

音声合成技術

任意の文章(テキスト)から音声を合成する技術を、テキスト音声合成技術あるいは単に音声合成技術と呼びます。情報源へのマルチアクセス・ユビキタス環境などの発展に伴い、音声での情報伝達の利便性が再認識され、音声合成技術に対するニーズが増大しています。NTTでは、このニーズに応え、音声対話エージェントや音声対話インタフェースを備えたロボットをはじめとして、電子書籍の読み上げや対話型のデジタルサイネージなど、さまざまな用途に利用できる、音声合成技術の研究開発を行っています。

ユーザの問いかけに応じてさまざまな情報を提供してくれる音声対話エージェントやロボットには、情報の内容を適切に伝えるだけでなく多様なユーザの好みに応じるためのさまざまなキャラクター性が求められます。このため合成音声には、ただ単にテキストの内容を読み誤らず自然性の高い音声で伝えるだけでなく、多彩なキャラクター性による豊かな表現力が求められるようになってきました。しかし、現在市場で主流の音声合成技術でユーザの好みに応じた新しい声質や口調を実現するためには、数時間以上の音声を事前に録音する必要があります。これではユーザの多様なニーズに対応することがコストの観点から難しくなります。そこで、より低コストで多様なユーザニーズに対応した合成音声を実現することを狙い、30分程度の音声で新しい声質や抑揚による合成音声を可能とする技術を開発しています。

●音声合成技術とその応用例

