

研究題目	屋外拡声音声品質の文章了解度をブラインド予測する手法の開発	報告書作成者	小林洋介
研究従事者	小林洋介		
研究目的	2011年の東日本大震災では、津波情報などの屋外拡声器による放送が聴き取りにくい事例が多く、屋外拡声システムの改善による災害時の QOL 改善が検討されている。しかし、既存システムの聴き取りやすさの調整には、多くの被験者による聴取実験が必要であり、技術者の勘と経験だけで行うには限界がある。このため、何らかの客観的な物理指標を用いた機器の設計開発が必要である。 音声の了解度や明瞭度を予測する物理的な指標の一つに主として屋内の残響のある空間での了解度・明瞭度指数の STI (Speech Transmission Index) が ISO と IEC で技術標準となっており、計測器も発売されている。しかし、STI は基準信号を放送し、その伝送特性による信号の変化を全て劣化として計算を行うため、イコライザ等による音声強調を利用した音声了解度の改善を計測することができない。言い換えれば、基の音声が入り込みの少ない聴き取りにくい音声であっても、無歪で伝送できれば STI は最も高いスコアになる。このため、音声信号処理による了解度改善を志向したシステムの評価のために、騒音計のように観測信号のみから音声の品質値をブラインド予測する計測器が必要である。 本研究では、様式-5 の図 1 に例示したような、音声了解度の予測器を開発する。具体的には、機械学習による非線形回帰を利用して、拡声システムから放送された音声をシミュレーションした評価信号の音響特徴量と主観評価値から了解度の予測モデルを作成し、その性能評価を行う。本研究で提案する予測モデルを将来の新システムの導入や現行のシステムの改良に利用することで、より聴き取りやすい屋外拡声システムの構築を補助し、国民生活全体の QOL 向上を目指す。		

研究内容

提案予測法の概要

本研究では図 1 の様に、拡声器から放送される通常放送などを利用して了解度を予測することが可能な計測器の予測モデル(非線形の予測回帰式)の作成を行う。予測モデルの内部処理は、録音した音声をセグメンテーション(時間分割)し、セグメントごとにブラインド音響特徴量(インパルス応答等の別の音源や、伝送前の原音声を利用しない音響特徴量)を計算し、セグメントごとに単語了解度の予測値を機械学習による単語了解度予測モデルより求め、予測単語了解度の時系列データを作成する。次に予測単語了解度の時系列データを非線形回帰により統合した文章了解度予測モデルより、文章了解度の予測値を得ることを目指す。

提案予測法のポイント

録音信号から求める特徴量と最終的に利用する文章音声了解度の間に単語了解度の予測値を入れることで、人間の音声知覚を模擬している点が STI 等の伝送経路の品質評価指標と決定的に異なり、より精度の高い予測法になると考えられる。また、屋外拡声系の評価では、ロングパスエコーと呼ばれる別の拡声器から遅れて聴こえてくる拡声音声が先行する拡声音声の聴取を妨害することが知られており、東日本大震災でも問題になっている。このエコーは、主音声とかなり近い成分を持った妨害音姓であり、音声信号処理において除去することは難しいが、エコーは必ず遅延して伝搬するため、時系列データに特徴が出ると予想され、この点からも提案法は予測精度が高くなり得ると考えられる。

本プロジェクトの内容

本プロジェクトでは、最終的な文章了解度予測に用いる 2 つのモデルのうち、図 2 に示す方式で単語了解度の予測モデルを構築した。年度前半に文章音声をういた屋外拡声器を模擬した主幹評価を行い、単一の劣化要因よりも複数の劣化要因が複雑に絡み合うことで了解度を低下させていることを明らかにした。一方で、最も聴き取りにくくさせているとされているロングパスエコーについては、機器の設計開発を行っているメーカー技術者との情報交換より、了解度評価を行うまでもなく悪いところは設計で改善していくという方向になっているとの話であった。本プロジェクトの最終的な了解度予測モデルは屋外拡声音声のあらゆる劣化に対応させることを想定してはいるが、予備実験の検討より、全てを考慮した予測モデルでは、主観品質との相関が高いモデルは構築できる見込みが薄くなった。そこで、若干方針を変更してより基礎的な検討を行うため、電話帯域への帯域制限」及び「単一の拡声器の伝送特性」の二つのみに絞った「単語」了解度評価値の予測モデルの構築を平成 27 年度の研究課題に設定した。

