

音声のエネルギー曲線における個人性への伝達系の影響*

◎網野加苗, 荒井隆行 (上智大・理工)

1 はじめに

音声の個人性は、音声信号中の様々な側面に現れる。例えば、話者内の平均的な声の高さは声帯の長さや厚さ、身長などと相関があり [1]、口腔、鼻腔、咽頭腔などの共鳴腔の形状の違いは、主に音声の周波数特性に現れる [2]。音声の個人性には、身体的な話者の特徴の他に、言語獲得過程で身に付けた話し方の癖や地域・社会方言が挙げられ、それによる発声発語器官の運動の違いにも話者の個人性が現れると考えられる。

著者らによる前研究 [3-5] では、計10名の男性話者の CV (子音-母音) 単音節発話におけるエネルギー遷移曲線をプロットし、その形状に個人差があることを示した。これらのエネルギー曲線は、子音から後続母音への調音運動を反映しており、子音による立上り時間の違いも表れる [6]。エネルギー曲線における話者間の違いは特に軟口蓋制御を含む鼻子音部に見られ、エネルギーの計算が可能な環境では、話者識別までは困難かもしれないが、話者の分類への活用は期待できる。

本稿では、エネルギー曲線に現れる話者の個人性に標準化周波数、符号化形式および残響などの伝達系が与える影響を調べた。また、前研究の単音節発話に加え、単語発話の分析も行った。

2 音響分析

2.1 分析条件

前研究 [3-5] に倣い、エネルギーの時間変化を分析パラメータとした。フレーム長 30 ms、シフト長 10 ms で、フレームごとのエネルギーを計算した。前研究 [3-5] では発話内の最大値によってエネルギーを正規化したものをそのまま利用したが、本研

究では人間の聴覚特性を考慮して [7]、正規化した値をさらに 1/3 乗したものをプロットした。

2.2 音声資料

分析対象の音声資料は、電子協日本語共通音声データ [8] より、成人男性 4 名分の発話を用いて作成した。Table 1 に一覧を示す。単音節は 6 種類の頭子音と日本語音素配列で許される後続母音から成る 24 語、単語発話には 10 語を選択した。各種各話者につき 3 回の繰り返し発話を用いた。

全ての発話に関して、標準化周波数が 16 kHz および 8 kHz のリニア PCM に加え、ITU-T の携帯電話通信規格 (G.729) の音声を作成し、分析対象とした。また、小部屋を想定した残響下の発話として、16 kHz のリニア PCM に残響時間が約 0.1 秒のインパルス応答を畳み込んだ音声も作成した。インパルス応答は、白色雑音に指数減衰関数をかけることによって人工的に作成し、残響時間の計算には、中心周波数 500 Hz, 1000 Hz, 2000 Hz のバンドから求まる EDT (Early Decay Time) の平均値を用いた。

3 結果

3.1 単音節発話

4 名の話者が発話した CV 単音節のエネルギー曲線を Fig. 1 に示す。上から 3 段は標準化周波数が 16 kHz のリニア PCM の単音節の曲線で、順に子音部が (a) 閉鎖音、(b) 摩擦音、(c) 鼻音の音節である。4 段目以下は、(c) を (d) 8 kHz にダウンサンプルしたもの、(e) G.729 に変換したもの、(f) 残響をかけたものを示している。黒の実線が有声音 (鼻音では両唇音)、赤の点線が無声音 (鼻音では歯茎音) である。

Table 1 Speech materials

CV Monosyllables	Consonants	/d/ /t/ /z/ /s/ /m/ and /n/
	Vowels	/e/ /a/ and /o/ for the stops, /e/ /a/ /o/ and /u/ for the fricatives, and /i/ /e/ /a/ /o/ and /u/ for the nasals.
Words	2 or 3 moraic	/migi/ /ido/ /do:zo/ /çidari/ and /ha:zime/
	4 moraic and more	/çturikomi/ /zandaka/ /katakana/ /bu:ŋmatsuu/ and /mo:itçido/

* Effects of the transmission systems on speaker individualities contained in the energy contours of speech, by AMINO, Kanae and ARAI, Takayuki (Sophia University).

