T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

# 論文 / 著書情報
# Article / Book Information

| | |
|---|---|
| 論題(和文) | 日本語話し言葉音声と読み上げ音声の音響的特徴の比較 |
| Title(English) | |
| 著者(和文) | 中村 匡伸, 岩野 公司, 古井 貞熙 |
| Authors(English) | Masanobu Nakamura, Koji Iwano, SADAOKI FURUI |
| 出典(和文) | 日本音響学会 2004年秋季講演論文集, Vol. , No. 2-P-25, pp. 411-412 |
| Citation(English) | , Vol. , No. 2-P-25, pp. 411-412 |
| 発行日 / Pub. date | 2004, 9 |

Powered by T2R2 (Science Tokyo Research Repository)

# 日本語話し言葉音声と読み上げ音声の音響的特徴の比較 *

◎中村匡伸　岩野公司　古井貞熙 (東工大)

## 1　はじめに

話し言葉音声の音響的特徴の分析は，話し言葉音声の認識性能の向上や，音声合成の品質向上に役立つと考えられ，非常に重要である．そこで本稿では，話し言葉音声を読み上げ音声と比較することにより，話し言葉音声の持つ音響的な特性を明らかにすることを目的とする．分析には，日本語話し言葉コーパス (CSJ) に収録されている学会講演音声と，同一話者による同一内容の書き起こしの読み上げ音声（以下，再読み上げ音声と呼ぶ）を用いる．両者には，話者・内容の相違がないため，より精密な特徴比較が可能であると考えられる．音響的特徴としては，音素のケプストラムと，発話速度の 2 つに注目し，分析を行う．

## 2　音声データ

本実験で用いた音声データは CSJ に収録されている男性・女性話者各 5 名による学会講演音声と再読み上げ音声である．

表 1 に，本実験で用いた音声データの ID を示す．10 名の話者の中には CSJ のコアに含まれる話者が男女 3 名ずつおり，これらの音声データには，人手によって音素ごとに時間ラベルが付加されている．

音声データは 16kHz でサンプリングされており，1 講演は約 10 分のデータである．実験に際して，まず転記ファイルをもとに講演音声を 400ms 以上の無音区間で区切り，区切られた区間を「発話単位」として定義した．発話単位が 1 秒未満の場合には，後続する発話単位と接続し，1 つの発話単位とみなした．

## 3　音響特徴量の抽出

### 3.1　ケプストラム

話し言葉と再読み上げという，発話タイプの異なる音声間のケプストラム特徴の相違について調べるため，それぞれの話者・発話タイプごとに，各音素の平均ケプストラムを求める．今回の分析対象とする音素は，表 2 のリストにある 31 種（母音 10 種・子音 21 種）とした．また，分析対象データは，表 1 にあげた全音声データ（20 種類）とした．

ここで各音素の平均ケプストラムは，以下のようにして抽出される．

(1) 音声データから MFCC 12 次元と対数パワー，それらの一次微分と二次微分成分の計 39 次元の音響パラメータを抽出する．分析周期は 10ms, 分析窓幅は 25ms とし，発話単位ごとに CMS 処理を行っている．

(2) 各話者・発話タイプごとに，分析対象データを用いて 28 混合 monophone HMM を学習する．全ての音素モデルは，3 状態の left-to-right 型 HMM とする．

(3) 出来上がった monophone HMM のうち，分析対象音素の HMM の第 2 状態から 12 次元 MFCC のベクトルを音素の平均ケプストラムとして取り出す．具体的には，MFCC の各次元について，mixture の平均値を混合重みで重み付けし，和をとることで，12 次元 MFCC を生成する．

**表 1.** 音声データの話者 ID

| | 学会講演音声 | 再読み上げ音声 | コア |
|---|---|---|---|
| 男性 | A01M0056 | R00M0187 | ○ |
| | A11M0369 | R00M0134 | ○ |
| | A11M0846 | R00M0036 | ○ |
| | A11M0469 | R00M0415 | – |
| | A01M0074 | R00M0132 | – |
| 女性 | A06F0128 | R00F0407 | ○ |
| | A01F0122 | R00F0178 | ○ |
| | A05F0043 | R00F0028 | ○ |
| | A01F0861 | R00F0149 | – |
| | A11F0703 | R00F0304 | – |

### 3.2　発話速度

発話タイプによる発話速度の違いを調べる実験には，表 1 のコアに含まれる 6 名の音声データを用いた．付与されている音素ラベルを用いて，各音素の音素長を求めた後，その逆数を取ることで各音素の発話速度を算出する．単位は音素/sec となる．分析対象の音素は，節 3.1 と同じく，表 2 に示す音素とした．

## 4　ケプストラムの縮小率の比較

各発話タイプによるケプストラムの違いを調べるため，再読み上げ音声に対する学会講演音声の音素 $p$ におけるケプストラムの縮小率 $red_p$ を求める．

$$red_p = \begin{cases} \dfrac{\|\boldsymbol{s}_p - \boldsymbol{s}_v\|}{\|\boldsymbol{r}_p - \boldsymbol{r}_v\|} & (p \text{ が母音のとき}) \\ \dfrac{\|\boldsymbol{s}_p - \boldsymbol{s}_c\|}{\|\boldsymbol{r}_p - \boldsymbol{r}_c\|} & (p \text{ が子音のとき}) \end{cases}$$

このとき $\boldsymbol{s}_p$ は話し言葉音声の，$\boldsymbol{r}_p$ は再読み上げ音声の音素 $p$ の平均ケプストラムとする．また，各発話タイプの母音と子音の平均ケプストラムを計算し，学会講演音声の母音及び子音の平均ケプストラムを $\boldsymbol{s}_v, \boldsymbol{s}_c$，再読み上げ音声の母音及び子音の平均ケプストラムを $\boldsymbol{r}_v, \boldsymbol{r}_c$ とする．話者ごとに $red_p$ を求め，その話者平均値を $\overline{red_p}$ とする．ノルムはユークリッド距離を用いる．