研究のポイント	新しく提案する図 2 の単語了解度予測システムの処理フローは、学習用の主観評価を複数被験者に対して行い、その正答と誤答のレスポンスを模擬する判別器を被験者の数だけ作成し、主観評価と同様の平均処理によって予測値を得る。判別器は SVM(support vector machine)とし、ハイパーパラメータを学習データで最適化する。個々の判別器に用いるノンブラインド音響特徴量は、ITU-T P.563 勧告の内部特徴量 12 次元と拡声器のインパルス応答を適応した音声認識システムの尤度スコアの計 13 次元を用いる。本稿では被験者 22 名分の評価結果を学習し、判別器を 22 個構築し、その出力を予測了解度とする。
研究結果	主観評価値と音声認識結果 提案システムの評価音声は、親密度別単語了解度試験用音声データセット 2007(FW07)の高親密度語女性 1 話者分用いた。FW07 の音源に、東日本大震災後に仙台市若林区荒浜小学校周辺で収集されたスピーカによる TSP 信号を 10 地点分畳み込み評価音源を作成した。以後、各地点を p01～p10 とする。また、実際の拡声音声システムでは、無線通信等で電話帯域に帯域制限されて伝送されているため、原音声を電話帯域に帯域制限した後に伝送系のインパルス応答を計算機上で畳み込んだのちに、音量のゲイン調整を行った。評価は予測モデルの作成用データとテスト用データに分けるため 2 セット行った。Set 1 は被験者 22 名分、Set 2 は Set 1 と異なる評価リストを用いた被験者 13 名分であり、被験者のほとんどは 10 代後半の高専生である。 音声認識は Julius を使い、ASJ-JIPDEC の音素バランス文にインパルスで 10 地点分畳み込んだ音源を学習させて適応モデルを構築した。適応モデルの比較対象には Julius の標準設定をベースラインモデルとして比較する。構築した認識システムの基本的な性能として、FW07 の認識率を正答率 Corr.と挿入誤差を考慮した Acc.で評価する。 屋外拡声器への適応を行った音声認識システム(Adap.)と適応を行っていないベースライン(Base.)の評価結果と主観評価のセット別の平均了解度(Intell.)を表 1 に示す。結果より、Corr.は適応により下がっているが、より実態に近い認識率指標である Acc.は拡声器への適応で大幅に改善している。適応モデルの Acc.と主観評価による了解度の相関係数を求めると、0.41 と若干の相関はあるが、人間の了解度評価による知覚モデルとの差は未だ大きく、今後の改良が必要である。 予測モデルの構築と精度検証 音声認識の出力結果の相関係数が 0.5 未満であり、現状の認識結果のみでは了解度の予測は困難であるため、音声認識計算に用いる対数尤度スコアを利用する。これは、認識率である Corr.と Acc.は正答のテキストが必要な指標であるため、観測された音声信号のみでは求められないためである。図 2 の判別器は SVM(support vector machine)とし、ハイパーパラメータを学習する set 1 の特徴量と評価値で最適化する。個々の判別器に用いるノンブラインド音響特徴量は、音声認識の対数尤度の他に、昨年度も用いた ITU-T P.563 勧告の内部特徴量 12 次元も用いる。学習する被験者数が多い方が予測精度が高くなると予想し、被験者の多い set 1 の結果を学習データ、もう一方の set 2 をテストデータとした。性能評価指標には、予測値と主観評価値との相関係数と平均二乗誤差 RMSE を用いる。 予測結果を表 2 に示す。学習したデータ(主観評価 Set 1 の予測)の結果を Train に、学習していないテストデータ(主観評価 Set 2 の予測)の結果を Test に示す。結果より、提案法はテストデータの相関が 0.724 と高く、RMSE も 3.5%と十分に小さい。図 3 に提案法の主観値と予測値をインパルス応答ごとにプロットした。図より、学習したデータはほぼ対角線上にあり、テストデータが対角線にマップされており、本予測法のロバストさがわかる。