目視によっても分かるように、リニア PCM の音声では、話者による曲線形状の違いが見られ、標準化周波数を変化させてもその違いは保たれている。特に曲線の左側の子音部は話者内で一貫しており、閉鎖音と摩擦音に関しては、有声音（黒実線）より無声音（赤点線）でばらつきが少ないことが分かった。

残響を付加した音声でも、若干形状は変化するものの、話者間の違いは保たれている。一方、非圧縮のリニア PCM から G.729 に変換した場合、各話者に元々見られた特徴はあまり保たれないことが分かった。

3.2 単語発話

次に単語発話のエネルギー曲線の例を Fig.2 および Fig.3 に示す。それぞれ「右 /migi/」「残高 /zandaka/」という発話であり、上から順に標準化周波数が 16 kHz と 8 kHz のリニア PCM, G.729, 残響下のデータである。同じ語を発話しても、話者によって形状が大きく異なり、どのデータ形式においても話者間の違いが保たれている。この違いを確認するために、話者間および話者内で曲線間の相関を求めた。相関係数にはピアソンの積率相関係数を用い、発話の始点をそろえて各単語の全組合せについて計算した。

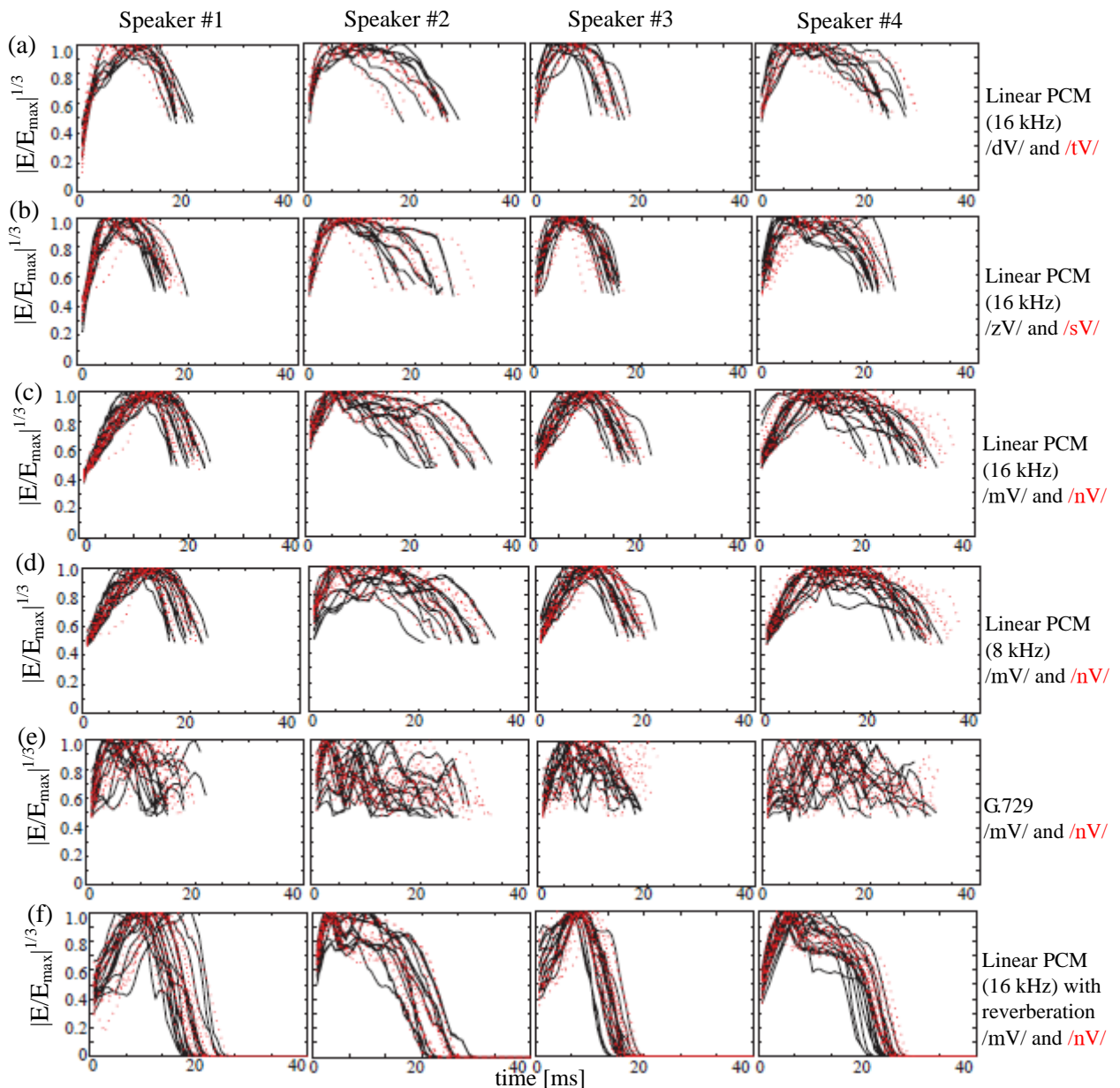


