

『日本語日常会話コーパス』で観察される日常会話音声のF0の多様性

著者	石本 祐一
雑誌名	言語資源活用ワークショップ発表論文集
巻	6
ページ	210-217
発行年	2021
URL	http://doi.org/10.15084/00003495

『日本語日常会話コーパス』で観察される日常会話音声の F0 の多様性

石本 祐一 (国立国語研究所コーパス開発センター) *

A diversity of fundamental frequency in everyday speech observed in the corpus of everyday Japanese conversation

Yuichi Ishimoto (National Institute for Japanese Language and Linguistics)

要旨

パラ言語情報や感情の影響により音声の F0 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はこれまで存在せず、日常話されている音声の F0 がどのようなものであるか大規模な分析を行うことが困難だった。国立国語研究所で構築を進めている『日本語日常会話コーパス』は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り、このコーパスを活用することで日常会話音声の様々な面を観察することが可能となる。本稿では、『日本語日常会話コーパス』モニター版（2018 年度版・2020 年度版）とこれまで日本語自発音声の大規模コーパスとして幅広い分野で活用されてきた『日本語話し言葉コーパス』のそれぞれ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 F0 の多様性を示す。分析の結果、『日本語日常会話コーパス』に収録されている音声は『日本語話し言葉コーパス』よりも発話の平均 F0 が幅広く分布していることがわかった。

1. はじめに

国立国語研究所では 2016 年より多様な種類の日常会話をバランス良く収録した大規模な日常会話コーパスである『日本語日常会話コーパス』(Corpus of Everyday Japanese Conversation, CEJC) の構築を進めており (Koiso et al. 2018, 小磯ほか 2019)、2021 年度末の本公開を予定している。自発音声はパラ言語情報や感情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、日常生活の多様な状況を網羅するような音声データはこれまでなく、CEJC により日常の自発音声の声の高さの多様性について大規模な定量的分析が行われることが期待できる。本稿では、日常会話音声における声の高さ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 (F0) に着目し、『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese, CSJ) (Maekawa et al. 2000) と比較しつつ CEJC モニター公開版に収録されている自発音声データから発話ごとにどの程度の F0 の違いが現れるのか調べた結果について報告する。

* yishi@ninjal.ac.jp

表1 インフォーマントの属性と会話場面数・発話時間

話者 ID	年齢	性別	職業	会話場面数	合計発話時間 (秒)
C001	40-44	女性	会社員・公務員等	9	6300.09
C002	55-59	女性	会社員・公務員等	14	6556.12
K001	35-39	女性	会社員・公務員等	12	6807.65
K002	50-54	女性	自営業・自由業	11	6201.68
K003	20-24	女性	大学生	14	6189.69
K004	40-44	女性	パート・アルバイト	12	7462.08
K005	35-39	女性	自営業・自由業	17	7655.34
K006	60-64	女性	自営業・自由業	14	7009.22
K007	25-29	男性	パート・アルバイト	14	6565.60
K008	50-54	女性	自営業・自由業	12	5181.36
K009	25-29	女性	会社員・公務員等	15	2972.47
S001	50-54	男性	会社員・公務員等	9	2863.24
S002	35-39	男性	会社員・公務員等	11	7796.11
T001	35-39	男性	自営業・自由業	11	6789.28
T002	40-44	男性	自営業・自由業	13	7679.76
T003	35-39	女性	専業主婦・主夫	12	5798.67
T004	60-64	女性	専業主婦・主夫	13	7579.69
T005	35-39	男性	会社員・公務員等	14	3303.99
T006	25-29	男性	大学院生	10	4544.10
T007	70-74	男性	無職・定年退職	14	10520.46
T008	30-34	女性	会社員・公務員等	22	5978.90
T009	20-24	女性	大学生	21	6320.67
T010	20-24	男性	大学生	10	6697.03
T011	40-44	女性	パート・アルバイト	10	5173.64
T013	65-69	男性	会社員・公務員等	17	6530.13
T014	45-49	女性	自営業・自由業	17	7817.85
T015	50-54	男性	会社員・公務員等	17	5646.89
T016	40-44	男性	会社員・公務員等	10	3063.25
T017	60-64	女性	会社員・公務員等	12	4945.04
T018	50-54	男性	自営業・自由業	9	5271.27
T019	40-44	男性	会社員・公務員等	11	3623.28
T020	40-44	男性	会社員・公務員等	23	6218.07
T021	65-69	女性	自営業・自由業	13	8860.50

