論文 / 著書情報
Article / Book Information

論題(和文)	フランス語における発声スタイルの違いがスペクトル特徴に与える影響の分析
Title(English)	
著者(和文)	別府 真由美, Jean-Luc Rouas, Martine Adda-Decker, 篠田 浩一, 古井 貞熙
Authors(English)	Mayumi Beppu, Koichi Shinoda, SADAOKI FURUI
出典(和文)	日本音響学会2010年秋季講演論文集, Vol. , No. 3-1-2, pp. 257-260
Citation(English)	, Vol. , No. 3-1-2, pp. 257-260
発行日 / Pub. date	2010, 9

フランス語における発声スタイルの違いがスペクトル特徴に与える影響の分析*

◎別府真由美 (東工大), Jean-Luc Rouas, Martine Adda-Decker(LIMSI-CNRS),
篠田浩一, 古井貞熙 (東工大)

1 はじめに

近年の音声認識では, 読み上げ音声やニュース音声であれば, かなり高い精度で音声認識できるようになった [1]. しかし話し言葉音声では, 現在最も性能の良い認識システムでも, 認識性能が大幅に下がってしまう. これは, 読み上げ音声と話し言葉音声が音響的にも言語的にも大きく異なるためである [2]. 話し言葉音声に対する認識性能を向上させるためには, 音声認識で用いる特徴量に対する発声スタイルの影響を分析することが重要である.

連続単語発声における音声スペクトルの広がりが, 孤立単語発声に比べて小さくなっていることはよく知られており, この現象はスペクトル縮小と呼ばれる. 読み上げ音声に対して話し言葉音声では, より自然で自発性の高い発声がなされることにより発声の怠けが生じると考えられ, このことはスペクトル縮小を引き起こすと考えられる. van Son らによる先行研究では, 話者 1 名に対して読み上げ音声と話し言葉音声を比較することで, 声道の共振周波数においてパラメータ空間上におけるスペクトル縮小が観測されている [10]. また, 中村らによる大規模なコーパスを用いた日本語における研究でも同様のスペクトル縮小が観測されている [3]. また話し言葉音声は自然性の高い発声であることから, 各音素に対しても多様な発声がなされると考えられる. 日本語で読み上げ音声と話し言葉音声における MFCC ベクトルの成分の分散に関する比較を行った結果, その分散の拡大が確認されている [3]. 本稿ではこの現象をスペクトル分散拡大と呼ぶ.

一方フランス語では, 読み上げ音声と話し言葉音声は構造的に異なることが知られている. 話し言葉音声では複雑な音節が単純化される傾向にあり, 語末の子音や強調が置かれない音節中の母音の脱落が頻繁に起こる [9]. このような現象が話し言葉音声の自動音声認識における認識誤りを引き起こす原因であると考えられる. そのため本稿ではフランス語の話し言葉音声と読み上げ音声に対し, スペクトル空間の縮小とスペクトル分散拡大に関して分析を行う.

2 音響的特性の分析

2.1 音響特徴量

サンプリング周波数 16kHz の音声データから 12 次元の MFCC 特徴量ベクトルとその一次微分, 二次微分成分, 対数パワーの一次微分, 二次微分成分の計 38 次元の音響パラメーターを抽出する. 分析周期は 10ms, 分析窓幅は 25ms とする.

2.2 スペクトル縮小率

読み上げ音声に対する話し言葉音声のスペクトル縮小率を算出する. スペクトル縮小率の算出にあたり, 比較基準となるコーパスを基準コーパスと呼ぶ. 基準コーパス (以下 R とする) は, ここでは読み上げ音声のコーパスとする. スペクトル縮小率は以下の式で表される.

$$\mathrm{red}_p(X) = \frac{||\mu_p(X) - \mathrm{Av}(\mu_p(X))||}{||\mu_p(R) - \mathrm{Av}(\mu_p(R))||} \quad (1)$$

$\mu_p(X)$ をコーパス X の音素 p における MFCC 特徴量ベクトルの平均とし, $\mu_p(R)$ を読み上げ音声の音素 p における MFCC ベクトルの平均とする. また Av は平均値を表す.