Fig. 1 Energy contours for the monosyllables uttered by the four speakers; the syllables contain: (a) a stop, (b) a fricative, or (c)-(f) a nasal; black solid lines show the syllables with /d/ /z/ or /m/, and red dotted lines show those with /s/ or /n/; speech data formats are: (a)-(c) linear PCM sampled at 16 kHz, (d) 8 kHz, (e) G.729 or (f) linear PCM with reverberation.

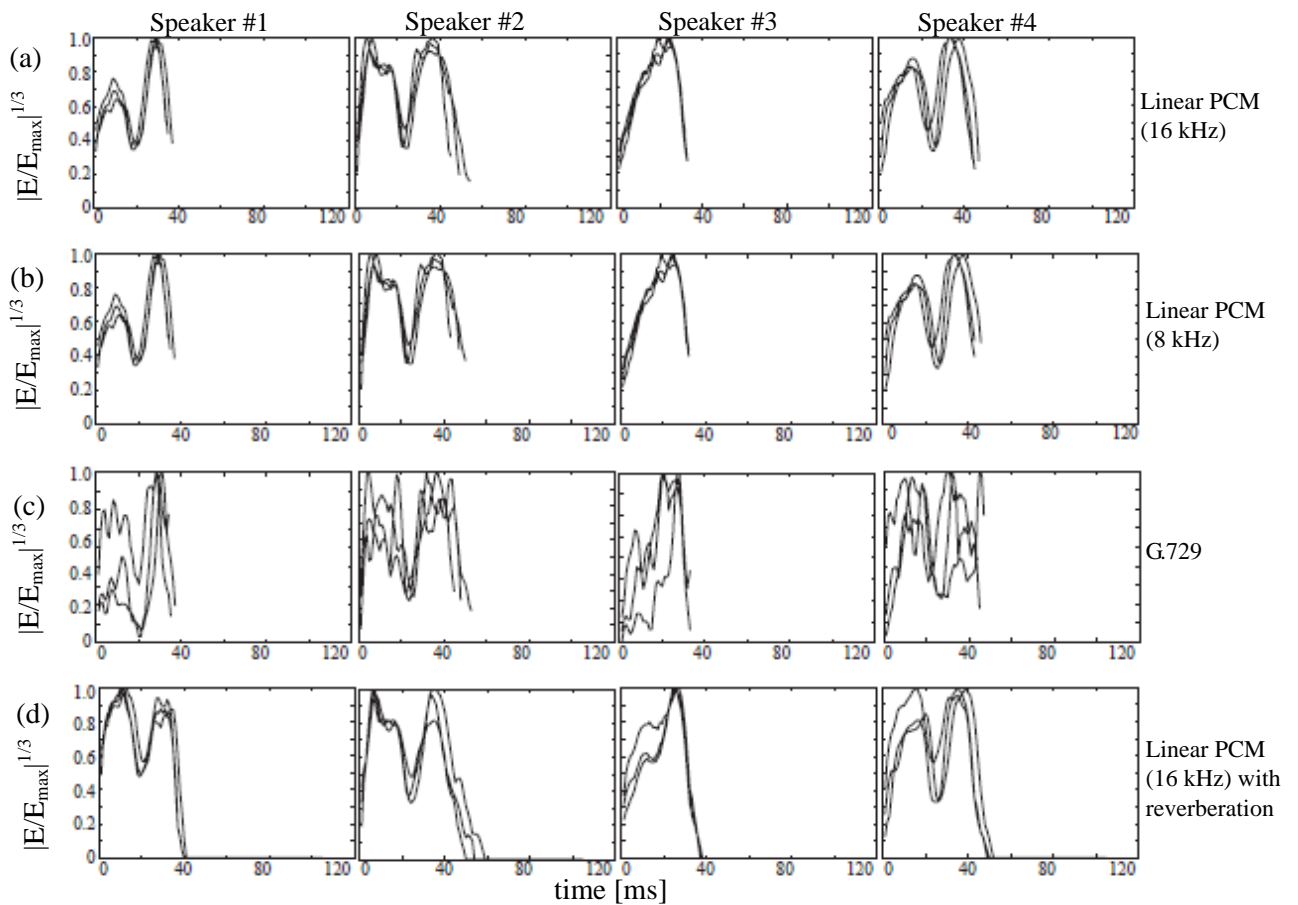


Fig. 2 Energy contours for the word /migi/ (right) uttered by the four speakers; speech data formats are: linear PCM sampled at 16 kHz (a), 8 kHz (b), G.729 (c) or linear PCM with reverberation (d).

