

声色の罫 <声のピッチ感の錯覚> : 疑似歌声・疑似ささやき声による検討

内田照久^{†1}

概要：声道長の縮小や拡大を模して、スペクトル包絡の周波数軸を伸長圧縮した声質変換音声では、基本周波数(f_0)の高低関係と声の高さの印象評価が逆転することがある。このピッチ感の錯覚が起こる条件の精査をするため、標準抑揚、平坦抑揚、逆相抑揚、疑似歌声、疑似ささやき声による評価実験を行った。

キーワード：ピッチ，声質，声道長，スペクトル重心，WORLD

Investigation of voice pitch illusion using quasi singing voice and quasi whisper

TERUHISA UCHIDA^{†1}

Abstract: When the speech is converted by expansion and compression of the fundamental frequency axis of spectrum envelopes simulating reduction and expansion of vocal tract length, the timbre of the voice changes systematically. In the converted voice, a reversal of the relationship between height of fundamental frequency and impression of voice pitch is often observed. In order to examine the conditions that cause this voice pitch illusion, evaluation experiments were conducted using standard intonation, flat intonation, antiphase intonation, quasi singing-voice, and quasi whisper.

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

1. はじめに

1.1 声の音色と声道長: スペクトル周波数軸の伸縮

発声器官を音響管と捉えた場合、声道の長さが変わると、声の音色は系統的に変化する。一般的に、声道長が短いと「細い声」となり、長いと「太い声」になる。この声道長の違いに起因する声色の変化は、音声信号処理の観点では、音声の平滑化スペクトルの周波数軸を伸長、圧縮することで模することができる[1]。このスペクトル周波数軸の伸縮によって、声道長(vocal tract length)の操作を模擬した声質の変換は、話者の秘匿などにも利用される[2]。

1.2 スペクトル重心とピッチ感への認知的バイアス

スペクトル周波数軸を伸長して、声道長の縮小を模した音声では、元の音声よりもスペクトル重心(spectral centroid)が上昇する。この高スペクトル重心となった音声は、たとえば基本周波数(fundamental frequency, f_0)が元の音声と同一であっても、「高い声」として知覚される。逆に、周波数軸を圧縮して、声道長の拡大を模した低スペクトル重心の音声では、「低い声」の知覚印象が生起する[2]。

このスペクトル周波数軸の伸縮に伴うスペクトル重心の移動は、単に声の音色を系統的に変えるだけに留まらず、声の高さの印象、声のピッチ感にも影響を及ぼす。そして、このピッチ感への認知的なバイアスは、基本周波数の高低の効果に重畳する形で、加算的に作用する[3]。

1.3 声のピッチ感の錯覚

さらに、その認知的バイアスは、基本周波数の高低関係と、声の高さの印象評価を逆転させる程、強い効果があることがわかってきた。実際に一定の条件の下では、「基本周波数が低い」音声であっても、「高い声」として誤って知覚されてしまう「声のピッチ感の錯覚(voice pitch illusion)」が報告されている[3, 4]。

具体的には、短い声道長を模した高スペクトル重心の音色で、基本周波数が低い「細くて低い声」と、長い声道長を模した低スペクトル重心だが、基本周波数は高い「太くて高い声」の2つの音声の対を聞いて比較をする。すると、基本周波数は低いにも関わらず、高スペクトル重心である「細くて低い声」の方が、「高い声」と評価される。単に、音色が異なるだけで、基本周波数の低い方が高い声として知覚されるという、直観的には理解しがたい現象であるが、かなり頑健に安定的に生起する[3]。

1.4 f_0 軌跡パターンの形状と認知的バイアスの変動

ところが、この不可思議な現象を引き起こす原因であるピッチ感への認知的バイアスの強度は、 f_0 軌跡パターンの形状によって、大きく変動することがわかってきた[3]。

まず、標準的な普通の発話のイントネーションの時は、認知的なバイアスが非常に強く効き、基本周波数の高低と声の高さの印象が逆転する。すなわち、ピッチ感の錯覚が生起する。

しかし、ロボット声のようなフラットな抑揚の時には、認知的バイアスの効果は非常に弱くなり、声の高さの印象

^{†1} 独立行政法人 大学入試センター
The National Center for University Entrance Examinations
e-mail: uchida@rd.dnc.ac.jp

