

UDC 534.78

話す速さの検出の基礎実験

鈴木誠史* 小倉俊彦**

PRELIMINARIES TO THE MEASUREMENT OF
THE SPEAKING RATE

By

Jouji SUZUKI and Toshihiko OGURA

It is necessary to monitor or control the speaking rate in the conversational test of a speech transmission circuit and the broadcasting service. But there is no suitable definition on the speaking rate.

This paper proposes a definition of the speaking rate by mora per second. A simple detector of mora is examined, based on the discrimination of a spectral feature of vowels and the syllabic nasal. The experimental results show that the proposed method is valid for measurement of quantity related to the speaking rate. The problems of normalization of the input level, measuring period, existence of the devocalized vowels and choked sounds, and the relation between the morae and quantity of sentence, are discussed for the construction of a practical equipment.

1. はじめに

音声通信回線の了解性についての通信テストや一般放送業務などで、話す速さを規定したりモニタの必要なことがある。このようなとき、単に普通の速さで話すように指示がなされたり、プロデューサによる経験的な計測や指示が行なわれている。これは、一つには発声速度（または話し言葉の速さ）に対しての定義が明確でなく、したがって計測の手法も確立されていないためである。ここでは、話す速さの定義をするとともに、物理的なその検出を試み、話す速さの測定装置の可能性を検討し問題点を明らかにする。

2. 話す速さの定義

話す速さを定義するためには、まず言語的な単位とし

て何をとるかを決定しなければならない。しかも測定装置の製作についての制約もある。ここではつきの諸点を考慮して検討を進める。

- (1) 音声言語の時間的な量を表わせる。
- (2) 一般に用いられる単位である。
- (3) 物理的な検出が容易である。
- (4) 対象は日本語に限定する。

ここで、(1)と(2)は矛盾する可能性がある。一般に用いられる言語の基本的な単位として、語、音韻、シラブルが第1に考えられる。英語などで量として言語をとり扱うとき、語数がもっともよく使われる。しかし語と発音の長さとの対応がないこと、音声言語では物理的な方法で語を検出することが、困難なことを考えると適当ではない。また日本語の場合には語数を単位とする習慣が少ない。

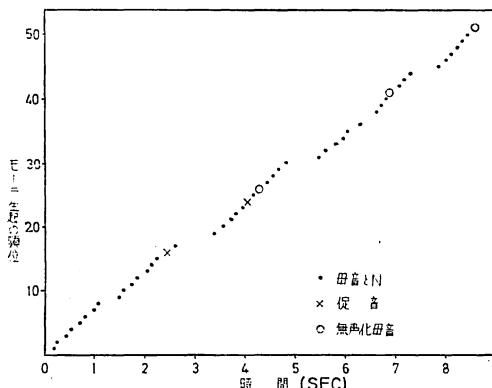
音韻は言語の基本的な音素であるが、音声との対応には複雑な関係があり、単位が小さいことと合わせてその検出にはいろいろな困難が予想される。シラブル(音節)

*通信機器部音声研究室 **千葉工業大学、現在千葉市役所

は、音声学的 (phonetical) な単位でだいたい上記の条件を満足できよう。シラブルと文字言語との対応が実用上必要になるが、幸いにも日本語には表音文字のカナがあり、言語学的に厳密なものではないが音節の考え方は比較的導入しやすい。

ところで、音節の長さを示す音形学的 (phonological) な単位としてモーラ (mora) がある。これは短母音の音節の長さを 1 単位とする。日本語のモーラは、単音節* と撥音 (N), 促音 (Q) とかなる⁽¹⁾。またカナでは 1 字が 1 モーラ (よう音では 2 字) に対応し⁽³⁾、いわゆるシラブルよりは文字との対応も簡単な関係にある**。

モーラの生起の時刻とその順位との関係を第 1 図に示



第 1 図 モーラ生起とその順位

す。これは天気概況の 1 節を資料としたスペクトログラムの観察により、モーラに対応したスペクトルパターンの中で、もっとも定常的なときをその生起の時刻としたものである。休止の時間を除くと、時刻とモーラの生起はきわめて直線的な関係にある。また休止時間を含めても、大局的には直線的といえよう。また、平均して 1 秒あたり数個の測定点 (モーラ) を期待できる。