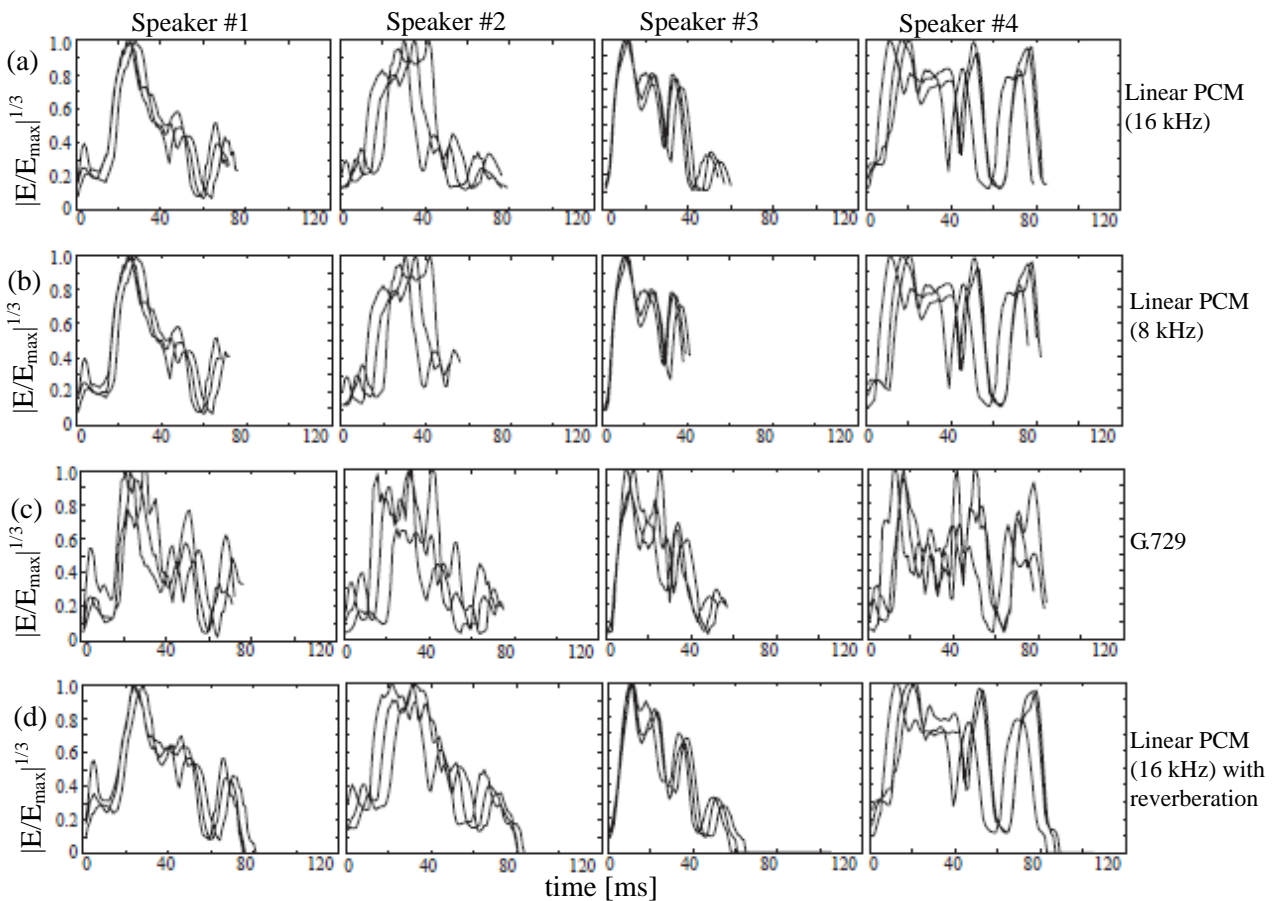


Fig. 3 Energy contours for the word /zandaka/ (balance; in a bank account) uttered by the four speakers; speech data formats are: (a) linear PCM sampled at 16 kHz, (b) the same at 8 kHz, (c) G.729 or (d) linear PCM with reverberation.