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インフォーマント（研究協力者）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。インフォーマント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、基本的にインフォーマントは必ず参加していることになるので、本稿ではイン

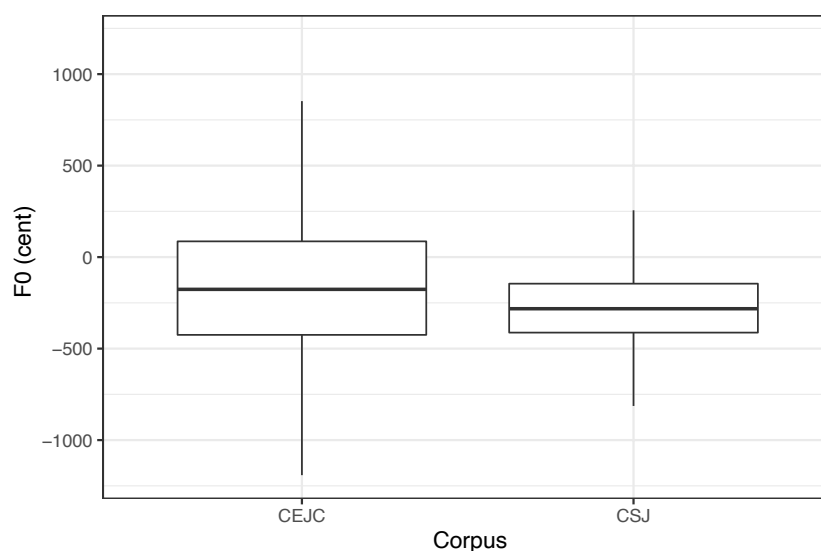


図1 発話の平均 F0 の分布

フォーマントの音声のみを対象として分析した。CEJC モニター公開版（2018 年度・2020 年度）の 100 時間分の中から、分析に用いたインフォーマント 33 名の属性、会話場面の数、発話時間の合計を表 1 に示す。すべてのインフォーマントにおいて 9 場面以上の異なる会話場面があり、平均してインフォーマント一人あたり約 102 分の発話時間となっている。

なお、音声の録音について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各参加者が IC レコーダを胸の位置に配置しているため、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場合もあり、常にクリーンな音声を得られているわけではない。

同じく自発発話を扱ったコーパスとして、比較のため CSJ も取り上げる。本稿では CSJ コアの学会講演・模擬講演・対話のデータから話者 219 名の発話を用いた。分析対象となる発話の総時間は約 40 時間である。

2.2 基本周波数抽出

CEJC モニター公開版には F0 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、モニター公開版の音声データから 1 ms 間隔で F0 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 F0 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を基準値としたセント値に変換した。セントとは、周波数を基準からの比率で現した単位で、100 セントで半音、1200 セントで 1 オクターブの違いがあることを示す。

CSJ には F0 情報が付与されているが、F0 推定エラーの影響をなるべく CEJC と同等にするために、STRAIGHT によって抽出された F0 を用いることとした。