2.3 スペクトル分散拡大率

スペクトル分散拡大の現象を分析するために, 読み上げ音声と話し言葉音声における 38 次元の共分散行列の対角成分ベクトル (以下分散ベクトルと呼ぶ) に関する比較を行う.

スペクトル縮小率と同様に基準コーパスを R として, スペクトル分散拡大率を以下の式で定義する.

$$\mathrm{ext}_p(X) = \frac{\sum_{k=1}^{K} \sigma_{pk}^2(X)}{\sum_{k=1}^{K} \sigma_{pk}^2(R)} \quad (2)$$

K は MFCC ベクトルの次元数 (ここでは $K = 38$),$\sigma_{pk}^2(X)$ はコーパス X における音素 p の分散ベクトルの k 番目の要素である.

3 日本語の発声スタイルの音響的特徴

中村らの研究 [3] では, 日本語の読み上げ音声と話し言葉音声の音響的特徴の違いを分析している.

ここでは, 日本語話し言葉コーパス (以下 CSJ) に含まれる数種類の発声スタイルを分析対象としている [4]. CSJ は 1999 年から 2004 年にかけて収録された 660 時間の音声で構成されており, モノローグ, 対話, 読み上げ音声など様々な発話形態を含んでいる. 実験にはその中でも代表的な 4 種類の発声スタイルを用い, 具体的には自発性の低い順に, 読み上げ音声 (学会講演音声の再読み上げ) , 学会講演音声, 模擬講演音声, 対話音声である.

スペクトル縮小率を測った実験で, $\mathrm{red}_p(X)$ の平均値は学会講演音声, 模擬講演音声, 対話音声の順に,

*An analysis of the influence of speaking style differences to spectral properties in French by Mayumi Beppu(Tokyo Institute of Technology) Martine Adda-Decker Jean-Luc Rouas(LIMSI-CNRS) Koichi Shinoda Sadaoki Furui(Tokyo Institute of Technology)

0.89, 0.92, 0.83 であり, ほぼ全ての音素において読み上げ音声に対するスペクトル空間の減少が見られた.

またスペクトル分散拡大率を測った実験で,$\mathrm{ext}_p(X)$ の平均値は学会講演音声, 模擬講演音声, 対話音声の順に, 1.08, 1.17, 1.23 であり, ほぼ全ての音素において分散の拡大が確認された.

スペクトル縮小, 分散拡大ともに, 対話音声において現象が顕著に現れている.

4 フランス語の音声データ

読み上げ音声, ニュース音声, 話し言葉音声の計 3 つのコーパスを用いた.

読み上げ音声として, フランス語の新聞読み上げ音声で構成された BREF を用いた [5]. BREF には 120 人の話者による 100 時間の音声が含まれる. 読み上げテキストはフランス語の新聞である LE MONDE から, 多くの語彙と様々な音素環境が含まれるように選ばれている. 以下全ての実験において基準コーパス R には BREF コーパスを用いた.

ニュース音声として,ESTER 2003-2004 コーパスと ESTER2 2007-2008 コーパスの合計 50 時間の音声を用いた [6]. コーパスには, フランス国内とその他のフランス語圏 (マグリブ, アフリカ) で放送されたラジオ音声から選ばれたデータが含まれる.

話し言葉音声として, 友人同士の会話が収録されている NCCFr コーパスを用いた [7]. NCCFr コーパスは 23 ペアの話者 (男性話者 : 24 名, 女性話者 : 22 名) による 36 時間の音声を含み, 各ペア 90 分の会話のデータが含まれる.

全てのコーパスには人手による書き起こし文が存在し, 自動アラインメントがされている.