は、基本周波数の高低関係に従うようになる。その場合、錯覚は消失することになる。

これらの知見からは、スペクトル重心の違いから生じる認知的バイアスは、常に安定している訳ではなく、 f_0 軌跡パターンの動的特性に応じて変動していることがわかる。中でも f_0 の変動幅の大きさは、認知的なバイアスの強度の変移を引き起こしている。すなわち、変動幅の拡大は認知的なバイアスの効果を促進し、縮小はそのバイアスを抑制する。

このふるまいについて、ピッチ感の評価に影響を与える音響属性の支配性の観点から捉え直す。 f_0 軌跡パターンがフラットな時は、「基本周波数の高低」がピッチ感への支配的な役割を担っている。しかし、 f_0 の変動幅が大きくなるにつれて、今度は「スペクトル重心の高低」が支配的な役割を担うように変わる。これは f_0 変動幅の大きさに応じて、ピッチ感を支配する音響属性が、緩やかに交替していると捉えることができる。

1.5 疑似歌声と疑似ささやき声による検討

しかし、この認知的バイアスやピッチ感の錯覚に関する詳細な生起条件や消失条件などの全体像は、まだほとんどわかっていない。 f_0 変動幅の要因についても、フラットな f_0 軌跡区間がどのくらいの時間長になったら、「基本周波数の高低」の支配性が優位になるのかといったことも、まだ明らかでない。

音楽としての歌声 一方で、歌声では音楽を成立させる上で、基本周波数が音高として支配的な役割を担っているのは明らかである。 f_0 軌跡パターンは、音楽的旋律として遷移する。声色の異なる別の歌い手が発する歌声であっても、基本周波数が同じであれば、楽音としての音の高さは等価であるとみなされる。すると、歌声では、「基本周波数の高低」がピッチ感の評価でも支配的な役割を担っていると想定される。

スペクトル包絡単独の効果量 また、 f_0 軌跡パターンの形状だけに原因をもとめるのではなく、平滑化スペクトルそのものの影響についても検討する必要がある。ピッチ感への認知的バイアスは、スペクトル周波数軸を伸縮させた結果として生じている。すると、純粋にスペクトル包絡の違いによって生ずる単独の効果量を測定する必要がある。そして、その測定のためには、 f_0 や調波構造を取り除いた音声を用いた検証が必要となる。

疑似的音声の利用 そこで本研究では、従来からの f_0 変動幅の要因に加えて、疑似歌声と疑似ささやき声を用いて、声のピッチ感に係る認知的バイアスについて検討する。

歌声については、 f_0 軌跡を操作して疑似的に模擬する。比較的短時間のフラットな f_0 が、音楽的音階のステップで遷移する疑似歌声(quasi singing-voice)を生成して用いる。

また、従来の研究では、スペクトル包絡の違いによる単独の効果量を検討するために、雑音駆動音声(noise-vocoded

voice)を用いてきた[3, 5]。しかし、雑音駆動音声は聴感上、どうしても不自然な印象が生ずる。そこで、基幹部分では雑音駆動方式を用いることで、 f_0 や調波構造を取り除くことができるメリットを生かしつつ、より自然な声に近い、疑似ささやき声(quasi whisper)を生成して用いる[6]。

1.6 聴覚実験の条件設定

本実験では、まず従来の研究[3]での音声(16kHz sampling)と比べて、より高品質な広帯域の音声(44.1kHz sampling)を用いて f_0 変動幅要因の追試を行う。それにより、ピッチ感の錯覚の生起、認知的バイアスの発現についての検証確認を行う。

ここではまず、(1) 標準的な抑揚、(2) フラットな抑揚、(3) その間となる中庸の f_0 変動幅であるが f_0 軌跡の高低が逆相となる抑揚、の条件を設定する。

また、歌声での、ピッチ感の評価への f_0 の高低の支配性の検証のために、(4) 疑似歌声の条件を設定する。さらに、ピッチ感へのスペクトル包絡単独の効果量の測定のために、(5) 疑似ささやき声を用いる実験条件を設ける。

そして、各条件の音声ごとに、スペクトルの周波数軸を伸縮する。そして、疑似ささやき声条件以外の条件では、それぞれの f_0 軌跡形状を維持したまま f_0 全体を昇降する。その上で、「低スペクトル重心・高 f_0 軌跡」と「高スペクトル重心・低 f_0 軌跡」となる音声の比較対を組む。そして、声の高さの印象評価を含む音声の印象評価による、対比較による聴覚実験を実施する。

