2—10—20 話者情報を担う変調周波数帯の調査 — 雑音駆動音声を用いた話者識別知覚実験 — *

◎高橋真保呂 荒井隆行 (上智大・理工)
金寺 登 高野友紀子 (石川高専) 村原 雄二 (上智大・理工)

1. はじめに

スペクトルや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を明らかにした。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 2~10 Hz) の変調周波数帯に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も確認している[4]。

一方、話者情報を担う変調周波数帯に関して、Van Vuuren ら[5]は自動話者認識実験によって 0.1~10 Hz に重要な話者情報が含まれていると報告している。人間が音声のどういった特徴を用いて話者を判断しているかを知ることは、自動認識への応用につながる。そのため金寺ら[6]は知覚実験を行い、MFCC (mel-frequency cepstrum coefficients) とピッチを用いた分析合成の結果、2~8 Hz の変調周波数帯に多くの話者情報が含まれていると報告している。そこでは駆動信号として元の音声から得られたピッチ情報を用いているため、ピッチ情報を含んだまま分析合成が行われている。ピッチは有用な物理的特徴の 1 つでありピッチ情報だけでもかなりの話者情報が含まれていることから[7]、実験結果にはピッチ情報による影響も含まれている可能性があると考えられる。

そこで本研究では、なるべくピッチ情報などの影響を受けずに話者認識において重要な変調周波数帯を調べるため、駆動信号としてもとの音声のピッチ情報を用いず、白色雑音のみを駆動源として音声を再合成し知覚実験を行った。

2. 実験条件

2.1 分析合成

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[8]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 MFCC を 5ms 毎に求めた。次に MFCC の時間軌跡を 511 点の FIR フィルタにより時間方向にフィルタリングし、特定の範囲の変調周波数成分のみを抽出した。さらに、フィルタリング後の MFCC に対して駆動信号として白色雑音を用いて合成した。最後に、文全体の音声の大きさを正規化した。

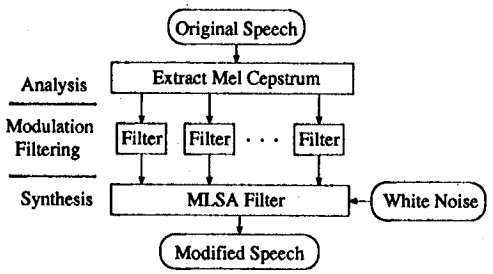

図 1. Block diagram of the speech-processing system.

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だけを持つ提示音声を作成した。使用した変調周波数帯は、0 , 0.5, 1, 2, 4, 8, 16 Hz, f_N(ナイキスト周波数) を遮断周波数とする 28 種類である。

識別対象話者は 上智大学 理工学部 電気・電子工学科の教員 5 名、被験者は 5 名の教員の声を日頃良く聞いている同学科の学生 28 名とした。

提示文には ATR 音声コーパス[9]から 「あらゆる現実をすべて自分のほうへねじまげたのだ。」を含む 12 文を選択し使用した。

ピッチ情報のない白色雑音駆動音声はささやき声のように聞こえ、慣れないと判断が難しい。そこで以下のような訓練をした。まず、識別対象となる 5 名の話者の原音声の確認を行った。次に 5 人の話者の原音声を確実に聞き分けることができているかをテストした。一方、フィルタリングせずに分析合成した 5 人の教員の白色雑音駆動音声を被験者が確実に判断できるまで訓練した。このとき原音声の確認、テスト、白色雑音駆動音声の訓練にはそれぞれ異なる音声サンプルを使用した。このような訓練を経て、訓練に合格したものだけを被験者として採用した。

実験では学習効果を避けるため、一人の被験者が同じ話者の同じ文を一条件だけしか聞かないように各被験者用の刺激音のセットをランダムに抽出した。

被験者には、1 文を聴取する度 5 人の中から 1 人の話者を必ず選択するよう指示した。また実験中は、

---

* Investigation of components of the modulation-frequency bands carring speaker infomation in speech.
— Human specker identification experiments with noise-excited speech —

