

1. 目的

音声信号に含まれる情報を言語情報・パラ言語情報 (意図・態度)・非言語情報 (個性・身体性・情動) に三分することがある[1]。筆者らはパラ言語情報の生成に関する研究をおこない、パラ言語情報の伝達に与する音響的特徴を抽出してきた[2-4]。本稿ではその過程で得た音声資料を利用しておこなったパラ言語情報の知覚実験結果を報告し、従来報告してきた音声特徴を知覚の観点から検討する。

2. 知覚実験

我々の先行研究では「～ですか」という文末を共有するテキスト十数文を、種々のパラ言語的意味で10回前後繰返し発話した資料を収録している。そこでとりあげたパラ言語意味は「Neutral 中立 (パラ言語情報を含めない棒読み)」「Focussed 強調 (同じく棒読みであるが、聴き手が遠方にいると想定して声を大きくした発話)」「Indifference 無関心 (相手の発話への相づち)」「Admiration 感心 (それは素晴らしい!)」「Disappointment 落胆 (ああ、嫌だ)」「Suspicion 疑い (嘘でしょう?)」の6タイプである。話者は男性2名 (ST, YS) と女性1名 (JH) である。この資料のなから、文1「そうですか」文2「山野さんですか」文3「あなたですか」を知覚実験の素材として選んだ。文1と2は3名全員の発話が収録されており、文3はSTの発話だけである。STは10回、他2名は11回各タイプの発話を収録したが、そこからパラ言語的意味や終助詞をとりちがえた発話を除外した発話数を表1に示す。各セル中の数字は左から文1, 2, 3に対応する。総計436個の発話を刺激音声として利用した。

Table 1. Number of utterances by speaker and paralinguistic information type.

Speaker	ST	YS	JH
Neutral	10 10 9	11 11	11 11
Focussed	10 10 10	11 10	11 11
Indifference	10 10 10	10 11	11 11
Admiration	10 10 10	11 10	10 11
Disappointment	10 10 10	11 10	11 11
Suspicion	10 10 10	11 9	11 11
Total	179	126	131

国立国語研究所の職員と東京都立大学の学生合計20名を被験者としてパラ言語情報の同定実験を実施した。実験ではまずパラ言語的意味について説明し、簡単な練習をおこなったのち、録音テープから再生される発話のパラ言語的意味を上記6タイプ中のひとつに同定することを要求した。刺激音声は発話者ごとに分けて提示することとし、STならば179個の刺激を1回ずつランダムに配した録音テープを作成した。被験者は都合3本のテープを聴取し判定することになるが、順序効果を避けるために実験ごとに再生するテープの順序を入れ替えた。

3. 多次元尺度構成法の適用

この実験で得た同定データをもとに任意の2刺激が同一のパラ言語的意味に同定される確率を計算すると刺激数×刺激数の大きさの類似度行列が得られる[5]。今回は話者3名に対する反応をすべてプールして436×436の行列を作成した。この行列に対してクラスカルの方法による多次元尺度構成法(MDS)を適用すると、

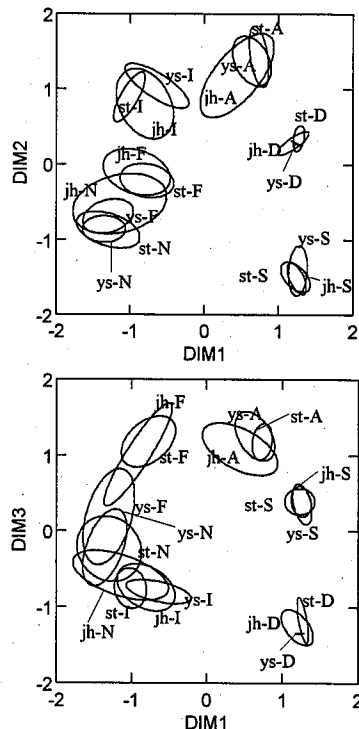


Figure 1. Probability elliptic representation of 436 stimuli in the three-dimensional MDS space.

2次元解ではstress=0.08, 3次元解では0.04となり、経験的にみて非常によいあてはまりである。図1に3次元解で得られた刺激の布置を話者毎タイプ毎に68%確率楕円として表示した。

DIM1はA, D, SとI, N, Fを分離している。DIM2の一端にはA, Iが分布し他端にはSが分布している。DIM3では一端にA, F, Sが他端にD, Iが分布している。

4. 音響量との相関

それでは図1における刺激の布置はどのような音響量と関係しているのだろうか。この問題を検討するために重回帰分析をおこなった。

我々が先行研究[2-4]でパラ言語情報の伝達に与するとして報告した音声特徴には、発話全体および句頭/末尾音節の持続時間長、発話のピッチレンジ、句頭のピッチ上昇の形状、末尾におけるピッチの上昇/下降、アクセントによるピッチ下降のタイミング、母音フォルマント周波数 (特にF2)、発話の冒頭と末尾における声質(phonation)の変化がある (図2参照)。

表2に今回の分析で独立変数としてとりあげた音響量を示す。先行研究で報告された特徴から持続時間とピッチに関するものを選び、定量化に問題のあるフォルマント周波数と声質は除外した。またRMS振幅の情報を新たに追加した。

図1の各軸における刺激の値を従属変数とした重回

* Perception of paralinguistic information: A Multidimensional scaling analysis.
Kikuo Maekawa (Natl. Lang. Res. Inst.) & Norimichi Kitagawa (Tokyo Metropolitan Univ.)

帰分析の結果を表3に示す。表中の値は各軸を従属変数としたときの標準偏帰係数であり、太字は5%水準で有意であることを示している。

Table 2. Independent variables used in the multiple regression analysis.

