

声のピッチ感の錯覚と話者の寸法知覚: スペクトル傾斜との関係性

内田 照久^{1,a)}

概要: 音声のスペクトルの周波数軸を伸長して声の音色を系統的に変化させると、基本周波数 (f_0) が低い声であっても、高い声だと誤って知覚されてしまう錯覚が生ずる。また、その周波数軸の伸長によって声道長の縮小を模した音色からは、体格の小さい話者が想起される。本研究では、スペクトル傾斜を操作することで、フォルマントの位置は固定したまま音色を変化させた場合には、どのような知覚が生ずるのかを検証した。WORLD を用いて声の音色と高さを制御した。(1) スペクトル周波数軸の伸縮 — スペクトル傾斜の増減、(2) 通常 f_0 パターンの昇降 — 平板化 f_0 の昇降、の各要因を組み合わせた音声の対比較による主観評価を行った。スペクトル傾斜による音色の違いは、声の高低判断に影響を与えたが、その効果は周波数軸の伸縮より小さい可能性が示唆された。話者の寸法知覚に関しては、スペクトル傾斜よりも f_0 が、その f_0 よりもスペクトル軸の伸縮が、すなわち声道長の長短の要因が最も支配的な役割を担っていた。

キーワード: ピッチ, 声質, 声道長, スペクトル重心, WORLD

The voice pitch illusion and perception of speakers' body size: Relationship with spectral gradient

Abstract:

When the frequency axis of the voice spectrum is extended, the timbre of the voice changes systematically. Then, even if the voice has a low fundamental frequency (f_0), the illusion that it is mistakenly perceived as a high voice occurs. Besides, the timbre that imitates the reduction of the vocal tract length due to the frequency axis extension reminds us of a small physique speaker. This study examined what perception occurs when the timbre changes while the formant position is fixed by manipulating the spectral gradient. WORLD was used to control the timbre and pitch of the voice. Subjective evaluation was performed by comparing speech pairs that combined experiment factors. Those are (1) expansion and contraction of the spectral frequency axis - frequency band emphasis and (2) standard f_0 pattern elevation - flattened f_0 elevation. The difference in timbre due to the spectral gradient affected the judgment of voice pitch, but it was suggested that the effect might be smaller than the expansion and contraction of the frequency axis. Regarding the speaker's body size perception, the spectral axis's expansion and contraction played the most dominant role than the other factor, and f_0 played a more dominant role than the spectral gradient. That is, the length of the vocal tract length played the most dominant role.

Keywords: Pitch, Tonal quality, Vocal tract length, Spectral centroid, WORLD

1. はじめに

1.1 声のピッチ感の錯覚

声道長の縮小拡大を模して、スペクトル包絡の周波数軸

を伸長圧縮させた音声では、基本周波数 (f_0) の高低関係と声の高さの印象評価の間に、反転現象が観測される。

たとえば、短い声道長を模してスペクトル周波数軸を伸長した音色で、かつ f_0 が低い「細くて低い声」と、反対に長い声道長を模して周波数軸を圧縮した音色で、 f_0 が高い「太くて高い声」を比較する。すると、 f_0 は低いのものに関わらず、周波数軸を伸長した「細くて低い声」の方が「高

¹ 大学入試センター 研究開発部
Research Division, National Center for University Entrance
Examinations, 2-19-23 Komaba, Meguro-ku, Tokyo 153-
8501, Japan

^{a)} uchida@rd.dnc.ac.jp

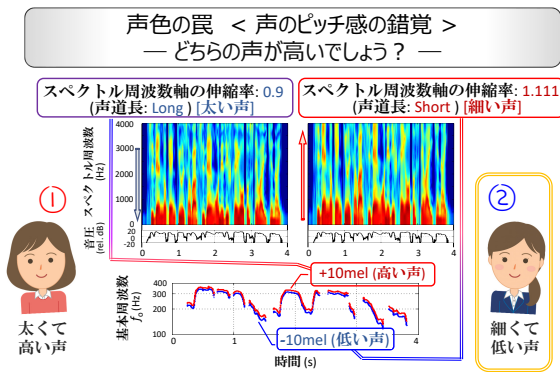


図 1 声のピッチ感の錯覚のイメージ

Fig. 1 Image of the voice pitch illusion.