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

フィルタリングしていない5人の話者の白色雑音駆動音声を常に確認できるように別の選択画面も同時に表示した。

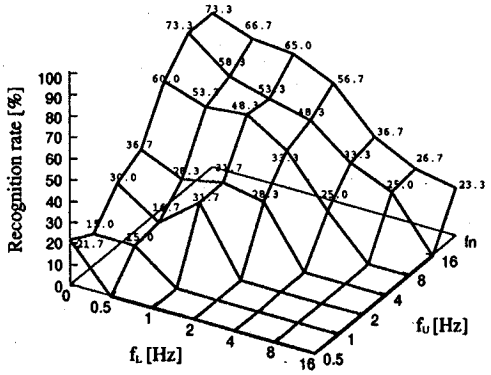

図 2. Recognition results for the band-passed time trajectories.

3. 実験結果

図2に、種々の変調周波数帯に対する話者識別率を示す。図中の 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3に示す。

図3を見ると、2~8 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[1]。また、文献[6]のピッチ情報を含んだまま分析合成した音声による話者識別知覚実験では2~8 Hz に、文献[5]の話者自動認識実験では0.5~8 Hz に多くの話者情報が含まれていると報告しているが、それらの結果とも一致している。本実験ではピッチ情報などを用いずに白色雑音駆動で音声を合成した。それにもかかわらず類似した結果が得られていることから、MFCC の時間変化にも話者情報が含まれていることが裏付けされた。

また、本実験の結果では 0.5~1 Hz の貢献度が低いのに対し0~0.5 Hz の貢献度が高くなっている。文献[6]の話者識別知覚実験では 0~ 0.25 Hz の貢献度が高く、0.25~1 Hz の貢献度が低い結果となっており、本実験の結果と類似している。一方、文献[5]の話者自動認識実験では0.125 Hz 以下は話者認識性能を低下させると報告している。話者識別知覚実験で0 Hz付近の変調周波数帯が重要であることから、人間の話者識別の拠り所として、フォルマントなどの静的な成分も重要であると推測される。

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雑音駆動音声に対する知覚実験により調査した。その結果、今回の実験環境ではピッチを含めた実験[6]と同様に 2~8 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。また、人間の話者識別の拠り所として、静的な成分も重要であると推測された。

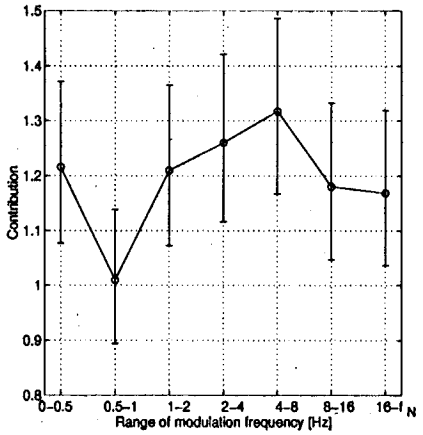

図 3. Contributions to recognition performance with 95% confidence intervals.

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、また実験に協力して下さった上智大学の教員ならびに学生の皆様に感謝いたします。

参考文献

[1] T. Arai, M. Pavel, H. Hermansky and C. Avendano, "Syllable intelligibility for temporally filtered LPC cepstral trajectories," J. Acoust. Soc. Am., Vol. 105, No. 5, pp. 2783-2791, 1999.

[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.

[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.

[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," 信学技報, SP98-51, pp.45-52, 1998.

[5] S. van Vuuren and H. Hermansky, "On the importance of components of the modulation spectrum for speaker verification," Proc. ICSLP, Vol.7, pp.3205-3208, Sydney Australia, 1998.

[6] 金寺 登, 高野友紀子, 荒井隆行, 高橋真保呂, "音声中の話者情報を担う変調周波数帯の調査," 音学講論, pp.361-362, Sep. 1999.

[7] 古井貞熙,"音響・音声工学," 近代科学社,1992.

[8] 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/

[9] 桑原尚夫, 匂坂芳典, 武田一哉, 阿部匡伸, "研究用 ATR 日本語データベースの作成," *ATR Technical Report*, 1989.