2004－ＨＩ－109　(17)
2004－SLP－ 52　(17)
2004／7／17

基本周波数パタンに見られる発話態度の分析

苔縄陽子[a]　津崎実[b]　加藤宏明[c]　匂坂芳典[d]

a)d) 早稲田大学国際情報通信研究科　〒169-0051　東京都新宿区西早稲田 1-3-10
b) 京都市立芸術大学音楽学部 〒610-1197　京都府京都市西京区大枝沓掛町13-6
c)ATR 人間情報科学研究所　〒619-0288　京都府相楽郡精華町光台 2-2-2

Email: a) yoko.kokenawa@toki.waseda.jp, b) minoru.tsuzaki@atr.jp, c) kato@atr.jp, d) sagisaka@giti.waseda.ac.jp

要旨： 発話態度に基づく基本周波数制御の手始めとして、基本周波数パタンと発話態度の依存関係の分析、主観評価実験を行った。種々の韻律を持つ一語発話“ん”の観察、印象記述に基づき、基本周波数の平均的高さ、時間変化形状によって言語情報として陽に表されない伝達韻律情報の大まかな分類を試みた。この分類の妥当性を確認するため、分類に従ったモデル音声を用いた主観評価実験を行った。多次元尺度構成法を用いた評定結果分析によって、基本周波数の平均的高さによって積極/消極、承服/不服、確信/不確かな発話態度が、また基本周波数の時間変化形状によって肯定/否定が伝達されることが確認できた。これらの実験から、これまで音声合成ではほとんど手がつけられていなかった、言語内容に表出されない発話態度・意図に対する韻律制御の可能性が示された。

An analysis of speaking attitude manifesting as fundamental frequency characteristics

Yoko Kokenawa[a]　Minoru Tsuzaki[b]　Hiroaki Kato[c]　Yoshinori Sagisaka[d]

a)d) GITS, Waseda University 1-3-10 Nishi-waseda Shinuku-ku, Tokyo, 169-0051, Japan
b) Kyoto City University of Arts 13-6 Kutsukake-cho, Oe, Nishikyo-ku, Kyoto, 610-1197, Japan
c) ATR Human Information Science Labs, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

Email: a) yoko.kokenawa@toki.waseda.jp, b) minoru.tsuzaki@atr.jp, c) kato@atr.jp, d) sagisaka@giti.waseda.ac.jp

Abstract: To control speech prosody expressing interlocuter's attitudes and intensions, F0 characteristics were acoustically and perceptually analyzed. Single word utterances “n” with different prosody were used to exclude linguistics content effects. F0 analysis showed the possibility of global categorization of prosodic characteristics by their height (high-low) and shapes (Rise, Flat, Fall and Rise&Fall). To confirm the consistency of this classification, perceptual experiments were conducted to identify impression of F0 height and shape differences using newly recorded controlled utterances of “n” Multiple Dimensional Scaling analysis showed the consistency between F0 characteristics and subjects' responses using 26 impression dimensional scaling, which supports the controllability of conversational speech prosody expressing. The speaker's attitude like “positive/negative”, “accept/ unsatisfied”, and “confident/doubt” could be implied depending on the degree of F0 height, while F0 shape, especially Fall and Rise&Fall differences convey the meaning of “positive/negative response”.

1. はじめに

コーパスベース音声合成の導入により、音声品質は、一昔に比べ大幅に向上し、適応領域は広がった。しかしながら、与えられたテキストの読み上げ音声としての品質はある程度満たされても、対話システムなどでの、双方向の情報伝達における音声の品質としては依然不十分である。とりわけ、音韻明瞭度などで考慮される音声品質が向上している分、対話音声としての韻律の不備はより顕著に認識され、その適切な制御が求められている。

この課題に対処するため、我々はこれまでに発話内容の語彙情報に直接依存した制御の可能性を示してきた[1]。これまでの研究では、程度副詞とそれらに後続する形容詞の語彙情報に基づき、発話内容から示唆される発話態度に対応した韻律の変化を生成面、知覚面、制御面から調べ、発話内容の語彙情報に基づく制御の可能性が明らかとなった。しかしながら、この限られた分析で得られた内容を一般化し、実際の合成へと展開するためにはさらに種々の検討を要する。中でも、発話者は実際の発話内容を包含したより広汎な発話態度・意図に基づいた意識的、無意識的な制御を行っており、対話音声そのものに対する発話情報記述、それらの情報に基づく韻律制御モデル化が必須である。

