昭和62年度厚生省心身障害研究
「家庭保健と小児の成長・発達に関する総合的研究」

ヒューマン・インタフェースにおける音声対話速度の適応化

（分担研究：相互作用と乳幼児の心理行動発達に
関する基礎的研究）

渡辺富夫 *

要約　個人の発声速度に対応して機械がその個人に最適な音声出力速度で応答できれば、円滑な音声対話が図られると期待されよう。本報告では、個人の発声速度を時間率［平均ＯＮ区間／（平均ＯＮ区間＋平均ＯＦＦ区間）］に基づいて推定し、その発声速度に音声応答速度を適応化させる手法を提案している。

見出し語：ヒューマン・インタフェース，発声速度，適応化

1. 緒言

人間機械間の音声対話においては、人間は機械へ音声入力する話し手であると同時に機械からの音声応答の聞き手である。人間にとって理解しやすい音声応答速度は、個人の発声速度（無声区間をも含めたスピーチ区間でのモーラ／秒）[注]と密接な関係があると考えられる。したがって、個人の発声速度に対応して機械がその個人に最適な音声出力速度で応答できれば、個人と機械との円滑なコミュニケーションが図られ，人間に適合したヒューマン・インタフェースの実現に役立つと期待されよう。

著者は、前報[1]において音声分析システムを開発し、官能検査により、人間一般について話し手が理解しやすいと考える発声速度と聞き手として理解しやすい音声応答速度との相関関係を明らかにし、機械からの音声応答速度を人間の発声速度に適応化させることの有効性を確認した。[2]

本報告では、留守番電話を想定した場合について、人間の発声速度を時間率[平均ON区間／（平均ＯＮ区間＋平均ＯＦＦ区間）］に基づいて推定し、機械からの音声応答のＯＦＦ区間を伸縮してその音声入力された時間率に一致させることにより、音声応答速度を適応化させる手法を提案している。

2. スピーチのON区間とOFF区間の特性分析

無声子音の前のＯＦＦ区間（無声区間）は、発声器官のメカニズム上生じる区間で、実際にはＯＮ区間（有声区間）と考えられる。ここでは、図１に示すようにＯＮ区間にハングオーバ（ある時間だけＯＮ区間を伸ばすこと）とフィルイン（ある値より小な継続時間のＯＦＦ区間のみをＯＮ区間に置換すること）を施して、上記ＯＦＦ区間をＯＮ区間に置換し、呼気段落区分でのＯＮ区間とＯＦＦ区間（ポーズ区間）を計測する。しかしこの呼気段落区分に適したハングオーバとフィルインの時間は、個人によって異なり、しかもポーズ区間が極端に短い場合には明確に決定できないので、任意のハングオーバとフィルインを施した場合について、人間

＊山形大学工学部
（Faculty of Engineering, Yamagata Univ.）

注）
１モーラは連続音声を仮名書きした場合の１文字に相当する。

一般に関する発声速度と時間率との相関関係を検討する。ここで時間率（speech activity)は、代表的なスピーチのＯＮ—ＯＦＦパラメータで、スピーチ区間内でＯＮ区間の占める割合をさす。しかしスピーチ区間の短い対話においては、スピーチ区間内でＯＦＦ区間の回数がＯＮ区間の回数よりも１つ少ないことは時間率の算出において無視しえないので、本論文では時間率を平均ＯＮ区間／（平均ON区間＋平均ＯＦＦ区間）と定義する。

２・１実験　　被験者は19才～23才の男性46人である。被験者は留守番電話を想定したメッセージを、表１に示すキーワードを参考にして自分なりの言葉で伝達するという条件で発声した。被験者46人のモーラ数の平均値±標準偏差は128.9 ± 17.0 モーラであった。

２・２測定　　分析システム構成図を図２に示す。被験者の音声は、ビデオテープにＰＣＭ録音した後、2.5KHzのローパスフィルタを通し、サンプリング周波数５KHzでＡ／Ｄ変換され、本システムに収録される。その音声信号のＯＮ区間とＯＦＦ区間の機械識別は、50msecの平均雑音レベルに12dB加えた値を臨界値として、ＯＮ—ＯＦＦの判定周期10msec（フレームと呼ぶ）についてこの値を越えた部分をＯＮ、越えない部分をＯＦＦとした。ハングオーバとフィルインを施した場合のＯＮ—ＯＦＦパターンの測定例を図１に示す。

