Vol.2017-SLP-119 No.22
2017/12/22

ホルマント平均に基づく声道長比推定法の検討

坂田 聡[1,a)]　上田 裕市[1]　渡邉 亮[2]

概要：先に，一様断面積音響管の長さとその共鳴周波数の関係性を基礎として，ホルマント軌跡から 2 話者間の声道長比を推定する方法を提案し，信頼性の高い結果を得た．本稿では，DP マッチング前後のホルマント軌跡の特徴から，従来法で用いたホルマントデータの代わりに，単語毎のホルマント平均や長文音声の累積ホルマント平均値を用いる簡易法を提案し，従来法である直接法と同程度の推定精度が得られることを比較・検討によって示す．

Estimation Method for Relative Vocal Tract Lengths using Means of Formant Frequencies

Sakata Tadashi[1,a)]　Yuichi Ueda[1]　Akira Watanabe[2]

1. はじめに

成人男性や女性，児童といった多様な話者を対象とする音声認識において，学習データの量的不足を認識パラメータに対する話者正規化で補える可能性がある．一般に，話者正規化は話者の声道長の差異に対する影響を除去することによって行われることから，話者間の相対的な声道長を推定する手法の確立は重要と思われる．

先に，われわれは，適正な次数による精度の高いホルマント推定法（IFC 法）を開発した．[1] さらに，2 人の話者が同一単語を発話するときのホルマント軌跡から声道長比を推定する方法を提案した．[2] この方法では，対象の話者が同一の単語群（数十語程度）を発話するデータを必要とし，かつ，単語内で音素毎の時間対応フレームを指定するためにホルマント軌跡の DP マッチングなどの処理が必要となる．

本研究では，従来法に近い精度の推定値がホルマント軌跡の累積平均値（収束値）から得られることを示す．それによって，発話単語群に対する制約の緩和と手法の単純化を図れることを示す．

[1] 熊本大学大学院先端科学研究部環境科学部門
Faculty of Advanced Science and Technology, Division of Environmental Science, Kumamoto University
[2] 熊本大学名誉教授
Professor Emeritus, Kumamoto Uniersity
[a)] tadashi@cs.kumamoto-u.ac.jp

2. ホルマント周波数による声道長比推定の原理

中性母音を発話するときの声道モデルを直管で表現すると，その共振周波数は声道長に逆比例する．[3]

$$F_n = \frac{(2n-1)c}{4L} \tag{1}$$

F_n は第 n ホルマント周波数 [Hz]，c は音速 [m/s]，L は声道長 [m] とする．この関係を 2 人の話者 A，B の声道（長さを L_A と L_B とする）とホルマント周波数（F_{nA}，F_{nB}）に適用すると，2 人の声道が長さが異なる相似形であるならば以下の式が成り立つ．

$$\frac{L_A}{L_B} = \frac{F_{nB}}{F_{nA}} \tag{2}$$

ここで，同一母音を発話する際の声道形状は発話者の努力によりほぼ相似形になっていると仮定され，その音響的特徴は F_1 と F_2 に現れることから，式 (2) に F_1 と F_2 を用いると 2 人の話者 A，B の間には，

$$\frac{F_{1B}}{F_{1A}} = \frac{F_{2B}}{F_{2A}} = \frac{L_A}{L_B} = \mu \text{ (Constant)} \tag{3}$$

の関係が成り立つ．$n > 2$ の高次ホルマントは，音素との対応が不明確になるので，必ずしも相似を保証するものではない．

ここで，2 話者 A, B を直交する x 軸，y 軸に割り当て，座標として F_1，F_2 を表示したのち，原点を通る直線に F_1，F_2 から下した垂線の距離の二乗和を最小にする直線の勾配を求める．このようにして求めた勾配は，式 (3) より，A の B に対するホルマント比（B の A に対する声道長比）μ を示すことになる．

3. 従来の声道長比推定法の概要

前節の原理に基づき，高精度の声道長比 μ を推定するのに，これまで，次のような声道長比推定法を提案した．[2]

