

『日本語日常会話コーパス』にみる 日常会話音声の基本周波数と談話行為の関係

石本 祐一 (ものづくり大学/国立国語研究所) *

Relationships between Dialog Acts and the Fundamental Frequencies of Everyday Speech on the Corpus of Everyday Japanese Conversation

Yuichi Ishimoto (Institute of Technologists/NINJAL)

要旨

パラ言語・非言語情報の影響により音声の韻律が様々に変動することが知られているが、日常生活の多様な状況を網羅するような音声データがこれまで存在せず大規模な分析を行うことが困難だった。しかし、多様な種類の日常会話をバランス良く収録した大規模コーパスである『日本語日常会話コーパス』の登場により、日常会話音声の様々な面を観察することが可能となった。そこで、『日本語日常会話コーパス』の音声の分析を通して、日常会話音声において観察される基本周波数の多様性を示す。先の研究において会話場面によって同一話者であっても発話の基本周波数が異なっていることを示したが、本稿では発話の談話機能に着目し、談話行為が基本周波数に与える影響を調べた。分析の結果、フィードバック肯定に分類される発話では基本周波数が低くなる傾向になることがわかった。また、談話行為によっては会話場面とは異なる影響を基本周波数に与える可能性が示唆された。

1. はじめに

2022年3月より本公開された『日本語日常会話コーパス』(Corpus of Everyday Japanese Conversation, CEJC)(Koiso et al. 2022)には多様な種類の日常会話がバランス良く収録されており、これにより日常会話の言語活動を多角的に研究できるようになった。例えば、自発音声は発話意図や感情を代表とするパラ言語・非言語情報の影響により韻律が様々に変動することが知られている(森ほか 2014)が、CEJCにより日常の自発音声の多様性について大規模な定量的分析を行うことができる。本稿では、日常会話音声における韻律の多様性を示すことを目的として、声の高さを表す音響特徴量である基本周波数に着目し、CEJCに収録されている自発音声データから発話の談話行為が基本周波数の違いとしてどのように現れるのか調べた結果について報告する。

* y_ishimoto@iot.ac.jp

表 1 対象調査協力者

協力者 ID	年代	性別	職業	会話場面数	合計発話時間 (秒)
C001	40 代	女性	会社員・公務員等	9	6300.09
C002	50 代	女性	会社員・公務員等	14	6556.12
K001	30 代	女性	会社員・公務員等	12	6807.65
K002	50 代	女性	自営業・自由業	11	6201.68
K003	20 代	女性	大学生	14	6189.69
K004	40 代	女性	パート・アルバイト	12	7462.08
S001	50 代	男性	会社員・公務員等	9	2863.24
T001	30 代	男性	自営業・自由業	11	6789.28
T002	40 代	男性	自営業・自由業	13	7679.76
T003	30 代	女性	専業主婦・主夫	12	5798.67
T004	60 代	女性	専業主婦・主夫	13	7579.69
T006	20 代	男性	大学院生	10	4544.10
T009	20 代	女性	大学生	21	6320.67
T010	20 代	男性	大学生	10	6697.03
T013	60 代	男性	会社員・公務員等	17	6530.13
T014	40 代	女性	自営業・自由業	17	7817.85
T015	50 代	男性	会社員・公務員等	17	5646.89

2. データ

2.1 コーパス

CEJC は日常生活で自然に生じる会話を収録することが念頭に置かれており、研究者の介入による影響を排除するために調査協力者自身による映像の録画および音声の録音が行われている。調査協力者の家族や友人等はそれぞれの会話場面に応じて登場するが、基本的に調査協力者は必ず参加していることになることから、調査協力者の音声のみを対象とした。さらに、発話相手の属性が一つに同定できる会話場面に限定し、本稿では表 1 に示す調査協力者 17 名の分析を行う。すべての調査協力者において 9 場面以上の異なる会話場面があり、平均して調査協力者一人あたり約 100 分の発話時間となっている⁽¹⁾。

また、CEJC には談話行為情報として発話単位ごとに談話行為タグ (Iseki et al. 2019) が付与されており、本稿では談話行為情報レベル 1 タグ (基本的な談話機能に関わる情報) を利用することとする。分析対象となる発話と談話行為との対応を表 2 に示す。談話行為ごとに発話数にはばらつきがあり、「注意獲得」や「注意獲得への対処」のように対応する発話が存在しない談話行為もある。