２・３スピーチのＯＮ—ＯＦＦパラメータ

任意のフィルインとハングオーバを施した場合の平均ＯＦＦ区間、平均ＯＮ区間、時間率はＯＦＦ区間の確率密度関数に基づいて以下のように計算される。[3]

2･3･1 任意のフィルインを施した場合

フィルインを施さない、即ちフィルイン時間が０でのフレーム数Ｋにおけるｏｆｆ区間の確率密度関数をｆ(K)とすると、フィルイン時間m

表１　キーワード表

“もしもし”
“山形大学の＿＿＿＿です。”
用件　講演の依頼
日　時　　11月３日（文化の日）
テーマ　　「情報の価値について」
“今週末までに＿＿まで連絡して下さい。”
“よろしくお願い致します。”

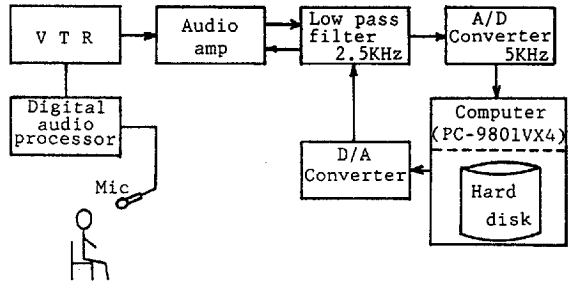

図２　分析システム構成図

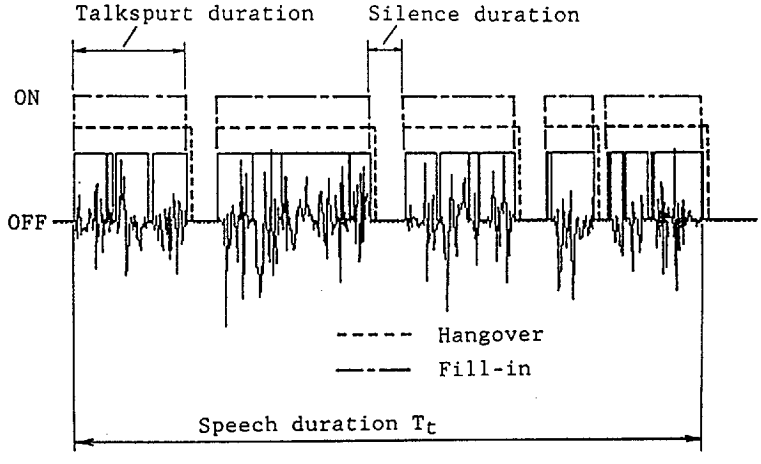

図１　フィルインとハングオーバを施した場合のＯＮ—ＯＦＦパターンの測定例

の確率密度関数fm(K)は、

$$fm(K) = f(K) / \sum_{j} f(j) \quad (1)$$

$$j,K = m+1, m+2, m+3, \cdots$$

$$m = 0, 1, 2, \cdots$$

である。したがってフィルイン時間mの平均OFF区間Sf(m)は、

$$Sf(m) = \sum_{k} (K \cdot fm(K)) \quad (2)$$

$$m = 0, 1, 2, \cdots$$

で算出される。一方、平均ON区間は以下のように求められる。m＝0でスピーチ区間TtのON区間の回数をN_0とすると、OFF区間の回数がON区間より1つ少ないことを考慮して、フィルイン時間mでのON区間の回数Nmは、