著者等はコーパスベース手法に基づく韻律制御計算モデルを念頭に置いているが、発話内容に直接表記されない対話時の韻律情報への対処は大きな課題である。特に、言語情報として陽に表現されない伝達内容と韻律制御との関係はほとんど分析されていない現在、その制御計算モデルを考える上でも現象の記述、把握が急務と考えられる。近年、実生活環境における音声データベースの作成なども進み[2]、音響面、知覚面からの分析も始められている[3]。しかし、合成を直接意識した韻律制御の解析は十分になされていない。

本稿では、言語内容として陽に表されない韻律情報の規定を目的として基本周波数（F0）パタンの分析を行った。発話内容情報による直接的な影響を除いた分析を行うため、会話に頻繁に用いられ、韻律情報による受け渡しに大きな役割を果たしていると考えられる一語発話”ん”を分析対象とした。言語内容として陽に表されない韻律情報の分析法としては多次元尺度構成法を用い、基本周波数パタンの平均的高さ・時間変化形状と、主観に基づく印象表現との対応関係を求めた。以下、次章では、実際の対話場面音声で観察された一語発話“ん”を対象としたF0の平均的高さと時間変化形状に基づく発話印象の分類について述べる。第3章では、この分類の妥当性を検証するために行う主観評価実験のための印象表現語についての検討を述べる。第4章では主観評価実験の詳細と結果を述べる。第5章では多次元尺度構成法を用いた実験結果の解析を行う。最後に、まとめ、今後の課題について触れる。

2. F0 の平均的高さと時間変化形状に基づく発話印象の分類

対話音声における韻律制御では、先の分析[1]で示した発話語彙そのものが内在的に持つ情報による制御だけではなく、会話状況に応じてあらわされる発話言語表現に独立な韻律制御が必要である。発話言語表現外の音声情報が伝達する情報の規定を目指して、友人同士の親しい関係である成人女性 4 名の 30 分間弱にわたる実際の対話を録音し、分析した。この対話中では一語発話“ん”が多用され（42 サンプル）対話を進めてゆく上での種々の情報を伝達していることが観察された。とりわけ、話者が聞き手に対して意識的・無意識的に示す“驚き”、“聞き返し”、“否定（いいえ）”、“了承（はい）”、

表 1.　基本周波数の平均的高さと時間変化形状に基づく発話印象の分類

dynamics / height	rise ↗	flat →	fall ↘	rise & fall ∿
higher ↑	驚き – 肯定的 聞き返し - 興味がある	相槌 - とても興味がある 相槌 - 興味がある 相槌 - 同意できる	了承（はい）- 興奮気味に 了承（はい）- 喜んで	否定（いいえ）-丁寧な
	驚き – 中立的 聞き返し – 中立的	肯定（はい）- 中立的	了承（はい）- 中立的 了承（はい）- 真剣な	否定（いいえ）- 中立的
↓ lower	聞き返し - あまり興味がない	躊躇 – 躊躇 了承（はい）- 真剣な 躊躇 – 疑い 躊躇 – 強い疑い	了承（はい）-あまり気が進まない	否定（いいえ）-丁寧でない

表 2.　主観評価実験に用いた発話”ん”の最高、最低基本周波数と時間変化形状

Height / Dynamics	High-range		Mid-range		Low-range	
	max	min	max	Min	max	min
Rise	354.55Hz	182.20Hz	282.58Hz	142.17Hz	194.21Hz	98.24Hz
Flat	264.55Hz	232.39Hz	213.23Hz	178.90Hz	162.77Hz	124.97Hz
Fall	305.50Hz	119.81Hz	234.22Hz	90.87Hz	155.94Hz	65.98Hz
Rise&Fall	363.46Hz	222.15Hz	273.08Hz	153.48Hz	163.17Hz	111.44Hz

“躊躇”、“相槌”といった、相手の発言に対しての返答、心的状況、問いかけなどを伝達しており、これらの情報は他の発話内容でも共通に用いられている。これらの伝達情報は共通に用いられる韻律的特徴により分類され、文脈や、次発話の言語内容から何らかの形で推察できることが期待される。42 サンプルの F0 を観察した結果、F0 の平均的高さ・時間変化形状によってこれらの情報が担われていることが判明した。以上の観察に基づく伝達情報の印象表現を、表 1 にまとめた。これらの観察結果の妥当性を確認し、より一般的、定量的な記述を目指して次章以降に示す検討を行った。

