

## フィラーの声質上の特徴に関する予備的分析\*

○前川喜久雄（国立国語研究所），森大毅（宇都宮大）

### 1 導入

本稿では日本語のフィラー（filled pauses, 以下 FP と略することがある）の声質上の特徴に関する予備的分析の結果を報告する。日本語には「えー」「えーと」「あー」等、多彩なフィラーがあり、学会講演などのモノログにおいては総語数の 6%前後がフィラーである。フィラーは、固有の語彙的意味が明瞭でなく、実際に用いられた文脈のなかで随時に意味が定まるという点で特殊な言語記号である[1]。フィラーには談話管理機能や談話・統語境界に関する情報を伝える機能があると考えられているが[2-3]、更に種々のパラ言語情報も多量に伝達していると予想される。

しかし、フィラーの音声生成上の特性は文献[4]以外にはほとんど分析されていない。そのため、例えば「フィラーのピッチは音韻的に指定されるのか?」「フィラーの母音の調音は通常の単語の母音と同一か?」等の基本的な問題に対しても我々は答えることができない。

本稿の第一著者は先に上記第一の問題を検討し、フィラーに対するピッチの音韻的な指定はなく、隣接語彙項目に対して指定されたトーンを補間することで、フィラーのピッチを高い精度で予測できることを示した[5]。

今回は、上記第二の問題の検討の手始めとして、フィラーに含まれる母音と通常の語彙項目（Lexical items, 以下 LX と略することがある）に含まれる母音との声質の特徴を比較検討した。声道の制御に関わる特徴としてフォルマント周波数、喉頭音源の制御に関わる特徴として振幅（intensity）とスペクトル傾斜を測定し、更に基本周波数の揺らぎの指標である jitter も試験的に測定した。

なお本稿では声質（voice quality）という語を単音よりも広い範囲で指定される調音上の特徴という意味（音声学的にはこれが本来の意味である）で用いており、発声（phonation）

の特徴という限定された意味では用いていないことに注意してほしい。

### 2 データ

データとして『日本語話し言葉コーパス』のコア部分（44 時間, 約 50 万短単位, 以下 CSJ-Core と呼ぶ）中のモノログ（学会講演と模擬講演）を利用する[6]。話者は男性 79 名、女性 58 名である。

CSJ-Core には多様なフィラーが多数含まれているが、今回はそのうち母音のみから構成される 5 種類のフィラーと通常の語彙項目の語頭に生じた長母音とを比較した。/eH/であれば後者には「例、映画、データ、経験」などの長母音が該当する。また持続時間が極端に短いサンプルを除外するために、持続時間が平均周期の 10 倍以上あるサンプルだけを分析対象とした。この制約下で得られる分析対象サンプル数を Table 1 に示す。以下本稿では、フィラーのサンプル数が過小なケースを除外して、男声の/aH/, /eH/, /oH/, 女声の/aH/, /eH/を分析の対象とする。

Table 1 Number of samples\*

| Speaker sex | Vowel | Filled Pause (FP) | Lexical Item (LX) |
|-------------|-------|-------------------|-------------------|
| Male        | /aH/  | 108               | 113               |
|             | /eH/  | 2411              | 764               |
|             | /iH/  | 16                | 281               |
|             | /oH/  | 66                | 2177              |
|             | /uH/  | 16                | 561               |
| Female      | /aH/  | 40                | 61                |
|             | /eH/  | 1049              | 529               |
|             | /iH/  | 2                 | 248               |
|             | /oH/  | 10                | 1910              |
|             | /uH/  | 12                | 500               |

\*Samples in the shaded cells are omitted.

### 3 分析

#### 3.1 フォルマント周波数

フォルマント分析は Praat[7]で実施した。CSJ-Core の分節音ラベルによって対象母音

\* Voice-quality characteristics of filled pauses: A pilot analysis. By MAEKAWA, Kikuo (NINJAL) and MORI, Hiroki (Utsunomiya Univ.).