5 特徴量に関する分析

まずそれぞれのコーパスにおける音素の継続時間長の分布を元に, 後の分析に用いる音素を選択する目的で, 3 種類の発声スタイルで, 自動的にアラインメントされた音素の継続時間長の比較を行った. 自動アラインメントには 3 状態の HMM を用いた.

結果を Fig.1 に示す. 図に示したように,BREF と ESTER の分布は,ESTER のグラフが若干左に偏っているものの非常に似ている. ニュース音声の発話速度は読み上げ音声よりも早いため, ESTER のグラフの偏りはその影響によるものと考えられる. しかし NCCFr コーパスの分布の形状は他とは明らかに異なっており, アラインメントの下限値である 30ms あたりにピークが見られる. これは, フランス語の話し言葉音声では単語が標準的に発音されずに一部の音素が脱落する現象が頻繁に起こるが, その際も最低の継続時間長でアラインメントがされてしまう影響によるものと考えられる [8].

音素の継続時間長の分布の形状はコーパスごとに異なっていたが, 中間値はいずれも 60ms から 70ms の間であり, ほとんど同じ値であった. よって続く第 5.1 節の実験では, 各コーパスに含まれる音素の継続時間長の平均値付近である 50~110ms の長さの音素を扱う.

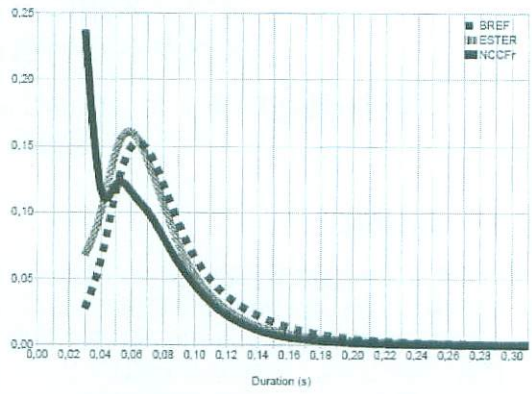

Fig. 1 発声スタイルによる音素の時間的長さの分布. 横軸:音素の長さ (ms) 縦軸:音素数の全体に占める割合.

5.1 スペクトル縮小率と分散拡大率

式 (1) と式 (2) に従い, フランス語の音声データでスペクトル縮小率と分散拡大率を計算した.

スペクトル縮小率を Fig.2 に示す. $\mathrm{red}_p(X)$ の平均値は ESTER, NCCFr の順に,0.80, 0.95 であり, ESTER においてスペクトル縮小が特に確認できる. /i/ と/y/を除くほとんど全ての音素においてスペクトル縮小が起こっている. しかし NCCFr ではスペクトル縮小があまり起こっていない. スペクトル空間の大きさは一部の音素を除いて, 読み上げ音声とあまり変わらないことが分かる.

スペクトル分散拡大率を Fig.2 の下 2 列に示す. $\mathrm{ext}_p(X)$ の平均値は ESTER, NCCFr の順に, 1.23, 1.31 である. スペクトル分散は ESTER, NCCFr ともに拡大しており, その度合は後者の方が大きい.

スペクトル分散は, より自発的な発声である会話音声の方がニュース音声よりも拡大しているが, スペクトル縮小については, 会話音声よりもニュース音声の方が現象が顕著に見られる. ただし本実験では, 話し言葉に多く含まれる 30ms 付近の短い音素は分析対象に含まれていない. そこで続く 5.2 節では, 短い音素と長い音素の分析結果の比較を主な目的とした実験を行う.

5.2 継続時間長の影響

本節では 5.1 節において分析の対象としなかった短い音素と長い音素について, スペクトル縮小と分散拡大への影響に違いがあるかどうかを調べた. 実験では 40ms 以下の継続時間長の音素を短い音素とし, 120ms 以上の継続時間長の音素を長い音素とした.