3. 韻律伝達情報を示す印象基本表現の規定

対話音声の韻律により言語表現として表出されない情報を伝達するためには、まず、韻律によって伝達される情報の規定が必要である。伝達内容のより一般的で精確な規定を行うため、聞き手が受ける発話態度・意図に関する主観に基づく印象表現を調べた。まず、先の分析でみられた F0 の平均的高さと時間変化形状だけを制御対象として考えるため、平均的高さと時間変化形状だけが異なる一語発話”ん”を用意した。音声発話は著者自らが行い、特定の発話状況を意識しない発話を心がけた。実験に用いた F0 の高さ、時間変化形状それぞれ 3 種類、4 種類計 12 種類の発話刺激の、最高、最低周波数、及び F0 の時間変化形状を表 2 に示す。

これらの F0 の高さと時間変化形状に基づいて制御された刺激を用いて評定実験を行った。評定は聴力レベルに問題のない、日本語を母語とする成人 5 名（男性 2 名、女性 3 名）が行った。評定実験では、これら 12 種類の刺激を評定者に聞かせ、表 1 作成時の経験を参考に、次に続く事が予想される句表現、またそれらから想定される発話者の発話態度についての表記を求めた。なお、一つの発話に対して、複数以上の回答が想定される場合、それらを全て記述させた。また、印象表現としては、極力、形容詞、または副詞で直感的に表現してもらうように指示した。

この結果、全部で 67 の印象表現が得られた。以降の分析では、この中から複数回答のあった、次の 26 の基本表現を用いることとした。“疑い”、“迷い”“否定”“疑問”“反論”“納得”“同意”“了承”“明るい”“暗い”“元気な”“弱々しい”“興味がある”“興味がない”“機嫌が良い”“機嫌が悪い”“軽い”“重い”“やさしそう”“ふてぶてしい”“わくわく”“面倒くさい”“嬉しい”“怒っている”“楽しい”“うざい”。

4. 印象基本表現による伝達情報ベクトル表示

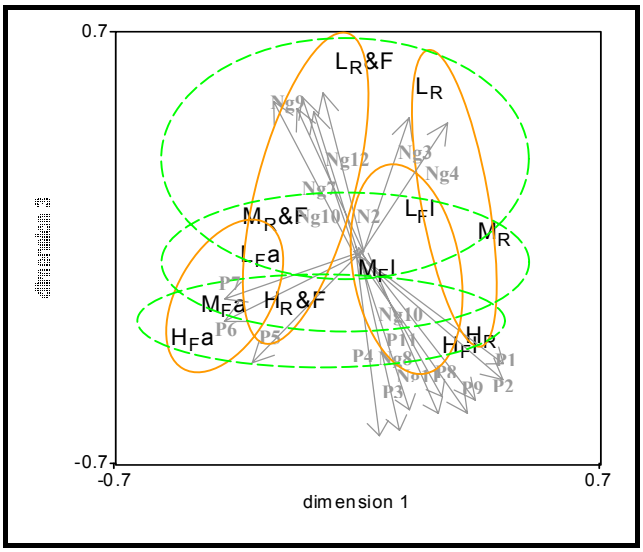
前章で得られた 26 の印象基本表現により、伝達内容を近似的にベクトル表示し、制御対象として考えた F0 の平均的な高さと時間変化形状との関係を求める。このため、前章で用いたものと同一の一語発話"ん"を用いた評定実験を行った。12 サンプルの一語発話"ん"の各々に対し、26 基本表現に、0（全く当てはまらない）～7（とても当てはまっている）の 8 段階評定、計 312 評定を求めた。被験者としては先の評定者とは異なる、聴覚レベルに問題のない、日本語を母語とする、成人 5 名（男性 1 名、女性 4 名）を用いた。評定に際しては、反復聴取可能な形で刺激呈示した。また、一人当たりの全判定時間は、30 分から 40 分程度であった。

