# プロミネンスと音源パラメータの関係について *

⊙丁 文、Nick Campbell (ATR音声翻訳通信研究所)

## 1 はじめに

音声合成における韻律データベースの作成や合成音声の自然な韻律の生成を目的とし、日本語と英語の対話文を用いて文中のプロミネンスと音源特性の関係について調べた。以前の研究[1]では、英語の音韻レベルのフォーカスによる音韻時間長とパワーの違いがあると指摘されている。また、声道特徴ホルマントと発話様式の相関関係も研究されてきた[2],[3]。今回は音源特性をRosenberg-Klatt(RK)モデルで近似し、基本周波数、声帯音源波形振幅、声門開放率、音源スペクトル傾斜の各パラメータと単語フォーカスを含んだ対話文のプロミネンスとの関係について考察した。実験結果により、これらのパラメータはプロミネンスとの相関関係が存在することが確認された。

## 2 音声資料

英語女性話者1名と日本語男性話者1名の会議登録の対話文(interactive dialogues)の音声を収録した。日本語98文は英語98文[1]を翻訳したものである。同じ文を単語フォーカスなし、単語フォーカスをシフトした3、4パターンとして収録した。日本語の例: (1) 案内書に 記載されている 口座番号に 振り込んで下さい。(2) 案内書に 記載されている 口座番号に 振り込んで下さい。(3) 案内書に 記載されている 口座番号に 振り込んで下さい。下線の部分はフォーカス単語である。

また、声門の開閉時点をより正確に図るために、日本語に対してはtwo-channel信号(音声と同期したeletroglottograph (EGG)信号)を記録した。

全ての音声データは手操作でphonemeラベルと単語プロミネンス位置を付与した。

## 3 分析手順

ARX分析法[4]を用いて音声信号から周期ごとのRK音源モデルパラメータとホルマントを同時に推定した。標本化周波数は12kHz、量子化bit数は15である。音源モデルパラメータは基本周波数$F_0$、声門波形振幅$AV$、声門開放率$OQ$、音源波形スペクトルの3kHzでの減衰量$TL$(dB)である。

ARX法での$F0$、$OQ$の推定の有効性を確認するために、日本語話者に対して音声信号と同時に録音したEGG信号を微分波形(differentiated EGG)に変換してから周期ごとの$F0$、$OQ$の値を計算した。

## 4 実験結果

ARX法で抽出された周期ごとの音源パラメータとその周期の声門閉鎖点の時刻を保存した。同様にDEGG波形から求めた$OQ$の値とDEGG波形の声門閉鎖点を保存した。次に、phonemeラベル、プロミネンスラベル及び分析結果に基づいて統計処理を行った。

プロミネンスと音源パラメータ($F0$, $AV$, $OQ$, $TL$)の関係を図1～図3に示す。"0"で示すのはプロミネンスなし、"1"はプロミネンスである。図1と図2により、各パラメータにはプロミネンスによる影響があり、$F0$に対する影響が一番大きいということが示された。

## 5 考察

音源振幅パラメータ$AV$は必ずしも音声波形のパワーと一致しない。これには以下の原因があると考えられる。狭母音/i/, /u/に対しては、音声パワーは小さいが、その音源振幅$AV$は音声パワーの大きい母音/a/などの$AV$より大きい場合がある。また、音声パワーを計算するとき、音声信号に含まれている雑音による誤差がある。

「プロミネンスあり」の場合の$OQ$は「プロミネンスなし」の場合の$OQ$より大きく推定されている。これをDEGGの$OQ$(図3)と比較すると、プロミネンスとの関係に対してはARX法とDEGGの結果と同じ傾向があった。これは音声を積極的に発声するとき、$AV$と$OQ$を大きくするためであると考えられる。

さらに、$OQ$の増加が$F0$に依存するかどうかを調べるために、図3(a)の$F0$の重なる部分の$F0$、$OQ$とプロミネンスの関係を調べた。その結果を図3(a')と

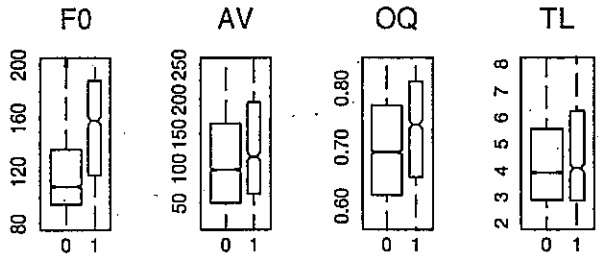


**図1: Relationship between prominence and voice source parameters for the Japanese speaker.**

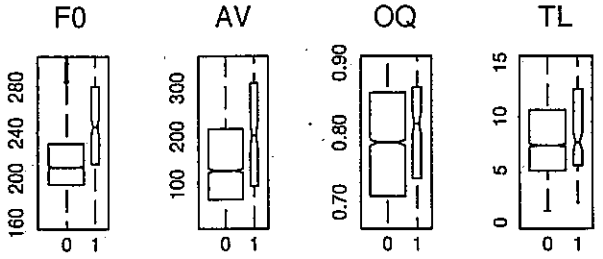


**図2: Relationship between prominence and voice source parameters for the English speaker.**

*On the correlation of prominence and voice source, By Wen Ding and Nick Campbell (ATR Interpreting Telecommunications Research Labs.)