2. 変換音声の対比較による評価実験

2.1 原音声

音声データベース SRV-DB[7]の ATR25 文の中から、男女各 2 名の計 4 名が発話した、異なる 8 種類のモノローグの音声を選んで、原音声として使用した(Table 1)。

ここで、原音声(F1-2)の音声波形とスペクトログラムを Figure 1 に示す。スペクトログラムをみると、基音とその倍音からなる縞模様の調波構造がみられる。この縞の一番下にある曲線が、声帯音源の基本周波数の f_0 軌跡パターンに相当する。

さらに、このスペクトログラムを垂直方向に見てみると、縞模様の中には濃淡があり、特定の周波数帯域で倍音成分が強かったり、弱かったりする様子が読み取れる。音声を発する時、口や顎、舌や唇などの動きによって、声道の中には、いくつか異なる大きさの空間が作られる。それぞれの空間は固有の共鳴周波数を持つため、それに近い倍音は強調されることになる。それが、縞模様の濃淡を作り出している。なお、倍音が強調される周波数を、低いものから順に、第 1、第 2、第 3 フォルマント(formant)と呼ぶ。

なお、この縞模様の細かい構造から、声道の共鳴特性を抽出して、より滑らかに表現したものは平滑化スペクトルと呼ばれる。

Table 1 Attributions of original speeches and assigned quasi singing-voice condition.

Speech wave data (ID)	Quasi Singing Voice Condition
F1-1 [†] : ATR_PF00_080_005	Major: Ascending
F1-2 [‡] : ATR_PF00_080_010	minor: Descending
F2-3 [†] : ATR_PF01_080_003	Major: Descending
F2-4 [‡] : ATR_PF01_080_008	minor: Ascending
M1-5 [†] : ATR_AM01_0800_000	Major: Ascending
M1-6 [‡] : ATR_AM01_0800_004	minor: Descending
M2-7 [†] : ATR_AM03_0800_011	Major: Descending
M2-8 [‡] : ATR_AM03_0800_018	minor: Ascending

Sampling frequency: 44.1 kHz, Resolution: 16 bit linear

[†][High Sp. - Low f_0] → [Low Sp. - High f_0]

[‡][Low Sp. - High f_0] → [High Sp. - Low f_0]

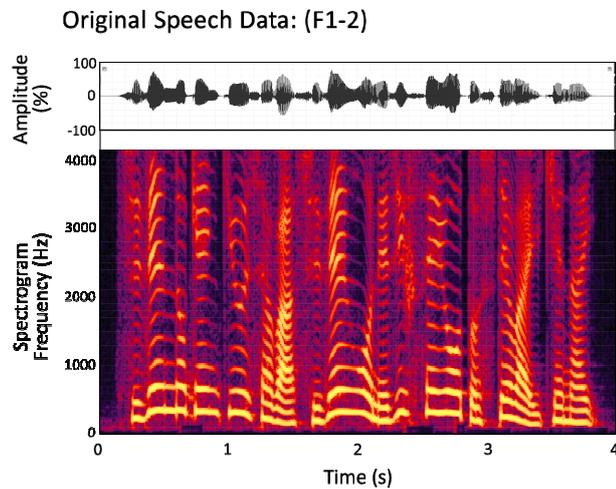


Figure 1 Example waveform and spectrogram of original speech data (F1-2: 自然の研究者は、自然をねじ伏せようとしてはいけない).

2.2 平滑化スペクトルと f_0 軌跡パターンの変換

声道長の操作を模したスペクトル周波数軸の伸縮、及び f_0 の操作には、WORLD v0.2.0_7[8, 9]を用いた。

WORLD は、vocoder ベースの分析合成方式として、音声を極めて高い品質で様々に変換することができる。まず、原音声を分析することで、音声を形作る音響的な要素に分解する。はじめに、声帯振動に相当する f_0 を抽出し、その後、声道の共鳴特性に相当する平滑化スペクトルをもとめる。さらに、スペクトルの帯域ごとに、調波構造によらない非周期性の成分が、どのくらい含まれているかを表す非周期性指標を算出する。

これらの要素は、それぞれ独立に操作することができる。そして、操作した要素を組合せて音声を再構成することで、品質を維持したまま、任意の変換音声を合成することができる。そこで、本実験では、後述する各種の変換処理に、この WORLD を使用した。

