

博士論文要旨

論文題名：

外乱環境における音声認識性能予測に関する研究

立命館大学大学院情報理工学研究科
情報理工学専攻博士課程後期課程

ふりがな ふくもり たかひろ
氏 名 福森 隆寛

近年、万人にとって使い勝手の良い理想的な情報機器の操作環境として、音声を利用して情報機器を操作するハンズフリー音声インタフェースが注目されている。しかしながら、実環境において音声認識システムを利用すると、雑音や残響などの外乱が音声に混入することの影響を受けて音声認識性能が著しく劣化する。

この問題の解決策として、外乱による音声認識性能の劣化を事前に予測できれば、その予測結果に基づいて音声認識性能を改善するための手法を前処理等に反映させることで、常に最大限の音声認識性能を発揮させることが可能となる。また音声認識性能を予測することは、これまでの音声認識評価が必要であった大規模音声収録や音声認識処理を大幅に削減できる側面も有する。そこで本論文では、外乱環境（具体的には雑音・残響環境）における高精度かつ簡便な音声認識性能の予測手法について検討する。

具体的には、第 1 にテレビ会議システムなどの屋内で利用する音声インタフェースを想定し、これまで音声認識性能の予測としては不十分であった残響環境での音声認識性能の予測指標を提案する。ここでは、初期反射音と後続残響音の関係を表す室内音響指標の中でも特に Definition (D 値) に着目し、事前に様々な環境で複数箇所計測したインパルス応答を基に算出した D 値と音声認識性能の関係に基づいて残響指標を策定する。そして、策定した残響指標と音声認識性能の予測位置におけるインパルス応答を基に残響下音声認識性能の予測を試みる。多種にわたる残響環境において、音声認識性能の予測実験を行った結果、提案手法の有効性を確認した。

第 2 に、スマートフォンなどの屋内外で携帯するような音声インタフェースを想定し、複数の外乱要因が混在する環境において、高精度かつ簡便に音声認識性能を予測するための手法を提案する。本論文では、雑音・残響下における音声認識性能の予測精度を向上させるために、PESQ、室内音響指標の D 値と残響時間を用いた音声認識性能の予測式を提案する。ここでは事前に計測した発話音声やインパルス応答を用いて算出した PESQ、D 値、残響時間、音声認識性能から雑音・残響指標を策定する。そして雑音・残響指標を用いて性能予測を行う発話位置におけるインパルス応答と発話音声から音声認識性能の予測を試みる。評価実験の結果、従来の雑音指標・残響指標を個別に用いて性能予測する手法よりも提案指標は頑健に雑音・残響下音声認識性能を予測できることを確認した。