い声」だと誤って知覚される (図 1)。

この現象を「声のピッチ感の錯覚」と呼ぶ^[1,2]。なお、ここで f_0 パターンが平板になると、今度は音色の影響が低下して、 f_0 の高低関係に対応した知覚判断に戻る。したがって、 f_0 の動的属性も、錯覚の生起と消失に関係する。

このスペクトル周波数軸を伸縮させる音色の操作では、言語音声の特徴づけるフォルマントの移動が伴う。もし、フォルマントの位置は固定したまま音色を操作した場合に、声の高さの知覚がどうなるかについては、検討の余地がある。

1.2 話者の寸法知覚

一方、聴覚系には、音声から直接、話者の体の大きさの寸法 ($\propto$ 身長) を推定する機構が備わっていると計算理論が提案されている。そこでは、話者の体の大きさの寸法の違いは、スペクトル周波数軸を伸縮させて声道長の変化を模擬した場合、周波数軸の伸縮率の差分にして 5% 程度で弁別できるとされる^[3]。

この寸法知覚については通常発声に留まらず、ささやき声についても検討がなされている。なお、ささやき声での寸法知覚は、被験者の個人差が大きい。そのため、ささやき声のピッチ感を仮説的に導入することで、被験者ごとの結果を説明するための計算モデルの構築も試みられている^[4]。しかし、この一連の研究では、スペクトル周波数軸の伸縮の他に、スペクトル傾斜の操作も試みているが、ピッチ感に直接的に関係する f_0 の操作を組合わせた検討は、通常発声においてもあまり行われていない。

1.3 研究の目的

声のピッチ感の錯覚の生起条件を精査するため、周波数軸は伸縮させずに、フォルマントは固定したままの条件でも、高低判断の反転が生ずるかを検討する。今回は、スペクトルの低域から高域への強調と抑圧によって、スペクトル傾斜を操作した音声でも、高低判断に影響が現れるか検

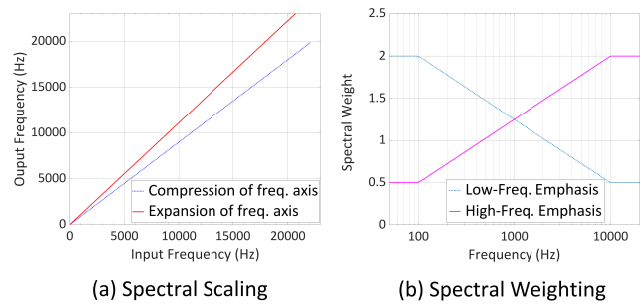


図 2 スペクトル包絡の操作

Fig. 2 Manipulation of spectral envelopes.

討する。

さらに、話者の体の大きさに関わる寸法知覚について、スペクトル周波数軸の伸縮、スペクトル傾斜の操作に加えて、 f_0 の昇降操作を組合わせた比較実験を行う。それによって、話者の寸法知覚への f_0 の影響を検討する。

2. 実験音声の生成

原音声は、ATR 音素バランス文から 8 スクリプト (S1–S8) を選定した。話者は、ナレーター事務所所属の男声・女声各 4 名 (M1–M4, W1–W4)、話者ごとに 2 スクリプトを割り当てた計 16 音声を用いた。

2.1 スペクトル包絡の操作

変換には WORLD^[5,6] を用いた。原音声からスペクトル包絡を抽出した上で、(1) スペクトル周波数軸の伸縮条件では、周波数軸を 0.9 倍に圧縮するものと、1.111(1/0.9) 倍に伸長するものを設定した。(2) スペクトル傾斜を操作するためのスペクトルへの直接重みづけ条件では、100 Hz–10 kHz の範囲を、対数軸上で線形に 1/2–2.0 倍の重みをかけて高域強調するものと、2.0–1/2 倍の重みで低域強調するものを設定した (図 2、図 3)。