(1) スペクトル周波数の伸縮操作

声の音色の操作のために、原音声の平滑化スペクトルの周波数軸を 0.9 倍、1.111(1/0.9)倍の 2 段階で伸縮を行った。これは、声道長の拡大と縮小に相当する。なお、音声変換時の品質維持のために、スペクトルの対数パワーを操作し、帯域の上限は 16kHz に制限した(Figure 2)。

(2) f_0 軌跡の昇降操作と形状変換

f_0 軌跡パターン全ての操作は、音高の心理的な知覚と対応する mel 尺度上で行った。具体的には、周波数 f (Hz) を(1)式によってメル値 m (mel)に変換した上で、 f_0 の操作を行った[10]。

$$m = 2595 \log_{10} (1 + f/700). \quad (1)$$

f_0 軌跡の昇降 比較対を作るための基本周波数の昇降のため f_0 軌跡パターンの高低位置を ± 10 mel で昇降させた。これは、後述する f_0 軌跡パターンの形状に係わる実験条件 1-4 で生成する全ての音声に適用した。

実験条件 1-3: f_0 変動幅と形状の操作 f_0 軌跡の mel 平均を中心軸とし、標準偏差(SD)を f_0 の変動幅として操作した。そして、(1) オリジナル 1.0 倍 (2) ほぼフラットな 0.02 倍、(3) f_0 を中心軸で高低反転した逆相 f_0 軌跡($r.p.$)での 0.4 倍、を設定した。

実験条件 4: 疑似歌声の生成 f_0 軌跡を音楽的な音階に変換した。長調(Major)と短調(minor)、それぞれの上行(Ascending: IL-II-III-V-IH)、下行(Descending: IH-V-III-II-IL)の各条件を、原音声ごとに割り当てた(Table 1)。音高を安定させた区間の f_0 の変動幅は、実験条件 2 と同じ標準偏差(SD)の 0.02 倍とした。また、比較対での f_0 軌跡の昇降の位置は、原音声の mel 平均を ± 10 mel した高さが、各音階での IV[#] となるように変換した。

この実験条件 1-4 で、 f_0 軌跡をメル値 m (mel)で計算してきた変換値を、(2)式で逆変換して周波数 f (Hz)に戻す。

$$f = 700 (10^{(m/2595)} - 1). \quad (2)$$

そして、得られた物理量である周波数 f (Hz)に基づいて、WORLD で変換音声を合成する(Figure 3)。

なお、実験条件 1-4 の最終的な変換音声は、「低スペクトル重心・高 f_0 軌跡」と「高スペクトル重心・低 f_0 軌跡」となる比較対を作るための組合せ条件で合成した(Figure 4)。

(3) 雑音駆動による疑似ささやき声の生成 (実験条件 5)

WORLD での音声合成時に、普通なら f_0 軌跡を描く声帯音源を模した音源信号を、白色雑音で差し換えて合成すると、調波構造が存在しない雑音駆動音声が得られる。それを元に、自然なささやき声を模して低域を取り除いた音声(truncated noise-vocoded voice)を生成して、疑似ささやき声として用いた[6]。

原音声のスペクトル周波数軸を圧縮、伸長して得られたスペクトル包絡の低域を減衰させて、雑音駆動で合成した疑似ささやき声で、比較対を作成した(Figure 5)。

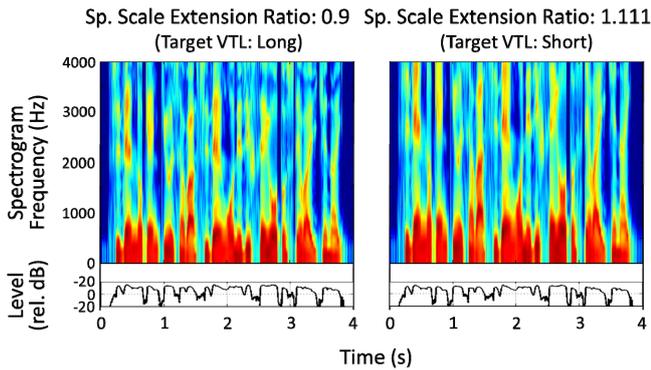


Figure 2 Example smoothed-spectra and level of converted speech (F1-2).