3.1 直接法（Direct method）

2 話者が同一の単語音声を発話するとき，声道形状がほぼ相似となる対応点がホルマント軌跡の DP マッチングで得られる．それらの対応点で与えられる，すべて F_1 と F_2 を話者のそれぞれを直交軸とする座標に表示する．この作業を多数単語で繰り返し，求められたホルマントを全て表示した相関図から，原点を通るという条件を付けた直線を主成分分析（条件付き主成分分析の第一主成分の直線）で求める．この直線の傾きは，

$$\mu = \frac{(\beta - \gamma) + \sqrt{(\beta - \gamma)^2 + 4\alpha^2}}{2\alpha} \qquad (4)$$

となる．[2] ここで，横軸話者のホルマントデータを x_i，縦軸のそれを y_i（$i = 1, 2, \ldots, N$:N は個数）とすると，$\alpha = \sum x_i y_i$，$\beta = \sum y_i^2$，$\gamma = \sum x_i^2$ である．声道長比は，式 (3) からホルマント比である μ の逆数として計算される．

図 1 に DP マッチング前後の単語音声のホルマント軌跡を示す．話者によってリズムや部分的な発話時間が異なるホルマント軌跡が，DP マッチングを適用することで非線形に伸縮し，話者間で音素系列の時間的な対応が取られていることがわかる．**図 2** の赤丸と青丸で示した分布は，単語音声群（ATR 音素バランス単語 216 語）を用いて得られた男性 M01 と女性 F03 のホルマント相関関係を示している，直観的には原点を通る直線が分布の関係性に当てはまる．また，全てのホルマントデータから求めた条件付き主成分分析の第一主成分の直線を同図 2 に示しているが，ホルマントデータはこの直線近傍に分布していることから，この直線の傾きを 2 話者間のホルマント比とすることが妥当性であると判断できる．なお，声道長比はホルマント比 μ の逆数として求められる．

3.2 厳密法（Strict method）

厳密法は，話速の変化が原因となる調音結合の変化がホルマントに与える影響を考慮して，2 人の話者に 3 話速の同一単語を発話させ，より厳密な形状相似の瞬間のみのホルマントデータから，直接法と同様な手法で声道長比を推定する方式である．すなわち，まず，異なる 2 話速のホルマント軌跡に DP マッチングを施し，話速の組み合わせの

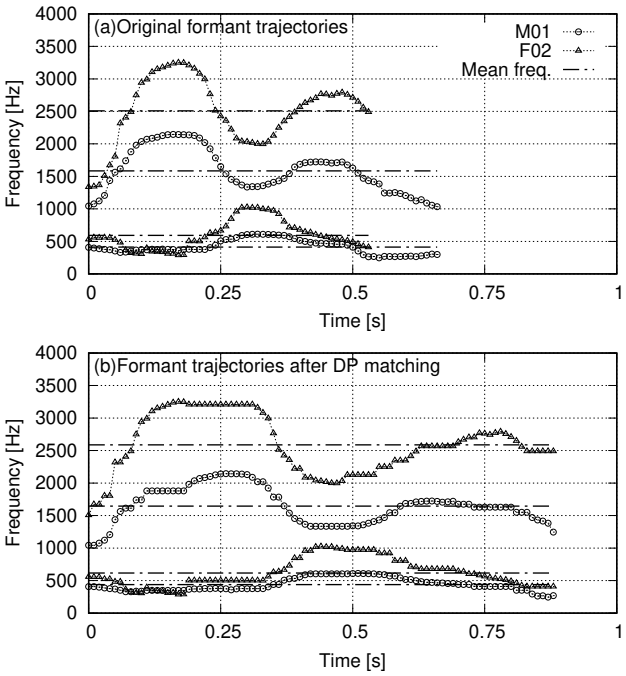


図 1 DP マッチング前後のホルマント軌跡
(男性 M01，女性 F02，単語音声/toriaezu/)