の開始時刻と終了時刻を求め、その区間における第1, 第2, 第3フォルマント(F1, F2, F3)の平均値を線型予測分析法で求めた。

### 3.2 振幅

振幅の分析も Praat で実施した。フォルマント分析と同様に対象母音区間の平均振幅を計算したうえで、話者ごとに FP, LX にわたる全母音を z スコア化した値を分析に利用する。

### 3.3 スペクトル傾斜

音声のスペクトルは喉頭音源のスペクトルと声道のスペクトルの積であるが、喉頭調節の違いによる音声のスペクトル傾斜の変化は、主として喉頭音源のスペクトル傾斜の変化を反映している。例えば、フィラーの声質に関わる要因の1つである声門閉鎖の不完全性は、声道壁の損失増加による F1 帯域幅の増大[8]などを通して声道のスペクトル傾斜にも影響を与えるものの、主として影響を受けるのは喉頭音源のスペクトル傾斜である。本稿では、自発音声では高精度な推定が難しい喉頭音源のスペクトル傾斜のかわりに、推定が容易な音声のスペクトル傾斜を分析する。

コーパスに含まれる大量の音声のスペクトル傾斜を短時間で分析するため、本研究ではケプストラムを用いてスペクトル傾斜を推定した。第1ケプストラム係数は、音声のスペクトル傾斜に関連した特徴量として、対話音声の声質の分析にも用いられている[9]。 $S(\omega) = |X(\omega)|^2$ を音声信号 $x(n)$ のパワースペクトルとすると、ケプストラム係数 $c_m$ は対数パワースペクトルのフーリエ係数であり、

$$\log S(\omega) = \sum_{m=-\infty}^{\infty} c_m e^{-jm\omega} \quad (1)$$

が成り立つ。 $\log S(\omega)$ の実対称性から

$$\log S(\omega) = c_0 + 2 \sum_{m=1}^{\infty} c_m \cos m\omega \quad (2)$$

であり、これを $|m| = 1$ までで打ち切って近似した対数パワースペクトル

$$\log S(\omega) \cong c_0 + 2c_1 \cos \omega \quad (3)$$

は元のスペクトルの概形の近似となっている。

Fig. 1 に CSJ-Core に含まれる女性話者のフィラー/eH/のパワースペクトル、および式(3)で近似したスペクトルを示す。

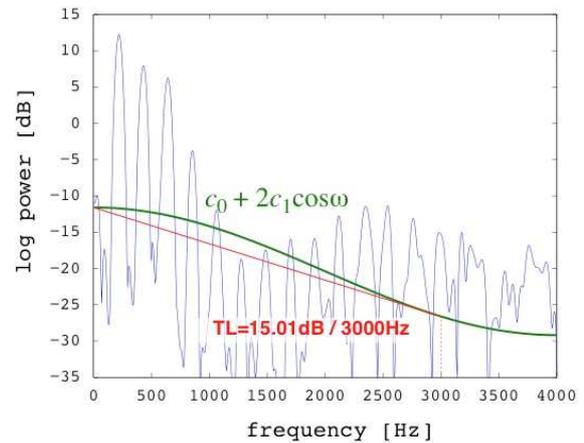


Fig.1 Spectrum approximation by the first cepstrum component

対数パワースペクトルを  $\cos$  関数で近似した時の  $\cos$  関数の振幅が $c_1$ に比例し、おおよそスペクトル傾斜に対応していることが分かる。

本稿では音声のスペクトル傾斜を $S(0\text{Hz})$ と $S(3000\text{Hz})$ の比(デシベル表示)で定義する。これは $c_1$ を用いて以下のように求められる。

$$\begin{aligned} & 10 \log_{10} \frac{S(0\text{Hz})}{S(3000\text{Hz})} \\ &= \frac{10}{\log_{10}} \left\{ 2c_1 \left( \cos 0 - \cos 2\pi \cdot \frac{3000\text{Hz}}{F_s} \right) \right\} \quad (4) \end{aligned}$$