2.2 f_0 パターンの操作

原音声の f_0 パターンを抽出した上で、(1) f_0 昇降条件では、 f_0 を mel 軸上で +10 mel 上昇させるものと、-10 mel 下降させるものを設定した。(2) 平板化 f_0 昇降条件では、 f_0 パターンを平板化した上で、 ± 10 mel で昇降させるものを設定した (図 4)。

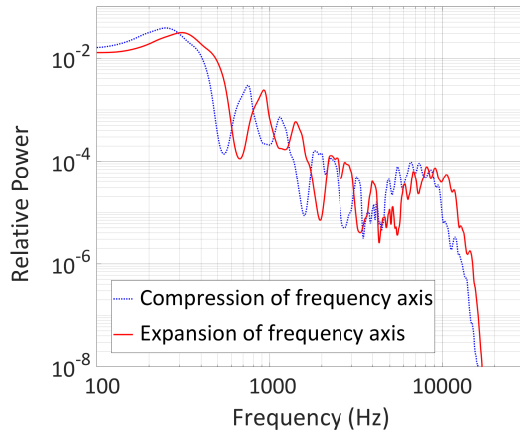
2.3 実験条件と比較音声対

2.3.1 条件 1：ピッチ感錯覚の追試条件

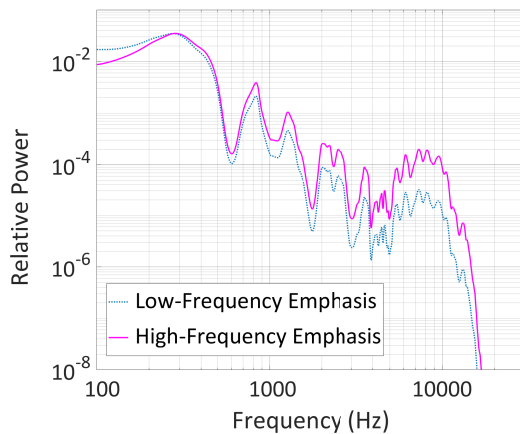
条件 1 の比較対は、スペクトルの周波数軸を圧縮、 f_0 は上昇させた音声と、周波数軸を伸長、 f_0 は下降させた音声の対である。この条件では、声のピッチ感の錯覚の生起が想定される。

2.3.2 条件 2：帯域強調の単独効果の検討

条件 2 は、スペクトルでの低域を強調した音声と、高域



(a) Spectral Scaling



(b) Spectral Weighting

図 3 時間平均スペクトル包絡 (M1S1).

Fig. 3 Time-averaged spectral envelope (M1S1).

を強調した音声の対である、 f_o は原音声のままで同一である。この条件では、声の高低判断への音色の影響の有無が評価される。

2.3.3 条件 3：帯域強調での錯覚の生起の検討

条件 3 は、低域を強調し、 f_o は上昇させた音声と、高域強調し、 f_o は下降させた音声の対である。この条件は、フォルマント固定時にピッチ感の錯覚が生起するかを検証するものである。

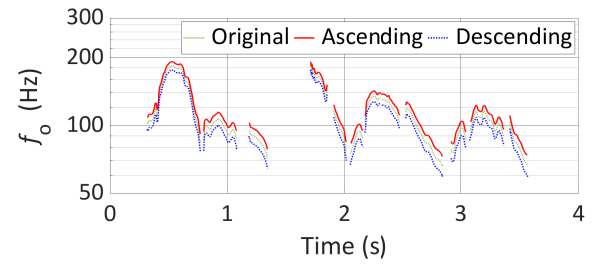
2.3.4 条件 4：帯域強調での平板化 f_o の検討

条件 4 は、低域強調して、平板化 f_o は上昇させた音声と、高域を強調、平板化 f_o は下降させた音声の対である。これは、 f_o の高低関係と知覚判断の対応が回復するかの確認、その検証に相当する。

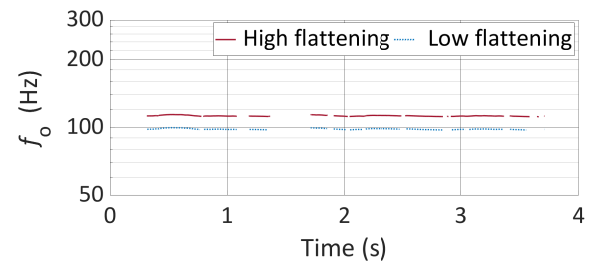
3. 対比較による聴覚実験

3.1 方法

首都圏の国公立 11 大学の 1 年生 123 名（男性：90 名・女性：32 名・未記入 1 名、18–21 歳）が参加した。参加者を 8 群に分けて、64 対の音声比較対も 8 系列に分割して配



(a) Standard f_o Contour



(b) Flattened f_o Contour

図 4 f_o パターンの操作 (M1S1).