$$Nm = 1 + (N_0 - 1) \cdot \sum_{k} f(K) \quad (3)$$

$$K = m+1, m+2, \cdots$$

である。

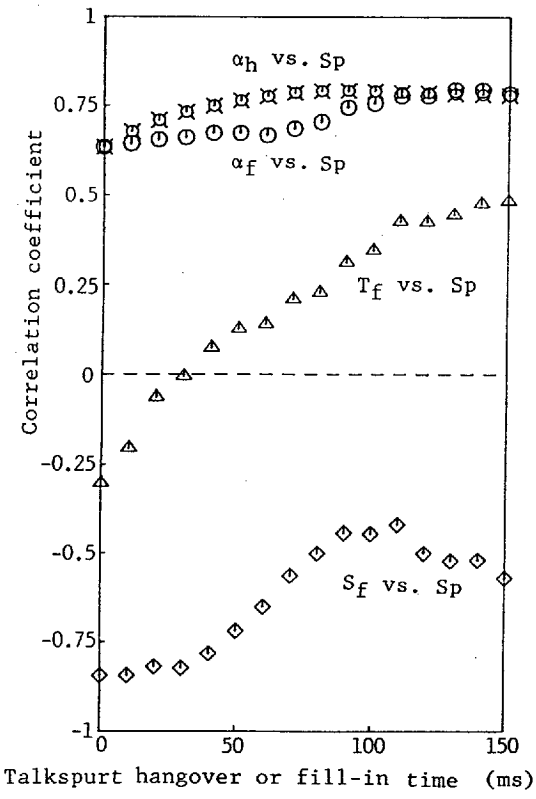

図3　フィルイン時間とハングオーバ時間の関数としてみた発声速度に対する時間率，平均ON区間及び平均OFF区間との相関係数

したがって平均ON区間Tf(m)は、

$$Tf(m) = Tt / Nm - (Nm-1) \cdot Sf(m) / Nm \quad (4)$$

$$m = 0, 1, 2, \cdots$$

で算出される。これは文献3)の近似計算式を修正したものである。時間率αf(m)は次式で定義される。

$$\alpha f(m) = Tf(m) / (Tf(m) + Sf(m)) \quad (5)$$

発声速度SpはTtとモーラ数Mrより次式で算出される。

$$Sp = Mr / Tt \quad (6)$$

2・3・2 任意のハングオーバを施した場合

ハングオーバの性質上、フィルインと比べて平均OFF区間Sh(m)はmだけ短く、平均ON区間Th(m)はmだけ長くなる。

$$Sh(m) = Sf(m) - m$$

$$Th(m) = Tf(m) + m \quad (7)$$

時間率αh(m)は次式で与えられる。

$$\alpha h(m) = Th(m) / (Th(m) + Sh(m))$$

$$= (Tf(m) + m) / (Tf(m) + Sf(m)) \quad (8)$$

2・4 分析　　発声速度を、早口、普通、遅口の3種に変えた場合、ON区間の伸縮に比べてOFF区間の伸縮の比率が大きく、早口ほど時間率が高くなる。人間一般についても発声速度を推定するのに、時間率が有効な指標になると考えられる。

図3に被験者46人について，フィルイン時間とハングオーバ時間の関数としてみた発声速度に対する時間率、平均ON区間及び平均OFF区間との相関係数を示す。[注] 発声速度と時間率の相関が比較的高いのは、フィルインが90msec以上、ハングオーバが80msec以上であった。最大の相関係数を示すのは、フィルイン時間が130msecで、ハングオーバ時間が90msecであり、その相関係数の値に差異はない。ハングオーバを施した場合、その時間だけ呼気段落区分でのOFF区間が短くなるので、以後はフィルインについて検討する。またフィルイン時間0msecでの発声速度と平均OFF区間との相関係数は－0.84であり、フィルイン時間130msecでの発声速度と時間率との相関と同程度に高い。しかし、音声対話速度の適応化においては、無声子音の前の短いOFF区間の伸縮は音声合成

上不適切であり、この呼気段落区分でのＯＦＦ区間が伸縮の対象になる。フィルインを施した場合の呼気段落区分での発声速度と平均ＯＦＦ区間との相関は、時間率との相関よりも低く、したがって発声速度と最も相関が高いのは時間率である。