線形予測係数から求められる LPC ケプストラムは、対数パワースペクトル包絡のフーリエ係数であり、 $c_1 = a_1$ が成り立つ[10]。よって、 $c_1$ は音声の線形予測分析により求めることができる。対象母音波形(標本化周波数 $F_s = 8000 \text{ Hz}$ )に対し、Praatにより8次の線形予測分析を実行し、得られた $a_1$ を全フレームにわたって平均し式(4)に代入することでスペクトル傾斜を求めた。

### 3.4 Jitter

Jitter は基本周波数の揺らぎの尺度であり、病的音声の評価のために利用される[11]。音声言語医学の領域では Jitter は通常、母音を単独で長く引き延ばしたサンプルを用いて測定されるが、ここでは連続音声への適用を試みた。

Jitter の定義には様々なものがあるが、ここでは連続する5周期の揺らぎを測定して評価する PPQ5 の値を利用する。PPQ5 は0~4(もしくは0~400%)の値をとる[12]。

## 4 結果

### 4.1 フォルマント周波数

3 個のフォルマント周波数のうち、興味深い結果が得られたのは F1 であった。測定結果を Fig. 2-3 に示す。男声の/eH/と/oH/、女声の/eH/の F1 平均値は、語彙項目(LX)よりもフィラー(FP)において有意に高い値を示した。図中の検定結果は Welch t-test による (以下同様)。なお今回分析対象から除いた女声の/oH/においても同じ傾向が観察された。

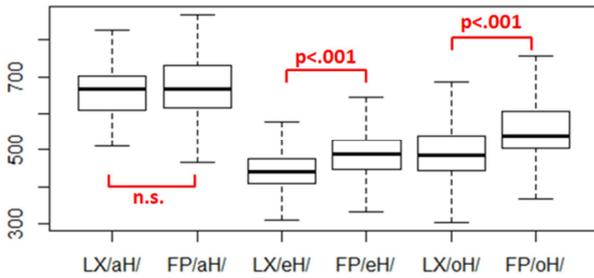


Fig.2 Mean male F1 [Hz]

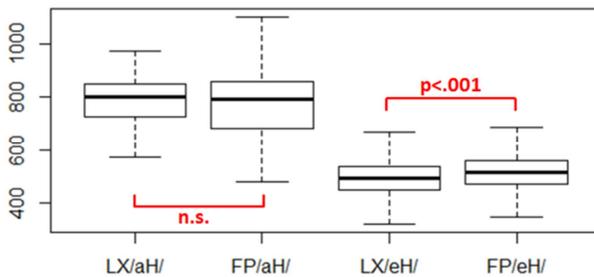


Fig. 3 Mean female F1 [Hz]

