自発音声中のフィラーの特性に関する予備的分析:位置と高さの分析

前川喜久雄（国立国語研究所） kikuo@ninjal.ac.jp

Preliminary study on the characteristics of filled pauses in spontaneous speech: Analysis of location and pitch height

Kikuo Maekawa (National Institute for Japanese Language and Linguistics)

１．はじめに

「あー」「えー」等のいわゆるフィラー(filled pauses)は自発音声の顕著な音声特徴のひとつである。フィラーは言語情報の伝達にはほとんど寄与していないので、情報処理研究の世界ではフィラーを「不要語」と呼ぶことがある。

しかし、フィラーがほぼ不要であると言えるのは言語情報の伝達に関してであり、パラ言語情報や談話の管理に関わる情報の伝達に関しては、フィラーはむしろ積極的な役割を果たしていること（定延・田窪 1995, 山根 2002）、また発話プラニングの研究にも重要な知見を提供することが知られている (Watanabe 2009)。これらの先行研究において十分に検討されていないフィラーの特徴がいくつかあるが、本発表ではフィラーの高さ（ピッチ）の問題をとりあげる。また議論のなかでフィラーの生起位置についても言及する。

フィラーの高さを問題にする背景には、自発音声の韻律アノテーション方式の問題がある。現在韻律アノテーションとして広く利用されている ToBI 方式では、韻律構造の階層構造を表現する韻律境界を参照しながら、発話の要所要所に指定された音韻論的なトーン（標準日本語では H と L の二段階）を補間することによって発話のピッチ形状が生成されると想定している。このようなイントネーションモデルに立脚して、フィラーを含む自発音声の韻律アノテーションを実施する際には、そもそもフィラーに対して何らかのトーンを指定すべきかどうかという根本問題を検討する必要がある。しかし X-JToBI の設計においては、時間的な制約からこの問題を十分に検討することなく、すべての韻律的フィラーに対して FH(filler high)ないし FL(filler low)のいずれかのトーンラベルを付与することにした (Maekawa et al. 2002)。今回の分析はこのアノテーション仕様の問題点を改めて検討しようとするものでもある。

２．データ

『日本語話し言葉コーパス』(CSJ)のうち X-JToBI による韻律ラベリングが施されているコア部分(CSJ-Core)を分析対象とした。CSJ-Core におけるフィラーの認定基準はやや複雑であるので簡単に説明する。転記作業では一定の基準を満たした要素をフィラーと認定して「(F えーと)」のようにタグを付与している（小磯他 2006）。一方 X-JToBI ラベリングでは、韻律特徴を認定基準に加味しており、要素全体が主観的にみて平坦に発話されているもの（言い換えれば句頭の上昇、アクセント核などが生じていない要素）のみをフィラーと認定している。両者は必ずしも一致しない。転記ファイル中の（F）タグと X-JToBI における韻律的フィラーの関係を表 1 に示す。今回の分析では X-JToBI アノテーションでフィラーと認定された 35,164 個のフィラーを一時的な分析対象とし、そこから種々の条件で絞り込んだデータを分析する。

CSJ の転記では可能な限り音声を忠実に表現しようとしているので、フィラーを転記した

文字列は多彩である。単純に分類すると異なりで 160 種以上のフィラーが存在することになる。これでは分析に不便なので、表 2 に示すように 2 段階の分類を施した。Class1 は母音の長短や促音、撥音などに配慮した細かな分類、Class2 は主に冒頭の分節音に着目した粗い分類になっており、両者は階層関係にある。なお頻度が 10 以下のフィラーはいずれのクラスにも分類していない。

以下ではフィラーの高さを分析する。そのために以下の 3 条件に適ったサンプルを抽出した。①フィラーの F0 値が抽出されており信頼できること、②フィラー直前のアクセント句末の L%トーンの F0 値が抽出されており信頼できること、③フィラー直後のアクセント句頭の%L トーンの F0 値が抽出されており信頼できること。4,892 個のフィラーがこの条件に適合した。これらのフィラーおよび前後のアクセント句に関わる F0 値は対数変換したうえで話者ごとに Z 変換した値を計算した。