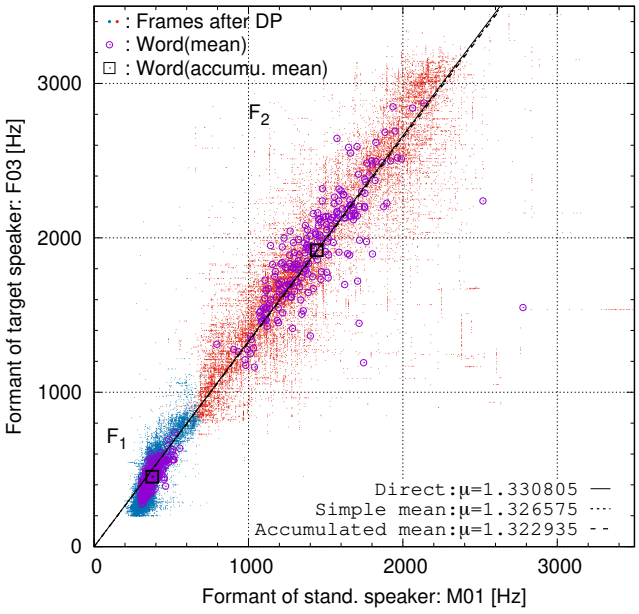


図 2 2 話者間のホルマント相関関係
(基準話者：男性 M01，対象話者：女性 F03)

異なったものから得た 2 つの相関図のそれぞれに主成分分析を行う．次に，求めた 2 つの第一主成分（直線）の交点として，声道形状が相似になる瞬間を求める．その時のホルマントデータのみを用いて声道長比を推定する．本方式を用いるためには，1 人の話者が発話速度を変えて発話した単語データが必要となるが，直接法との比較において，通常の話速であれば直接法の結果は厳密法の結果に近く十分な精度が得られることが示されている．

4. ホルマント平均値を用いた推定法の提案

先に述べたように，声道長比推定の従来法は，各話者による同一単語群の発話（直接法）やそれに加えての発話速度の異なる語群（厳密法）を必要とする．ここでは，そのような発話語群に対する制約を外して，読み上げ文や自由発話から声道長比を推定する方法について検討する．

図 1 のようなホルマント軌跡の視察によれば，発話速度の変化によってホルマント軌跡の上下幅は変化するが，平均値の変動は小さいと思われる．この現象は，DP マッチング処理の有無にはほとんど依存しない．さらに，それらのホルマント平均比は，DP マッチング後の対応フレームのホルマント比に近いことが予想される．

したがって，直接推定に用いたホルマント軌跡の DP マッチング後の時間対応点の代わりに，単語単位での原ホルマント平均を算出し，多数の単語ホルマント平均値より式 (4) を用いて声道長比を推定する方法を提案する．ホルマント平均値の算出は非常に簡便であるため，従来法と比較してこの方法による推定精度が同程度であることが示されれば，声道長比推定手法の単純化が可能になる．同一単語や同一短文にこの方法を適用する場合をここでは「単純平均法（Simple mean method）」と呼ぶ．

さらに，個々の単語（または，短文）について求められたホルマント平均の全単語の全フレームに関する平均値（累積ホルマント平均値）は全ホルマント軌跡の重心を示し，語群の音韻がバランスしている場合には，断面積均一音響管で近似される声道によって作られる中性母音のホルマントに近い値となる．累積平均は，最終的には F_1 と F_2 各 1 点となるので，その相関から式 (4) を用いて，声道長比を推定することになる．この累積ホルマント平均値は，単語や文の順序には依存しないので，音韻分布に大きな偏りがない長文の朗読音声や自由発話に適用できる．したがって，多数の同一単語対を用いる従来法の音声資料に対する制約を緩和することができる．