今後の課題	<u>文章了解度への拡張</u> 当初検討した文章了解度の予測への適応を見送り、単語了解度のみとした。これは現実的に研究成果を上げていく上では必要なスケールダウンであったが、実際のシステムへの適応を考慮するとあまり良い判断とは言えない。現在は予備実験として取得した文章了解度の主観評価値を用いて図1のシステムを構築している。しかし、動的な特徴量の変化について未考慮であるため、P.563と音声認識の尤度の時系列変化をより考慮する必要がある。 <u>実音源の利用</u> 本プロジェクトではインパルス応答こそ実測のデータを利用しているが、屋外録音した音声を利用していない。しかし、本プロジェクトの成果をもとに拡声器メーカーとの共同研究計画が進展しており、実音源を用いた提案法の評価の可能性が出てきた。今後はシミュレーションによる予測モデルの構築と、実音源を用いたシステムの評価の両立を目指して研究活動を続けていきたい。 <u>計測器としての利用</u> 本プロジェクトは最終的に了解度を計測する計測器の開発を目指している。このため、提案法の実時間動作を目指した組み込み処理が必要である。少なくとも本助成で完成した単語了解度予測モデルについてはマイクロホンから入力した音声について(制度は別にして)評価値を出すところが現実的には可能であるため、計測器のプロトタイプとして利用可能な形に進展させたい。
-------	--

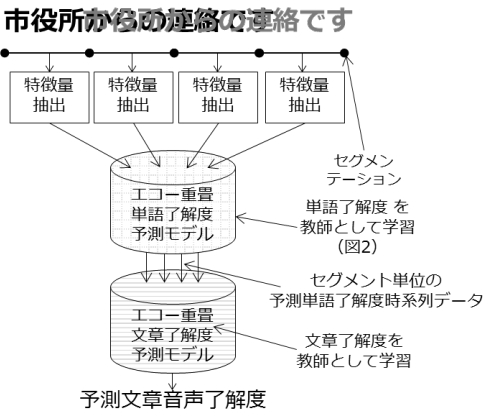


図 1:提案する理解度予測モデル

表 1:単語理解度と音声認識率

	Intell.(%)		Base.(%)		Adap.(%)	
	Set1	Set2	Corr.	Acc.	Corr.	Acc.
p01	92	91	48.8	-4.6	41.3	32.5
p02	79	74	44.0	0.6	38.1	27.8
p03	97	93	46.4	-13.4	38.7	31.1
p04	88	78	47.7	-5.3	39.4	28.6
p05	91	83	46.3	-6.9	37.9	29.7
p06	90	92	46.9	-14.4	37.8	26.4
p07	92	83	46.1	-11.9	36.4	28.4
p08	94	80	46.4	-10.6	36.5	25.7
p09	87	91	45.2	-8.1	37.8	29.4
p10	85	87	49.1	-4.5	37.7	28.0

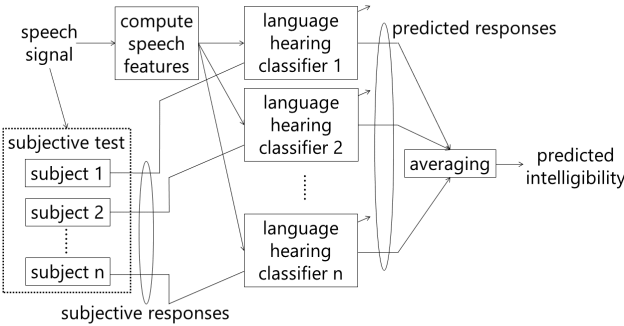


図 2:単語理解度予測モデル

表 2:予測値

	RMSE	<i>r</i>
Train	0.040	0.940
Test	0.035	0.724

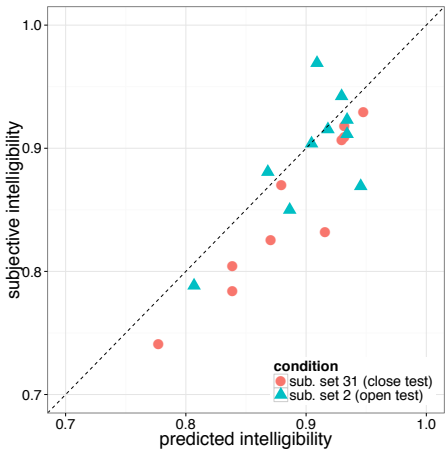


図 3:予測性能

(注:フローチャート図, ブロック図, 構成図, 写真, データ表, グラフ等 研究内容の補足説明にご使用下さい。)