⁽¹⁾ CEJC では日常生活の自然な振る舞いを妨げるヘッドセットマイク等の利用は避けて各参加者が IC レコーダを胸の位置に配置して音声を録音しているため、収録状況や参加者間の位置関係によって他の話者の音声や雑音が入り込んでしまうことがあり、すべての発話時間が分析に利用できるわけではない。

表 2 談話行為情報（レベル 1 タグ）と対象発話数

タグ	談話機能	発話数	
T_Inform	タスク系	情報提供	1830
T_Question		情報要求	291
T_Answer		返答としての情報・確認提供	418
T_Request		依頼系	191
T_A-Request		依頼系への対処	167
T_Offer		申し出	19
T_A-Offer		申し出への対処	35
T_Attention		注意獲得	0
T_A-Attention		注意獲得への対処	0
T_SelfSpeech		独り言	158
S_Greeting		社会的付き合い管理系	挨拶
S_Apology	謝罪		20
S_A-Apology	謝罪への対処		7
S_Thanking	感謝		29
S_A-Thanking	感謝への対処		2
FB_Positive	フィードバック系	フィードバック肯定	3193
FB_Acceptance		フィードバック了承	7
FB_Repetition		フィードバック反復	126
FB_Lexical		フィードバック語彙的反応	31
O_NV	その他	非言語行動	14
O_Cue		合図	89
O_Undet		不明	67

2.2 基本周波数抽出

CEJC には基本周波数 ($F_0^{(2)}$) の生データの情報は付与されていない。そこで、音声分析合成システム STRAIGHT の音源情報分析関数 (Kawahara et al. 2017) を用い、音声データから 1 ms 間隔で F_0 推定を行った。なお、有声/無声判定は STRAIGHT から得られる非周期性情報を利用した。さらに、推定エラーの影響を低減させるために、推定された F_0 のうち上位 5% と下位 5% の値を取り除いた。最終的に、性差を正規化するため各話者ごとの平均値と標準偏差を用いた z-score に変換した。

3. 日常会話音声の多様性

3.1 会話相手と基本周波数の関係

これまでに調査協力者から見た会話相手の属性と調査協力者の発話の平均 F_0 との関係調べており、詳細は Ishimoto and Koiso (2019) を参照されたいが図 1 に示すような関係が観察された。すなわち、発話の平均 F_0 は、子ども・配偶者・父母といった近親者に対して低くなる、友人に対しての発話では高くなる、同僚・取引先・客に対しては友人よりは低く近親者よ

(2) 近年ではフォルマント周波数 (F_1, F_2, F_3, \dots) との意味的な違いを明確にするために基本周波数を F_0 と表記することもある (Titze et al. 2015) が、本稿では従来から主に使用されてきた F_0 という表記を用いる。

