3—2—1　音声中の話者情報を担う変調周波数帯の調査 *

○金寺 登　高野友紀子 (石川高専)　荒井隆行　高橋真保呂 (上智大・理工)

1. はじめに

対数スペクトルの時間軌跡またはケプストラムの時間軌跡のフーリエ変換は変調スペクトルと呼ばれている。言語情報を担う変調周波数帯に関して、荒井ら[1]は、知覚実験により、明瞭度を保持するために必要なほとんどの情報が1~16 Hzの変調周波数帯に存在することが明らかにした。また、金寺ら[2, 3]は、ASR(Automatic Speech Recognition)実験により、言語情報のほとんどが1~16 Hz (特に2~10Hz)の変調周波数帯存在することを確認した。さらに言語情報を担う変調周波数帯を効率よく表現することで、認識性能が向上することを確認した[4]。

一方、話者情報を担う変調周波数帯に関して、Vurenら[5]は、話者識別実験によって、0.1~10 Hzに重要な話者情報が含まれると報告している。

本報告では、どのような変調周波数帯が音声中の話者情報を抽出する上で、どの程度重要であるかを知覚実験により調査した結果について報告する。

2. 実験条件

2.1 分析条件

様々な変調周波数成分を持つ音声を生成し、話者識別知覚実験を行うために、信号処理ツール[6]を用いて、図1に示す分析合成を行った。まず原音声より、窓長25msのブラックマン窓を用いて12次のMel-Frequency Cepstrum Coefficients(MFCC)を5ms毎に求めた。次にMFCCの時間軌跡を1023点のFIRフィルタによりフィルタリングし、ある範囲の変調変調周波数帯の成分のみを抽出した。さらにフィルタリング後のMFCCとピッチを用いて音声を合成し、最後に文全体の音声の大きさを正規化した。

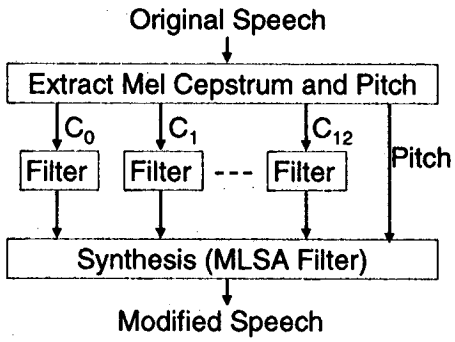

図 1. 分析合成方法

2.2 話者識別知覚実験

まず話者識別の対象音声に前節の分析合成を施し、特定の変調周波数帯のみを含む提示音声を作成した。使用した変調周波数帯は、0 Hz, 0.25, 0.5 1, 2, 4, 8,

表1. 話者識別率 [%]

f_L [Hz]	f_U [Hz]							
	0.25	0.5	1	2	4	8	16	f_N
0	1.3	2.5	2.5	20.0	45.0	60.0	76.3	82.5
0.25		5.0	5.0	18.8	36.3	58.8	63.8	72.5
0.5			12.5	16.3	36.3	55.0	58.8	63.8
1				13.8	22.5	41.3	56.3	67.5
2					22.5	36.3	38.8	46.3
4						16.3	13.8	32.5
8							16.3	3.8
16								6.3

* Investigation of components of the modulation-frequecy bands carring speaker infomation in speech.

By Noboru Kanedera, Yukiko Takano (Ishikawa National College of Technology), Takayuki Arai, and Mahoro Takahashi (Sophia University)

16, f_N(ナイキスト周波数) を遮断周波数とする 36 種類である。

識別対象話者は 石川高専 電子情報工学科の教官 5 名、被験者は 5 名の教官の声を日頃良く聞いている同学科の学生 16 名 (男女各 8 名) とした。提示文には「あらゆる現実をすべて自分のほうへねじまげたのだ。」を用いた。

実験の前に 5 名の教官の声を別の文「青い植木鉢」で 2 度、識別対象となる話者の音声を確認してもらった。次に、各被験者に 180 文 (36 種類×5 名分) を提示した。提示順序は、判別が難しいと判断されるフィルタリング条件の順とし、各条件ごとに乱数で提示話者順を決定した。1 文を聴取する度にいずれの話者であるか または わからないかを番号で答えるよう指示した。