第 1 図からも示されたように、モーラが時間を考慮した単位だけに、これを話す速さの定義に用いることは有効と考えられる。また日本語のモーラが、カナ表記、音節ときわめて関係が深いことも好都合である。さらに後述するように、モーラの検出と母音の検出とはほとんど同じことを意味し、物理的な検出も容易であると考えられる。したがって、ここではモーラ／秒を話す速さとし、その検出を目標に検討を進める。

3. 音 声 資 料

ここで実験に使用する音声資料は、男性アナウンサに

*東京方言ではア、イ、ウ、エ、オ、……キヤ、キュ、キョ、……など 108 種類である⁽¹⁾。

**たとえば“学校, /gaQkoo/”は 2 音節であるが、4 モーラである。

よって話された天気概況と“アナウンス教室”⁽³⁾の一部である。前者は漢字混り文で原稿用紙 1 枚に相当する。そのモーラの構成を第 1 表に示す。この原稿は 3 種類の速さ (56 秒, 82 秒, 110 秒; 以下資料 F, M, S とする)

第 1 表 資料 A (天気概況) のモーラの種類

NO.	I	E	A	O	U
MORAE	11	2	12	22	4
VOWEL	0	0	18	0	0
W	18	8	5	4	3
R	0	0	1	7	0
Y	0	0	4	0	5
P	10	6	8	17	0
T	14	1	22	7	2
K	0	0	4	5	1
B	0	0	2	4	2
D	0	0	15	1	7
G	0	4	6	1	7
H	8	1	9	0	2
Z	2	1	0	0	1
M	4	1	13	10	1
N	5	1	15	0	0
Y	10	1	0	14	0
CH	0	0	0	1	0
KY	0	0	0	0	1
BY	0	0	0	2	0
J	0	0	0	1	0
GY	0	0	0	0	0
SH	0	0	0	0	0
HY	0	0	0	0	0
ZY	0	0	0	0	0
MY	0	0	0	0	0
NY	0	0	0	0	0
RY	0	0	0	0	0
SUM	82	46	134	106	58
N	26				
Q					
TOTAL (SUM+N)	452				

で読まれている***。アナウンス教室はやさしい解説でモーラ数は 1438、所要時間は 210 秒である。これを資料 A とする。

4. モーラ検出に対する基本的態度

話し言葉の中で、モーラを区切って発声することはなく、連続的な音声信号からモーラを区分しなければならない。厳密に考えると、完全なモーラの区分は連続音声識別と同じような困難があると想像される。⁽⁴⁾しかし、ここでは完全な区分、検出でなく、話す速さの弁別ができる程度に一定の割合で検出する立場をとる。

日本語のモーラの構成は、V*⁴, CV, CjV (V: 母音音素, C: 子音音素, j は半母音 /j/) と N, Q である。ここで Q の区間は物理的な信号のない時間であり、前後の V と C の調音上の拘束条件と強度などから認識されるもので、簡単な装置による検出は困難である。したがって、ここでは Q の検出を除外する。その結果、モーラの検出 (厳密にはモーラ数の検出) は、母音と N の検出に帰着する。音声信号としては、母音と N は周期性のある波形の連続あるいは、比較的定常なスペクトルパターンとして観察される。特に母音の場合は、この

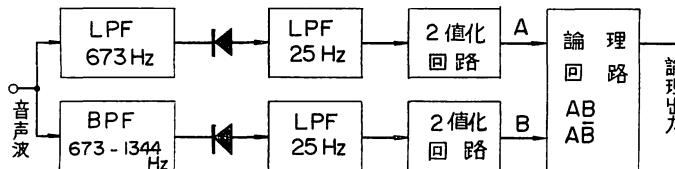
***この資料は電気通信研究所、第 4 特別研究室の好意によるものである。

*4 声帯音源音素⁴を考えて、いわゆる母音は CV の 1 種 'V' であるとされるが、ここでは単に V として扱う⁽⁴⁾。

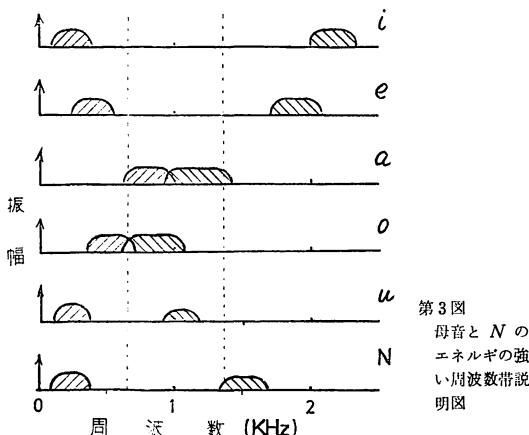
ような特性を利用した識別の実験が多く行なわれている^(6,7)。ここではその知識を活用し、できるだけ簡単な方法で検出を試みる。なお、対象は音声資料として用意したように、成人男声に限る。