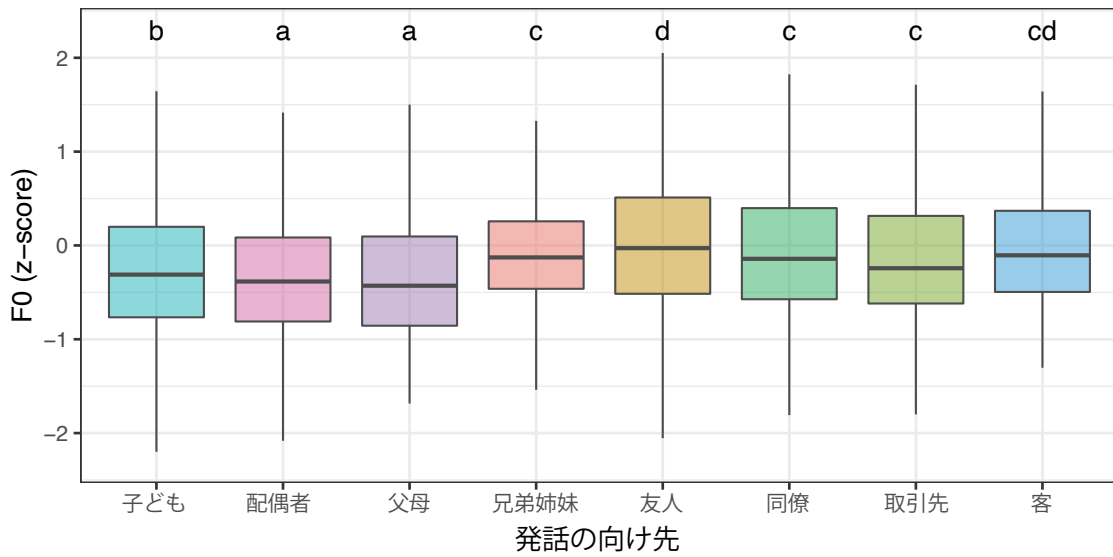


図1 会話相手の属性に対する発話の平均 F0 (Ishimoto and Koiso (2019) より再掲)