3. 実験結果

表 1 に、種々の変調周波数帯に対する話者識別率を示す。表中の f_L は低域遮断変調周波数、f_U は高域遮断変調周波数を表している。これらの話者識別率を基に文献[2, 3]の方法で各変調周波数帯の話者識別に対する貢献度を 95%信頼区間付きで求めたものを図 2に示す。図中の貢献度は、対応する変調周波数帯を含めることで、話者識別誤り率が 1/(貢献度) になることを表している。従って、貢献度が 1 より大きければ大きほど対応する変調周波数帯に多くの話者情報が含まれていることになる。

図 2より、2 Hz ～ 8 Hz に多くの話者情報が含まれていることがわかった。この範囲の変調周波数帯は、言語情報が多く含まれる範囲と一致している。文献[5]の話者自動識別実験では、0.5 Hz ～ 8 Hz に多くの話者情報が含まれているのに対し、0.125 Hz 以下はかえって話者識別性能を低下させると報告している。文献[5]と比較して、今回の知覚実験では、0 Hz ～ 0.25Hz の貢献度が高く、0.25 Hz ～ 1 Hz の貢献度が低い結果となった。

4. まとめ

音声中のどの変調周波数帯にどの程度の話者情報が含まれているかを知覚実験により調査した。その結果、今回の実験環境では 2 Hz ～ 8 Hz の変調周波数帯に多くの話者情報が含まれるという結果を得た。

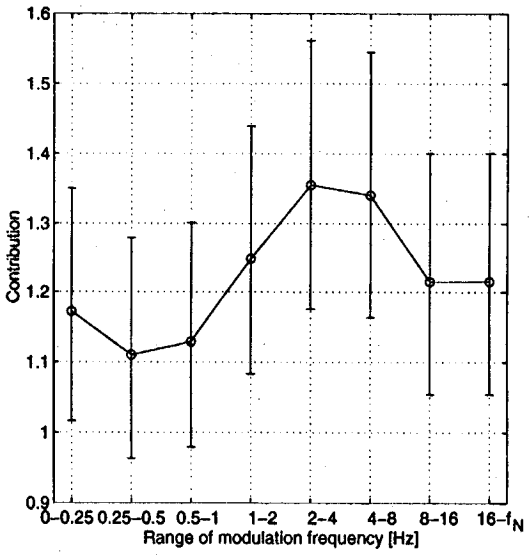

図 2. 各変調周波数帯の話者識別に対する貢献度

謝辞

音声信号処理ツールキットを公開して下さった名古屋工業大学の徳田恵一先生をはじめ、開発に携わった多くの方々に感謝いたします。

参考文献

[1] T. Arai, M. Pavel, H. Hermansky and C. Avendano, "Syllable intelligibility for temporally filtered LPC cepstral trajectories," The Journal of the Acoustical Society of America, Vol. 105, No. 5, pp. 2783 - 2791, (1999.5).

[2] N. Kanedera, T. Arai, H. Hermansky, and M. Pavel, "On the relative importance of various components of the modulation spectrum for automatic speech recognition," Speech Communication, Vol.28, pp.43 - 55 (1999.5).

[3] 金寺 登, 荒井隆行, 船田哲男, "音声中の言語情報を担う変調スペクトル特性の検討," 音学講論, pp.3 - 4, (1999.3).

[4] 金寺 登, 荒井隆行, 船田哲男, "複数の変調スペクトル解像度を用いた音声認識の耐雑音性," SP98-51, pp.45-52 (1998.7).

[5] S. van Vuuren and H. Hermansky, "On the importance of components of the modulation spectrum for speaker verification," In Proc. of the ICSLP, Sydney Australia, (1998.11).

[6] 徳田恵一ほか, "音声信号処理ツールキット," http://kt-lab.ics.nitech.ac.jp/~tokuda/SPTK/