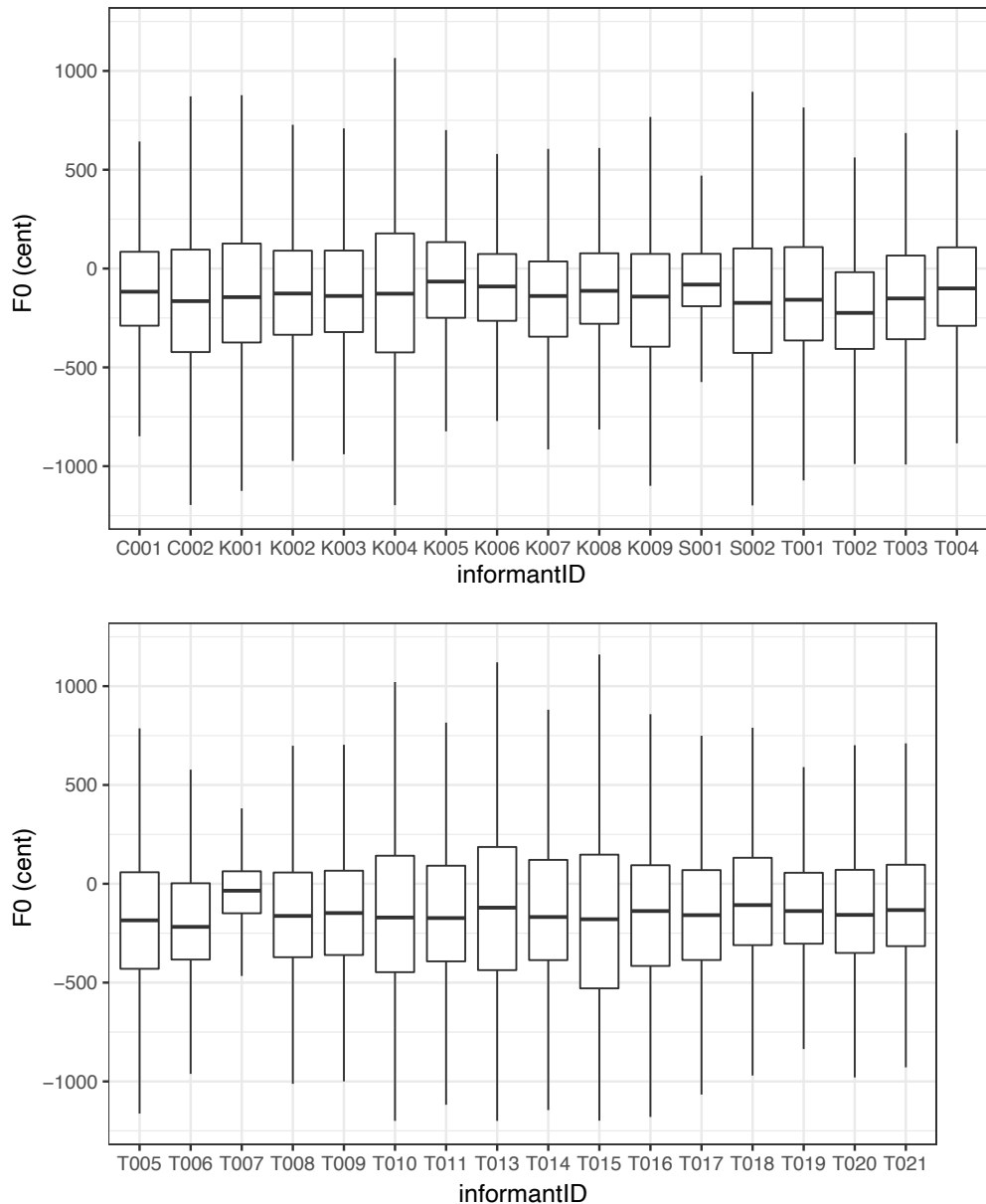


図2 CEJCの各インフォーマントの発話の平均F0の分布

3. 自発発話のF0

3.1 発話の平均F0の分布

コーパス全体として発話ごとにF0がどの程度異なっているかを調べるために、CEJCの各発話の平均F0を求めた。さらに、比較のためにCSJに対しても同様に発話の平均F0を求めた。それぞれのコーパスの結果を図1に示す。なお、CEJCの発話単位としてはコーパスに付与されている「長い発話単位」を用い、CSJは節単位強境界以上を発話区切りとして扱うこととした。

図1の箱ひげ図において、CEJCの第1四分位点は-453.3セント、第3四分位点80.1セン

トで、533.4 セントの幅を持つのに対し、CSJ の第 1 四分位点は-413.5 セント、第 3 四分位点-145.3 セントで、268.2 セントの幅となっており、CEJC の方が CSJ に比べて 2 倍近く広い分布を示している。また、CEJC と CSJ では第 1 四分位点よりも第 3 四分位点の方が違いが大きいことから、CEJC は特に高い F0 を持つ発話が CSJ よりも多く現れていると考えられる。