ESTER と NCCFr のスペクトル縮小率を, 短い音素と長い音素で比較した結果を Fig.3 に示す. 短い音素での $\mathrm{red}_p(X)$ の平均値は ESTER, NCCFr の順に, 0.67, 0.72 である. 標準的な長さの音素における $\mathrm{red}_p(X)$ の平均値は ESTER, NCCFr の順に, 0.80, 0.95 であるため, 短い音素では特に NCCFr コーパスにおいて, 標準的な長さの音素のスペクトル縮小と比べると, より顕著に縮小していることが分かる. ここで扱っている音素は 40ms 以下と非常に短いが, NCCFr コーパスでは 40ms 以下の長さの音素が非常

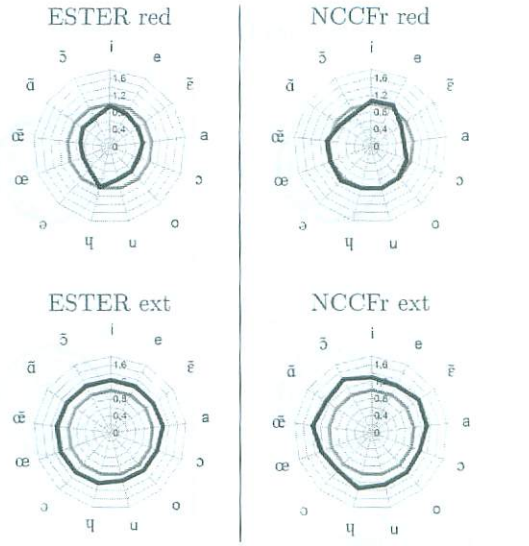

Fig. 2　50~110ms の母音を分析対象としたときのスペクトル縮小率 (上段) とスペクトル分散拡大率 (下段).

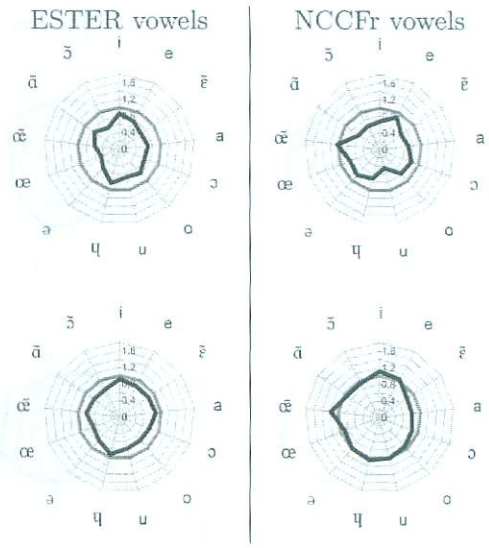

Fig. 3　40ms 以下 (上段) と 120ms 以上 (下段) の長さを持つ母音を分析対象としたときのスペクトル縮小率

に多いため, この現象は無視できない. 標準的な長さの音素と同様, 会話音声よりもニュース音声の方がスペクトル縮小が起こる傾向にあるが, 会話音声では短い音素ほどスペクトル縮小が顕著に起こっている. また会話音声はコーパスで頻出する短い音素により特徴づけられると考えられるため, 長い音素でのスペクトル縮小はここでは重要ではない.

分散拡大率を音素の継続時間長によって比較した結果を Fig.4 に示す. どちらの長さにおいても音素の分散は拡大している.

5.3　語彙コンテキストの影響

本節では単語の役割を考慮したときに現れると予想される, スペクトル縮小と分散拡大への影響の違いを分析する.

