

令和元年6月7日現在

機関番号：10103

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K21584

研究課題名（和文）了解度計を実現するための時系列解析を利用した文章了解度予測モデル

研究課題名（英文）Sentence intelligibility prediction model using time series analysis for hand-held intelligibility meter

研究代表者

小林 洋介（Kobayashi, Yosuke）

室蘭工業大学・大学院工学研究科・助教

研究者番号：10735103

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：音声の主観的音声品質評価は、public-address (PA)システム、音声マスキングシステム、補聴器、携帯電話などの開発に広く使用されています。しかし屋外のPAシステムなど、評価の難しい環境で利用される場合はその客観評価が難しい。そこで、機械学習を含む客観的了解度評価システムを提案する。本研究で提案したシステムを屋外で録音した信号による実験でこのアプローチの有効性を検証することができた。

研究成果の学術的意義や社会的意義

駅や空港、学校など音声を放送する拡声システムは様々なところに実装されているが、必ずしも聴き取りやすいとは言い難い。本研究ではより聴き取りやすい拡声システムの設計に利用可能な音声の了解度（聴き取りやすさ or 聴き取りにくさ）の計測器を機械学習技術を利用して開発する。これにより身近な音声システムが聴き取りやすく最適化されることで国民生活の向上を目指す。

研究成果の概要（英文）：Subjective speech quality assessment has been used widely for the development of public-address (PA) systems, speech masking systems, hearing aid, mobile-phone, etc. However, as this assessment is difficult in many cases, we propose an objective speech intelligibility evaluation system that includes a machine learning technique. Through outdoor-field recorded signal experiments, the effectiveness of these approaches could be verified.

研究分野：音声情報処理

キーワード：音声了解度 音声明瞭度 機械学習 屋外拡声システム PAシステム 予測モデル

1. 研究開始当初の背景

ICT技術の発展により、あらゆるところで音声を提供できるようになり、これまでにない音声システムの利用形態が拡大している。特に暗騒音や残響があり静かではない屋外での利用が増え、その聴き取り易さが重要な品質指標となっている。音声の聴き取り易さについては Fletcherらの主観評価（聴取実験）による単音節明瞭度試験[H. Fletcher and J.C. Steinberg, 1929]にはじまり、その後単語音声を利用した単語理解度試験が開発され、表1のように利用されている。理解度試験は被験者実験であり、計測にコストがかかるため、図3のように基準音信号の伝送特性から理解度を予測する STI (Speech Transmission Index)[H.J.M. Steenken, 1980]が標準化された。STIは屋内の残響環境での音声品質指標だが、様々な環境での音声品質評価にも利用されている。しかし、図1のロングパスエコー問題（以下、LE評価系）では、基準音信号の計測に雨風や環境騒音などの影響で自由に計測できず[栗栖他, 2013]、図2のスピーチプライバシー保護システム（以下、SP評価系）の性能評価では、主音声とマスキングの話者特徴の近さによって聴き取りにくくする情報マスキング[N. I. Durlach et al., 2003]であるため、物理指標だけではそもそも十分な予測ができないといった問題がある。

これらの問題を解決するため、騒音計と同様に計測器が観測した音のみから理解度予測を行う必要がある。先行研究[Y. Takano, K. Kondo, 2010, T. Arai, 2013]では、伝搬音声の音声認識結果から理解度予測を試みたが、適応条件が厳しく現状では十分な予測性能は得られていない。

応募者はこれまで、音声システムの理解度評価及び機械学習による回帰分析を用いた騒音下での単語理解度予測を研究してきた。これらの経験を基に、LE評価系の文章理解度予測への応用を試みたが、予測値は主観値に対し無相関となった。しかし、別の実験で拡声音声自体の単語理解度予測は高精度であることを確認した。この結果より、音響特徴量から高精度に予測できるのは単語理解度であり、文章理解度予測のためには、単語理解度予測値の時系列分析による人間の知覚の模倣が必要であると考えた。そこで、図4に示す文章音声を短時間のセグメントに分割した予測理解度の時系列データ解析を着想した。