３．位置の分析

３．１　節中の生起位置

まずフィラーの生起位置を検討する。CSJ には節境界情報（丸山他 2006）が付与されているので、これを利用して節中のどのような位置にフィラーが生じているかを検討した。結果を表 3 に示す。行は節の冒頭から語（短単位）を単位に数えたフィラーの生起位置であり、列は Class1 で分類したフィラーである。最長 15 単位までを分析対象とし、その範囲に生じたフィラーの総数に占める各位置の生起数を示した。ただし 15 単位までの生起総数が 10 未満のフィラーは除外した。太字は各フィラーの最頻値を、網掛けは中央値を示す。表 3 からはクラス DE のフィラーがもっぱら節の冒頭に生じている点でそれ以外のフィラーとは顕著に異なっていることがわかる。

表 1: 転記における（F）タグと X-JToBI におけるフィラー認定の関係

転記の(F)タグ	X-JToBIでのフィラー認定	
	有	無
有	29,880	6,426
無	5,284	---

表 2: フィラーの分類

Class1	Class2	Example
A	A	(F あ)　(F あっ)
AH		(F あー)
AN		(F あの)　(F あのー)　(F あーの)　(F あーのー)等
DE	D	で　て　んで
E	E	(F え)
EH		(F ええ)　(F えー)　ええ
ET		(F えーと)　(F えーとー)　(F えーっと (F えっと)等
KN	K	この
KO		こう
M	M	(F ま)
MH		(F まー)
MO		もう
N	N	(F ん)
NH		(F んー)
NT		(F んーと)　(F んと)　(F んーとー (F んーっと)等
UN		(F うーん)　(F うん)　うん
SN	S	(F その)　(F そのー)　その
U	V	(F う)
UH		(F うー)
I		(F い)
IH		(F いー)
O		(F お)
OH		(F おー)

３．２　生起タイミング

次に直前および直後のアクセント句（フィラーはアクセント句とみなさない）との関係での生起タイミングを検討する。当該フィラー直前のアクセント句の末端時刻を T1、直後のアクセント句の始端時刻を T2、そして当該フィラーの開始時刻を T3 としたとき、(T3-T1)/(T2-T1)によってフィラー開始時刻の相対位置である relPosit を定義する。relPosit は 0.0 以上 1.0 未満の値をとる。0.0 のときは直前のアクセント句の直後にポーズを介することなくフィラーが生起している。フィラーの前にポーズが生じると relPosit>0.0 となり、フィラーの生起タイミングが後続アクセント句に接近するほど（T2-T1 が一定ならポーズが長くなるほど）relPosit は 1.0 に近づ

くがフィラー自体の持続時間長があるので 1.0 になることはない。

図 1 は class1 の各フィラーごとに relPosit の分布状態を箱ひげ図で表示している。ここでも DE は relPosit の 1.0 よりに分布が集中している点で特徴的な分布を示している。これと類似した分布を示すフィラーに N がある。反対に AH, AN は relPosit の 0.0 近くに分布が集中している。そしてこれらのフィラーとは対照的に A, M, MO, O などは relPosit の広い範囲にわたって分布している。

表 3: 節中の位置ごとの class1 フィラーの生起頻度

Location	A	AH	AN	DE	E	EH	ET	KO	M	MH	MO	N	SN	U	UH	IH	O	OH
1	0	0	9	**90**	78	84	10	0	6	1	2	0	1	1	0	0	0	0
2	4	10	75	0	38	131	31	3	12	4	**4**	2	4	2	1	0	2	2
3	**14**	**16**	**99**	0	**93**	**241**	**37**	3	17	1	1	**5**	**21**	0	2	**3**	2	11
4	**14**	8	73	1	83	164	29	0	**24**	5	**4**	4	13	2	1	0	**5**	**17**
5	4	9	67	2	47	130	22	5	17	6	2	3	15	0	1	1	4	8
6	5	3	33	1	65	133	18	**6**	**24**	**7**	3	4	11	1	2	0	1	10
7	6	9	41	2	41	110	18	5	10	6	3	0	13	1	**5**	0	2	11
8	3	3	36	3	54	118	9	5	20	3	1	0	4	1	2	1	3	6
9	3	4	43	4	47	105	8	1	13	3	3	3	6	1	0	**3**	2	8
10	4	5	35	4	41	98	9	5	21	6	1	1	6	**3**	0	0	0	3
11	3	2	21	1	41	93	7	1	17	2	3	2	6	0	0	0	2	5
12	2	0	17	0	36	73	7	4	11	1	1	4	3	0	0	1	3	3
13	2	2	19	2	32	75	2	2	10	5	1	1	6	0	0	0	1	5
14	2	3	13	0	29	55	7	3	7	4	1	3	2	1	0	2	2	2
15	2	2	18	1	16	51	9	2	6	0	1	3	2	0	0	0	2	4
Total	68	76	599	111	741	1661	223	45	215	54	31	35	113	13	14	11	31	95

