有名声優の合成音声を利用可能な音声合成サービス「FutureVoice Actors」を始めとした NTT グループ内外の様々なサービスに利用されるようになってきている。

音声合成技術とは？

音声合成技術とは、入力された文章、文字を聞かせたい音声に変換するメディア変換技術である（図 1）。

これまでは、ナレーションやアナウンスといった淡々とした文章テキストの

音声変換がメインであったが、デジタル技術の進歩と音声サービスの拡大によって、アニメやゲーム、CM・広告などにシナリオや演出を加味したサービスへの応用が進んできている。

また、グローバル化が進む中、音声を使った多言語サービスや音声合成アナウンスの活用が盛んである。

人間研では、そういった用途の拡大に向けて、日本語話者の声質を保ったまま、他言語の合成音声を生成可能な技術を開発し、日本語の収録音声から英語、中国語等の音声アナウンスを実現している（図 2）。

音声合成技術の新たな価値創出に向けて

これまでに説明した音声合成技術では、主に音声を作る側（作り手・話し手）をどのように再現するかという点に主眼を置いていた。一方で人が話



NTT 人間情報研究所
思考処理研究プロジェクト
特別研究員 井島 勇祐氏

すという行動を考えた時には、一人のみで話すことはほぼなく、音声を聞く側（受け手）の存在を考慮することが重要になると考えられる。こうした課題に対して、人間研では、音声の聴取による受け手の内面への影響（感情や心理状態の変化）に着眼した研究に取り組み始めている。この取り組みでは、音声を聴取