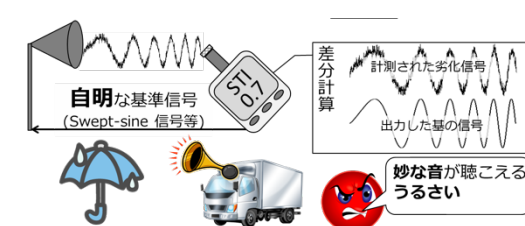
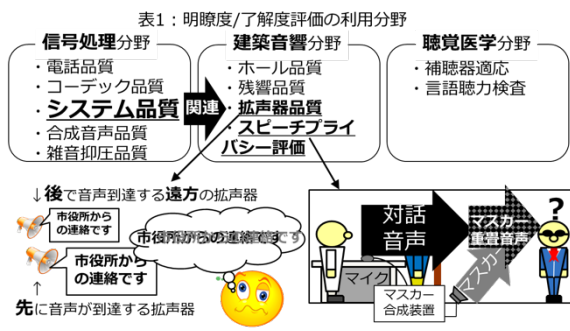


図1 屋外拡声器のロングパスエコー問題：東日本大震災で放送が聴き取れない問題が発生した。

図2 スピーチプライバシー保護システムの例：自己音声を含む音を合成し、情報漏えいを防ぐ。

図3 従来の理解度予測：原音が既知のノンブラインド（信号の劣化が全てわかる条件）予測であり、基準音の放送が必要で、気象条件や、大型車の走行音など突発性の外乱に弱く、住民の理解も必要。

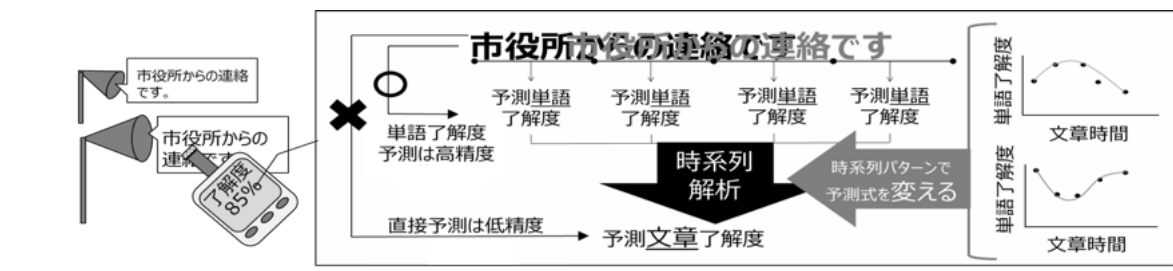


図4 提案する「理解度計」のイメージ：原音がわからなくても良いブラインド予測で、予測単語理解度の時系列データをさらに非線形回帰して予測文章理解度を求める(LE評価系での例)。

2. 研究の目的

本研究では、実際に放送される音声に近い文章音声を用いた理解度計に相当する機器の開発を目的とした。申請時は文章理解度による評価を想定していたが、その後の屋外拡声機器を設計開発している企業研究者らとの情報交換の結果、主観的な聴き取りにくさ指標である LDR(Listening Difficulty Rating) [M. Morimoto et al., 2003]での評価をし始めているとのコメントを得た。このため、マスキングシステム、屋外拡声システム共に単語理解度と文章理解度は LDR の評価と予測モデル構築を行うこととした。このため、文章理解度評価法の設計に関する時間が、図4に示した理解度計測器のプロトタイプの開発も行う。

3. 研究の方法

本報告では、目的に示した理解度計測器のプロトタイプ的设计について詳細を報告する。具体的にはハンドヘルドな計測機器とするため、小型計算機でも計算可能なモデル構築を行う。具体

的には聴覚特性を考慮した音響特徴量である MFCC (Mel Frequency Cepstrum Coefficients) の時系列データを利用し、計算コストが線形時間で設計しやすい機械学習法の 1 つである RF (Random Forest) を利用したモデルとした。ここでの目的変数となる主観評価指標は文章理解度ではなく LDR を用いた。LDR は先行研究より、単語理解度が十分であっても聞き取りにくい範囲を計測可能であることが示されており、定義が曖昧な文章理解度を利用するよりも良いと考えた。さらに、リアルタイム動作時には、今解析している信号が音声かどうかの判別が必要である。そこで、VAD (Voice Activity Detection : 音声区間判別) も RF で実装し、音声区間の LDR を求めることとした。