---

∗ Comparison of acoustic characteristics between spontaneous speech and reading speech in Japanese
By Masanobu Nakamura, Koji Iwano, and Sadaoki Furui (Tokyo Institute of Technology)

**表 2.** 音素のリスト

| | |
|---|---|
| 母音 | /a,i,u,e,o,a:,i:,u:,e:,o:/ |
| 子音 | /w,y,r,p,t,k,b,d,g,j,ts,ch, z,s,sh,h,f,N,N:,m,n/ |

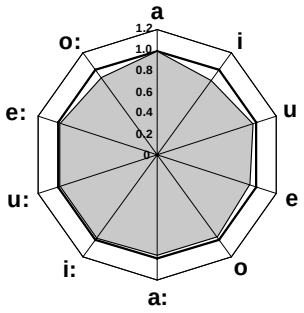


**図 1.** 母音におけるケプストラムの縮小率

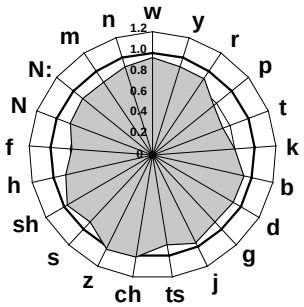


**図 2.** 子音におけるケプストラムの縮小率

## 5 実験結果

### 5.1 ケプストラム

図 1 に母音におけるケプストラムの縮小率 $\overline{red_p}$ を示す．同様に図 2 に子音における縮小率 $\overline{red_p}$ を示す．また $\overline{red_p} = 1$ を太線で表記する．図 1 より，ほとんどの母音における $\overline{red_p}$ が 1 に近いことが分かる．これは，母音によるケプストラムの広がりに発話タイプによる差が無いことを意味している．逆に図 2 では大半の音素における $\overline{red_p}$ が 1 よりも小さくなっており，その平均は 0.89 である．これは，再読み上げ音声に比べ，話し言葉音声における子音のケプストラムの広がりが縮小していることを意味している．

図 3 は 12 次元の MFCC を 2 次元の主成分ベクトル空間に射影した結果を示している．各点は各音素に対応している．左図が母音，右図が子音の分布を示している．両図とも第 1，第 2 主成分ベクトルをそれぞれ $x$ 軸，$y$ 軸としている．学会講演音声の各音素の平均ケプストラムを黒点で表し，その分布を実線の楕円で近似している．それに対して再読み上げ音声では白点と破線の楕円で表している．母音では，発話タイプによってケプストラムの分布に大きな違いは見られないが，子音では，再読み上げ音声に対して学会講演音声のケプストラムの分布の広がりが小さくなっており，縮小率に関する結果と一致している．

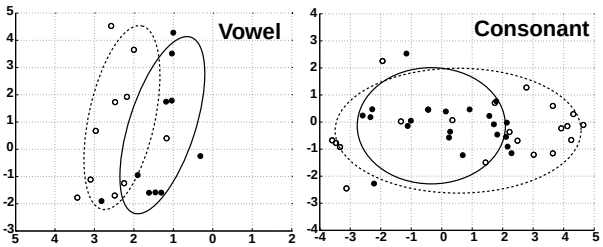


**図 3.** 発話タイプによる平均ケプストラムの分布の違い

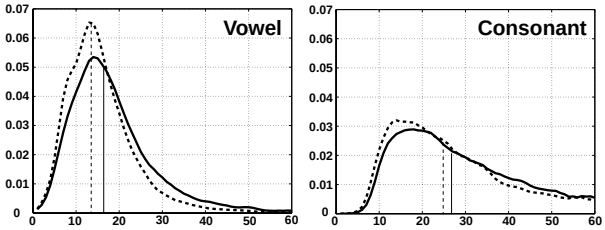


**図 4.** 発話タイプによる発話速度の分布の違い

### 5.2 発話速度

図 4 に発話タイプによる発話速度の分布を示す．左図は母音，右図は子音の分布を表す．両図とも学会講演音声を実線で，再読み上げ音声を破線で表し，発話速度分布の中位値を付記した．ここで各分布の中位値は，母音では，学会講演音声が 16.0 音素/sec，再読み上げ音声が 13.9 音素/sec，子音では，学会講演音声が 27.8 音素/sec，再読み上げ音声が 25.0 音素/sec となった．また，ウィルコクソンの順位和検定を行ったところ，母音・子音とも有意水準 1%で読み上げ音声よりも話し言葉音声の方が発話速度が速いという検定結果が得られた．

## 6 まとめ

本実験では話し言葉音声と読み上げ音声の音響的な特性の違いを明らかにするため，話し言葉音声と再読み上げ音声のケプストラムと，発話速度の違いの 2 つに注目して分析した．

ケプストラムの違いに関しては，母音よりも子音の方が読み上げ音声に対する話し言葉音声のケプストラムの縮小度合いが大きく，また，ケプストラムの分布の広がりが小さくなっていることが分かった．

発話速度の違いに関しては，母音，子音とも読み上げ音声よりも話し言葉音声の方が速くなる傾向が確認された．この結果は，音声データとして CSJ の学会講演音声と ATR の読み上げ音声を比較した研究報告 [1] とも一致している．

## 謝辞

本実験を進めるに当たり，貴重なご助言をいただいた前川喜久雄氏 (国語研) に感謝いたします．

### 参考文献

[1] K.Maekawa, “Corpus of Spontaneous Japanese: Its Design and Evaluation,” *Proc. SSPR 2003*, pp.7-12 (2003).