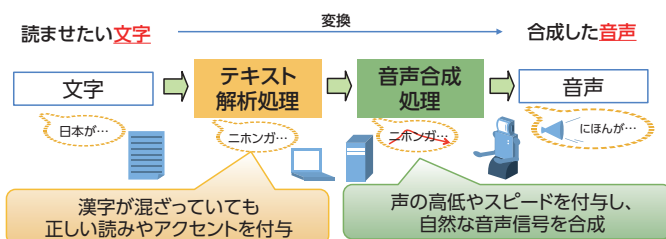


図 1 音声合成技術の音声処理フロー

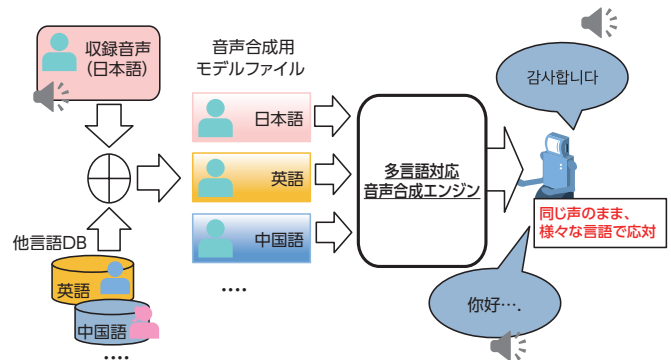


図 2 クロスリンガル音声合成技術

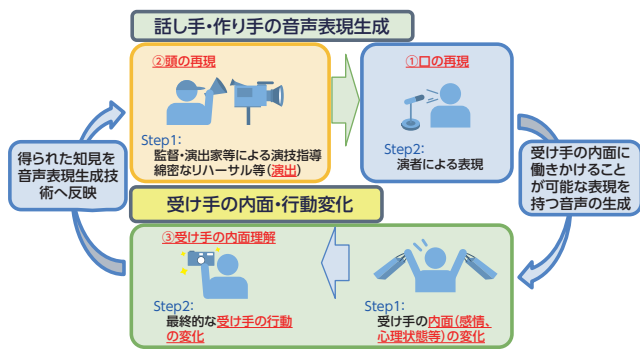


図3 音声合成による豊かなコミュニケーション実現の取り組み

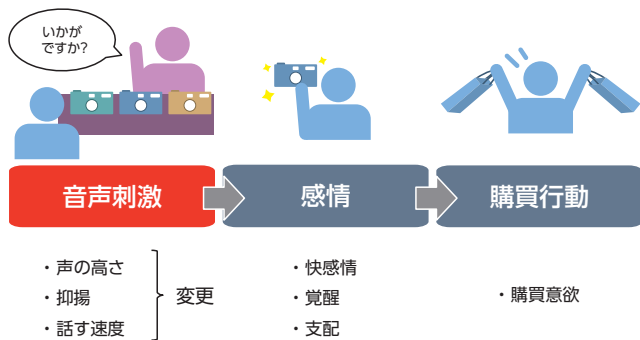


図4 音声刺激を用いた行動モデルの検証

した受け手がどのような感情、行動変化を起こすのかといった「受け手の内面・行動変化」の理解に加えて、得られた知見を音声合成へと反映させる「作り手・話し手の音声表現生成」を含め一体で取り組むことで、音声合成技術の新たな価値創出を目指している(図3)。以下に、「受け手の内面・行動変化」に関する研究について説明する。

音声刺激による消費者行動モデル

災害時の避難警報・指示の音声ガ

れる。この課題に対して、人間研では消費者行動を、「外部刺激⇒感情⇒行動」の3ステップで表す感情を媒介とする消費者行動モデルに着目し、音声の特徴と購買意欲との関係の分析について検証を実施している(図4)。

実験は、クラウドソーシングで募集した日本語を母語とする男女202名に、話し方を変えた電化製品に対する「宣伝文」を聴取してもらい、それを聞いた際に感じた「感情」と「購買意欲」についての評価項目に回答をしてもらうことで、音声の特徴と

ンダンスや、選挙候補者の演説において、話者の性別、声の高さ等の音声刺激の違いによって、受け手の行動等が変化することが知られている。

一方、これらの研究では、音声と受け手の行動変化のみを対象としており、音声の聴取から最終的な行動の変化までの不明瞭なままであるため、行動に至るまでのプロセスを段階的にとらえる必要があると考えら

感情、購買意欲の関係を分析している。分析の結果、広告文の読み上げによる音声刺激においても外部刺激⇒感情⇒購買意欲のモデルが成立すること、発話速度が最も快感に影響し、快感が最も強く購買意欲に影響することが分かった。また人の購買意欲に与える影響の70%を感情によって説明できることが分かった。

今後の展開について

これまでの研究によって、購買行動に置いて音声の特徴の違いが、受け手の感情や購買意欲に影響を与えているということが分かっている。一方で、「受け手の内面・行動変化」の理解に向けた課題は多くあると考えている。今後は、前節で示した消費者行動モデルが購買以外の行動に適用できるか検証が必要であると考えている(図5)。また、同一の音声刺激を聴取したとしても、全ての受け手が同じ内面・行動変化を引き起こすとは考えにくい。受け手の性別、年齢等に加え、趣味嗜好等を加味することで、「受け手の内面・行動変化」の理解の高度化に向けた研究開発を進めていきたい。

また得られた知見等を音声合成技術へ活用することで、現在の音声合成技術では難しい多様な表現力を持つ合成音声の生成の実現に向けた研究開発も進めていく予定である。

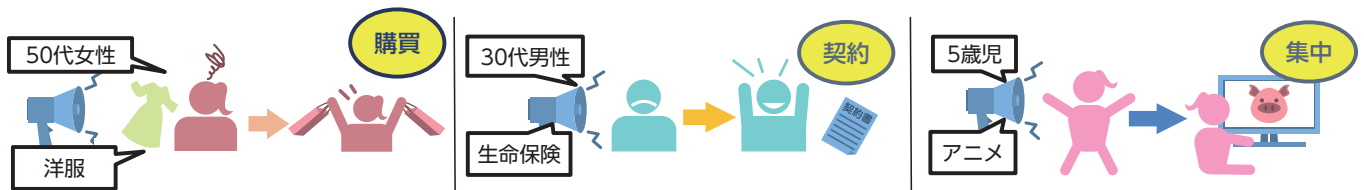


図5 今後の研究テーマの展開