図４にフィルイン130ｍｓｅｃでの発声速度と時間率との関係を示す。時間率αより、発声速度S_pは次式の回帰直線で推定される。

$$Sp = 10.5\alpha - 0.279 \quad (9)$$

フィルイン時間130ｍsecでのモーラ当りのＯＮ区間とＯＦＦ区間との関係を図５に示す。両者の相関は低く、早口の人が必ずしもＯＦＦ区間が短いわけではない。したがって、最適な音声応答速度の制御は、厳密にはＯＮ区間と、ＯＦＦ区間の両方について行なう必要がある。しかし、図５に示すようにモーラあたりのＯＮ区間との標準偏差／平均値の比は2:9で、ON区間の変動はＯＦＦ区間に比べて小さく、OFF区間のみの制御によってかなりの適応化が図られると考えられる。事実、フィルイン時間130ｍsecでのモーラ当りのＯＮ区間と時間率との相関係数が－0.11とほとんど0に近いのに対し（図６）、モーラ当りのＯＦＦ区間と時間率との相関係数は－0.95と極めて負の相関が強い（図７）。

(注)

平均ＯＮ区間及び平均ＯＦＦ区間の発声速度に対する相関係数は、各々ハングオーバ時間とフィルイン時間の関数として等しい。

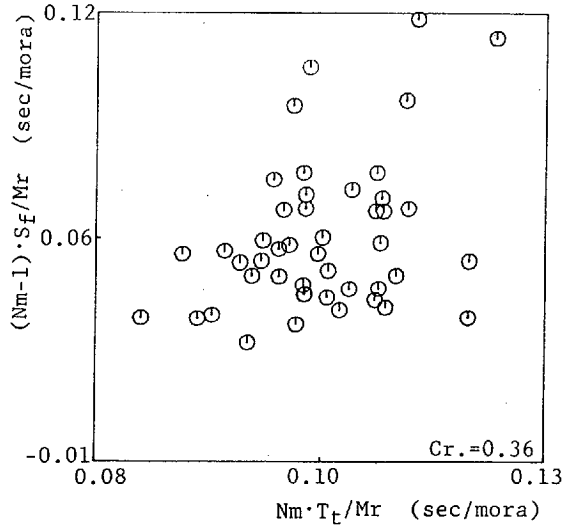

図５　モーラ当りのＯＮ区間とＯＦＦ区間との関係（フィルイン130ｍsec）

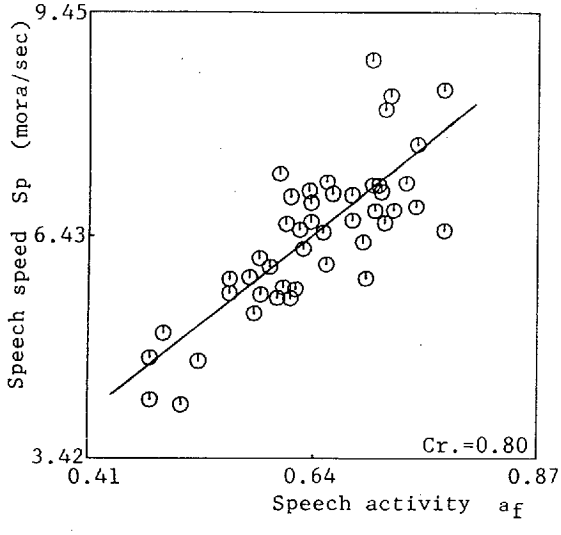

図４　発声速度と時間率との関係（フィルイン130ｍsec）

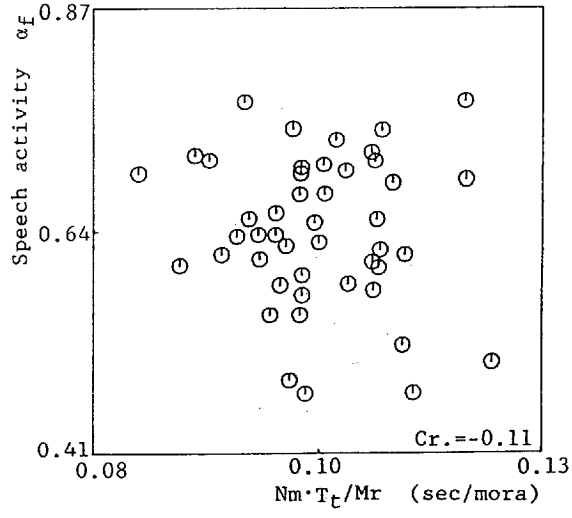

図６　時間率とモーラ当りのＯＮ区間との関係（130ｍsec）

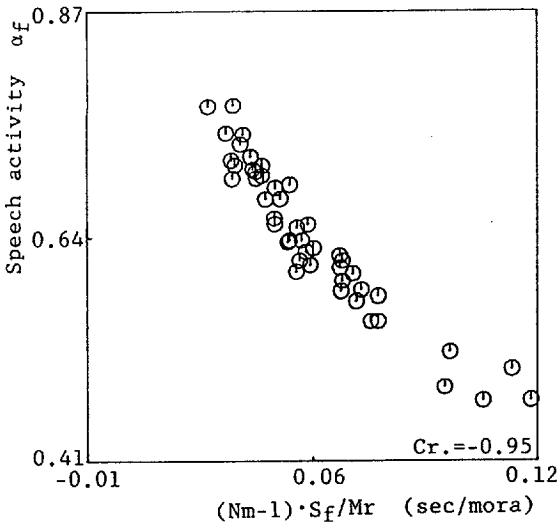

図7　時間率とモーラ当りのＯＦＦ区間との関係（フィルイン130msec）

3. 音声対話速度の適応化

2章での分析結果より、音声応答速度の適応化は主として、ＯＦＦ区間について行なえばよいことが判明した。また呼気段落区分での発声速度との相関は時間率が最も高く、音声応答速度を入力された発声速度に適応化させるには、音声応答のＯＦＦ区間を単にその入力音声の平均ＯＦＦ区間に置換するのではなく、時間率に一致するように平均ＯＦＦ区間を伸縮する必要がある。時間率を合わせることは、そのＯＮ－ＯＦＦパターンを合わせることであり、ＯＮ区間とＯＦＦ区間の繰り返しのリズムに適応化させることである。