#### 4. 研究成果

図 5 左に提案する計測器のフローを示す。この計測器は、2 台のシングルボードコンピュータを用い、屋外に設置されている拡声器等から放送される音声を録音し、リアルタイムに LDR 予測結果を表示する。実機システムには、マイクロホンに入力された音源をオーディオインタフェースを通して AD 変換し、Board 1 で 1 秒ごとに録音し、フレーム長を 100 ミリ秒として 12 次元の MFCC とパワーおよび、delta パラメータを算出する。次に、Board 2 で学習済みモデルを使用して、VAD と LDR 予測を行う。最後に、Board 1 の LCD ディスプレイに VAD の結果と予測 LDR を表示する。開発したシステムの外観を図 5 右の写真に示す。

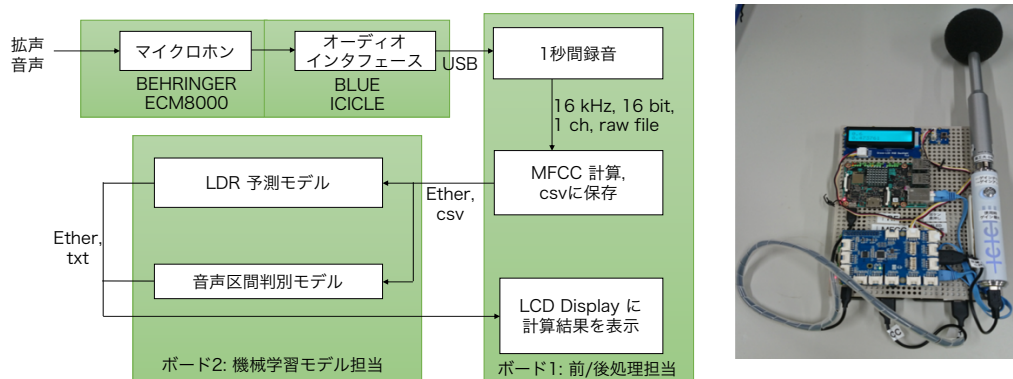


図 5 提案システムのフローと写真

VAD と LDR 予測を行うため、機械学習を用いた。機械学習を行う際、多数の音源が必要である。また、本計測器で用いるマイクロホンの特性を考慮した音源も必要である。そこで、LDR 学習音源と VAD の学習音源を作成するために、室蘭工業大学 V/R 棟前広場でインパルス応答を取得した。このインパルス応答を畳み込んだ主観評価音源は 160 音作成した。主観評価は防音ブース内でラップトップマシンに接続したオーディオインタフェースからヘッドホンを用いてダイオティックに被験者へ提示した。被験者は日本語話者 20 代 19 人である。本実験は室蘭工業大学研究倫理審査委員会の承認のもと行なわれた。

主観評価音源を 1 秒ごとに切り出し、音響特徴量を 100 ミリ秒ごとに求めて、LDR を予測する。よって、1 秒の音源より 260 次元のモデル入力特徴量が求まる。全音源の約 70% をランダムに選びトレーニングデータとしてモデルの学習に、残りの約 30% を予測精度のオープンテストに用いた。シングルボードコンピュータに実装すること、音源数が 1136 と少なく、RF による VAD モデルより決定木の深さが浅くなることを考慮し、RF の決定木の数を 1000 とした。作成したモデルの RMSE は 0.20、相関係数は 0.88、決定係数は 0.74 とまだ誤差は残るものの、聞き取りにくさの傾向は概ね掴むことができるモデルが構築できた。今後は入力特徴量の見直し、深層学習の利用などによるモデルの高度化が必要である。

#### 5. 主な発表論文等

[雑誌論文] (計 2 件)

