

音声認識技術とその応用

岩崎知弘*
花沢利行*

要 旨

音声は人間にとって最も自然な情報伝達手段であり、音声インタフェースは、直感的で分かりやすく利便性も高い。この音声インタフェースとして実用化されている代表的分野には、CTI(Computer Telephony Integration)サーバ、カーナビ、音声ワープロなどが挙げられる。

CTIサーバは公衆向け情報サービスに用いられ、最近、ボイスポータルとしての応用が注目を集めている。このCTIサーバの場合、音声は、電話網を通ることによって帯域制限され、騒音やひずみ加わる。このような電話特有の劣化に対応するための電話音声認識技術について述べる。

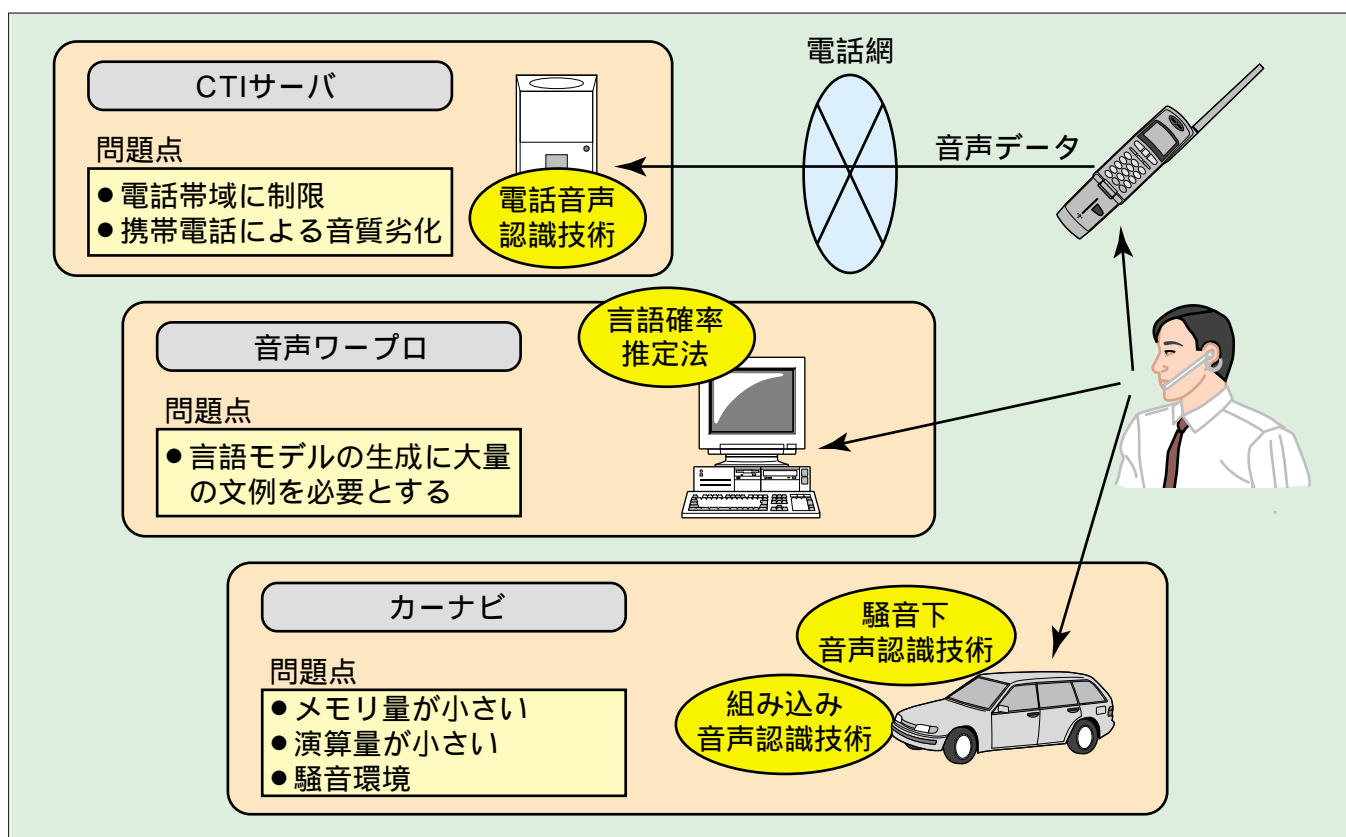
車の運転中は目と手が拘束されるため、カーナビ操作を音声だけで実現できる音声認識への期待は高い。しかし、カーナビは、目的地(住所)入力のための大語彙(ごい)の音声認識、高速走行騒音下の音声認識機能を、演算量やメモ

リが小さい環境で実現する必要がある。ここでは、これらの問題を解決する小さな演算リソースで動作する組み込み用途の音声認識技術と騒音下音声認識技術について説明する。

また、音声入力による文書作成への期待は依然として高い。この音声ワープロで不可欠の言語モデル作成のためには大量のコーパス(文例)が必要であり、従来は特定業務用に対応できなかった。

本稿では、少量のコーパスから容易に業務専用の言語モデルを作成できる言語確率推定法などのディクテーション技術について述べる。

三菱電機は、IT機器の高機能化に合わせて、今後、更に使いやすい音声インタフェースを提供していく予定である。



音声インタフェースの実用例

近年、様々な機器で音声インタフェースを備えるものが増加している。CTIサーバでは、電話音声認識技術を用い、電話網を通して送られてくる音声を認識することによって情報検索が可能となっている。最近増加しているカーナビでは、組み込み用音声認識技術と騒音下音声認識技術を用い、機器制御や目的地設定のために音声認識が用いられている。パソコン上でも音声認識機能によって音声ワープロを実現しているが、少量のコーパスから言語モデルを生成する言語確率推定法により、業務特化の音声ワープロを実現できる。