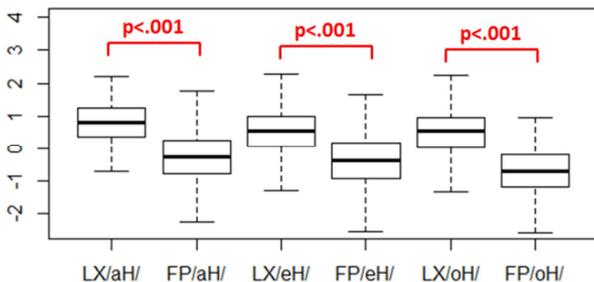


Fig. 4 Mean male z-normalized intensity

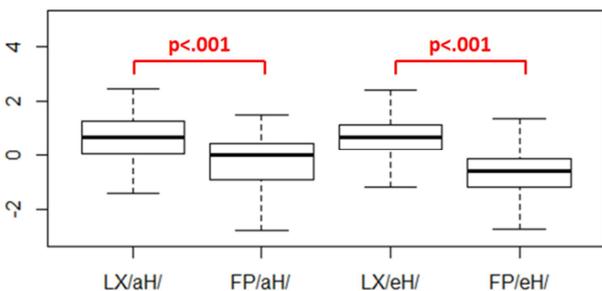


Fig. 5 Mean female z-normalized intensity

### 4.2 振幅

振幅平均値の測定結果を Fig. 4-5 に示す。すべての母音において FP の振幅平均値は LX よりも有意に小さい値を示した。

なお今回分析から除いた女声の/oH/においても同じ傾向が観察された。

### 4.3 スペクトル傾斜

スペクトル傾斜平均値の測定結果を Fig. 6-7 に示す。男声ではすべての母音において FP は LX よりも有意に高い値を示した。一方、女声ではいずれの母音においても TL に有意差は認められなかった。

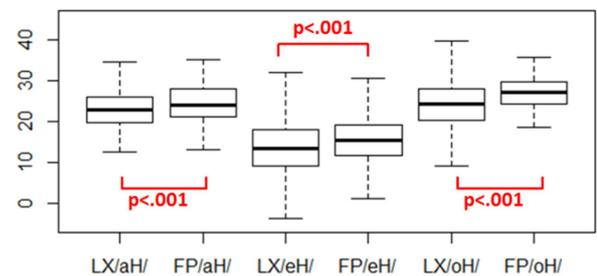


Fig.6 Mean male spectral tilt [dB]

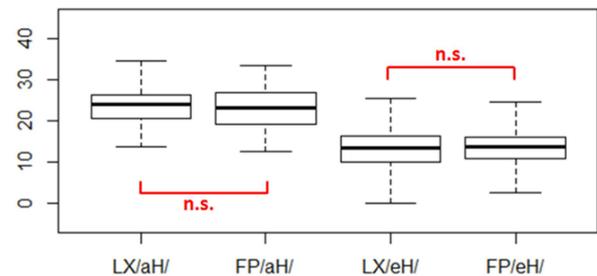


Fig.7 Mean female spectral tilt [dB]

### 4.4 Jitter

Fig. 8-9 に Jitter の平均値を示す。縦軸の値は百分率で表された Jitter (PPQ5) を 10 を底とする対数尺度に変換した値である。Jitter は男声女声ともに LX よりも FP の方が一貫して有意に高い値を示した。

これはフィラーの方が基本周期 (基本周波数) の揺らぎが大きいことを示唆する結果である。ただし Jitter の値は、基本周期の測定に誤りが含まれていると意味をなさないことに注意が必要である。今回の測定値のなかには、種々の原因で基本周期の測定に失敗しているサンプルも含まれているので、Fig. 8-9 は参考値とみなす必要がある。

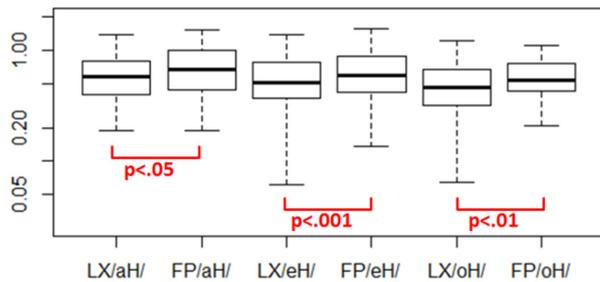


Fig. 8 Mean male jitter [%] in log10 scale

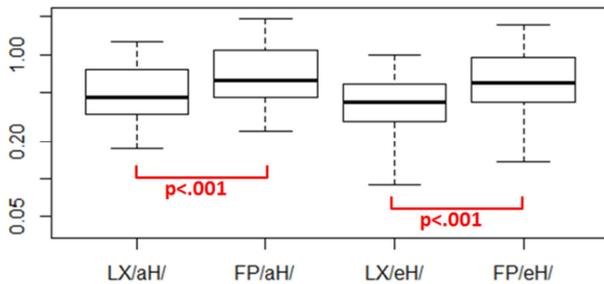


Fig. 9 Mean female jitter [%] in log10 scale

## 5 議論と今後の課題

フィラーと通常の語彙項目とは母音の音響特徴に組織的な差異がある。例えばフィラーの/eH/, /oH/は、統計的にみると、語彙項目に比べて開口度が大きく（F1 が高く）、振幅が小さく、基本周期の揺らぎが大きい。更に男声ではスペクトル傾斜が増大して「弱い」発声になる。