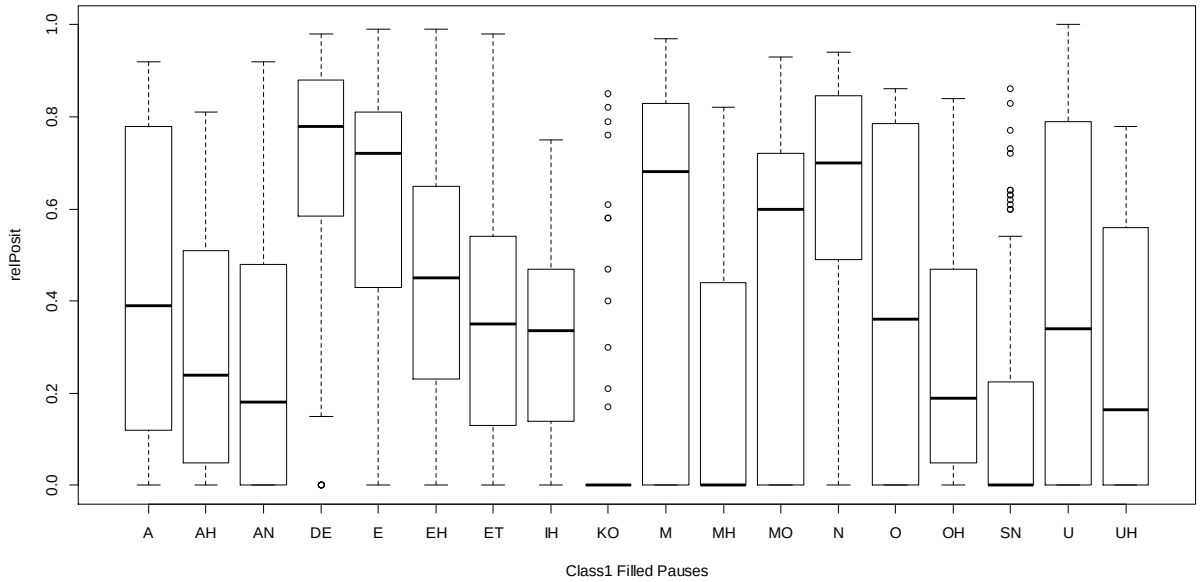

図 1: Class1 フィラー（N>9）の生起タイミング(relPosit)の分布

４．高さの分析

４．１　高さと生起タイミングの関係

フィラーの高さと前節で検討した relPosit の間には緩やかな相関関係が認められる。図 2 は relPosit>0.0 であるサンプルを対象に、relPosit を 10 等分して各時間区分に属するフィラーの対数正規化 F0 値(F0Logn)の分布を示したものである。relPosit（横軸）が小さい間は F0Logn（縦軸）はほとんど変動していないが、relPosit が 0.7 以上になると両者間に正の相関が生じている。

図 3 は図 2 と同様の分析を Class2 のフィラーのうち生起頻度の高いもの(N>50)に施した結果である。A, D 以外のクラスには図 2 と同様の関係が認められる（A も relPosit>0.9 では上昇している）。