いま、留守番電話のように予め録音されている音声を人間の発声速度に適応化させる場合を考える。機械応答音声と入力音声の時間率が各々$\alpha 1 = T1/(T1+S1)$と$\alpha 2 = T2/(T2+S2)$のとき、両者の時間率が一致するように、

$$T1/(T1+S') = T2/(T2+S2) \quad (10)$$

$$\therefore\ S' = T1 \cdot S2/T2$$

S1をT1・S2／T2に伸縮することによって、即ち、フィルインを施した（呼気段落区分での）各ＯＦＦ区間を（T1・S2）／（T2・S1）の比率で伸縮すれば、適応化が図られる。このとき、機械応答速度Sp1=Mr／（N_0・T1+（N_0-1）・S1)は、

$$Sp1' = Mr/(N_0 \cdot T1 + (N_0-1) \cdot T1 \cdot S2/T2) \quad (11)$$

Mr：スピーチ区間内のモーラ数

になる。

全被験者46人中、時間率が最大値，平均値，最小値の被験者ａ，ｂ，ｃ各々について、本手法を用いて本人以外の被験者45人の発声速度に適応化させた結果を図８に示す。適応化させた発声速度Sp1’と時間率$\alpha 2$には、

$$Sp1' = [Mr/(N_0 \cdot T1)] \cdot [T2/(T2+(N_0-1) \cdot S2/N_0)]$$
$$\fallingdotseq [Mr/(N_0 \cdot T1)]\alpha 2 \quad (12)$$

の関係があり、図中の＋印の傾きはMr／（N_0・T1)、即ちａ，ｂ，ｃ各々のＯＦＦ区間を含まないＯＮ区間のみに基づく発声速度である。いま発声速度の適応化率λを次式で定義する。

$$\lambda = 1 - \sqrt{\Sigma \varepsilon_o^2 / \Sigma \varepsilon_n^2} \quad (13)$$

