3B－2

コミュニケーションにおける引き込み —音声のON-OFFに基づく頷き反応モデル—

○渡辺 富夫・大久保 雅史
岡山県立大学情報工学部

1．はじめに

対話においては対話者相互に音声と動作･表情が引き込み、コミュニケーションを円滑にしている。とくに頷き反応はノンバーバル情報の調整子と呼ばれ、インタラクションに重要な役割を果たしている。著者らは既に音声のON-OFFに基づく頷き反応モデルを提案し、システムを試作して有効性を示してきた[1]。このモデルは、呼気段落区分での音声のON-OFFパターンを特徴づける時間率の線形結合で頷きを区間推定するマクロ層と、その区間で頷きの開始点を音声のON-OFF時系列の線形結合で推定するミクロ層からなる階層モデルで、いずれの階層もMA(Moving-Average)モデルを用いている。しかし、二つのサブシステムからなる線形フィードバックシステムの解析では、MAモデルよりも自己にも基づくARMA（Auto-Regressive Moving-Average）モデルの有効性が指摘されている[2]。本論文では、ミクロ層を対象にして、MAモデルとARMAモデルを比較検討し、ヒューマンインタラクションにおけるMAモデルの妥当性を示している。

2．対面コミュニケーション実験

対話者は予め話し手と聞き手が設定され、一方から他方への話の伝達を対象とした。実験は、一対一でテーブルを囲んでまず対面で3分間、次に対話者の視界を衝立で遮った非対面で3分間を2セット繰り返して行い、対話者の各々の表情を2台のカメラで画面を2分割して音声とともに収録した。話し手は学部4年男性3人で、聞き手は学部3年男性1人である。

音声については、サンプリング周期1msで50msの平均雑音レベルに12dBを加えた値をON(1)とOFF(0)の臨界値として、50msごとに150msのフィルイン（150ms以下のOFFをONに置換する操作）を施し、音声のON-OFFのデータとした。頷きについては、ビデオフレーム単位（1/30秒）で目視により頷き開始から終了までをON、それ以外をOFFとして50msの時間単位で2値化した。これら時系列のデータ間の関係については、以下の相互相関関数C(τ)で評価した。

$$C(\tau)=\frac{\sum_{i=1}^{n-\tau}\{x(i)-\mu_x\}\{y(i+\tau)-\mu_y\}}{\sqrt{\sum_{i=1}^{n}\{x(i)-\mu_x\}^2}\sqrt{\sum_{i=1}^{n}\{y(i)-\mu_y\}^2}}$$

μ_x：音声x(i)の平均値，μ_y：頷きy(i)の平均値
n：データ数　　τ：時間遅れ

分析対象時間はコミュニケーション中央時間の120秒間（データ数n=2,400）、50 msのずれ時間で最大5秒に設定した。

対面及び非対面での話し手の音声のON-OFFに対する聞き手の頷きの分析結果の一例を図1に示す。対面では-0.1秒に顕著な負の相関がみられる。これは、対面では話の区切りを予測して頷き、両者が引き込んでいることを示している。一方、非対面では頷き回数が少ないために両者に有意な相関はみられず、引き込みが起きにくいことがわかる。

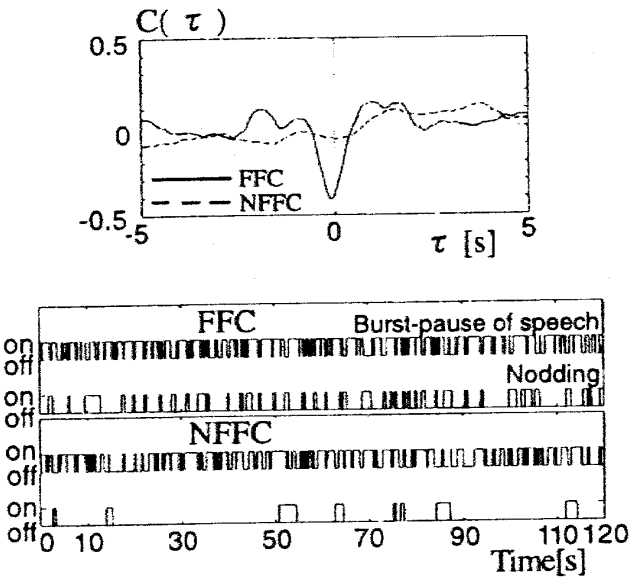

図1 対面(FFC)及び非対面(NFFC)での話し手の音声のON-OFFに対する聞き手の頷きの分析結果

3．音声のON-OFFに基づく頷き反応モデル

分析結果より、話の区切りを検出してから頷かせたのでは反応が遅く、頷きを予測する必要がある。ここでは、頷きx(i)を音声y(i)の線形結合で予測するMAモデルと自己の推定した頷き$\hat{y}$(i)にも基づくARMAモデルを用いて頷き反応の推定を行った。

Entrainment in Communication — A Voice-Nodding Interaction Model

Tomio Watanabe and Masashi Okubo
Faculty of Computer Science and System Engineering, Okayama Prefectural University
111 Kuboki, Soja, Okayama, 719-1197 JAPAN

$$\hat{y}(i) = \sum_{j=1}^{J} a(j)x(i+1-j) - \sum_{k=2}^{K} b(k)\hat{y}(i+1-k)$$

MA モデルでは b(k)=0

a(j),b(k):予測係数　　K,L:予測次数

各モデルの予測次数については、AIC (Akaike's Information Criterion)を基準に選定した。測定値の誤差の絶対的な大きさが未知の場合には、