- [1] Keita Noguchi, Yosuke Kobayashi, Jay Kishigami nad Kiyohiro Kurisu, “Listening difficulty estimation model using short-time objective intelligibility measure for outdoor public address systems,” *Acoustical Science and Technology*, 2 pages, accepted, 査読有
- [2] Yosuke Kobayashi and Kazuhiro Kondo, “A Case Study of Boot-strap Masker Quality Assessment for Speech-Privacy Protection,” *Journal of Information Hiding and Multimedia Signal Processing*, Volume 8, Number 6, pp. 1460-1469, November 2017, 査読有

[学会発表] (計 22 件)

- [1] Keita Noguch, Yosuke Kobayashi, Jay Kishigami and Kiyohiro Kurisu, “Listening

- difficulty estimation model using short-time objective intelligibility measure for outdoor public address systems,” Abstract of International Symposium on Universal Acoustical Communication 2018 (UAC2018), p.46, Oct.22-24, 2018, Sendai, 査読有
- [2] Keita Noguch, Yosuke Kobayashi, Jay Kishigami and Kiyohiro Kurisu, “Listening Difficulty Meter Using Machine Learning for Assessment of Public-Address Systems,” Proc. The 7th IEEE Global Conference on Consumer Electronics (GCCE 2018), pp. 471 - 472, Nara, Japan, 2018. DOI: 10.1109/GCCE.2018.8574812, 査読有
- [3] Hirokazu Akadomari, Kosuke Ishikawa, Yosuke Kobayashi, Kengo Ohta and Junichi Kishigami, “HMM-based Speech Synthesizer for Easily Understandable Speech Broadcasting,” Proc. The 7th IEEE Global Conference on Consumer Electronics (GCCE 2018), pp. 714 - 715, Nara, Japan, 2018. DOI: 10.1109/GCCE.2018.8574710, 査読有
- [4] 赤泊寛和, 石川耕輔, 太田健吾, 小林洋介, 岸上順一, “「よく聴こえる」拡声システムのための特定話者に適合した HMM 音声合成システムの評価,” 日本音響学会 2018 年秋季研究大会, 1-R-40(Poster), pp.1183-1184, 大分県大分市, (2018 年 9 月 12-14 日), 査読無
- [5] 野口啓太, 小林洋介, 岸上順一, 栗栖清浩, “屋外用聴き取りにくさ計プロトタイプ of 改善,” 日本音響学会 2018 年秋季研究大会, 1-Q-25(Poster), pp.437-440, 大分県大分市, (2018 年 9 月 12-14 日), 査読無
- [6] 小林洋介, 野口啓太, 栗栖清浩, “拡声音声への環境情報フィードバックシステムの基礎検討,” 聴覚研究会資料, vol. 48, no. 6, H-2018-106, pp. 591-597, 仙台市 (2018 年 8 月 23-24 日), 査読無
- [7] 小林洋介, “雑音下音声了解度と客観的音声了解度指標との関係,” 電子情報通信学会技術研究報告, vol. 118, No. 149, EA2018-1, pp. 1-6, 札幌市 (2018 年 7 月 24-25 日), 査読無
- [8] 赤泊寛和, 石川耕輔, 太田健吾, 小林洋介, 岸上順一, “HMM 音声合成と発話構文解析を利用した「よく聴こえる」拡声システム,” 情報処理学会研究報告, Vol.2018-MUS-119 No.51, 4 pages, 東京都文京区, (2018 年 6 月 16-17 日), 査読無
- [9] 野口啓太, 小林洋介, 岸上順一, 栗栖清浩, “屋外拡声システムの主観的聴き取りにくさの客観計測器の提案,” 情報処理学会研究報告, Vol.2018-MUS-119 No.38, 4 pages, 東京都文京区, (2018 年 6 月 16-17 日), 査読無
- [10] 野口啓太, 小林洋介, 岸上順一, 栗栖清浩, “屋外用聴き取りにくさ計のプロトタイプ開発,” 日本音響学会 2018 年春季研究大会, 2-P-3(Poster), pp.659-660, 埼玉県南埼玉郡宮代町, (2018 年 3 月 13-15 日) 学生優秀発表賞受賞, 査読無
- [11] Keita Noguch, Yosuke Kobayashi, Jay Kishigami and Kiyohiro Kurisu, “Listening Difficulty Meter Using the ITU-T P.563 Feature Set for Public-Address System,” Abstract of Joint Seminar on Environmental Science and Disaster Mitigation Research 2018 (JSED2018), pp.73-74, Mar. 2, 2018, Muroran, 査読無
- [12] Yosuke Kobayashi, Kosuke Ishikawa, Kengo Ohta and Junichi Kishigami, “Semantic Parser for Easy Understandable Speech Broadcasting,” Proc. The 6th IEEE Global Conference on Consumer Electronics (GCCE 2017), pp. 28 - 29, Nagoya, Japan, 2017. DOI: 10.1109/GCCE.2017.8229362, 査読有
- [13] 児島悠太, 小林洋介, “音環境センサネットワークの基礎検討,” 日本音響学会 2017 年秋季研究大会, 1-P-40(Poster), pp.749-750, 愛媛県松山市, (2017 年 9 月 25-27 日), 査読無
- [14] 児島悠太, 小林洋介, “音環境センシングシステムの開発,” 電子情報通信学会技術研究報告, vol. 117, No. 138, EA2017-22, pp. 125-130, 札幌市 (2017 年 7 月 20-21 日), 査読無
- [15] 石川耕輔, 太田健吾, 小林洋介, 岸上順一, “発話構文解析を利用した「よく聴こえる」拡声システムの基礎検討,” 情報処理学会研究報告, Vol.2017-MUS-115 No.47, 6 pages, 東京都文京区, (2017 年 6 月 17-18 日), 査読無
- [16] 小林洋介, 近藤和弘, 坂本修一, “実環境を模擬した少数サンプルによる了解度推定,” 1, 1-P-28, pp.583-586, 神奈川県川崎市, (2017 年 3 月 15-17 日), 査読無
- [17] Yosuke Kobayashi, Kengo Ohta, Kazuhiro Kondo and Shuichi Sakamoto, “Speech intelligibility prediction method using machine learning for outdoor public address systems,” 5th Joint Meeting Acoustical Society of America and Acoustical Society of Japan, Journal of Acoustical Society of America, Vol. 140, No.4, Pt. 2 of 2, p. 3192, (Published: Oct. 2016), Presentation: 28 Nov.-2. Dec. 2016, Invited, 査読無
- [18] Kazuya Taira, Yosuke Kobayashi and Kazuhiro Kondo, “Estimation of binaural speech intelligibility based on better ear model,” 5th Joint Meeting Acoustical Society of America and Acoustical Society of Japan, Journal of Acoustical Society of America, Vol. 140, No.4, Pt. 2 of 2, p. 3215, (Published: Oct. 2016), Presentation: 28 Nov.-2. Dec. 2016, 査読無
- [19] 小林洋介, 近藤和弘, “主観・客観評価値を用いたスピーチマスキングシステムの品質推