以上のことから、CEJC はこれまで CSJ では十分に観察できなかった種類の発話を多く含んでおり、自発発話の F0 の多様性が反映されたコーパスとなっているといえる。

次に、CEJC の発話の平均 F0 のインフォーマントごとの分布を図 2 に示す。インフォーマントによって差はあるものの、ほとんどのインフォーマントの発話が図 1 に示された全体の分布と同様の傾向を示しており、CSJ よりも幅広い平均 F0 を持つ発話であることがわかる。すなわち、インフォーマント個人においても F0 が高い発話と低い発話が混在しており、一人の話者に着目したときの自発発話の多様性が CEJC で観察できるといえる。

3.2 F0 が高い発話と低い発話の傾向

ここではインフォーマント 2 名 (K004, T015) の発話に着目し、平均 F0 が高い発話と低い発話の例を観察する。

図 3 は、図 2 のインフォーマント K004 の分布の第 3 四分位点付近に該当する平均 F0 を示す発話「すごいいい匂いだよ」の F0 軌跡を示している。この発話は自宅で家族とお菓子作りをしている際に発せられたものであり、喜びの感情が込められていることから全体的に F0 が高くなっている。特に、発話前半が話者の平均 F0 よりも高く、発話後半も F0 がそれほど落ち切らないという特徴がある。

図 4 は、図 2 のインフォーマント K004 の分布の第 1 四分位点付近に該当する平均 F0 を示す発話「あの面接すんのに二十年前のスーツはまずいだろうと思って」の F0 軌跡を示している。この発話は友人宅で次男のサッカー部のママ友とお茶会をしている際に発せられたものである。通常、友人と話をする際は F0 が高くなる傾向にあるが、この発話では発話全体が話者の平均 F0 を越えない高さであるとともに発話中の F0 のレンジが狭くなっている。なお、インフォーマント K004 はこの会話場面全体で常に F0 が低い発話を行なっているわけではなく、発話内容に対する話者の心境が低く平板な F0 として現れているといえる。

図 5 は、図 2 のインフォーマント T015 の分布の第 3 四分位点付近に該当する平均 F0 を示す発話「どうせ一位になるってわかってたらお父は一走やるんだよ」の F0 軌跡を示している。この発話は飲食店で近隣の町会のメンバーと飲み会をしている際に発せられている。平均すると高い F0 となる発話だが、発話前半の F0 は高い一方で終盤の F0 はかなり低くなっており、ほぼ 2 オクターブ (2400 セント) の広いレンジでの発話となっている。

図 6 は、図 2 のインフォーマント T015 の分布の第 1 四分位点付近に該当する平均 F0 を示す発話「生姜ここにあるよ」の F0 軌跡を示している。この発話は自宅で友人・配偶者とキャンプの打ち合わせをしている際に発せられたものである。この会話場面ではインフォーマント T015 は楽しく話すことがほとんどで高い F0 の発話が多く、この発話の前までやはり高い F0 の発話が続いていたが、この発話は直前の家族の質問を受けてその家族だけに向けて発声した

