# 発話音声の聞き取りやすさ向上のための音声特徴量解析

佐賀 圭真[1,a)]　井村 誠孝[1,b)]

**概要**：動画配信などで，一般人が自身の音声を多数の人間に届ける機会が増えている．しかし，話者の声が聞き取りにくいと，聴者は何をしゃべっているのかに集中して聞く必要があり，話者の魅力が損なわれてしまう．本研究では, 発話された音声の特徴量と人の評価する聞き取りやすさに着目し，音声の聴き取りやすさを向上するシステムの構築を行うことを目的としている．システム構築のために，音声特徴量と聞き取りやすさの評価との間の関係性を調査した．主観評価によって音声の聞き取りやすさの評価値を得て，音響特徴量と評価値との相関係数を見たところ，高い相関のある音響特徴量の存在を確認できた．

## 1. はじめに

一般人がメディアを通して自身の音声を多数の人間に届ける機会が増えている．「YouTuber」という言葉がよく見られるようになったことからも，動画配信サービスや SNS などを利用した動画配信や生放送が活発に行われるようになってきていることがわかる．一般人が配信する動画を視聴する人も増えており，多くの人が楽しむものとなっている．しかし，話者の声が聞き取りにくいものであると，聴者は何をしゃべっているのかに集中して聞く必要があり，コンテンツ本来の魅力が損なわれてしまう．配信の環境では，直接対面しての会話と違い，聞き取りにくいことを伝える際に時間がかかることや，他の聴者の存在もあり，聞き返すことは難しい．話者の声の聞き取りやすさを改善することで声の聞き取りに対する聴者の負担が軽減され，より多くの人が動画の配信，視聴を楽しめるようになる．

聞き取りやすさに関する要因は，環境による外部要因と発話者による内部要因の二つに分けられる．外部要因は，環境音や残響音など，聞きたい音に外から別の音が加わることによる聞き取りやすさの要素であるのに対して，内部要因は，話し方や滑舌など，発話された音声そのものが持つ聞き取りやすさの要素である．本研究では，内部要因に重きを置き，発話特徴に着目した聞き取りやすさ向上システムを設計し実装することで，聞き取りにくい音声を聞き取りやすくして，コミュニケーションの円滑化を目指す．

## 2. 関連研究

聞き取りやすさに関する従来研究としては，音声信号の聞き取りやすさを評価する手法の確立を目指した基礎研究と，聞き取りやすさの向上のための音声信号処理に関する研究がなされている．

聞き取りやすさを評価する手法は，SD（Semantic Differential）法や一対比較法による直接的な主観評価が一般的であるが，聞き取りの正解率に基づく文章・単語了解度 [1] や，物理的指標を用いる方法 [2] も提案されている．元の音声信号が存在する場合は，SN（signal-to-noise）比などの信号間の比較により得られる指標が有用である [3] が，元の音声信号にどれだけ余計な信号が上乗せされているかを測る指標にすぎず，元の音声信号の必要性を排除できない．元の音声信号が存在しない場合でも，変調エネルギーと評価実験で得られたラベルを用いた LSTM ネットワークによる音声品質予測 [4] などにより信号の明瞭度を求めることができるが，どの程度雑音や残響の影響を抑えられているかの指標である．本研究では元の音声信号そのものの聞き取りやすさの評価をしたいため，いずれも本研究で適用できる聞き取りやすさの指標とはなりえない．

音声信号処理によって聞き取りやすさを向上するための取り組みとしては，時間領域での処理である話速の変換 [5] や，周波数領域での処理であるスペクトルの時間変化強調 [6] など，様々なアプローチがなされている．

## 3. 提案システム

提案システムは，音声信号を入力とし，聞き取りやすさを向上させる処理を施したのち，処理後の音声信号を出力する．入力された音声信号から聞き取りやすさにかかわる音響特徴量を抽出し，音響特徴量のパラメータを適切に変更し，パラメータの変更が適用された音声信号を出力することで，音声の聞き取りやすさを向上させる．

[1] 関西学院大学
[a)] egj93497@kwansei.ac.jp
[b)] m.imura@kwansei.ac.jp

聞き取りやすさを向上させる処理の実装にあたり，どの音響特徴量が聞き取りやすさにかかわるものなのかを知る必要がある．本研究では，主観評価実験を通じて，必要な知見を収集する．