Table 2 Average intra- and inter-speaker correlations of the energy contours

Correlation coefficients		Linear PCM (16 kHz)	Linear PCM (8 kHz)	G.729	Reverberation
/migi/	Intra-speaker	0.837	0.854	0.520	0.857
	Inter-speaker	0.112	0.143	0.037	0.023
/zandaka/	Intra-speaker	0.702	0.686	0.560	0.726
	Inter-speaker	0.371	0.317	0.212	0.398

相関係数の平均値を一部 Table 2 に示す。話者の識別や分類を目的とする場合、話者内相関が高く、話者間相関が低いことが望ましい。Table 2 から、2つの語を比較すると /migi/の方が目的に適っていることが分かる。紙面の都合上、他の語の相関係数は掲載しないが、分析を行った10語のうち、/migi/ /ido:/ /do:zo/ /çidari/ /hazime/ /bunmatsu/ /mo:içido/ では話者内相関が0.8以上であり、/migi/ /çurikomi/ /katakana/ では話者間相関が0.15未満であった。

音声の形式別には、ダウンサンプリングもしくは短い残響を加えた場合には結果に影響を与えないが、G.729に変換して圧縮を行うと、話者内相関が著しく低下した。同時に話者間相関も低下したが、話者内相関の低下を補完できるほどではなかった。

4 まとめと考察

4名の男性話者の発話に関して、エネルギー曲線に見られる話者の特徴への伝達系の影響を調べた。標本化周波数が16 kHzのリニアPCMの音声では、CV単音節と単語発話の両方において話者による曲線形状の違いが見られたが、G.729に変換した音声では話者間の違いがあまり保たれないことが分かった。本研究の条件下では、エネルギー曲線に見られる話者間の違いは残響に対して比較的頑健であったが、残響時間を変化させた場合については今後調べる必要がある。

CV単音節では、母音部より子音部で話者内のばらつきが少なかった。これはエネルギーの立上りというパラメータそのものが子音の弁別に寄与していることから妥当な結果だと言える。さらに、子音が無声音の場合に有声音と比較して話者内で一貫していた点については、声帯振動を伴わない無声音では、声帯筋の弛緩が起らず、口腔内圧を一定に保ちやすいことが理由として考えられる。

単語発話では、話者内で若干の時間的なずれ

はあるものの、エネルギーのピークや谷の相対的な位置関係は一貫しており、話者間では大きく異なっていた。話者によっては特定の音素の発音に一貫した特徴が見られることもあり、例えば Fig. 2 に示した /migi/ という語では、Speaker #3 のみがガ行音を鼻濁音として発話していることが単峰性の曲線からも見て取れる。

本研究では、曲線間の相関係数を求める際に、各繰り返し発話における発話速度の違いや発話間の動的な伸縮、発話開始のタイミングの違いなどを考慮していないため、話者内相関が実際よりも低くなった可能性がある。時間的なアラインメントについては、さらに検討の余地がある。

エネルギー曲線に現れる話者の個人性は、調音動作における話者特性を反映しているが、音声の個人性知覚を考える際、聞き手がこれを利用していかどうかは明らかでないため、今後実験によって確かめる必要がある。また、エネルギー曲線は呼吸や発声の個人性も反映すると考えられるため、今後は空気力学的な観点からも話者間の違いを調査したい。

謝辞

この研究は文部科学省私立大学学術研究高度化推進事業 上智大学オープン・リサーチ・センター「人間情報科学研究プロジェクト」の助成を得た。

参考文献

- [1] Bachorowski and Owren, JASA, 106(2), 1054-1063, 1999.
- [2] Ladefoged, "A Course in Phonetics (4th ed)," Heinle, 2000.
- [3] 網野, 荒井, 音講論(秋), 359-360, 2008.
- [4] Amino and Arai, Proc. Interspeech 2008, 1917-1920, 2008.
- [5] Amino and Arai, to be published in Forensic Science International, 2009.
- [6] Raphael *et al.*, "Speech Science Primer: Physiology, Acoustics, and Perception of Speech (4th ed.)," Lippincott Williams & Wilkins, 2002.
- [7] Hermansky, JASA, 87(4), 1738-1752, 1990.
- [8] 電子協日本語共通音声データ
http://www.sunrisemusic.co.jp/dataBase/fl/voicebase01_fl.html