AIC=n・log s + 2p

n: データ数　s: 残差二乗和　　p : パラメータ数

で表される。図1の対面について、AICをMAモデルとARMAモデル（J=Kとした）で最小二乗解析により算出した結果を図2に示す。MAモデルでは予測次数J=40で、ARMAでは80（J=K=40）でAICが最小であり、この次数が最適であると判断される。この次数で図1について各モデルの予測係数を算出し、頷き（平均頷き時間0.85秒に設定）を推定した結果を図3に示す。またこの予測係数を用いて自己（W2:話し手Wの対面2セット目）及びW1と他の2組の対面での頷きの推定値$\hat{y}$(i)と実測値y(i)との相互相関関の最大値C(τ)max（時間遅れ0.5秒以内）を図4に示す。ここで時間率はスピーチ区間でのON区間の占める割合で、発話速度と密接な関係があり、典型的な音声のON-OFFパラメータである[3]。MAモデルとARMAモデルで自己W2の頷きの推定結果には差異はなく、W1と他の2組(R,M)に適用した場合にはMAモデルの方が推定がよいことがわかる。この傾向は他の組に基づいても同様であった。

線形フィードバックシステムであれば、理論的にはARMAモデルの方が推定が良いはずであるが、結果は逆でMAモデルの方がロバスト性が高く、著者らが提案した音声-頷き反応モデルでMAモデルを用いた妥当性を示している。もちろんこのMAモデルだけで頷きを推定したのでは頻繁に頷きが生起するので、ON区間とOFF区間を単位ユニットとしてユニットでの時間率のMAモデルにより、各ユニットに頷きがあるか否かを推定するマクロ層を設けることで、頷きの精度が高められる。

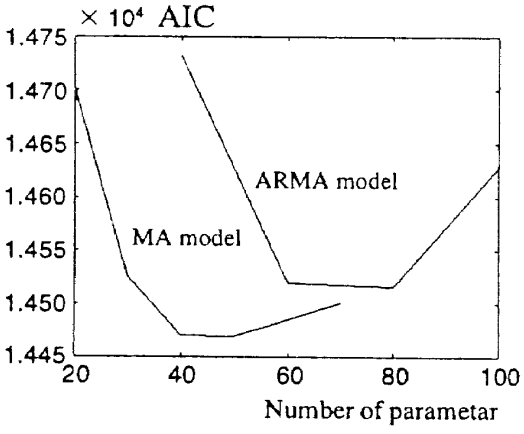

図2　各モデルに対するAIC

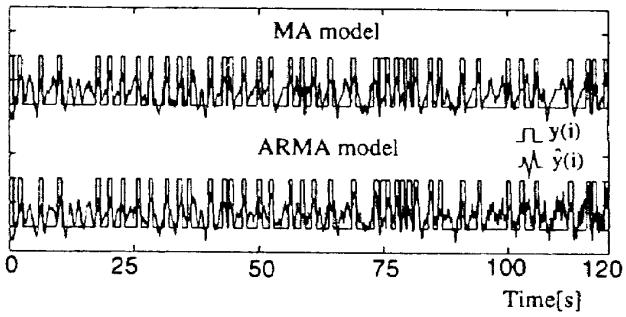

図3　各モデルでの頷きの推定結果

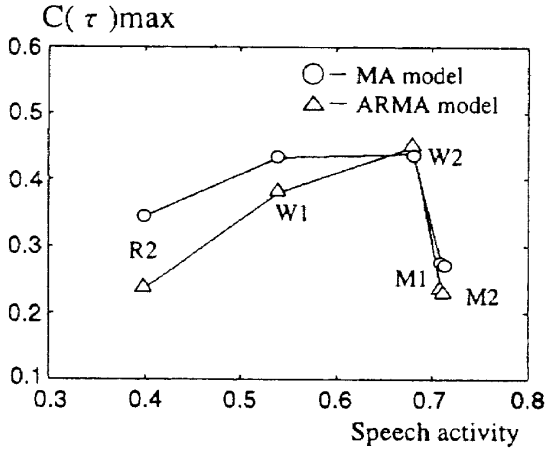

図4　自己及び他の2組の対面での頷きの推定値と実測値とのC(τ)max

4．おわりに

本論文では、対面コミュニケーションにおける呼気段落区分での音声のON-OFFに基づく頷き反応の推定モデルとしてMAモデルとARMAモデルを選定し、比較検討した。まず対面及び非対面での話し手の音声のON-OFFに対する聞き手の頷きを分析し、対面では話の区切りを予測して頷いていることを示した。次に各モデルでの最適予測次数をAICに基づいて決定し、この予測次数でそれ自身及び他の対面での頷きを推定した結果、MAモデルの方がロバスト性が高かった。これは、ヒューマンインタラクションにおけるMAモデルの妥当性を示すもので、著者らの提案する音声-頷き反応モデルの有効性を示している。

参考文献

[1] T.Watanabe and A.Higuchi: Facial Expression Graphics Feedback for Improving the Smoothness of Human Speech Input to Computers, Advances in Human Factors/Ergonomics, Vol.18A, pp.491-497 (1991).

[2]赤池弘次、中川東一郎：ダイナミックシステムの統計的解析と制御、サイエンス社 (1972)

[3] T.Watanabe: The Adaptation of Machine Conversational Speed to Speaker Utterance Speed in Human-Machine Communication, IEEE Trans. on Systems, Man, and Cybernetics, SMC20, 2, pp.502-507 (1990).