## 4. 聞き取りやすさと音響特徴量の関係の調査

本節では，提案システムの実装に必要な，聞き取りやすさと音響特徴量との関係に関する調査について述べる．

本研究では，主観評価によって音声データに対する聞き取りやすさの評価値を決定し，音声データから得られる音響特徴量と聞き取りやすさの評価値との間の相関をとることで音響特徴量と聞き取りやすさの対応関係を調べる．

### 4.1 音声データ

調査に用いた音声は，発声に不自由を感じていない男性4名に，定型文を読み上げてもらい録音することで準備した．読み上げる文章はATR音素バランス503文[7]Aセットより以下の3文を選択した．

- 救急車が十分に動けず救助作業が遅れている．
- 言論の自由は一歩譲れば百歩も千歩も攻めこまれる．
- ちょっと遅い昼食をとるためファミリーレストランに入ったのです．

発話者の感覚での「普通」,「速め」,「遅め」の3種類の話速で上記3つの文章をそれぞれ発話してもらい，録音した．録音時のサンプリング周波数は44.1 kHzで，量子化ビット数は16 bitであった．1つの文につき12音声，合計36音声を得た．

### 4.2 聞き取りやすさの主観評価

実験協力者に対して，基準となる音声（基準音声）と評価対象の音声（対象音声）の二つを聞いてもらい，複数の評価項目に対して対象音声が基準音声と比べてどうであったかを5段階で評価してもらった．

評価項目は「声の高さ」,「声の速さ」,「声の大きさ」,「声の抑揚」,「声のこもり具合」,「聞き取りやすさ」,「滑舌」,「メリハリ」,「間の取り方」,「流暢さ」の10項目とした．-2から2までの5段階で評価してもらうにあたり，評価項目に対して対象音声が基準音声と同程度に感じた場合には0と評価するよう指示した．基準音声，対象音声ともに同じ文章を読み上げたものであり，評価の決定まで自由に聞き直すことが可能とした．音声の評価は，文章ごとに分けて行った．提示順はランダムとし，対象音声1つにつき1回ずつ評価を行ってもらった．実験協力者は聞こえに関して不自由を感じていない男女10名であった．各音声について，10名の評価値を平均し，各項目の評価値を得た．

### 4.3 音声信号からの特徴量抽出

#### 4.3.1 音声データの分割

入力した音声データを，開始時刻をずらした窓関数によってフレーム列に分割し，音響特徴量を計算した．フレーム幅は25msとし，開始時刻は10msずつずらすものとした．

#### 4.3.2 音響特徴量

本研究では以下の音響特徴量を使用した．

- 振幅の二乗平均平方根値（RMSenergy）
- メル周波数ケプストラム係数（MFCC）（12次元）
- 零交差率（ZCR）
- 基本周波数（F0）

以上15種類の特徴量にそれぞれの特徴量の1次微分したものを加えて，合計30種類の特徴量を使用した．

#### 4.3.3 統計量の計算

得られた特徴量を3つの連続するフレーム列で移動平均により平滑化した後，以下の12種類の統計量を計算をした．

- 算術平均，標準偏差，歪度，尖度
- 最大値，最大値を出力した位置，最小値，最小値を出力した位置，最大値と最小値の差
- 線形近似の勾配度，線形近似のオフセット，線形近似の二乗誤差

以上30の特徴量に12種類の処理が適用され，「特徴量-微分したものか-処理」の組み合わせにより合計360個の特徴量を各音声から抽出した．

### 4.4 結果

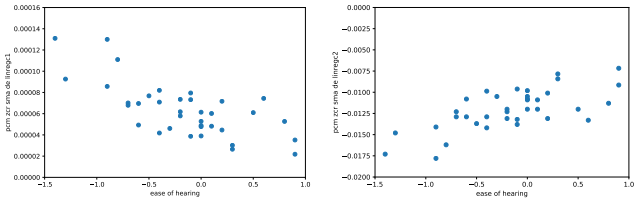
音響特徴量の抽出にはオープンソースソフトウェアであるopenSMILE[8]を用いた．各音声データに対応する音響特徴量と聞き取りやすさの評価値で相関をとった．表1に抽出した音響特徴量と聞き取りやすさの評価値との相関係数の一部を示す．大きな相関を得たのはZCR-微分-線形近似の勾配度，ZCR-微分-線形近似のオフセット，ZCR-標準偏差，ZCR-線形近似の二乗誤差，MFCC（1次）-微分-算術平均，MFCC（10次）-微分-歪度などである．