Fig. 4 Manipulation of f_o pattern (M1S1).

表 1 主観評価のための形容詞項目

Table 1 Adjective items for subjective evaluation.

Trait	Adjectives
Voice Quality (12) —	High , Manly, Husky, Placid, Powerful, Young, Deep, Forceful, Nasal, Falsetto, Whisper, Fast-talking
Naturalness (4) —	Natural, Comprehensible, Clear, Listenable
Body Size (4) —	Large, Tall, Fat, Robust

置した。

音声の比較評価には、声の高さの評価を含む、声質表現語 [7] の項目などを用いた (表 1)。

携帯型の CD プレーヤー (Sony: D-EJ002) から、ステレオ・イヤホンで音声比較対を提示した。音声のラウドネス・レベルは -24 LKFS とした。参加者は 2 つの声について、項目ごとにどちらの声がその項目の内容にあてはまるかを 10 段階で評定した。比較対ごとの評定時間は 60 秒であった。

3.2 評価指標

評定結果は、各項目の内容が、比較対で f_o の低い方の音声にあてはまるとしたものが正数、対比音声にあてはまるとしたものが負数になるように、 $-4.5 \sim +4.5$ 点に換算した。いずれも、絶対値が大きい程、内容がよりあてはまることを示す。音声対に割り当てられた参加者による、項目別の評定値の平均を、各比較対での項目ごとの評価指標

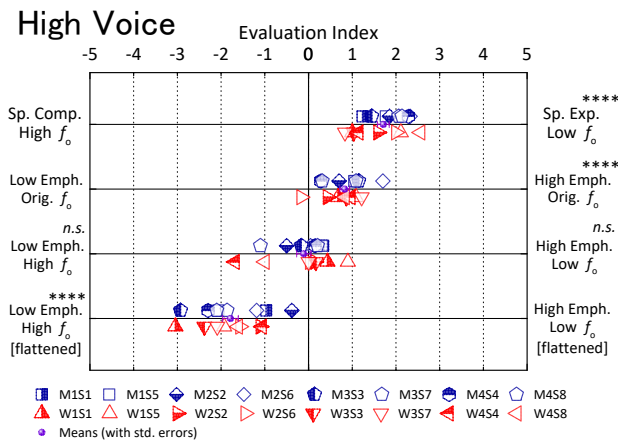


図 5 声の高さの実験条件ごとの評価指標得点
Fig. 5 High voice score for each exptl. condition.

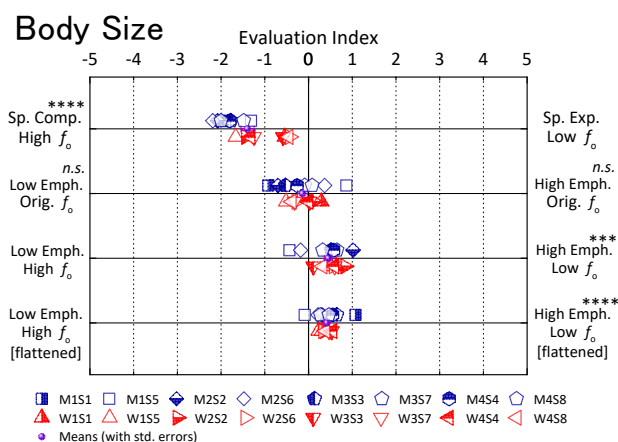


図 6 話者の体の大きさの実験条件ごとの評価指標得点
Fig. 6 Body size score for each exptl. condition.