Exp. Condition 5: Quasi Whisper [Truncated Noise-Vocoded Voice]

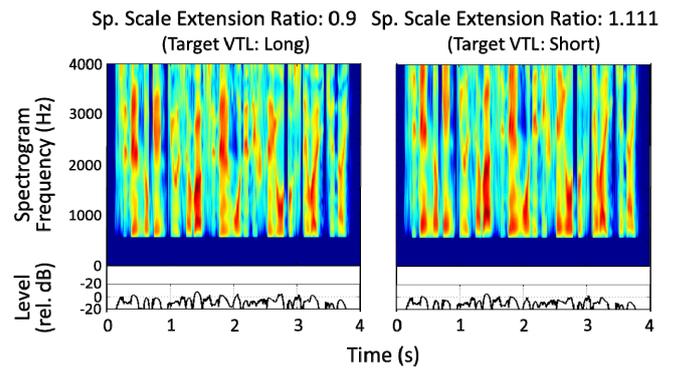


Figure 5 Example truncated-spectra and level of converted quasi whisper.

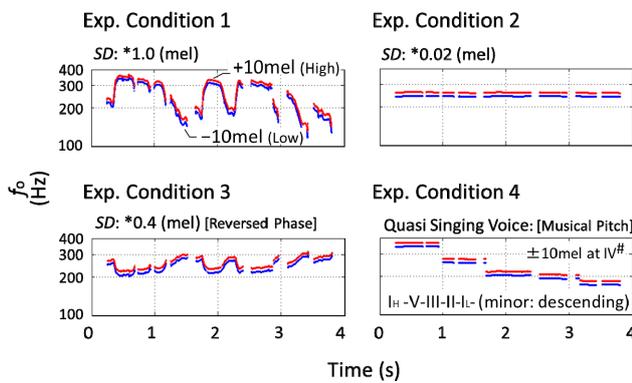


Figure 3 Example f_0 contours of converted speech (F1-2).

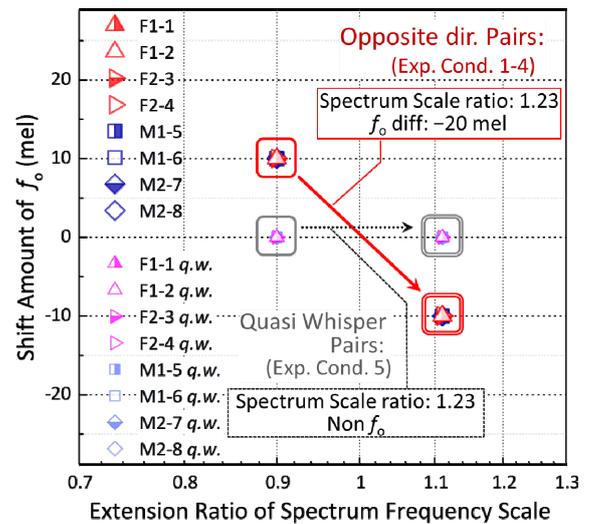


Figure 6 Example assignments of comparison pairs.

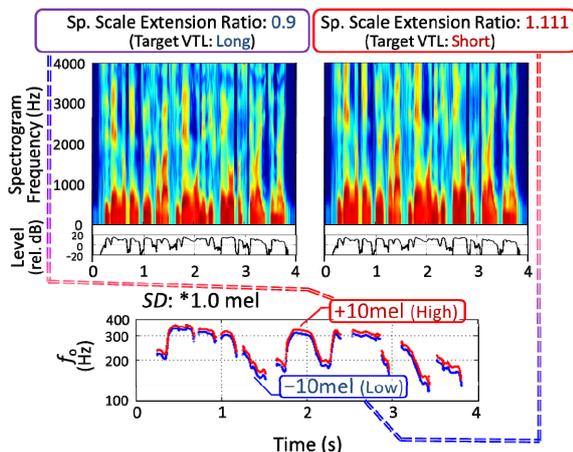


Figure 4 Combination of smoothed-spectra and f_0 contours.