### 4.5 考察

零交差率に関するいくつかの特徴量で聞き取りやすさの評価値との相関がみられたことから，零交差率と聞き取りやすさには関係があると考えられる．最も相関係数が大きかった特徴量と次に相関係数が大きかった特徴量の散布図を**図 1** に示す．ZCR-微分-線形近似の勾配度は負の相関，ZCR-微分-線形近似のオフセットは正の相関があることから，零交差率の変化量が最初は高く，後になるにつれて低くなっていく方が聞き取りやすい音声になると考えられる．また，ZCR-標準偏差に負の相関があることから，零交差率のばらつきを抑えたほうが聞き取りやすい音声にな

表 1　特徴量と相関係数

| 抽出した音響特徴量 | 相関係数 |
|---|---|
| ZCR-微分-線形近似のオフセット | 0.695 |
| MFCC（1 次）-微分-算術平均 | 0.553 |
| ZCR-微分-線形近似の勾配度 | -0.709 |
| ZCR-標準偏差 | -0.623 |
| ZCR-線形近似の二乗誤差 | -0.591 |
| MFCC（10 次）-微分-歪度 | -0.510 |




[1]ZCR-微分-線形近似の勾配度と聞き取りやすさの評価値　[2]ZCR-微分-線形近似のオフセットと聞き取りやすさの評価値

図 1　相関のあった特徴量の散布図

ると考えられる．

## 5. おわりに

本研究では，発話特徴に着目した音声の聞き取りやすさ向上を目指すシステムを提案し，提案システムの設計に必要な音響特徴量と聞き取りやすさの対応関係の調査を行った．調査の結果，聞き取りやすさに相関のある音響特徴量の存在を確認し，零交差率が聞き取りやすさに関係する音響特徴量であることが分かった．今後は，今回得られた相関のある特徴量のパラメータを所望の値に変化させる信号処理を実装し，効果の検証を行う．

### 参考文献


[1] 坂本修一, 鈴木陽一, 天野成明, 小澤賢司, 近藤公久, 曽根敏夫. 親密度と音韻バランスを考慮した単語了解度試験用リストの構築. 日本音響学会誌, Vol. 54, No. 12, pp. 842–849, 1998.

[2] 翁長博. 残響音場の音声了解度に対応する物理指標の提案. 日本音響学会誌, Vol. 66, No. 3, pp. 97–104, 2010.

[3] 洲脇志麻子, 立入哉. Sn 比と残響時間が文章了解度，主観的評価に及ぼす影響. *AUDIOLOGY JAPAN*, Vol. 49, No. 1, pp. 86–92, 2006.

[4] B. Cauchi, K. Siedenburg, J. F. Santos, T. H. Falk, S. Doclo, and S. Goetze. Non-intrusive speech quality prediction using modulation energies and lstm-network. IEEE/ACM Transactions on Audio，Speech and Language Processing, Vol. 27, No. 7, pp. 1151–1163, 2019.

[5] 渋谷徹, 渡辺瞳, 小林洋介, 近藤和弘. 音韻特徴を捉えた音声伸長・短縮の了解度への影響と適応話速変換方式の提案. 映像情報メディア学会誌, Vol. 66, No. 10, pp. J377–J384, 2012.

[6] 川原竣介, 平川凜, 中藤良久. 音声スペクトルの時間変化強調による明瞭性改善手法の提案. 産業応用工学会論文誌, Vol. 6, No. 1, pp. 51–54, 1991.

[7] 匂坂芳典, 浦谷則好. Atr 音声・言語データベース. 日本音響学会誌, Vol. 48, No. 12, pp. 878–882, 1992.

[8] Florian Eyben, Martin Wollmer, and Bjorn Schuller. opensmile – the munich versatile and fast open-source audio feature extractor. In *Proceedings of the ACM Multimedia 2010 International Conference on multimedia*, pp. 1459–1462, 01 2010.