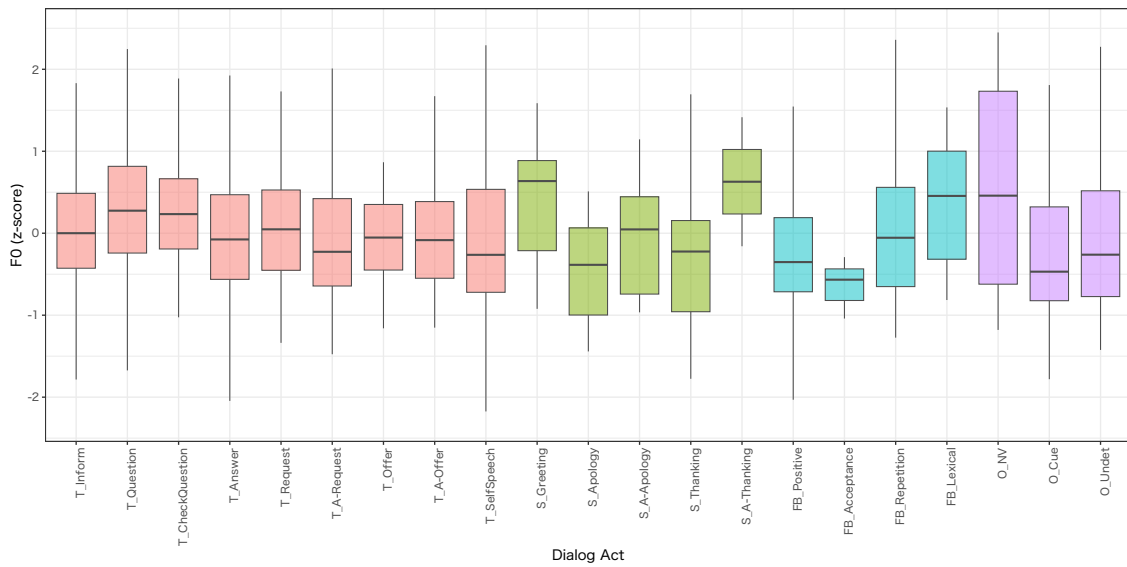


図2 談話行為情報に対する発話の平均 F0

りは高くなる、といったように会話相手によって異なっていた。このように、会話場面や発話対象によって発話の F0 に変動が生じていることが示された。

そこで次節では、さらに発話の韻律の多様性を示すため、談話行為と F0 との関係について調べる。

3.2 談話行為と基本周波数の関係

発話の平均 F0 と表 2 に示した談話行為情報との関係を図示したものが図 2 である。

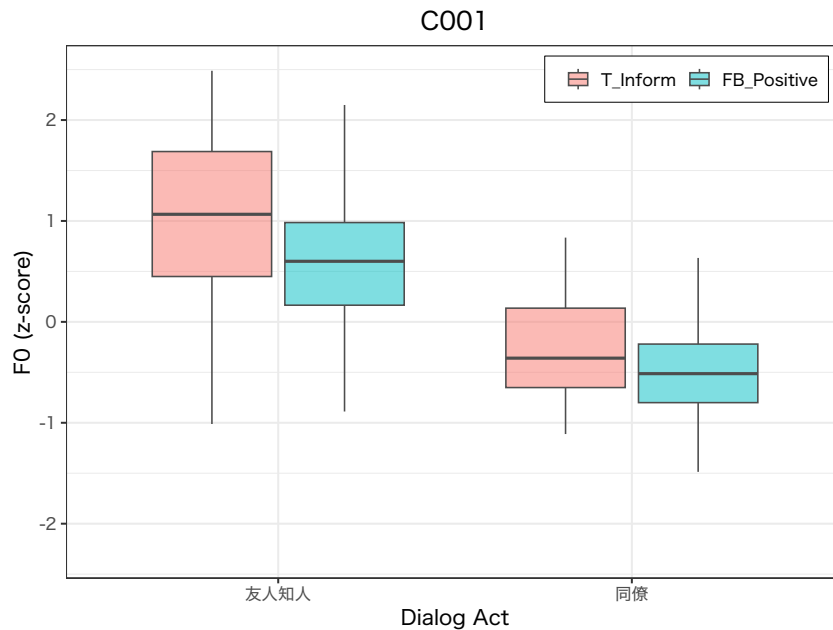


図3 C001 の談話行為情報と F0 との関係

談話行為を「タスク系」や「社会的付き合い管理系」、「フィードバック系」といった大きな括りで見たととき、F0 の違いに定性的な傾向は見られなかった。また、談話行為タグを個別に見ると、S.Greeting（挨拶）や S_A-Thanking（感謝への対処）は他よりも高く、FB_Acceptance（フィードバック了承）は他よりも低い、というような分布になっているが、これは対象となる発話数が少ないことが原因である可能性が大きい。

そこで、発話数が十分にある談話行為に着目し、T_Inform（情報提供）と FB_Positive（フィードバック肯定）に限定して、個人ごとに違いを観察する。加えて、図 1 にて示されたような会話相手の属性による違いも合わせて調べることにした。

調査協力者 C001 における T_Inform と FB_Positive の発話の平均 F0 を図 3 に示す。C001 では、T_Inform よりも FB_Positive の方が F0 が低くなる傾向が見られた。また、同じ談話行為でも会話相手の属性によって高さが変わっているが、T_Inform と FB_Positive の違いはそれぞれの属性で同程度であった。これは、会話場面により発話の F0 のベースラインが変わる可能性を示唆していると言える。

次に、調査協力者 K004 における T_Inform と FB_Positive の発話の平均 F0 を図 4 に示す。K004 においても C001 と同様、T_Inform よりも FB_Positive は低くなる傾向が見られた。しかし、「子ども」に対する発話では T_Inform と FB_Positive に大きな違いは見られず、談話行為による違いが誰が相手でも生じるとは言えない結果となっている。

最後に、調査協力者 T002 における T_Inform と FB_Positive の発話の平均 F0 を図 5 に示す。T002 では、「取引先」に対する発話では談話行為による F0 の違いが見られるが、「配偶者」に対する発話ではあまり F0 は変わらない。これも K004 同様、会話相手の属性により談

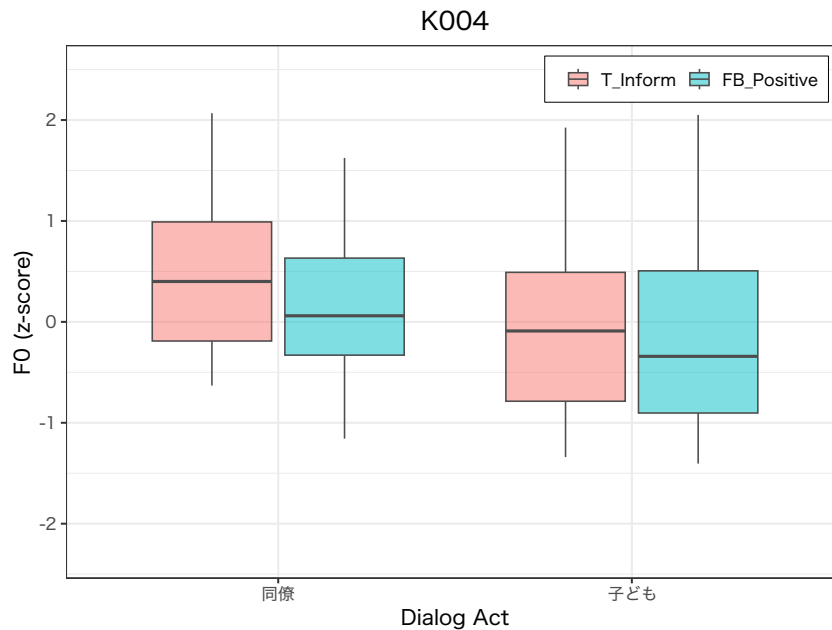


図 4 K004 の談話行為情報と F0 との関係

話行為による違いが生じない場合の存在を示唆している。また、先の研究（図 1）で見られた「配偶者」に対する発話の F0 が「取引先等」相手よりも低くなるという傾向が、T002 の FB_Positive の発話では見られないという結果になっている。