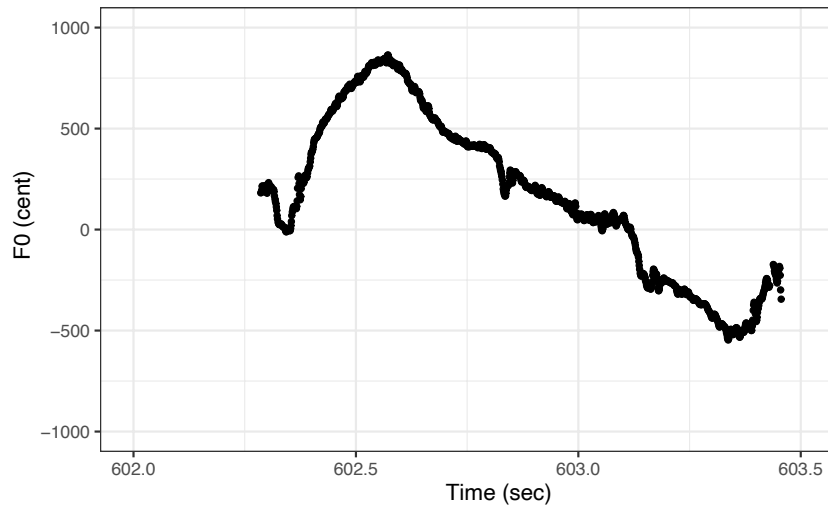


図3 インフォーマント K004 の発話「すごいいい匂いだよ」の F0

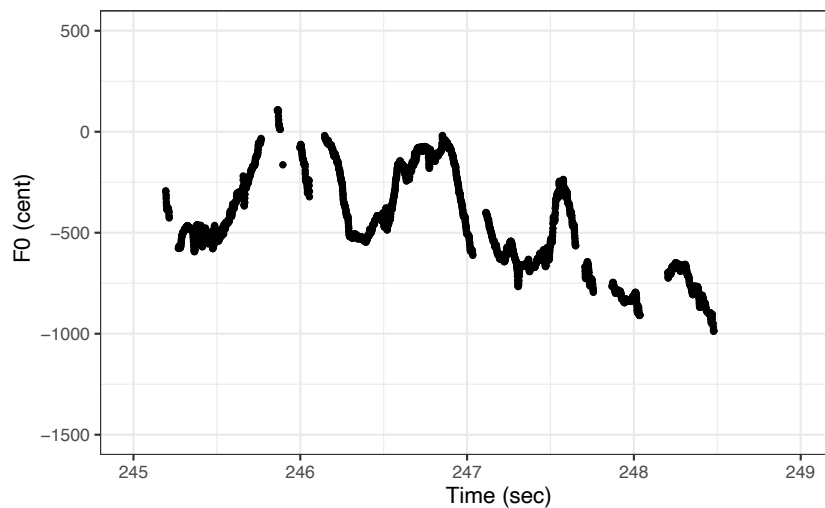


図4 インフォーマント K004 の発話「あの面接すんの二十年前のスーツはまずいだろうと思って」の F0

ことにより、F0 が低い発話となった。発話全体として話者の平均 F0 を越えていないという点では図4の例と同じではあるが、発話終盤の F0 は十分落ちておりレンジが狭くなってはいないという点で異なる。

3.3 議論

読み上げ音声や独話音声では、どの発話もほぼ同じ高さで始まり、発話中に徐々に F0 は下がっていき、最終的に発話末で話者の下限となる高さまで下がりることが多い。一方、3.2 節に示したように、日常会話音声では同一話者でも会話場面や発話対象によって発話の F0 が大きく異なっていることがわかる。例えば、発話の開始時の F0 の高さを見ると、高い F0 で始まらない発話 (図3、図6) が見られた。さらに、発話の終わりで F0 が下がり切らない場合