5. 実験

母音識別実験を参考に予備実験を行ない、最終的に第2図の検出回路を構成した。ここでフィルタリングされた音声信号は整流・平滑され、さらに2値化される。2値化のレベルは入力信号のレベルに換算するとそれぞれ -8dBm である（音声入力は 0dBm を基準とする）。2値信号によるモーラ検出の論理は簡単に、 AB , $A\bar{B}$ とする。なお、その後の処理はソフトウェアによって行なう。

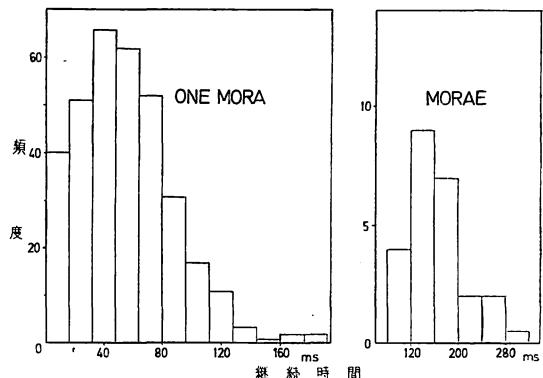


第2図 モーラ検出回路の構成

第3図
母音とNのエネルギーの強い周波数帯説明図

日本語の各母音とNのスペクトルの特徴は第3図のように示されるが、この論理によると原理的には*/a/*, */o/*と*/u/*の一部は*AB*で、*/i/*, */e/*, */N/*と*/u/*の一部が*A\bar{B}*で検出される。なお*AB*の論理出力の前後で生じる*A\bar{B}*は、Bチャネルの信号に付随したものとして、約10msの間は無視している。

母音やNに対応して得られる論理出力はある継続時間をもつ。これは、1個のモーラに対応するときでも、発声の仕方、モーラの種類、前後のモーラの種類により



第4図 論理出力の時間長の分布の例 (資料M)

変化する。また、2重母音や隣接するCが有声音のときなどでは2モーラ（まれには3モーラ以上）が、一つの包絡の論理出力として観測されることがある。1モーラと2モーラ以上に対応して観測された論理出力の継続時間の分布を第4図に示す。話す速さによって時間軸の伸縮はあるが、分布の形は他の資料でもほとんど変化しない。

ここで、論理回路出力に時間閾値を設け、これ以上に継続時間が長いときは2モーラとみなすことにする。この閾値は1モーラを2モーラに誤る数と、2モーラを1モーラに誤る数が等しくなるように選定すると、F, M, Sについて136, 144, 200ms, Aについては136msになる。この処理が理想的に行なわれるならば、長時間の観測によってはほとんど誤差を生じない。

すべての音声資料について、その最適の閾値を設定した場合と、普通速度Mの閾値を使用した場合の検出率を第2表に示す。SにMの閾値を適用したときは100%を超すが、これは1モーラを2モーラと誤って判定することが多いためである。Aの検出率が他に比較してやや低いことは、この資料が天気概況に比べると感情を含み、レベルの変動も大きいことに起因している。

母音別の検出率を第3表に示す。*/a/*, */o/*はほとんど100%検出できるが、*/i/*, */u/*の検出率はやや低い。これは、*/i/*, */u/*は*/a/*, */o/*よりも強度が低いことと、無声化の傾向があり、この検出法では無声化母音が検出できないことによる。

モーラ検出結果は時系列の信号である。その生起の時間間隔の分布の例を第5図に示す。資料Fは第5図Aとほとんど同じ形であり、Sの中央値はMよりも長い時間の間隔に移る。図に示されてはいないが、休止時間のためにモーラ間隔は2秒付近まで分布している。速さ