まずそれぞれのコーパスに含まれる単語を頻度に基づいて機能語と内容語に分類した. 機能語を内容語から分離する方法は以下の通りである. コーパスに出現する単語を頻度の降順に並べ, 上位 100 単語から単語の累積数がコーパスに含まれる全単語の 50%を占めるまで, 機能語の一覧に加える. このようにして自動的に得られた機能語の一覧から, 紛れた一部の内容語を手動で省いた. 例えば我々は今回フランス語のニュースのコーパス (BREF ESTER) を扱っているため, 「France (フランス)」, 「Monsieur (~氏)」, 「président(首相)」といった内容語が省かれる対象となった. さらに得られた機能語は限られているため, 各音素の頻度に大きなばらつきがある. 従って頻度が 1000 回以下の音素は分析対象から除いた. 除いた母音は/ɔ o ə/である.

スペクトル縮小率の分析結果を Fig.5 に, 分散拡大率の結果を Fig.6 に示す. $\mathrm{red}_p(X)$ の機能語での平均値は ESTER, NCCFr の順に,0.81, 0.98 であり, 内容語では, 0.80, 0.94 である. また $\mathrm{ext}_p(X)$ の機能語での平均値は ESTER, NCCFr の順に,1.26, 1.32 であり, 内容語では, 1.22, 1.30 である. スペクトル縮小は, 機能語よりも内容語において顕著に現象が起こっている. 一方, スペクトル分散については, 内容語よりも機能語の方が分散拡大率は大きい.

フランス語では, 一般的に機能語は内容語よりもはっきりと発音されない傾向にある. このためスペクトル縮小は機能語において顕著に現象が起こることが予想されたが、実験では内容語においてより顕著なスペクトル縮小が観測された. しかしスペクトル分散は機能語の方が拡大しており, このことは機能語が内容語よりも曖昧に発音される傾向にあることとの一貫性がある.

5.4　日本語との比較

日本語において中村らが示した結果では, 発声スタイルが自発的であればあるほどスペクトル縮小, スペクトル分散拡大の傾向が見られた [3].

一方フランス語では, スペクトル分散は日本語と同様に自発的な発声であるほど拡大していたが, スペクトル縮小は会話音声よりもニュース音声の方が現象が顕著に現れた. この原因については現段階では不明であり, これを明らかにすることは今後の課題である. しかし話し言葉音声に多く含まれる継続時間長の短い音素を対象とした実験において, 話し言葉音声では短い音素ほどスペクトル縮小が起こりやすいことが示された. これは話し言葉音声において, 短い音素ほど発音が曖昧になる傾向を示していると考えられる.

6　おわりに

本稿ではスペクトル縮小とスペクトル分散拡大の現象に関する, フランス語の話し言葉音声と読み上げ音声の違いの分析を行った. その結果, スペクトル分散は自発的な発声ほど拡大していることが分かった. 一方スペクトル縮小については, より自発的である会話音声よりもニュース音声で顕著に現象が現れた. し