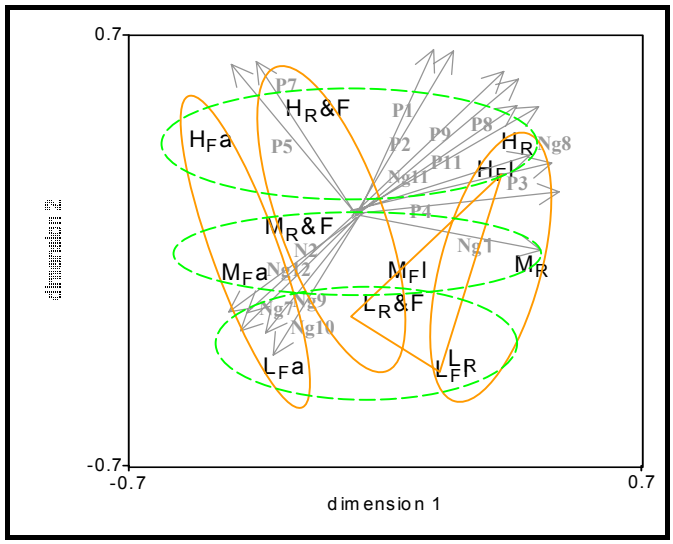
表3.　各次元に対する寄与率（VAF）

	Dimension			
	1	2	3	4
VAF	0.7398	0.8036	0.816	0.5952

表4　個人評定間のばらつき (３次元への寄与度)

Raters	Dimension		
	1	2	3
SY	0.9459	0.7651	0.8470
CA	0.4933	0.7668	0.7051
YY	0.6428	0.7821	0.8761
FY	0.7063	0.4332	0.6148
KK	0.5833	0.7868	0.6851





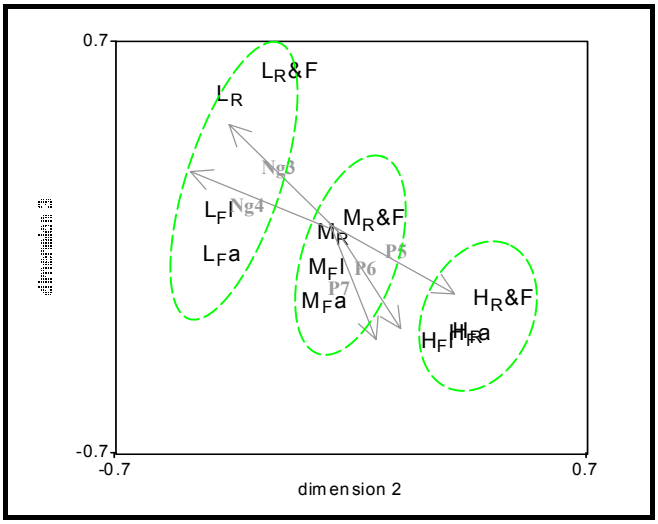


図１．ＩＮＤＳＣＡＬによる３次元空間における印象基本表現の投影（同一のＦ０高さ・時間変化形状毎に実線・破線で囲んである）。

凡例　図１中における印象基本表現(26)とＦ０(高さ3×時間変化形状4)のシンボル表記

F0 Height	Symbol
High	H
Mid	M
Low	L

Symbol	F0 Shape
R	rise
Fl	flat
Fa	fall
R&F	Rise and fall

Symbol	Expression Word (Positive)	Symbol	Expression Word (Neutral)	Symbol	Expression Word (Negative)
P1	元気な	N1	否定	Ng1	怒り
P2	明るい	N2	重い	Ng2	反論
P3	興味がある	N3	軽い	Ng3	疑い
P4	わくわく			Ng4	迷い
P5	同意			Ng5	邪魔臭い
P6	納得			Ng6	興味がない
P7	了承			Ng7	機嫌が悪い
P8	嬉しい			Ng8	面倒くさい
P9	楽しい			Ng9	ふてぶてしい
P10	機嫌が良い			Ng10	うざい
P11	優しい			Ng11	弱々しい
				Ng12	暗い

5. F0 の平均的高さ・時間変化形状と韻律伝達情報の MDS 分析

表現データに対して多次元尺度構成法(Multidimensional Scaling: 以下、MDS)を用いて分析を行った。MDS を用いることにより、距離や類似度を表すデータを基に独立な次元を求め、各サンプルが従う構造や制約の多次元表現・理解を期待した。ここでは 26 の印象基本表現に対して、各刺激間の評定値差によって得られる距離行列を入力データとした。また、ここでは、複数の評定者による評定データを用いた分析を行っているため、個人間の評定の差異も合わせて調べることとした。このため分析には、個人差を考慮した MDS 手法である INDSCAL（Individual Differences multidimensional Scaling）[4,5]を用いた。この方法では、MDS で得られる空間は個人間で共通であり、個人による対象相互の非類似度の差異は、MDS 刺激空間軸に対する、各個人のウェイトによって表される。

本分析では表 4 に示す個人毎の評定の分析結果例（3 次元）から分かるように、各次元ともに適当な重みで重み付けがなされており、評定者間に特に際立った差異は見られなかった。従って以下の分析では評定者 5 名全員のデータを合わせて分析に用いた。また、INDSCAL の次元数は経験則的判断に基づく場合が多いが、ここでは、表 3 に示す分散の割合（VAF）を参考にして 3 次元とした。