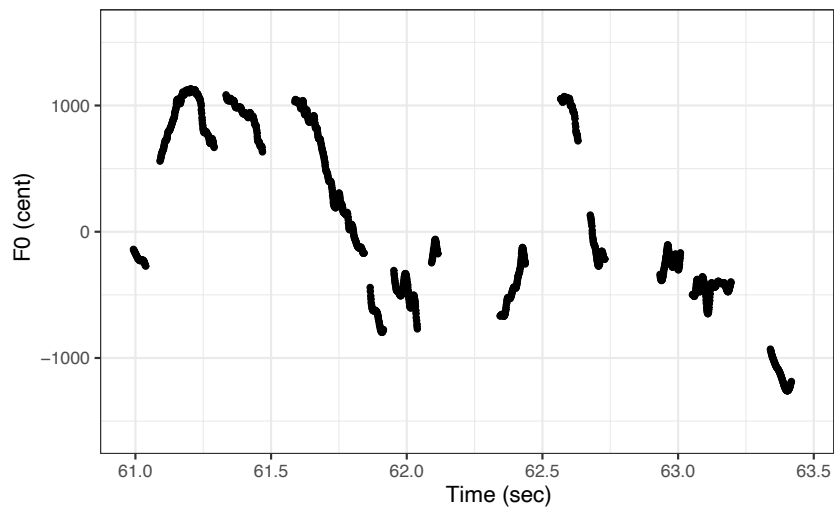


図5 インフォーマント T015 の発話「どうせ一位になるってわかってたらお父は一走やるんだよ」の F0

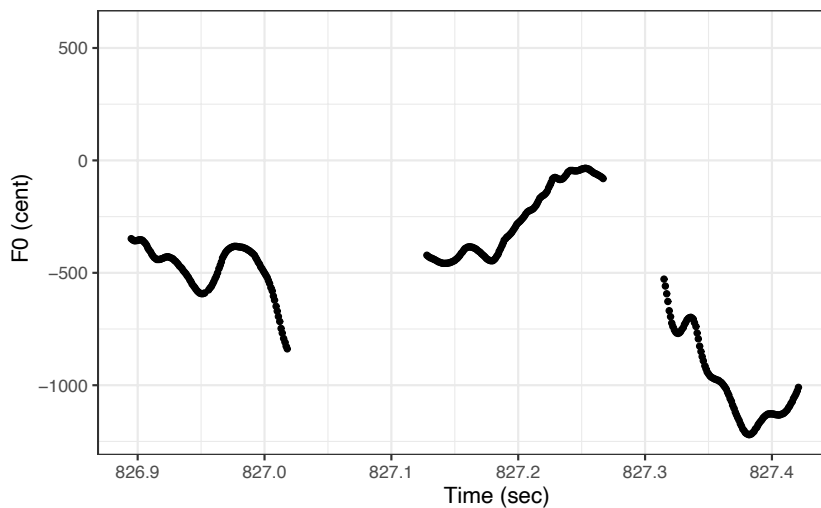


図6 インフォーマント T015 の発話「生姜ここにあるよ」の F0

(図4)もある。また、発話の終わりで F0 が下限まで下がりきっていても、開始時の高さの違いから発話の F0 レンジが異なることとなる (図5, 6)。

図2から明らかなように、このような個人内の F0 の違いが大きい発話が CEJC には数多くある。CEJC データを基にして発話の F0 の多様性についてさらに詳細な分析を行うことが今後の課題である。

4. おわりに

日常会話音声における F0 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 の分布を『日本語話し言葉コーパス』(CSJ)と比較して調べ

た。その結果、CEJCの発話のF0はCSJよりも2倍近い広さの分布を示し、これまで自発発話の分析に用いられてきたCSJの音声とは異なる特徴を日常会話音声を持つことが示唆された。また、CEJCの日常会話音声のF0の多様性は個人間だけではなく個人内にも現れていることが示された。

謝 辞

本研究は、国立国語研究所 機関拠点型基幹研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」、コーパス基礎研究「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」、およびJSPS 科研費 18K11514の助成を受けたものです。

文 献

- Hanae Koiso, Yasuharu Den, Yuriko Iseki, Wakako Kashino, Yoshiko Kawabata, Ken'ya Nishikawa, Yayoi Tanaka, and Yasuyuki Usuda (2018). "Construction of the Corpus of Everyday Japanese Conversation: An Interim Report." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pp. 4259–4264.: European Language Resources Association (ELRA).
- 小磯花絵・天谷晴香・石本祐一・居關友里子・白田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉 (2019). 「『日本語日常会話コーパス』モニター公開版の設計と特徴」 言語処理学会第25回年次大会発表論文集, pp. 367–370.
- 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.
- Kikuo Maekawa, Hanae Koiso, Sadaoki Furui, and Hitoshi Isahara (2000). "Spontaneous Speech Corpus of Japanese." *Proceedings of LREC-2000 (Second International Conference on Language Resources and Evaluation)* Vol. 2., pp. 947–952.
- Hideki Kawahara, Ken-Ichi Sakakibara, Masanori Morise, Hideki Banno, and Tomoki Toda (2017). "Accurate estimation of f0 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.

関連 URL

『日本語日常会話コーパス』

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

『日本語話し言葉コーパス』

<https://ccd.ninjal.ac.jp/csj/index.html>

音声分析合成システム STRAIGHT

https://github.com/HidekiKawahara/YANGstraight_source