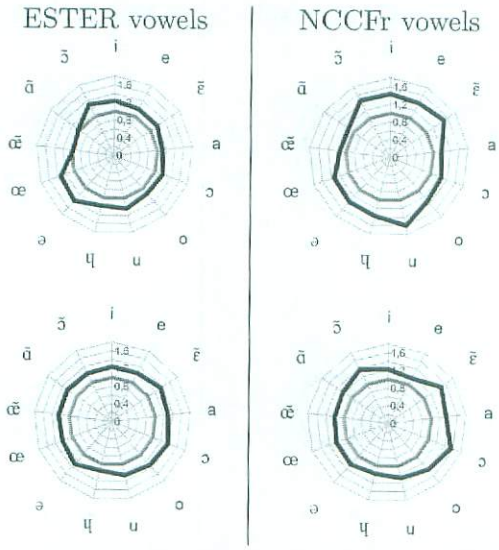

Fig. 4　40ms 以下 (上段) と 120ms 以上 (下段) の長さを持つ母音を分析対象としたときのスペクトル分散拡大率

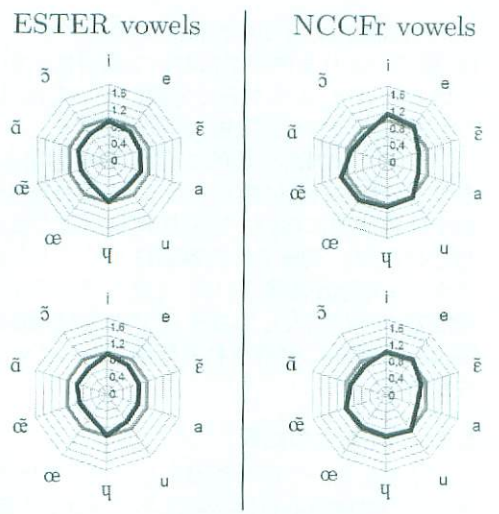

Fig. 5　機能語 (上段) と内容語 (下段) に含まれる母音を分析対象としたときのスペクトル縮小率

かし話し言葉音声コーパスに頻繁に現れる短い音素と, 標準的な長さの音素を比較すると, 自発的な発声では短い音素でより顕著にスペクトル縮小が起こっていることが確認された. また単語の役割を考慮したとき, スペクトル縮小は機能語よりも内容語において顕著に現象が観測されたが, スペクトル分散については内容語よりも機能語において分散拡大率が大きいことを示した.

今後は今回得られた知見をフランス語の話し言葉音声の認識性能改善に役立てることができるかどうかを検討する必要がある.

参考文献

[1] P. Fousek et al. "Transcribing Broadcast Data Using MLP Features" InterSpeech'08 pp. 1433-1436 Sept. 22-26 2008.

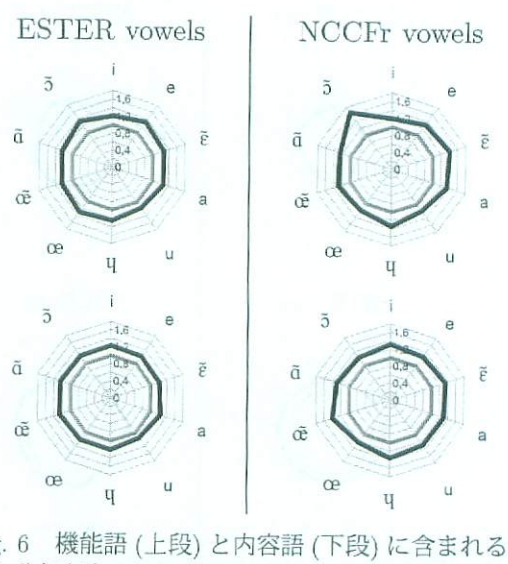

Fig. 6　機能語 (上段) と内容語 (下段) に含まれる母音を分析対象としたときのスペクトル分散拡大率

[2] S. Furui "Recent advances in spontaneous speech recognition and understanding" ISCA & IEEE workshop on Spontaneous Speech Processing and Recognition (SSPR) 2003.

[3] M. Nakamura et al. "Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance" Computer Speech and Language 22:171-184 2008.

[4] K. Maekawa "Corpus of spontaneous japanese: its design and evaluation" ISCA & IEEE workshop on Spontaneous Speech Processing and Recognition (SSPR) 2003.

[5] L. Lamel et al. "Bref a large vocabulary spoken corpus for french" Eurospeech 1991.

[6] S. Galliano et al. "Corpus description of the ester evaluation campaign for the rich transcription of french broadcast news" Language Evaluation and Ressources Conference 2006.

[7] F. Torreira et al. "The nijmegen corpus of casual french" Speech Communication in press.

[8] M. Adda-Decker et al. "Contributions du traitement automatique de la parole à l'étude des voyelles orales du français" Traitement Automatique des Langues 49 2008.

[9] M. Adda-Decker et al. "Investigating syllabic structures and their variation in spontaneous french" Speech Communication 46(2):119-139 2005.

[10] R. J. J. H. van Son et al. "An acoustic description of consonant reduction" Speech Communication 28(2):125-140 1999.