第2表 時間閾値とモーラ検出率の関係

閾 値		音 声 資 料 (%)			
時間長(ms)	(最適資料)	S	M	F	A
200	(S)	93.7	—	—	—
144	(M)	100.7	91.6	93.2	84.1
136	(F, A)	—	—	93.7	85.5

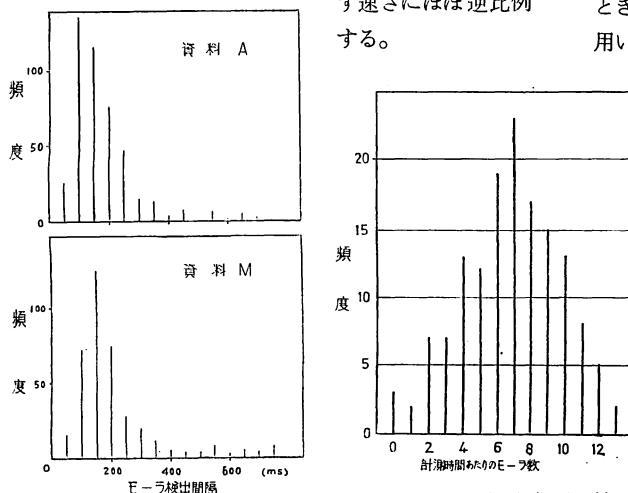
第3表 母音(Nを含む)別のモーラ検出率

母音とN	音 声 資 料 (%)			
	S	M	F	A
i	87.2	90.2	89.7	67.8
e	97.4	100.0	100.0	97.0
a	98.3	99.2	100.0	97.2
o	96.2	96.8	100.0	96.0
u	70.2	80.4	75.0	65.0
N	90.5	95.5	100.0	96.4

時間閾値はそれぞれの資料の最適値による。

を指示する間隔に相当するモーラ計数時間の推定は、第5図のような分布からでも可能であるが、確率として利用するには資料数が少ないので、直接的な方法による。

計数時間を増加しながら、この時間内に生起するモーラの分布を求める。分布の形は、計数時間が短かいとポアッソン分布に似た形を示し、長くなるにつれて正規分布に近づく。この方法で正規分布に近くなったときを实用上の最短の計数時間と考える。その1例を第6図に示す。この時間は1秒(F)から2秒(S)とみられ、話す速さにはほぼ逆比例する。



第5図 検出されたモーラの時間間隔の分布の例

第6図 モーラ速度の頻度分布の例
(資料 A) 計測時間: 1.44秒, 平均:
4.75 モーラ/秒

6. 実験結果の検討

6.1. モーラ検出方法の改善

ここでは簡単なことを第1に装置を考えたが、实用化的段階では二、三の点で改善の必要がある。

この装置で処理される音声信号は、比較的に一定のレベルであると考えられる。しかし、安定した能力をもたらせるためには音声識別装置で必要とされるように、入力レベルの変化の影響を受けないモーラ検出回路が必要である。その方式としては、(a)振幅基準化方式⁽⁷⁾、(b)相関係数の利用⁽⁴⁾が考えられる。文献(4)では、スペクトルパターンの相関を利用したが、波形領域で計算しても同じ効果が期待できる。また、IC、LSIなどの発達により比較的簡単に金物の実現が可能であろう。

モーラの存在の検出だけでなく、母音とNの識別機能を付加すれば、論理出力に対する時間閾値の重要性が減少しよう。しかし長音の存在もあり時間閾値は残す必要がある。

6.2. 促音と無声化母音

促音の検出を行なうためには、単語の識別と同じレベルの検出能力が必要と考えられる。しかし、約12500のモーラの統計を求めたところ、Qの存在は約1.3%であった。文の種類で多少この比率は異なる。しかし、検出率の精度と装置の簡単さからみて、Qの検出を無視することが得策である。

もう一つの問題は第1図にも例がある無声化母音の存在である。これは文の終り、あるいはアクセントのある音節につづく CV(Cが無声子音, Vが/i/, /u/)のとき、Vが無声化する現象である。したがって、ここで用いたような抽出法では検出できない。この検出も子音

識別、母音識別、さらに上に述べたような文の構造、音韻結合まで考慮しなければならない。ここでは、話す速度をモーラ/秒で定義したが、実際にはこれにある定数を乗じたものが測定されると考えればよい。この定数が音声資料にかかわらず一定ならば実用上はじゅうぶんである。なお、ここで用いたサブジェクトMに関しては、スペクトログラムの観察によると約5%のモーラで無声化がおこっている。したがって、第2表の結果は、無声化母音以外は非常によく検出が行なわれたことを示しているといえよう。