F0	F0 average of the whole utterance
F0_I	F0 average of the 1 st syllable
F0_F	F0 average of the last syllable
PR	Pitch range of the whole utterance
PR_I	Pitch range of the 1 st syllable
PR_F	Pitch range of the last syllable
RMS	Maximum RMS amplitude of the whole utterance
RMS_I	Maximum RMS amplitude of the 1 st syllable
RMS_F	Maximum RMS amplitude of the last syllable
DUR	Duration of the whole utterance
DUR_I	Duration of the 1 st syllable
DUR_F	Duration of the last syllable
PT	Timing of accentual pitch fall relative to the end of the accented syllable

Table 3. Results of multiple regression analysis. Each cell shows the value of the standard partial-regression coefficient. Values in bold face are significant at the 0.05 level. The last row shows the multiple coefficients of determination.

Independent Variable	Dependent Variables		
	DIM1	DIM2	DIM3
F0	0.290	0.181	0.611
F0_I	-0.307	-0.066	-0.159
F0_F	0.023	-0.160	-0.359
PR	-0.008	0.394	0.199
PR_I	0.058	0.042	0.168
PR_F	-0.060	-1.036	0.067
RMS	-0.075	0.036	0.592
RMS_I	0.056	0.057	-0.213
RMS_F	0.059	0.106	0.174
DUR	0.583	0.019	0.174
DUR_I	0.088	-0.094	0.087
DUR_F	0.425	0.368	-0.149
PT	0.259	0.155	-0.216
R ²	0.877	0.641	0.532

表3中の決定係数(R²)はDIM1, 2, 3(従属変数)すべてについて0.5を越えており、重回帰分析による予測の精度が高いことを示している。

偏帰係数値をみるとDIM1では発話全体の持続時間(DUR)、末尾音節の持続時間(DUR_F)、冒頭音節の平均F0の寄与が大きい。DIM2では末尾音節のピッチレンジ(PR_F)が顕著に大きく、次いで発話全体のピッチレンジ(PR)と末尾音節の持続時間(DUR_F)が寄与している。DIM3では発話全体のF0平均(F0)と発話全体のRMS振幅(RMS)の寄与が大きく、次いで末尾音節のF0平均(F0_F)の寄与が認められる。

5. MDSの各次元の解釈

MDSによって構成された三つの次元の意味を重回帰分析の結果をふまえて解釈すると以下のようになる。

まずDIM1。A, D, SはN, F, Iよりもパラ言語的意味がはっきりしていると考えられるから、「パラ言語的意味の濃淡」に関する軸と解釈できる。重回帰分析の結果はこの判断がもたらした持続時間の特徴によって下されていることを示唆している。実際A, D, SとN, F, Iの間に持続時間の顕著な差が認められることは図2に示した典型例からも読みとれる。

次にDIM2の一端に位置するSは問いかけという形で聴き手に対する「働きかけ」をおこなっている発話で

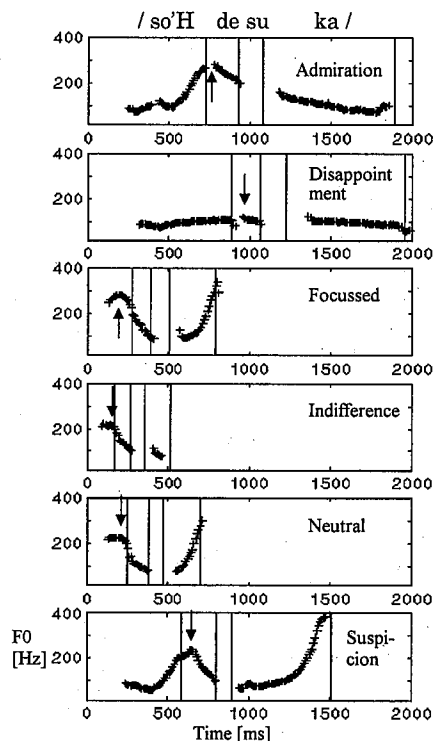


Figure 2 Typical duration and pitch patterns of sentence 1 (/so'H de su ka/) under six paralinguistic information conditions. Utterances are aligned at the beginning of the consonant /s/. Vertical lines stand for the syllable boundaries. Arrows denote the timing of accentual fall. Speaker ST.

あり、他端に位置するA, Iは「働きかけ」のない発話であることから「働きかけ」の有無に関する軸と解釈できる。重回帰分析では最終音節のピッチレンジと持続時間の寄与が大きい、これは発話末尾におけるイントネーションの上昇/下降を反映している。

最後にDIM3は大きな声で発話される強い発話であるA, F, Sと弱い発話であるD, Iを分離していることから「強弱」の軸と解釈できる。重回帰分析の結果においても発話全体のRMS振幅とF0が大きく寄与している。

6. 結論

従来報告してきたパラ言語情報の生成に関する音声特徴のうち表3にとりあげたものは、すべて何らかの形でパラ言語情報の知覚に寄与していることが確認された。表3ではDIM1からDIM3にかけてR²が漸減してゆが、今回分析の対象から除外したフォルマント周波数や声質の変動を表現する変数を重回帰分析の独立変数に追加すればより高精度の予測が可能になると期待される。

ただし今回の知覚実験は6種類のパラ言語的意味に限定された実験であるから、より広範囲の意味を対象とした場合にも、今回と同一の次元が抽出されるかどうかは今後検討すべき課題である。

参考文献

- [1] 藤崎. 音講論集 (平成6年秋) 287-290, 1994.
- [2] 前川. 音講論集 (平成9年秋) 381-384, 1997.
- [3] Maekawa. *Proc. ICSLP98* (CD-ROM), Paper No. 0997, 1998.
- [4] 前川. 国語研50周年記念研究発表会資料, 93-98.
- [5] 高根. 多次元尺度法. 東大出版会, 1980.