とした。

条件ごとに、どちらの音声が高いと評価されたかの結果を図 5 に、どちらの話者の体が大きいと評価されたかの結果を図 6 に示す。指標が正值で大きい程、 f_0 の低い音声の方がより高い、と評価されたことを表す。帰無仮説 H_0 を $\mu = 0$ とおいた両側検定の t 検定の結果を、比較条件ごとの音声種の右肩に示す (*: $p < .05$, **: $p < .01$, ***: $p < .001$, ****: $p < .0001$)。

3.3 声の高さの主観評価の結果

声の高さの知覚の特徴について、図 5 の結果に基づいて整理する。まず条件 1 では、スペクトル周波数軸の伸長による細い音色で、 f_0 の低い音声の方が、高い声だと評価されていた。 f_0 の高低関係と、知覚判断が反転しており、ピッチ感の錯覚の生起が再確認された^[1,2,8]。

条件 2 では、 f_0 パターンは同一であっても、高域強調した音声の方が高いと評価された。フォルマント固定時にも、音色の違いが高低判断に影響を与えていることを示す結果と言える。

条件 3 では、高域を強調した音色で f_0 が低い音声は、低

域が強調されて f_0 が高い音声と、同等の高さの声だと評価された。この条件でも高低判断への音色の影響が見られるが、評価の反転にまでは至らなかった。

条件 4 では、低域が強調された音色でも、平板化 f_0 を上昇させた音声は、高域を強調して、 f_0 を下降させた音声よりも高いと評価された。フォルマント固定時にも、 f_0 の平板化によって、高低関係と対応した知覚判断が回復することが示された。

3.4 話者の体の大きさの主観評価の結果

話者の体の大きさとして評定した寸法知覚について、図 6 の結果に基づいて整理する。条件 1 では、スペクトル軸を圧縮した「太い声」の話者の方が、「体が大きい」と評価された。その一方で、その音声の f_0 は高かったことから、寸法知覚では声道長に関連する音色の要因の方が支配的であることが想定される。

次に条件 2 の結果を見ると、体の大きさについては帯域強調による有意な影響はみられなかった。すなわち、低域を強調しても、その話者の体が大きいとは必ずしも感じられなかったということになる。

条件 3・条件 4 では、高域が強調されていても、「 f_0 が低い」話者の方が「体が大きい」と知覚されていた。ここでは、スペクトル傾斜よりも、 f_0 の方が支配的であったと考えることができる。なお、 f_0 パターンが通常音声の動的なおものでも、平板化 f_0 の静的で動かないのもであっても同様の結果であった。その点で声の高さの評価とは様相が大きく異なっていた。

4. まとめと課題

4.1 音の高さの知覚と 2 つの音響特性

本実験では、音声のスペクトルの周波数軸を伸長して声の音色を系統的に変化させると、基本周波数が低い声であっても、高い声だと誤って知覚してしまうピッチ感の錯覚の生起を再確認した。この声色の操作では、周波数軸の伸長によってスペクトル重心も上昇する。

石田・赤木 [9, 10] は、調波複合音のスペクトル包絡の形状と基本周波数をそれぞれ独立に操作した音刺激を作成し、それらの音同士の類似性を判断させたデータを多次元尺度構成法 (MDS) で分析し、直交する 2 軸を抽出した。その結果から、音の高さの知覚には、物理的な音響特性として、直交する 2 つの要因が想定されるとした。その上で、音の高さは、基本周波数のみで知覚されているのではなく、スペクトル包絡の重心も、音の高さの知覚に影響を及ぼしている可能性があるとしている。

声のピッチ感の錯覚は、そのどちらか片方が極端に支配的になる条件が揃った場合に生起する現象だと考えることもできる。通常の音声のように、 f_0 がアクセントやイントネーションを形づくりながら動的に変化する時には、声

の音色の属性としてのスペクトル重心の要因が、ピッチ感については支配的な役割を担っていると考えられる。一方、平板化 f_0 のように、 f_0 が静的で動かない時には、声の音色は前と同じであっても、今度は f_0 の要因がピッチ感に対する支配的な役割を担うようになって考えられる。