- 定, ” 電子情報通信学会技術研究報告, vol. 116, no. 302, EA2016-59, pp. 67-72, 大分市 (2016年11月17-18日), 査読無
- [20] 小林洋介, 近藤和弘, 坂本修一, ” 雑音下での少数サンプルによる了解度推定, ” 聴覚研究会資料, vol. 46, no. 7, H-2016-42, pp. 471-476, 七尾市 (2016年10月14-15日), 査読無
- [21] 小林洋介, 近藤和弘, 坂本修一, ” 少数サンプルによる了解度推定法の詳細分析, ” 日本音響学会 2016年秋季研究大会, 2-P-30, pp. 627-630, 富山県富山市, (2016年9月14-16日), 査読無
- [22] 小林洋介, 太田健吾, 近藤和弘, ” 機械学習と音声認識による拡声音声品質予測, ” 情報処理学会研究報告, Vol.2016-MUS-111 No.43, 4 pages, 東京都品川区, (2016年5月21-22日), 査読無

## 6. 研究組織

### (1) 研究分担者

なし

### (2) 研究協力者

研究協力者氏名：近藤和弘

ローマ字氏名：KONDO Kazuhiro

研究協力者氏名：坂本修一

ローマ字氏名：SAKAMOTO Shuichi

研究協力者氏名：太田健吾

ローマ字氏名：Ohta Kengo

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。