CSJ-Coreを用いたパラ言語情報の研究*

○前川 喜久雄（国立国語研究所）

1 はじめに

パラ言語という術語は研究者によって、また学派によって大きく異なる意味で用いられる。筆者は藤崎博也氏に従って、話者が意図的に表出する情報で、範疇的な意味を表しているが範疇内で程度差が存在するものをパラ言語情報と呼んでおり[1]、1997 年から 2000 年にかけて、生成と知覚の両面からパラ言語情報に関する実験的研究を進めた。

生成面については、パラ言語情報が音声の持続時間、イントネーション、ホルマント周波数等に顕著な影響を及ぼすこと[2]、ホルマント周波数の変化を生み出す舌と唇の変移(displacement)は母音セグメントだけでなく発話全体にわたって生じており、或る種の声質(voice-quality)の問題として把握すべきであること[3]、さらに発声様式にも顕著な差異が生じていること[4]などを報告した。

知覚面については、明瞭に生成されたものであれば、基本的なパラ言語情報は音声信号を聴取することによって 90%前後の精度で知覚（forced-choice）されうること、MDS によって構成されたパラ言語情報の知覚空間は日本語母語話者の場合 3 次元であること、知覚空間と生成面の研究で見出された各種音響特徴の間には組織的な対応付けが可能であること、日本語学習者の知覚空間は母語話者のそれに酷似しているが、日本語に全く触れたことのない非母語話者はパラ言語情報を正確に知覚することができず、特に発話末の上昇イントネーションの差異によって表明されるパラ言語的意味の知覚に困難が大きいこと[5,6]を報告した。この過程で得た実験データの一部は、宇都宮大学（当時）の粕谷英樹氏によっても分析が行われている[7,8]。

2 「口調」

これら一連の研究で筆者があつかったパラ言語情報は「疑念」「落胆」「感心」等々、発話の意図にかかわるものであった。しかし、言うまでもなく、それだけがパラ言語情報ではない。筆者は最近、『日本語話し言葉コーパス』（CSJ）を利用して、「口調」(tone of voice)の研究を始めた。国語辞典のなかには「口調」を「声の出し方、ことばの選び方などにあらわれる特色」（三省堂国語辞典第 5 版）と説明しているものがあるが、これを個人に属する特徴とみてはならない。「教師口調」「演説口調」「冗談口調」「命令口調」などの定形表現が存在することから或る程度社会化された音声の型の存在が窺われる。

今回は手始めとして、CSJ に記録されている 4 種類の音声タイプの別を、定性的な韻律特徴のみを用いて予測した結果を報告する。

3 データ

CSJ は筆者らが 2004 年に公開した 750 万語規模の自発音声コーパスである[9]。そのうち Core と呼ばれる約 50 万語分のデータには X-JToBI 方式による音声アノテーションが施されている[10]。

CSJ-Core には「学会講演」が 70 講演（実発話時間で約 840 分）、「模擬講演」が 107 講演（900 分）、対話が 18 発話（179 分）、再朗読（すなわち多少とも自発的な学会講演ないし模擬講演の転記テキストを同一話者が朗読した音声）が 6 講演（82 分）含まれている。

これらに付与された X-JToBI による分節音・韻律ラベル（対話は主話者の発話のみラベリング）の総数は 3,364,285 個である。そのうち今回は、X-JToBI の Tone 層、BI 層、PRM 層、MISC 層で利用する 23 種類のラベルの生起頻度をファイル毎に集計したデータを作成して分析した。

4 結果

上記データに対して R 言語の lda 関数を用いて線形判別分析を施した。結果を Fig.1 に示す。LD1,2,3 は判別得点、図中の記号「A」は学会講演、「S」は模擬講演、「D」は対話、「R」は再朗読の音声である。全データによ

* Study of paralinguistic information using the CSJ-Core, by Kikuo MAEKAWA (National Inst. Jap. Language)

る分析の正判別率は 89.6%、LD1,2,3 間での分散の比率は 0.64, 0.31, 0.05 であった。また Leave-one-out 交互検証における平均正判別率は 81.6%であった。

この分析では本質的に定性的な韻律特徴（韻律境界の強さのランクとその下位分類、句末イントネーションの種類とその下位分類など）だけを利用している。これに連続量データとして、アクセント句単位の平均発話速度と平均モーラ長の情報を追加して再分析すると、判別率はデータ全体で 95.5%、交差検証で 88.1%まで上昇する。