2.3 実験音声の比較対の配置

実験条件 1-4 の比較対は「低スペクトル重心・高 f_0 軌跡」と「高スペクトル重心・低 f_0 軌跡」となる音声の組合せである。これは、スペクトル重心の高低によるピッチ感への影響の方向と、基本周波数の高低による影響の方向とが、互いに逆向きの音声の組合せである。ピッチ感の錯覚は、この組合せの条件の時にみられる。この比較対の相互関係を、スペクトル軸の伸縮率と f_0 昇降のシフト量に基づいて

Figure 6 に示した。

実験条件 5 の疑似ささやき声では f_0 が存在しないので、「低スペクトル重心」と「高スペクトル重心」の組合せとなる(Figure 6)。したがって、この音声比較対でのピッチ感へのバイアス量は、スペクトル重心の違いによって生ずる効果量であるとみなすことができる。

実験全体では、80 の変換音声からなる、40 対を用いた。提示順序は、同条件での同性話者による対間で反転させた。

2.4 実験の手続き

参加者 東京都内の 6 つの国立大学の 1 年生 140 名(男性: 106 名・女性: 34 名, 18~23 歳)が参加した。参加者を 10 群に分割して、音声比較対も 10 系列に分けて配置した。

評価項目 音声の比較評価には、声の高さの印象評価を含む、声質表現語[11]の項目など[2, 3]を用いた(Table 2)。

手続き 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。参加者には、2 人の声の印象について、項目ごとに、どちらの声がその項目にあてはまるかを、10 段階で評定するように指示した。対ごとの評定の制限時間は 60 秒であった。

Table 2 Adjectives to express subjective impression

Trait	Adjectives
Voice Quality: 声質表現語	高い声, 男性的な声, かすれた声, 落ち着いた声, 迫力のある声, 若い感じの声, 太い声, 張りのある声, 鼻声, ささやき声
Naturalness: 自然性	自然な, わかりやすい はっきりした, 聞き取りやすい
Speaker's Body Size: 話者の体格	体の大きい, 背の高い 太った, がっしりした

3. 比較評定実験の結果と考察

3.1 評価観点ごとの評価指標と尺度得点

スペクトル重心の影響の方向を固定して捉えるために、比較対の中で「高スペクトル重心・低 f_0 軌跡」の声を評価音声として、もう一方を対比音声として扱った。

評定結果については、各項目の内容が、評価音声にあてはまるものが 0.5 ~ 4.5 点、対比音声にあてはまるものが -0.5 ~ -4.5 点になるよう換算した。いずれも、その絶対値が大きくなる程、それぞれがよりあてはまることを示す。そして、各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、各比較対の項目ごとの評価指標とした。なお、自然性と話者の体格については、各 4 項目からなる合計尺度得点から算出した。

3.2 f_0 軌跡パターンの形状の違いによるピッチ感の変動

スペクトル重心の高低と、 f_0 軌跡の高低によるピッチ感への影響の方向が相反する条件の下で、 f_0 の変動幅や形状、時間的安定性や音楽的音階での遷移が、声の高さの評価にどのような影響を与えたか分析した。

また、疑似ささやき声条件での結果から、 f_0 や調波構造が存在しないスペクトル包絡単独での効果量と、 f_0 や調波構造が存在する普通の音声での結果を比較した。

声の高さの評価指標の平均について、話者の性別の要因 (2 水準: 男声, 女声) と、 f_0 の軌跡パターンの要因 (5 水準: 変動幅 (SD): 1.0, 0.02, $r.p.$ 0.4, 疑似歌声, 疑似ささやき声) の 2 要因分散分析を行った。

まず、性別要因の主効果が有意だった ($F_{(1,30)}=4.56, p<.05$)。全般に声の高さの評価は、女声>男声で高かった。つまり、スペクトル重心の高低による影響、ピッチ感への認知的なバイアスは、女声でより強く見られた。

次に、 f_0 軌跡パターンの主効果が高度に有意であった ($F_{(4,30)}=11.54, p>.0001$)。 f_0 軌跡全体の高低差は同じでも、 f_0 の軌跡パターンが異なると、ピッチ感に差異が見られた。なお、性別と f_0 軌跡パターンの交互作用に有意差は見られなかった ($F_{(4,30)}=0.23, n.s.$)。そこで、これ以降は、 f_0 軌跡パターンの条件間での多重比較の結果に基づいて検討する。については、声の高さの評価指標の、実験条件ごとの結果を Figure 7 に示す。

High Voice: 高い声: (pitch score)