４．２　高さの予測モデル

前節の知見に基づいてフィラーの予測を試みる。五つの予測モデルを考案して予測精度を比較する。

Model 1（図 4 参照、以下同様）は先行アクセント句末の L%トーンの値が relPosit に関わらずすべてのフィラーにコピーされるモデルである。

Model 2 は反対に後続アクセント句頭の%L トーンの値がすべてのフィラーにコピーされるモデルである。Model 3 は Model1 と 2 の折衷で、relPosit<0.7 であれば Model 1 を、それ以外では Model 2 を適用する。

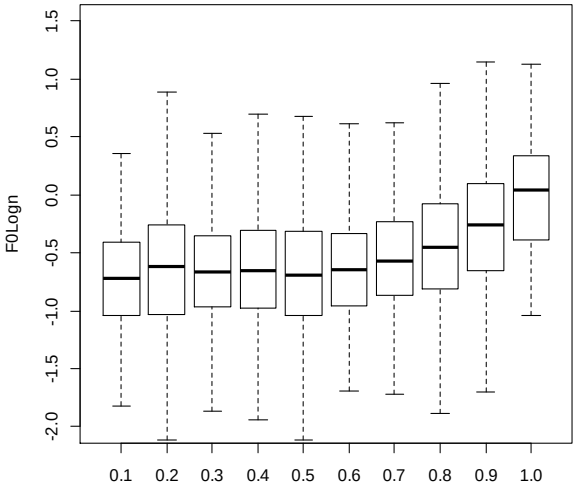

図 2: relPosit と対数正規化 F0(F0Logn)の関係

Model 4 は L%と%L を直線補間して、フィラーの開始時刻(relPosit)に対応する F0 値をフィラーの高さとする。最後に Model 5 は relPosit<0.7 では Model 1 を適用し、その後の時間区間においては relPosit に応じて L%から%L を直線補間するモデルである。

予測結果を表 4 にまとめた。各モデルの予測値の RMS 誤差を F0 計測値(F0Hz)と対数正規化 F0 値(F0Logn)の両方について示す。予測精度上位のモデル 2 種に網掛けを施し、誤差最少のモデルを太字で表示した。Model 4 と 5 の成績は伯仲しているが、DE に関する Model 5 の予測誤差が大きい。全般的に見ると Model 4 の方が問題が少ないといえる。

表 4: 予測モデルの評価

Filled Pause	N	Model 1		Model 2		Model 3		Model 4		Model 5	
		F0Hz	F0Logn	F0Hz	F0Logn	F0Hz	F0Logn	F0Hz	F0Logn	F0Hz	F0Logn
AN	411	26.9	0.64	37.0	0.87	28.6	0.70	25.7	0.65	**26.1**	**0.62**
DE	118	76.6	1.83	28.7	0.54	38.9	0.77	**28.5**	**0.52**	46.2	1.02
E	769	26.5	0.76	27.3	0.71	27.0	0.69	20.6	0.58	**18.4**	**0.50**
EH	1790	24.4	0.72	31.4	0.88	25.1	0.72	22.1	0.69	**22.1**	**0.64**
ET	221	30.0	0.79	35.5	0.86	30.5	0.77	**24.4**	**0.65**	28.0	0.73
M	194	31.6	0.91	22.7	0.57	23.0	0.60	**18.0**	**0.49**	21.3	0.59
ALL	3975	27.6	0.77	30.7	0.81	26.2	0.70	22.2	0.64	23.0	0.63