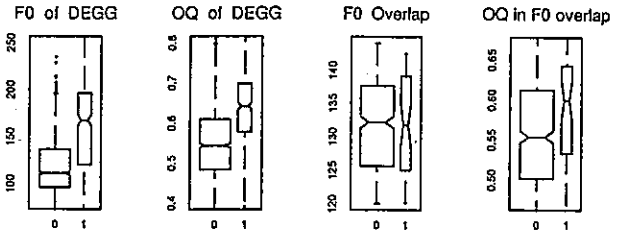


図3: Relationship between prominence and $F0$, $OQ$ with EGG signal for the Japanese speaker.

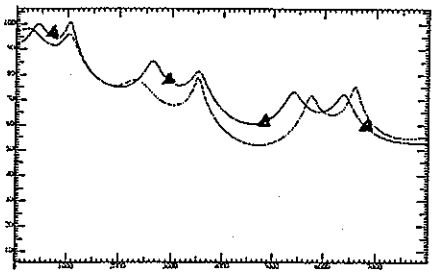

図4: LPC Spectral of /a/ in a prominent and the corresponding non-prominent context. ("- ▲ -" line means prominence.)

(b') に示す。明らかにプロミネンスによる $OQ$ の違いが $F0$ に依存しないことが分かる。

一方、Cummings の研究 [5] では、男性話者の母音を対象として、angry, loud, soft, normal などの音声における音源の違いを調べている。その結果は (1) loud 音声の声門開放区間が normal 音声より長い；(2) angry, loud 音声の声門閉鎖の傾き (glottal closing slope) が normal 音声より大きいということが示されている。前者は loud 音声の OQ の値が大きくなるということを示しており、今回の実験の $OQ$ の推定結果と一致する。後者は loud 音声の $TL$ の値が normal より小さいということを意味している。

しかし、今回の実験の $TL$ の推定値は二人の話者ともプロミネンスありの方がやや大きい。その原因を調べるために、強調した単語と強調していない同じ単語中の母音 /a/ の音声スペクトルを比較した。両者の LPC スペクトルを図 4 に示す。図から、強調された単語のスペクトルの第 2,3,4 ホルマント辺りの成分は大きく、高域成分はそれほど大きくない、スペクトルの傾きは強調なしの場合よりやや大きい、という現象が見られた。また、両者音声の逆フィルタ波形のスペクトルを観察したところ、スペクトルの傾きの差が小さいということが分かった。これは今回の実験において「プロミネンスあり」の $TL$ がそれ以外の部分の $TL$ よりやや大きく推定されている原因の一つであると思われる。したがって、(1) 音源信号スペクトルの違いが主に $AV$、$OQ$ に依存し、$TL$ による影響が小さい; (2) 調音状態によるホルマントの変化はプロミネンスを表現する一つの要因である、と考えられる。

表 1: Prominence prediction linear model for the Japanese speaker.

formula : prom = $F0$*19.88 + $AV$*0.98 + $OQ$*1.34 + $TL$*0.10 - 18.24

F value :

$F_{F_0}(1, 8296) = 718.24$, Pr(F)=0.00
$F_{AV}(1, 8296) = 39.83$, Pr(F)=0.00
$F_{OQ}(1, 8296) = 8.22$, Pr(F)=0.00
$F_{TL}(1, 8296) = 0.24$, Pr(F)=0.62

表 2: Prominence prediction linear model for the English speaker.

formula : prom = $F0$*29.57 + $AV$*1.95 + $OQ$*1.57 + $TL$*0.75 - 30.3

F value :

$F_{F_0}(1, 3454) = 48.52$, Pr(F)=0.00
$F_{AV}(1, 3454) = 13.52$, Pr(F)=0.00
$F_{OQ}(1, 8296) = 1.64$, Pr(F)=0.20
$F_{TL}(1, 8296) = 1.07$, Pr(F)=0.30

次に、プロミネンスと音源パラメータの相関関係及び各パラメータの貢献度を調べた。音源パラメータをアイテムとして線形予測モデル (generalized linear model) を使って予測式を求めた。まず、予備実験を行ない、正規分布に近似するように、パラメータを以下のように変換した:

$F0$=log($F0$), $AV$=sqrt($AV$),
$OQ$=exp($OQ$), $TL$=log($TL$).

そして、各パラメータを最大値で (0,1) の間に正規化した。日本語話者と英語話者の線形予測式と F value をそれぞれ表 1 と表 2 に示す。予測式による音源パラメータの貢献度は $F_0$, $AV$, $OQ$ と $TL$ の順である。

## 6 むすび

本稿は単語フォーカスのレベルで、プロミネンス文を発声するときの声門音源パラメータの変化を調べた。日本語と英語話者各 1 名で分析した結果にプロミネンスによる $F0$, $AV$, $OQ$ の増加が確認されたが、$TL$ の推定値には大きな変化は現れなかった。また、今回考察していない phoneme 持続長もプロミネンスに対する重要な要因の一つである。今後は、これらの関係に基づいてプロミネンスの検知、生成などの応用に利用していきたい。

参考文献

[1] N. Campbell, ICPhS 95, Stockholm, pp.676-679 (1995).
[2] 阿部, 音響学会誌, 51, pp.929-937(1994).
[3] 前川, 音講論, 1-P-22, (1996/3).
[4] W. Ding, et al., IEICE Trans. Inf. & Syst., Vol. E78-D, pp.738-743(1995).
[5] K. E. Cummings and M. A. Clements, J. Acoust. Soc. Am., 98, pp.88-98(1995).