ε_o：適応化後の発声速度と目標の発声速度との差

ε_n：適応化前の発声速度と目標の発声速度との差

λは、ａ，ｂ，ｃについて各々61.6％,40.0％，59.4％であった。ｂのλが、ａ，ｃのλよりも低

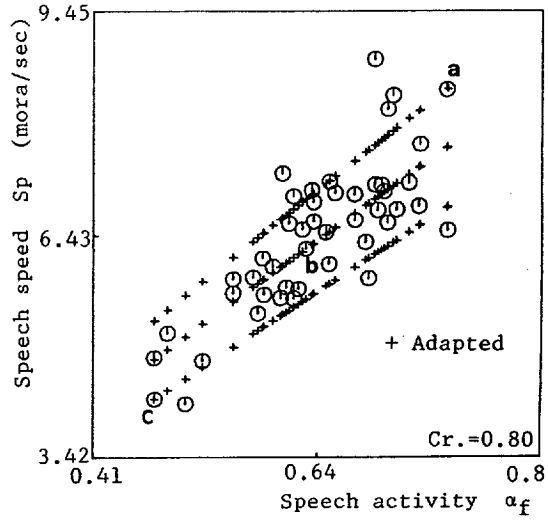

図８　全被験者46人中、時間率が最大値，平均値，最小値の被験者ａ，ｂ，ｃ各々について、本人以外の被験者45人の発声速度に適応化させた結果

いが、これは$\Sigma\varepsilon_n^2$がもともと小さいためで、$\sqrt{\Sigma\varepsilon^2/45}$は最も小さく0.61モーラ／秒であった。

4. 結言

本報告では、人間一般についての発声速度と時間率との関係を明らかにするとともに、個人の発声速度を時間率に基づいて推定し、その発声速度の時間率に一致するように機械からの音声応答のＯＦＦ区間を伸縮して音声応答速度を適応化させる手法を提案した。本手法で、音声応答の平均ＯＦＦ区間を単に発声速度の平均ＯＦＦ区間に一致させるのではなく、時間率に一致するように平均ＯＦＦ区間を伸縮させることの根拠は、1)時間率とＯＦＦ区間とは強い相関関係があり、音声応答速度の適応化は主としてＯＦＦ区間について行なえばよいことが判明したこと、2)呼気段落区分での発声速度との相関は、時間率の方が平均ＯＦＦ区間よりも高かったことによる。時間率を合わせることは、そのＯＮ－ＯＦＦパターンに、即ち話し手のリズムに適応化させることである。本手法を人間と機械との音声対話に適用することにより、個人に適合した円滑な対話が図られると考えられる。

文献

1）渡辺富夫ら：乳幼児の泣き声収録分析システムの開発、厚生省心身障害研究「家庭保険と小児の成長・発達に関する総合的研究」昭和61年度研究報告書, 51-53, 1987.

2）渡辺富夫：マン・マシン・インタフェースにおけるスピーチの理解しやすい速度に関する研究, 日本機械学会論文集, 53（496), 851-856, 1987.

3）Gruber, J.G.: A Comparison of Measured and Calculated Speech Temporal Parameters Relevant to Speech Activity Detection, IEEE Trans. on Communications, 3(4), 728-738, 1982.

Abstract

Adaptation Method of Machine Conversational Speed to Speaker's Speed for Man-Machine Communication

Tomio WATANABE

This paper first discusses the relationship between the speech speed and the speech activity [mean talkspurt duration／(mean talkspurt duration＋mean silence duration)] in 46 subjects for any value of "hangover" and "fill-in". Secondly, based on this relationship, an adaptation method of machine conversational speed to speaker's speed is proposed.The method controls the mean silence duration in the machine's speech activity in the same way as the speaker's speech activity. The simulated adaptation rate to speaker's speed in this method is approximately 60%. This method is effective in adaptation for the great differences between machine's speech activity and a speaker's speech activity, and is useful in realizing a man-machine interface with smooth information exchange.

要約 個人の発声速度に対応して機械がその個人に最適な音声出力速度で応答できれば、円滑な音声対話が図られると期待されよう。本報告では、個人の発声速度を時間率[平均 ON 区間/(平均 ON 区間+平均 OFF 区間)]に基づいて推定し、その発声速度に音声応答速度を適応化させる手法を提案している。