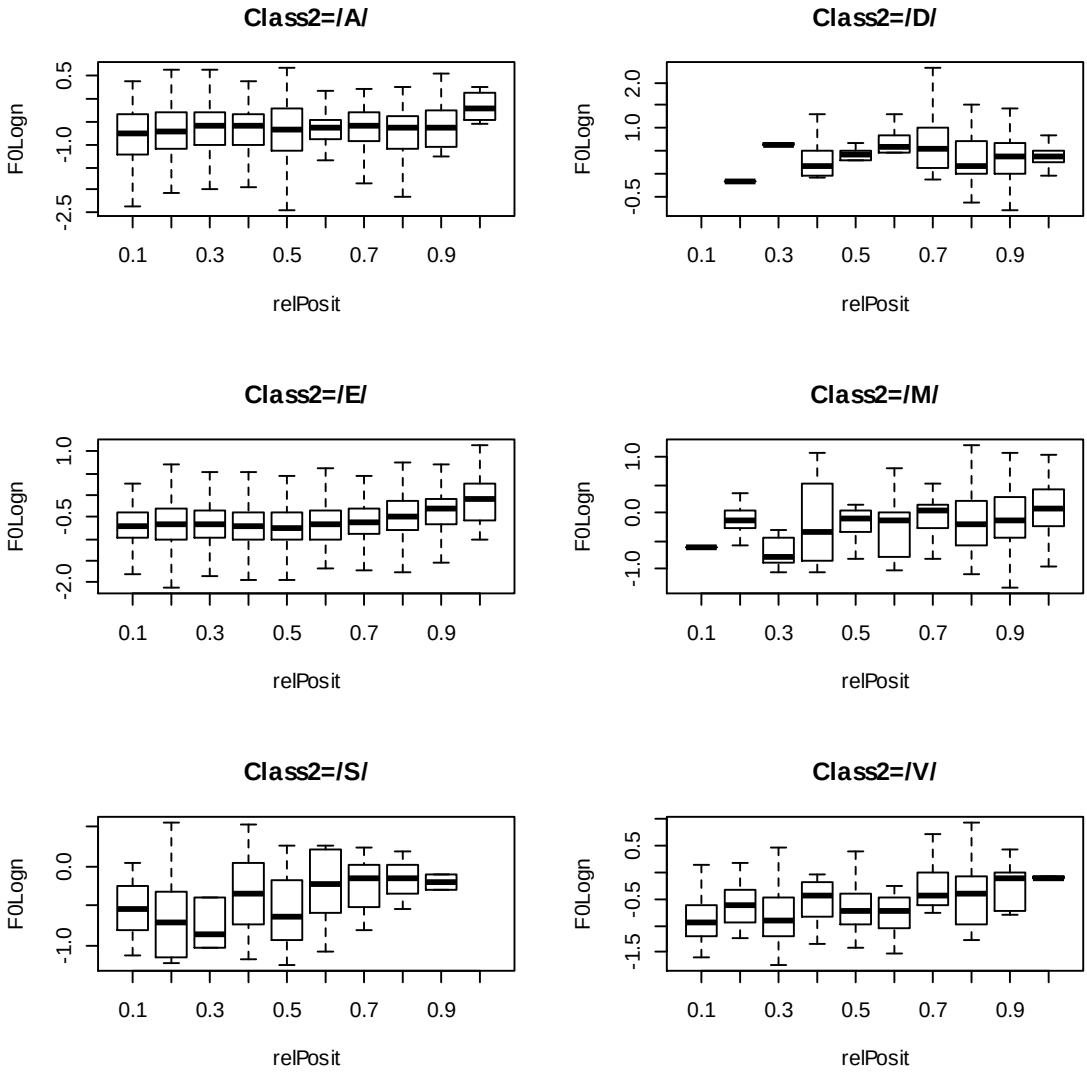

図 3: Class1 フィラーにおける生起タイミング(relPosit)と対数正規化 F0(F0Logn)の関係の例

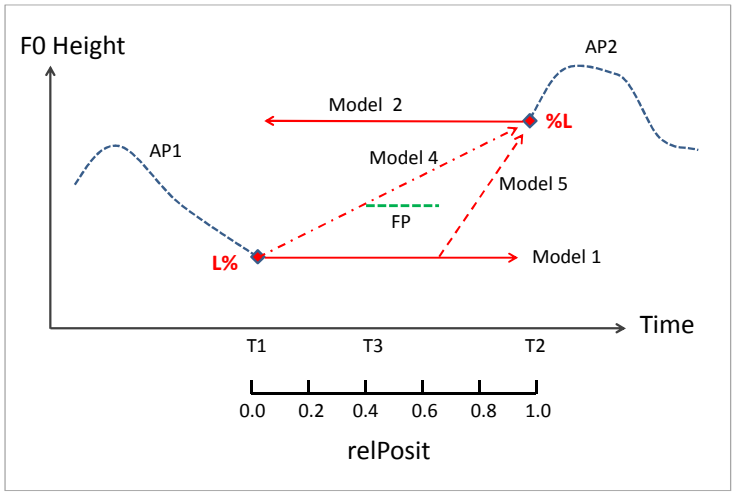

図 4: フィラーの高さの予測モデル（AP はアクセント句、FP はフィラー）

５．議論と暫定的結論

我々が日常耳にする自発音声中のフィラーは様々な高さで発音されている。またフィラーによって高く発音されやすいものと低く発音されやすいものがあるという印象がある。しかし今回の分析によって前後のアクセント句のトーン情報とフィラーの生起タイミングの情報があればフィラーの高さはかなりの程度まで予測可能であることが明らかになった。先にのべたフィラーの多様性は、フィラーそのものの特性ではなくそれがおかれた環境の特性によってもたらされている可能性が高い。

例えば Class1 の DE（Class2 の D）はしばしば節の冒頭に生じており（表 3)、生起タイミングは後続アクセント句の直前であることが多い（relPosit が 1.0 に近い、図 1)。ところで DE が生じやすい種類の節境界ではピッチレンジがリセットされることが多い。[1]リセットが生じると先行アクセント句末の L%よりも後続アクセント句の%L の方が高くなるので、結果的に DE は高い F0 を伴うことになる。ここで重要なのは当該フィラーの発話中での生起位置とタイミングに関する特性が結果的にこのフィラーを高く発音させているのであって、当該フィラーに高いピッチが直接指定されているわけではないという点である。

このような現象をアクセント句に由来するトーンの spreading とみる人もあるだろうが、この解釈は適当でない。今回検討したモデルのうち Model 1, 2, e は spreading を想定したモデルであるが、これらはおしなべて予測性能が芳しくなかった。フィラーはレキシコンのレベルでピッチの指定をうけていないだけでなく、表層的にもピッチを指定されておらず、その高さは隣接する（アクセント句由来の）トーンの補間によって受動的に決定されていると考えるのが今回の分析結果に対する妥当な解釈であろう。このような F0 計算メカニズムは Pierrehumbert and Beckman (1988)において夙に提案されているものであるが、それがフィラーにもあてはまること、そしてフィラーの生起位置とタイミングの特性が表層的にフィラーの高さの著しい多様性をうみだしていることのふたつが本研究の新しい知見である。