各軸の解釈を行うために、重回帰分析を用いて、それぞれの表現語に対する平均評価値を 3 次元空間に射影させた。図 1 に、重相関係数、また回帰係数の大きかった印象基本表現を、3 次元空間の座標軸に射影させたものを示す。

分析の結果、各軸は積極/消極、承服/不服、確信/不確か、また肯定/否定といった発話者の態度、心的状況を示す事が判明し、それらは F0 の高さと時間変化形状を用いて組織的に制御されている。

図 1 に示されるように、次元 1 では、“興味がある”、“明るい”、“元気な”、“弱々しい”、“機嫌が良い”、“やさしそう”、“わくわくする”、“面倒くさい”、“嬉しい”、“楽しい”、と、“暗い”、“機嫌が悪い”、“重い”、“ふてぶてしい”、“うざい”が対をなす形で現れている。“面倒くさい”、“弱々しい”を除いては、積極/消極を示す軸となっている。次に、次元 2 は“同意”、“了承”、に対して、少しのずれは見られるが“怒り”が対をなしており、相手に対して、承服/不服を表す返答を示す軸となっている。次元 3 は、“同意”、“了承”、“納得”に対して、“疑い”、“迷い”らが対をなしており、確信/不確かな返答が伺える軸となっている。

音声サンプルの分布については、次元 1 では、正の印象をもつ発話者態度の方向に、より F0 の高い発話刺激が配置されている。また、この軸と同じ向きで、Rise、Flat の F0 時間変化形状をもつ F0 の発話が分布しており、この順で、より消極的な発話となっている。次元 2 の分布からは、F0 が低くなるにつれて不服な返答をしている事が分かる。次元 2 の軸は、Rise&Fall、Fall の F0 の時間変化形状をもつ F0 発話が同じ向きで現れ、また、この順で、より不服な返答が表現されている。最後の次元 3 では、F0 の刺激が高い刺激が、より確信をもった態度を示しており、次元 2 の軸同様に、F0 の時間変化形状が Fall、Rise&Fall が関わっている事が伺える。

以上の結果を、各々の基本周波数の制御の観点から見直すと、F0 の高さは発話の積極/消極的な発話態度により規定する事が考えられる。図 1 の結果が示すように、全ての次元に共通して、F0 が高いほど、積極、承服、確信などを示しておる事から、高さは、これらの情報によって制御される。一方、F0 の時間変化形状は、“返答”のような、肯定/否定の発話意図によって制御される。次元 2、3 の軸が示すように、承服/不服を表す返答、または確信/不確かな返答という否定/肯定に呼応して、Fall、Rise&Fall を制御できる。

6. まとめ

本稿では、対話音声の韻律制御の第一歩として、F0 パタンと発話態度との依存関係を分析した。実際の対話に出現した一語発話“ん”の分析から、F0 パタンの平均的高さ、時間変化形状によって発話態度、が制御されている。この対応関係の妥当性を確認するために、モデル音声を用い、印象基本表現に基づいた主観評価実験を行った。多次元尺度構成法を用いた分析の結果、F0 の平均的高さが、“積極/消極”、“承服/不服”、“確信/不確か”な発話態度を示し、F0 の時間変化形状が、肯定/否定とよく呼応することが分かった。

以上の分析結果により、これまでに調べた言語的内容だけでなく、発話態度や発話意図に基づいた F0 制御の可能性が示唆された。今後、音声合成の韻律制御モデルの実現を目指して、他の韻律情報の制御とも合わせた制御計算モデル化を進めてゆく。

文献

[1] Sagisaka,Y; Yamashita, T; Kokenawa, Y. "Speech Synthesis with Attitude" Speech Prosody 2004, pp.401-404, March 2004.

[2] Campbell, N. “Speech & Expression; the value of a longitudinal courpus”, Processing of the Language Resources & Evaluation Conference, Lisbon 2004

[3] Nick Campbell, Donna Erickson “What do People Hear? A study of the Perception of non-verbal Affective Information in Conversational Speech” Journal of the Phonetic Society of Japan, Vol.8, no1, pp.9-28, April 2004

[4] 津崎実，河井恒，“音質の長期的変動に関する印象空間－方法論と実験結果－”，日本音響学会講演

論文集, 2000 年 9 月, 237-238

[5] Borg, I., Groene, P. “Modern Multidimensional Scaling: Theory and Application” 1997, Springer, N.Y.