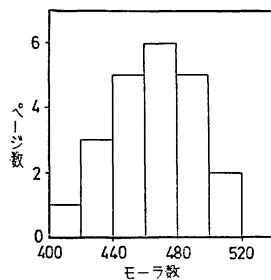
6.3. 指示装置

話す速さを指示するための計数時間については、第6図によって一つの考え方を示した。この時間が長くなれば正しい速さを示し、指示の変動も少なくなる。しかし変化に対しての応答が悪くなる。したがって、実用機としては、移動平均による指示、時間による荷重平均（時定数をもった指示回路）などが有効であろう。この点に関しては、最終的に人が情報を受けることから、人による評価試験が必要である。また、単なるモーラ数の計数もモニタとしては有効であろう。

6.4. モーラと原稿の量の関係

第2章の(2)で、ここで用いる言語の量の単位は一般的なものであることが望ましいと述べた。モーラは単音節、カナなどとの対応で比較的に入りやすいとはいえる。われわれが言語を量的に表わす習慣は、原稿用紙あるいは電報文にしかない。後者は特殊な用法なので、一般的な原稿用紙（400字詰）数とモーラの関係を調べておくと好都合である。

第7図に新聞の社説の原稿用紙21枚分のモーラ数の分



第7図
原稿用紙1枚あたりのモーラ数の分布例、原稿内容：社説、1枚あたり平均466モーラ

布を示す。第3章で述べたように天気概況はほぼこれと同じくらいであり、資料Aでは380くらいになる。この差は原稿に含まれる漢字の量によって支配されるが、文章の種類により平均値が異なる。実用に先立って原稿の内容別に第7図のような統計資料を求めておく必要がある。

6.5. 話す速さと休止時間

ここでは、休止時間を含めて実験を進めてきた。ところで、発声の速さを変えたときには、休止時間が音声区間よりも大きくこれに応じ、しかも伸縮比の偏差も少ないことが調べられている⁽⁹⁾。ただし、休止時間の分布は一般にきわめて広く、顕著なピークも存在しない。したがって休止時間を計測して話す速さと関連させるためには、長い計測時間を必要とし、速さを指示する目的には適合しない。しかし、この方法は物理的な測定が簡単なことから、モーラ/秒のように厳密ではないが、長い文

（音声）の速さを指示する量として使用できよう。

7. む　　す　　び

本報告の成果はつぎに要約される。

- (1) 話す早さをモーラ/秒で定義した。
- (2) 日本語モーラの特質から、モーラの検出を母音と撥音の検出に帰着させた。
- (3) 簡単なスペクトル分析装置を用いて、(2)の考えに立脚したモーラ抽出実験を行ない、この方法でも話す速さに対応した量が測定できることを示した。
- (4) 実用機として考慮しなければならない問題（計数時間、促音、無声化母音、原稿用紙の枚数との関係など）の検討を行なった。

測定装置の実用化には追加実験、実用試験による改良が必要であるが、基本的な問題はここにつくされたと考えられる。また、モーラ/秒による速度の定義は、いままでばくぜんとしていた音声言語の速さを明確にするとともに、正確な話す速さの記述や制御に役立つものと期待される。最後に、この問題を提起された川上前通信機器部長と、いろいろと討論された音声研究室の各位に感謝する。

参 考 文 献

- (1) 服部四郎；言語学の方法，pp. 362，東大出版会，1960。
- (2) 同上，pp. 361。
- (3) NHK 放送教育研究会編；アナウンス教室，NHK サービスセンタ，昭43。
- (4) 中津井，鈴木；連続音声識別の検討，電波研季報，11, No. 53, 1965.
- (5) 文献(1), pp. 289.
- (6) 鈴木，中田；母音の識別，電波研季報，7, No. 29, 1961.
- (7) 鈴木，角川；単音節中の母音の識別，電波研季報，8, No. 39, 1962.
- (8) 服部四郎；音声学，pp. 157，岩波全書，1951。
- (9) 此企，金森，大泉；連続音声中の各音韻の持続時間の性質，信学会，制御 65.1—19, 情報 65.1—20, 1966。