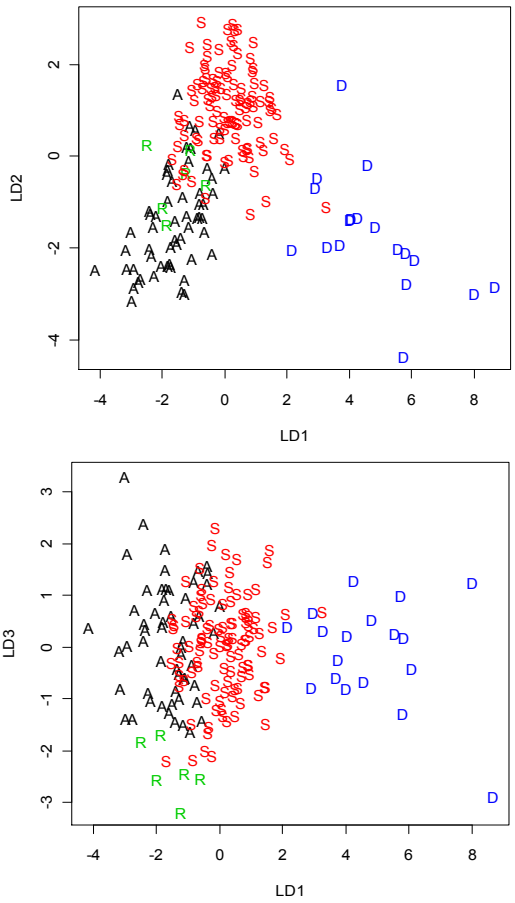

Fig.1 Result of linear discriminant analysis. Upper and lower panels show respectively distributions of the 201 speech files on the LD1-LD2 and LD1-LD3 planes, respectively. Plotting symbols of “A”, ”S”, ”D”, and “R” stand respectively for academic presentation speech, simulated public speaking, dialogue, and reproduction speech.

5　議論と今後の課題

国語辞典の説明にあったように「口調」には「ことばの選び方」すなわちテキスト特徴が関与しているはずである。しかし今回の分析は、CSJ に限れば、音声特徴だけで高い判別率が得られることを明らかにした。特に、定性的な韻律特徴だけでも交差検証で 8 割以上の正判別率が得られることは、発話の韻律構造がパラ言語に関して豊富な情報を提供していることを示しており、非常に興味深い。

詳細な要因分析は今後の課題であるが、顕著な要因を指摘しておく。上昇調イントネーションの頻度は学会講演で高く模擬講演で低い。反対に、上昇下降調イントネーションは模擬講演で高く学会講演で低く、そして再朗読では一層低い。アクセント句末に上昇調ないし上昇下降調イントネーションが生じているにも係らずカタセシス（アクセント核が惹起するピッチレンジの狭め現象）がリセットされないアクセント句境界(2+b ないし 2+bp)の頻度は、学会講演において高く模擬講演と対話では低い。

今後は詳細な要因分析とともに、テキスト情報と韻律情報の寄与を比較検討する予定である。また定性的な韻律特徴と非言語情報（性別、年齢など）の関係も分析する予定であるが、予備分析では、性別はかなり高い精度で判別できることがわかっている。

参考文献

[1] Fujisaki, “Prosody, Models, and Spontaneous Speech,” in Sagisaka et al. eds. *Computing Prosody*, 27-42, Springer, 1996.

[2] 前川，北川，認知科学, 9(6), 46-66, 2002.

[3] Maekawa and Kagomiya, *Proc. ICSLP2000*, Beijing, 349-352, 2000.

[4] Fujimoto and Maekawa, *Proc. ICPhS 2003*, 2401-2404, 2003.

[5] Maekawa. *Proc. Speech Prosody 2004*, Nara, 367-374 , 2004.

[6] 前川，広瀬編『韻律と音声言語情報処理』pp.24-34, 2006.

[7] Kasuya, Kiritani, Maekawa, *Proc. ICPhS 99*, 2505-2512, 1999.

[8] Kasuya, Yoshizawa, Maekawa, *Proc. ICSLP2000*, Beijing, pp.345-348, 2000.

[9] Maekawa, *Proc. ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, 7-12, 2003.

[10] Maekawa, et al., *Proc. ICSLP 2002*, Denver, 1545-1548, 2002.