ここでは，この方法を「累積平均法（Accumulated mean method）」名づける．

図 2 に，2 話者が発話した同一の単語群に対して，「直接法」「単純平均法」「累積平均法」のそれぞれを適用し，算出した 2 話者間のホルマント分布図と声道長比を示す．図によれば，「直接法」による分布の中に「単純平均法」による分布があり，さらに，その分布のほぼ中心に「累積平均法」のデータがある．（正確に述べれば，単語ごとにフレーム数が異なるので，各単語のホルマント平均にフレーム数の荷重をかけてさらに全平均をとったものが中心になる．）3 個の分布は，いずれも直観的に原点を通る直線が当てはまるように見える．実際に，各ホルマントデータに条件付き主成分分析を適用して得た直線を示すと，近接したものになり，式 (4) から得た声道長比 μ は，従来法（直接法）を真値とした場合，単純平均法による誤差は 0.32%，累積平均法による誤差は 0.59%と極めて小さく，推定値はよく一致しているといえる．したがって，単純平均と累積平均による声道長比の推定は妥当と思われる．

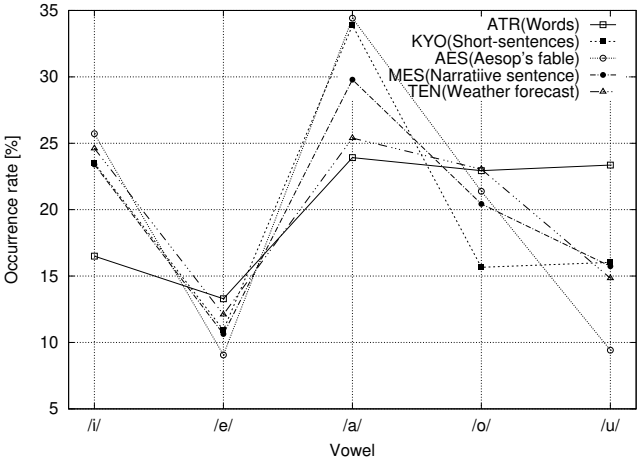


図 3 音声資料における母音の出現頻度

5. 数種の語群を用いた声道長比推定実験

前節で提案したホルマント平均値を用いた声道長比推定法が数種の異なる音声資料の発話に対して，どの程度普遍的に成立するかを調べる．

5.1 音声資料とホルマント推定法

音声資料には，重点研究「音声言語」・試験研究「音声 DB」連続音声データベース（PASL-DSR）を用いた．その中から，トータルの有声音時間長が 20 s を越える 5 資料を選択した．それらは，単語音声として ATR 音韻バランス単語 216 語（以降 ATR と表記する），短文として日本語教育用リスト 70 文（KYO），長文としてイソップ童話「北風と太陽」（AES）とナレーション文章（MES），天気予報文章（TEN）の 3 種を用いる．

各資料の母音個数としての出現頻度は，**図 3** のようになっている．母音のバランスは，ATR 単語群が最もよく，次が天気予報文（TEN），他は/e/，/u/に比べて/a/の頻度が高い．

また，単語（ATR）と短文（KYO）には，他の長文に比べて，感情表現が入りにくいと考えられる．

話者数は 12 名（男女各 6 名）である．記録フォーマット（RAW 形式，16kHz-16bit）を 12kHz-16bit に変換後，IFC 法により分析窓長 20 ms，フレームシフト 10 ms でホルマント分析する．

ホルマント周波数の推定には，逆フィルタ制御法（IFC 法）を用いる．[1] IFC 法では，音声信号を逆フィルタ制御によって単共振信号に分解して，分解信号の零交差周波数分布の荷重平均からホルマントを求める．このとき，零交

表 1　音声資料毎の有声音区間長

時間 [秒]	ATR(単語)	KYO(短文)	AES(長文)	MES(長文)	TEN(長文)
平均	129.21	65.23	28.19	25.07	28.40
最大	110.43	57.96	24.32	21.46	24.58
最小	150.19	80.05	33.46	30.40	33.27

差周波数分布の集中度に最適分析次数の情報があり，それを利用して分析次数を定めることができるため，最適な次数判定と高精度のホルマント推定が可能となる．[4]

各話者の音声資料毎の有声音区間長を平均値，最大値，最小値について**表 1** に示す．同表から，発話速度に関して，話者間に大きな差は認められない．

上記の音声資料を用いて，従来法である直接法と本稿で提案する単純平均法，累積平均法を用いて声道長比（ホルマント比）を推定し，それらの推定精度を比較する．その際，基準話者は男性話者 M01 として，データベース内のその他の話者 11 名についてホルマント比（声道長比）の推定を行う．