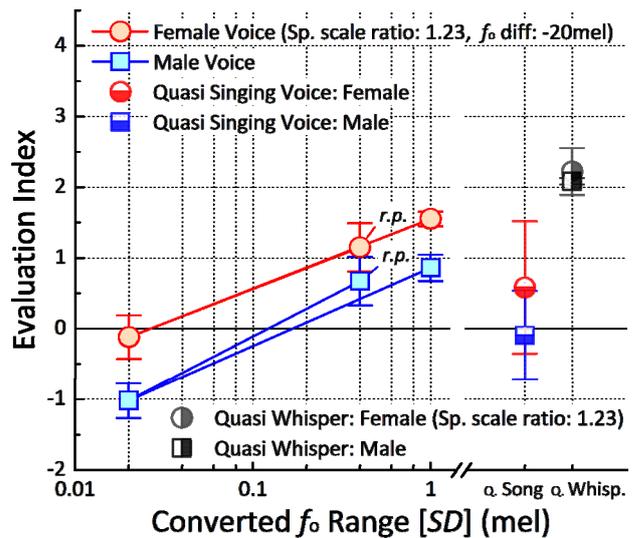


Figure 7 Means and standard errors of pitch scores for converted f_0 range, quasi singing-voice and whisper.

(1) 実験条件 1-3: f_0 変動幅とピッチ感の錯覚

原音声を同じオリジナルの抑揚の変動幅 (SD: 1.0 倍) と、それよりやや小さい逆相 f_0 軌跡 ($r.p.$) での変動幅 (SD: 0.4 倍) では、実際には f_0 軌跡が低い評価音声の方が、「高い」と評価されており、ピッチ感の錯覚が生じていた。

しかし、 f_0 軌跡がフラット (SD: 0.02 倍) になると逆転し、 f_0 軌跡の高低に従う評価に変わって、錯覚が消失した。

なお、ピッチ感の評定指標は、変動幅の対数値に対して線形に上昇していた。そこでは f_0 軌跡パターンの高低反転による輪郭形状 (contour) の違いは、必ずしもピッチ感には影響を与えていないと考えられる。すると f_0 の変動幅が、ピッチ感の認知的なバイアス量の変動を引き起こす主たる原因である、と考えるのが合理的であろう。

これらの結果は従来知見とも合致する [3]。しかるに、広帯域の高品質音声においても、ピッチ感の錯覚の生起と消失が確認されたと言ってよい。

(2) 実験条件 4: 疑似歌声でのピッチ感の評価の不安定性

疑似歌声では、1 オクターブに及ぶ f_0 の変動範囲がある。それに関わらず、疑似歌声で f_0 軌跡が低い評価音声は、オリジナルの抑揚の変動幅 (SD: 1.0 倍) の時よりも、低い声として評価されているように見える。すなわち、 f_0 軌跡がフラットな音声の場合と同じように、 f_0 の高低に呼応する方向に向かって、評価が変移しているように見える。

しかし、この疑似歌声条件では、 f_0 変動幅の実験条件 1-3 に対して有意な差は見られなかった。標準誤差を見ると、他の条件よりも大きく、条件内でのバラツキがかなり大きかったことが伺える。この結果からは、「歌声では、 f_0 の高低が、ピッチ感への支配的な役割を担っているはずである」、とする当初の想定は、必ずしも完全には支持されなかった。

ここで、ピッチ感の評定指標のバラツキが大きくなった原因を検討してみる。まず、今回の疑似歌声の生成法では、変換音声は真に音楽的な歌声に至っていない可能性がある。そのため、参加者ごとに、「言葉としての音声」として聞く者と、「音楽としての歌声」として聞く者が、分かれてしまっている可能性がある。

また、音楽の歌声として聞いた者であっても、その者が相対音感保有者か、絶対音感保有者かによって、絶対音高での高低判断の精度が異なっている可能性もある。

したがって、歌声を念頭においたアプローチに関しては、上記の点もふまえた上で、さらなる検証が必要である。

(3) 実験条件 5: 疑似ささやき声の高低判断

疑似ささやき条件では、スペクトル重心の高い評価音声でのピッチ感の評定指標が、他のどの条件よりも高かった。疑似ささやき声には f_0 が存在しないので、スペクトル包絡の違いによる認知的バイアスの効果量が、直接ピッチ感の評価に反映していると考えられる。また、この結果からは、たとえ f_0 や調波構造が存在しなくても、音声の高低判断は安定してなされていることがわかる。