3.3 議論

3.2 節の結果から、十分多い発話数がデータから得られた談話行為に関してのみの検証にとどまったが、FB_Positive では発話の平均 F0 が低くなりやすいという傾向が見られた。

また、先の研究（図 1）において会話相手の属性（会話場面）によって発話の F0 が異なることを示していたが、今回の調査協力者 K004 や T002 の結果からは発話の談話機能によっては会話相手の属性が発話の F0 に与える影響が小さいことが示された。これは、会話相手の属性よりも談話機能の方が発話の F0 に与える影響が強いことを示しているともいえるが、一方で調査協力者 C001 のように会話相手の属性と発話の談話機能の影響を両方ともそのまま反映した発話となっている場合もある。このような個人による違いがどのような要因によるものかはさらなる検証が必要である。加えて、T_Inform と FB_Positive 以外の談話行為に関しては、今回対象とした発話の数では明確な傾向を述べるのが難しく、今後さらにデータ数を増やした検証を行う予定である。

4. おわりに

日常会話音声における基本周波数の多様性を示すために、『日本語日常会話コーパス』(CEJC)に含まれている音声の発話の平均 F0 を、発話の談話機能と会話相手の属性の面から調べた。その結果、フィードバック肯定に該当する発話の基本周波数が低くなりやすい傾向が見られた。



図 5 T002 の談話行為情報と F0 との関係

また、個人によって談話機能が基本周波数に与える影響の大きさが異なり、先の研究によって示されていた会話相手の属性による影響も発話の談話機能によって現れない可能性が示唆された。しかし、情報提供やフィードバック肯定以外の多くの談話機能の影響は今回の分析においてはデータ数が少なく明確な傾向を掴むことができなかったため、今後の課題としたい。

謝 辞

本研究は、国立国語研究所 共同研究プロジェクト「多世代会話コーパスに基づく話し言葉の総合的研究」、および JSPS 科研費 18K11514 の助成を受けたものです。

文 献

- Hanae Koiso, Haruka Amatani, Yasuharu Den, Yuriko Iseki, Yuichi Ishimoto, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, Yasuyuki Usuda, and Yuka Watanabe (2022). "Design and Evaluation of the Corpus of Everyday Japanese Conversation." *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 5587–5594. Marseille, France: European Language Resources Association.
- 森大毅・前川喜久雄・粕谷英樹 (2014). 『音声は何を伝えているか：感情・パラ言語情報・個人性の音声科学』音響サイエンスシリーズ/日本音響学会編:12 コロナ社.
- Yuriko Iseki, Keisuke Kadota, and Yasuharu Den (2019). "Characteristics of everyday conversation derived from the analysis of dialog act annotation." *2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, pp. 1–6.

Ingo R Titze, Ronald J Baken, Kenneth W Bozeman, Svante Granqvist, Nathalie Henrich, Christian T Herbst, David M Howard, Eric J Hunter, Dean Kaelin, Raymond D Kent, Jody Kreiman, Malte Kob, Anders Löfqvist, Scott McCoy, Donald G Miller, Hubert Noé, Ronald C Scherer, John R Smith, Brad H Story, Jan G Švec, Sten Ternström, and Joe Wolfe (2015). “Toward a consensus on symbolic notation of harmonics, resonances, and formants in vocalization.” *The Journal of the Acoustical Society of America*, 137:5, pp. 3005–3007.

Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, and Tomoki Toda (2017). “Accurate estimation of f_0 and aperiodicity based on periodicity detector residuals and deviations of phase derivatives.” *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 1556–1564.

Yuichi Ishimoto, and Hanae Koiso (2019). “Prosodic diversity according to relationship among participants in everyday Japanese conversation.” *Proceedings of LPSS 2019*, pp. 62–66.

関連 URL

『日本語日常会話コーパス』

<https://www2.ninjal.ac.jp/conversation/cejc.html>

音声分析合成システム STRAIGHT

https://github.com/HidekiKawahara/YANGstraight_source