このような差が生じる原因のひとつとしてパラ言語情報の関与が考えられる。先述のとおりフィラーには言語情報が希薄であり、主に文献[13]が広義のパラ言語情報と呼ぶ情報を伝達していると予想される。パラ言語情報は連続性に富み多様であるが [13-17]、それを伝達するフィラーの音形（語形）は高々数十タイプに限定されている。そこで声質を制御することでパラ言語情報を詳細化する必要が生じるのではないかと考えられる。

パラ言語情報の影響が分節音の調音に及ぶことは夙に指摘されており[14-15]、パラ言語情報の伝達に発声上の特徴が利用されていることも指摘されている[15-17]。これらの先行研究は主に語彙項目を対象としたものであるが、同じ影響がフィラーには一層顕著に観察されるであろうことは想像に難くない。

ただし、今回の分析では多様なパラ言語的意味を担ったフィラーを区別せず、一括して分析している。そのため使用頻度の高いパラ

言語的意味の伝達に利用される音響的特徴を把握したに留まっている可能性が高い。

今後は発声様式のアノテーションを施して、様式ごとに音響特徴を分析するなどの工夫が必要である。現在、そのために機械学習によるアノテーションの準備を進めている。

謝辞: 本研究の一部は科研費基盤研究（課題番号 26284062）に依るものである。Jitter 分析にコメントを頂いた粕谷英樹氏に感謝する。

## 参考文献

- [1] 田窪「感動詞の言語学的位置づけ」月刊言語, 34 (11), 2005.
- [2] 定延・田窪「談話における心的操作モニター機構」言語研究, 108, 1995.
- [3] Watanabe, Michiko. *Features and roles of filled pauses in speech communication*. Hituzi, 2009.
- [4] 渡辺和希「課題志向型コーパスにおける感動詞の音質」第 26 回日本音声学会全国大会予稿集, 2012.
- [5] Maekawa. "Prediction of F0 height of filled pauses in spontaneous Japanese". *Proc. 6<sup>th</sup> DiSS*, 2013.
- [6] 前川『日本語話し言葉コーパス』の概要」日本語科学, 15, 2004.
- [7] <http://www.fon.hum.uva.nl/praat/>
- [8] Hanson. "Glottal characteristics of female speakers: Acoustic correlates", *JASA*, 101 (1), 1997.
- [9] Arimoto et al. "Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment," *Acoust. Sci. Tech.*, 33 (6), 2012.
- [10] Rabiner & Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, NJ, 1993.
- [11] 日本音声言語医学会『新編声の検査法』医歯薬出版, 2009.
- [12] [http://www.fon.hum.uva.nl/praat/manual/PointProcess\\_Get\\_jitter\\_ppq5\\_.html](http://www.fon.hum.uva.nl/praat/manual/PointProcess_Get_jitter_ppq5_.html)
- [13] 森・前川・粕谷『音声は何を伝えているか』コロナ社, 2014.
- [14] Maekawa & Kagomiya. "Influence of paralinguistic information on segmental articulation", *Proc. ICSLP*, Beijing, 2000.
- [15] Maekawa. "Production and Perception of 'Paralinguistic' Information". *Proc. Speech Prosody*, Nara, 2004.
- [16] 石井・石黒・萩田「韻律および声質を表現した音響特徴と対話音声におけるパラ言語情報...」情報処理学会論文誌, 47(6), 2006.
- [17] 石井・石黒・萩田「Breathy/Whispery 発声の音響特徴と音声コミュニケーションにおける役割」信学技報 SP2008-42, 2008.