この結果は、従来の研究で雑音駆動音声を用いた実験での結果と、ほぼ同じ傾向を示す結果であった[3]。この疑似ささやき声は、雑音駆動音声の低域除去によって、本来は備えていた情報の一部が欠落したスペクトル包絡から生成されたものである。しかし、自然なささやき声を模擬する範囲内での情報の欠損であれば、雑音駆動音声と同程度のかかり強い認知的バイアスが生じることが見出された。

3.3 f_0 軌跡パターンのピッチ感への影響の選択的特異性

スペクトル周波数軸の伸縮による音色の系統的な変化は、声の太さや、話者の体格の印象評価に影響を与える[1-3]。そこで、ピッチ感に強い影響を与えていた f_0 軌跡パターンの変動幅や形状が、声の太さや話者の体格にはどんな影響を与えているか検討した。

ピッチ感と同様の 2 要因の分散分析の結果、スペクトル重心が低いと、声が太く、体格が大きいと評価されていた。しかし、 f_0 軌跡パターンの違いによる有意な差は見られなかった。また、話者の性別との交互作用も見られなかった。したがって、 f_0 軌跡パターンの違いは、ピッチ感に対してのみ、特異的に影響を及ぼしているとみられる。

4. まとめと課題

この実験から、広帯域の高音質な音声でも、声の音色が異なると、 f_0 軌跡が低い音声でも高いと知覚される錯覚が生起することが確認された。しかし、音色が高さの知覚に影響を与えること自体は、無限音階、シェパード・トーンの例からもわかるように、特段に新しい発見ではない[12]。むしろ、抑揚がフラットになるだけで、声の高さの印象が f_0 の高低関係に従うようになり、錯覚が消失してしまうという現象の方が、より不可思議であるとも言える。

今回、声のピッチ感の錯覚の生起と消失の頑健性を示すことができた。しかし、この錯覚のより詳細な生起条件、消失条件はまだわかっていない。さらに、聴覚処理のどの段階で起こっているのか、そもそも、この現象はどうして起こるのか、また、この現象は聴覚や音声認知の上でどのような意味を持つのか、といったより本質的な部分は全くわかっていない。

検討の一つの方向として、声を“動的特性を持つ複合音”として捉え直してみるのも一考の価値があるかも知れない。「ピッチ感の錯覚の生起と消失」は、音声の場合に限ってだけ起こる現象なのか、それとも、より一般の複合音全般でみられる現象なのか、検討する余地があると思われる。

今回の実験の疑似歌声では、ピッチ感の安定した評価が得られなかった。今後、歌声や、楽器の楽音などを用いた検証と共に、より厳格な統制が可能な“動的な複合音”を用いた研究が進めば、聴覚の認知的特徴としての一般化を視野に入れた、検証の糸口になるのではないかと考える。

付記 本研究は、JSPS 科学研究費補助金(基盤研究(C) JP15K04103)の援助を受けました。

参考文献

- [1] D. R. R. Smith and R. D. Patterson.: The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age, *Journal of Acoustical Society of America*, Vol. 118, No. 5, pp. 3177–3186, (2005).
- [2] 内田照久: 話者の匿名性の確保を目的とした声道長の制御を模した声質変換の評価, *日本音響学会誌*, Vol. 73, No. 3, pp. 151-162, (2017).
- [3] 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).
- [4] 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).
- [5] Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M.: Speech recognition with primarily temporal cues, *Science*, Vol. 270, No. 4234, pp. 303–304, (1995).
- [6] 内田照久: 雑音駆動音声による疑似ささやき声の簡易生成と評価, *日本音響学会 2017 年春季研究発表会講演論文集*, pp. 353-354, (2017).
- [7] 高橋弘太: 高橋弘太研究室 音声データベース, <http://www.it.cei.uec.ac.jp/SRV-DB/> (参照 2019-05-16).
- [8] 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).
- [9] 森勢将雅: 音声分析合成, *日本音響学会(編) 音響テクノロジーシリーズ 22*, コロナ社, (2018).
- [10] O’Shaughnessy, D.: *Speech Communication: Human and Machine, 2nd edition*, Wiley-IEEE Press, New York (1999), pp. 109–139.
- [11] 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).
- [12] Shepard, R.N.: Circularity in judgements of relative pitch, *Journal of Acoustical Society of America*, Vol. 36, No. 12, pp. 2346–2353, (1964).