ただし今回の分析は言語学的にはかなり粗いものである。個々のサンプルを仔細に検討した場合、フィラーが表層音韻論のレベルで例外的にピッチ指定を受けていると考えるべきケースがないとは断言しにくい。この問題については今後詳細な分析を進める予定である。

謝辞: 本研究は日本学術振興会科学研究費補助金 K23520483（「自発音声データの定量的解析による日本語韻律構造理論の再構築」代表者：前川）および国立国語研究所共同研究プロジェクト「コーパス日本語学の創成」の援助によって実施された。共同研究プロジェクトメンバーのコメントは非常に有益であった。

文献

小磯花絵・西川賢哉・間淵洋子(2006).「転記テキスト」『国立国語研究所報告 124 日本語話し言葉コーパスの構築法』, pp.23-132.

定延利之・田窪行則(1995).「談話における心的操作モニター機構」言語研究, 108, pp.74-93.

丸山岳彦・高梨克也・内元清貴(2006).「節単位情報」『国立国語研究所報告 124 日本語話し言葉コーパスの構築法』, pp.255-322.

山根智恵(2002).『日本語の談話におけるフィラー』くろしお出版.

Maekawa, K., H. Kikuchi, Y. Igarashi and J. Venditti (2002). “X-JToBI: An extended J_ToBI for spontaneous speech.” *Proc. ICSLP2002*, Denver, pp. 1545-1548.

Pierrrehumbert, J. and M. Beckman (1988). *Japanese Tone Structure*. The MIT Press.

Watanabe, M. (2009). *Features and Roles of Filled Pauses in Speech Communication*. Hituzi Shobo Publishing.

---

[1] そのことを明らかに示す分析結果もあるが紙幅の関係で本稿に掲載することができなかった。口頭発表時に補充する予定である。