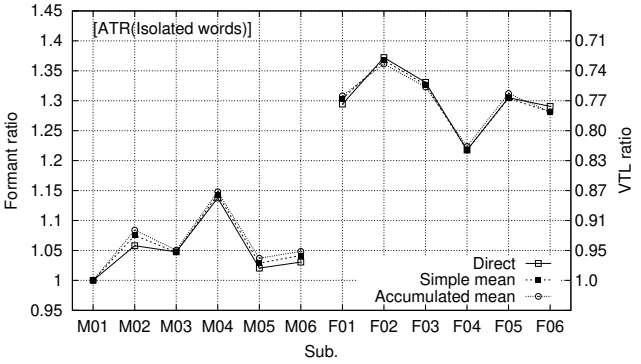
5.2　単語音声および短文を用いた分析結果の比較

図 4 は，単語 ATR および短文 KYO を用いてホルマント比（声道長比の逆数）を推定した結果である．直接法（Direct）と単純平均法（Simple mean）及び累積平均法（Accumulated mean）を適用した結果を比較すると，提案された 2 つの平均法による結果は，直接法の結果に近接しているとともに，音声資料が異なってもほとんど変わらない推定値になることがわかる．また，音声資料が異なっても，推定方法の違いによる推定値のばらつきは，話者間の値の変化に比べてはるかに小さい．推定値の分散分析の結果は**表 2** に示すように，基準話者 M01 を除いた男性話者群（M02〜M06）と女性話者群（F01 F06）の F 値は，有意水準 0.5%の棄却域をはるかに超えるので，推定結果の信頼性が証明できる．

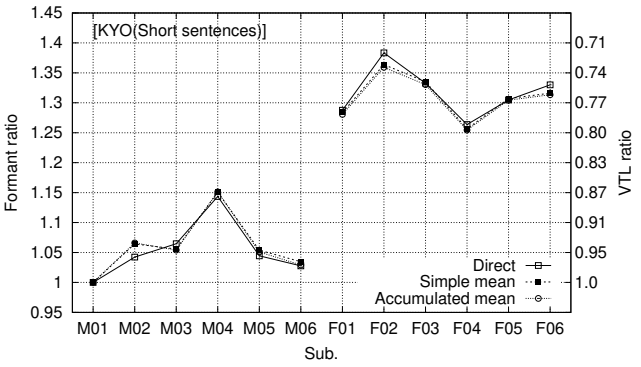
5.3　累積平均法による長文音声からの声道長比推定

単語と短文を用いた推定から，累積平均法の有効性が証明されたので，ここでは，3 種の長文音声を用いて声道長比の推定を行い，音声資料の違いの推定結果への影響を調べた．**図 5**(a) は，3 種の長文音声（AES，MES，TEN）の累積平均法による結果と，比較のため先の実験で使用した単語及び短文を連結して得られた音声の累積平均法による結果を示している．

図より，分析に使用する音声試料の違いは，男性話者群にはほとんど影響しないが，ナレーション文章（MES）を用いたときに一部の女性話者の推定値にばらつき生じている．このことは**表 3** に示す分散分析の結果において，女性群の級内分散の増加に現れている．それによって，F 値は減少するにもかかわらず，有意水準 0.5%棄却域より級間



(a) 単語 (ATR) を用いた結果



(b) 短文 (KYO) を用いた結果

図 4　推定されたホルマント比 (声道長比) の比較 (基準話者：男性 M01)

分散ははるかに大きく，6 人の女性の声道長の違いは明らかである．資料の違いによる推定結果のばらつきの原因には，資料内の文章の長さが不足で累積平均が収束値に達しない場合や，母音出現頻度のアンバランス，パラ言語情報の影響などが考えられる．試しに 3 種の長文を連結することで有声音区間を延長して累積ホルマント平均を求め，その値から声道長比を推定してみると，図 5(b) のようになる．結果は，連結単語や連結短文に近い推定値が得られ，資料差の影響がかなり抑えられる．