このような解釈を検証するためには、厳格な統制が可能な“動的な調波複合音”を用いた研究がもてらる。さらにそれは、聴覚の認知的特徴としての一般化を視野に入れた検証の糸口になるものと期待される。

次に、フォルマントの位置を固定した形でのスペクトル傾斜の操作による低域強調・高域強調の音色の違いは、声の高低判断に影響を与えることが示された。これは、上記のスペクトル重心が音の高さの知覚に影響を与えるとする仮説とも矛盾しない結果である。しかし、今回のスペクトルの重みづけ条件の範囲では、その効果はスペクトル周波数軸の伸縮条件と比べると小さい。また錯覚と呼ぶことができるような、高低評価の反転にまでは至らなかった。ただし、今回使用した帯域強調の条件は、あくまでも一つの事例に過ぎない。今後、重みづける対象の範囲、重みづけ関数の形状などについて、さらに精査していく必要がある。

4.2 寸法知覚とスペクトル傾斜・ f_0 ・声道長

本実験では、 f_0 が高くて、スペクトル周波数軸を圧縮して長い声道長を模した太い声の話者は、体が大きいと評価された。一方、低域を強調した声であっても、その話者の体は、必ずしも大きいとは評価されなかった。さらに、高域を強調した声であっても、 f_0 が低いと、体が大きいと評価された。

これらを総合して考えると、話者の寸法知覚に関しては、帯域強調よりも f_0 が、その f_0 よりもスペクトル軸の伸縮が、支配的な役割を担っていると考えられる。すなわち、声道長の長短の要因が、今までのところ最も支配的であると考えられる。なおここで、 f_0 がスペクトル傾斜の効果を上回る影響を与えている点は興味深い。

さて、スペクトル傾斜による帯域強調の影響については、聞き手である被験者ごとに系統的な傾向性が存在する可能性はある。しかし、全体の平均として捉えた場合は、必ずしも明確な影響は見られなかった。ただし、今回の帯域強調の条件はあくまで一事例に過ぎないので、今後の精査が不可欠である。

参考文献

- [1] Uchida, T.: Reversal of relationship between impression of voice pitch and height of fundamental frequency: Its appearance and disappearance, *Acoustical Science & Technology*, Vol.40, No.3, pp.198–208 (2019).
- [2] 内田照久, 森勢将雅: 声のピッチ感の錯覚と疑似歌声・疑似ささやき声による検討, 情報処理学会論文誌, Vol.61,

- No.4, pp.807–816 (2020).
- [3] 山本航大, 入野俊夫, 岡本江美, 松井淑恵, 西村竜一, 河原英紀: スペクトル傾斜の異なる音声の寸法知覚と聴覚モデルによる説明, 日本音響学会 2016 年春季研究発表会講演論文集, pp.481-484.
- [4] 上村怜央, 入野俊夫, ロイ D. バターソン: ささやき声の寸法知覚におけるピッチ感を導入したモデル化日本音響学会 2020 年春季研究発表会講演論文集, pp.461-464.
- [5] Morise, M., Yokomori, F. and Ozawa, K.: WORLD: a vocoder-based high-quality speech synthesis system for real-time applications, *IEICE transactions on information and systems*, Vol. E99-D, No. 7, pp. 1877-1884, (2016).
- [6] Morise, M.: D4C, a band-aperiodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol. 84, pp. 57-65, (2016).
- [7] 木戸博, 粕谷英樹: 通常発話の声質に関連した日常表現語—聴取評価による抽出—, 日本音響学会誌, Vol. 57, No 5, pp. 337-344, (2001).
- [8] 日本基礎心理学会: 錯視・錯聴コンテスト 2016 (第 8 回), <http://www.psy.ritsumei.ac.jp/~akitaoka/sakkon/sakkon2016.html>, (参照 2019-05-16).
- [9] 石田舞, 赤木正人: 音色的高さに影響を及ぼす物理量の検討, 日本音響学会 2010 年春季研究発表会講演論文集, pp.635-638.
- [10] Ishida, M., & Akagi, M.: Pitch perception of complex sounds with varied fundamental frequency and spectral tilt, *Proceedings of 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing*, pp. 480-483, (2010).