6.　まとめ

単語ごとのホルマント平均や文章などの累積ホルマント平均から声道長比を推定する簡易的な方法を提案し，従来法である同一単語のホルマント軌跡の DP マッチングによって得られる多数のホルマント対応点から声道長比を推定する直接法の結果と比較した．その結果，有声音区間が十分に長い音声資料を用いることで従来法と同程度の推定値が得られることが示された．信頼性の高い推定値が得られるために必要な発話時間長の検討が必要であるが，本手法により簡便で正確な声道長比の推定が可能になるため，音声認識における話者正規化などへの応用が期待される．

謝辞　本研究は JSPS 科研費 17K01568 の助成を受けたものです．

表 2 推定方式毎の分散分析の結果

(a) 単語音声 (ATR) を用いた結果

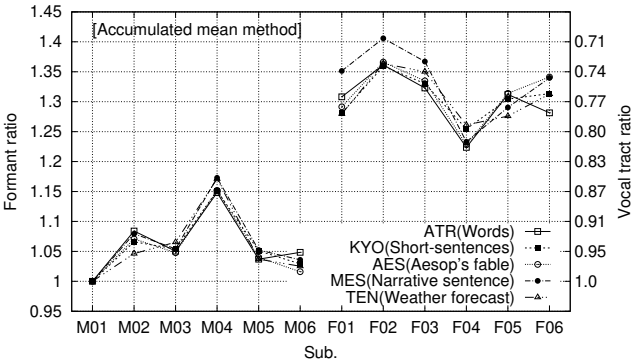
	級内分散	級内自由度	級間分散	級間自由度	F 値	0.5%棄却限界
男性群	0.000709	10	0.025001	4	88.17	7.34
女性群	0.000303	12	0.035810	5	283.81	6.07

(b) 短文音声 (KYO) を用いた結果

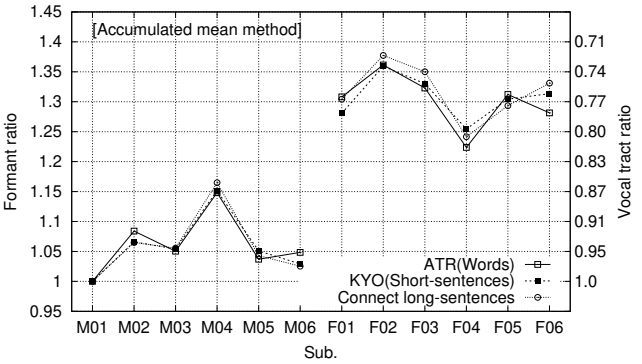
	級内分散	級内自由度	級間分散	級間自由度	F 値	0.5%棄却限界
男性群	0.000524	10	0.025276	4	120.70	7.34
女性群	0.000576	12	0.022272	5	92.84	6.07

表 3 音声資料毎の分散分析の結果

	級内分散	級内自由度	級間分散	級間自由度	F 値	0.5%棄却限界
男性群	0.002430	20	0.051974	4	106.92	5.17
女性群	0.010832	24	0.049019	5	21.72	4.49



(a) 音声資料毎の比較



(b) 連結長文音声と単語，短文の比較

図 5 音声資料の違いによるホルマント比 (声道長比)(基準話者：男性 M01)

参考文献

[1] A. Watanabe, "Formant estimation method using inverse-filter control", IEEE Transactions on Speech and Audio Processing, 2001, Vol. 9, no. 4, pp. 317-326.

[2] Watanabe, Akira and Sakata, Tadashi, "Reliable Methods for Estimating Relative Vocal Tract Lengths From Formant Trajectories of Common Words," IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, 2006, Vol. 14, no. 4, pp. 1193-1204.

[3] Pickett, J.M.: The Sounds of Speech Communication: A Primer of Acoustic Phonetics and Speech Perception, University Park Press (1980).

[4] 渡邉 亮, "逆フィルタ制御 (IFC) ホルマント推定における分析次数自動決定法," 日本音響学会誌, 2013, Vol. 69, no. 4, pp. 168-178.