

分割された日本語単音節の知覚実験的研究¹⁾

竹 内 義 夫

Perceptual Study of Segmented Japanese Monosyllables

The purpose of this study was to investigate the possible influence of a certain method of segmentation of speech sounds upon their identification.

Procedure: 1) A male utterance of forty Japanese discrete CV syllables, which included fifteen consonants [p, t, k, b, d, g, s, h, dz, d₃, r, w, j, m, n] and five vowels [a, i, u, e, o], was recorded on magnetic tape. 2) Positional matching between a syllable's recorded location on the tape and its indicated position on a spectrogram was done by means of a technique employing a sound spectrograph. 3) A boundary between consonants and vowels was determined at the point where the vowel formants began to enter a quasi-stable state on the spectrogram. 4) The tape was cut vertically at two points each of which respectively bisected the vowel duration and the consonant duration (in the case of the voiced consonant, for this study, the buzz bar was not considered to be within the consonant portion). 5) The initial half of consonant segment tapes and the latter half of vowel segment tapes were joined together to make new syllables named "initial consonant syllables" from the consonant portion weighting and referred to as "IS" hereafter (Figure 1). 6) The remaining halves of the consonant segments and vowel segments were then left to form "transitional syllables", referred to as "TS" hereafter (Figure 1). Thus two kinds of syllables having equal duration in details were synthesized from one original syllable (Figure 2). 7) These syllables were edited in random order and re-recorded to make a test tape. 8) This test recording was then played to a listening panel of twenty students in psychology.

Result: 1) Syllable intelligibilities for the IS were significantly higher in general than for the TS (Table 1). 2) Vowel intelligibilities for both IS and TS

註 1) 本研究の要旨は、筆者が京都大学文学部在職中、1958年7月、日本心理学会第22回大会において、京都大学教育学部宇阪良二、今栄国晴、本吉昌枝との共同研究、「視聴覚系における情報伝達の研究」の第4部として、筆者によつて発表された。

were affected only slightly from bisecting the durations (Table 2). The influence of bisecting, however, was somewhat more marked in the vowels [i] and [u] than in the other vowels (Table 3). 3) Consonant intelligibilities were approximately as high as for whole syllables, because vowels in both IS and TS were highly intelligible (Tables 1 and 4).

Aside from the intelligibilities of consonants, a study of consonant perception using a confusion matrix revealed the following results (Table 4). Stop consonants among IS were highly intelligible, but for TS they were frequently not heard, especially the consonants [p] and [b]. Fricatives and affricatives in IS were also highly intelligible. For TS; [s] sounds were heard most frequently as [t], while identifications of [k] for [h] stimulus sounds and of [d] for [d₃] stimulus sounds were dominant among error responses. In the hearing of liquids, misidentifications in IS increased as much as in TS. In both IS and TS, mishearing of [r] and [b] was marked. For other liquids, the observation of misidentification tendencies was not reliably established. For IS the perception of nasals as nasals was most frequent, while in TS correct identifications and [b]-hearings were dominant.

For each consonant sound the possible perceptual cues were analyzed for this paper by means of spectrographic studies of consonant sounds and by studies of confusion matrices. From this data, it was concluded that initial durations contained a higher concentration of phonetic information about consonants than did final durations.

Yoshio TAKEUCHI

1. 序

音声は物理的には連続的な流れであるにもかかわらず、不連続な分節した単位（単音、又は音素）の継時的出現として認知せられる。これは音声の最も著しい特徴の一つに数えられている。音声の実験的研究には、この音声連続体を区分、または分割する操作がその研究過程に含まれるのが常である。こゝで音声連続体を区分し、切片化する操作を分割と定義するならば、従来行われてきた音声分割の研究目的は次の二つに大別することができよう。

その一つは音声の時間軸上における冗長性の研究である。Miller ら²²⁾は音声を周期的に断続して、その了解度と断続周期および一周期内的での on と off の時間比との関係を調べた。彼等の研究によると音声は約半分が失われても、10~100cps 位の断続周期では単語の了解度は約90%に達し、音声は時間軸上においても、音声の他の構成次元におけると同様に過剰な認知的手掛りを持っていることが明らかにされた。

Garvey⁹⁾は Miller らの断続において、音声の切断されている空白の時間の部分を取り去り、残りの on の時間の部分を接ぎ合わせることによって、周波数変化を起さずに speed up された音声を作った。このようにしても音声は高い了解度を示したが、さらに Miller らの実験との比較によって、断続による音声の一部の欠損の効果と Speed up による効果を分離することができた。Garvey は録音テープを同じ長さの切片にし、その一部を捨て残りを再結合したのであるが、Fairbanks ら⁴⁾はテープを切断することなく、通常のテープを再生するのと同様の操作で、Garvey の chopping-splice 法と同じ操作を自動的に行う装置を考案し、time compression の数多くの研究を出している。

以上の分割音声の認知実験では、音声を周期的に分割したので、音声のどの部分がどう切断され失われたかという点では全くランダムで再現性を持たない。必然的に単語または文章の段階における平均的な了解度が問題とされた。

Steinberg ら²⁷⁾は無意味音節の始めの部分を取り除き、切断された部分が多くなるにつれて単音の認知がどうなるかを調べた。この実験の動機として、電話通信において実際に起る、いわゆる話頭切断の実験的検討が考えられていたのであろう。

分割の研究の第二の問題は、物理的な音声連続体を、言語学的単位²⁾に対応するように分割しようとするもので、分節的分割とすることができる。スペクトログラフの発達した現在では、この問題は、一見容易に思われる。すなわち、音声のスペクトログラフを取り、そのパターンの上で不連続が起っている所、および空隙の所をその音声を構成する単音の境界線と考えるのである。しかし、一つの単音がその継続時間を通じて、一様な性質を保持することは、むしろ例外であって、一つの単音が二つ以上の音響的下部構造から成り立っているのをみれば、この安易感はたちまち消失するであろう¹⁹⁾³⁾。結局、分節的分割は、分割する単音について、何らかの音響的特徴を知った上で行われるのである。また逆に、単音の音響的特徴は分節的分割によって対象的に固定された音声信号を音響分析することによって得られるのであるから、分割と分析は、相互依存的または弁証法的関係にあるとすることができる。分節的分割は単音の音響的構造とその時間的変化を知ること他に他ならぬのであって、(他の単音からは)弁別的な(同じ単音にとっては時間的、場所的に異なっても)固有的な音響的特徴の定量化を目指す音声分析に発展的に吸収されるべき問題で

註 2) 音声学的には単音¹⁹⁾、音韻論的には音素であろうが、今後は単音によつて代表させる。

3) 例えば無声破裂音音節において、burst と aspiration, aspiration と trransitionの始めに子音と母音の境界を引く間違つた分割を想像せよ。

ある。強いていえば、音声分析が音声の時間軸に垂直な次元を主としたのに対して、分割はむしろ時間軸そのものに重点を置くものということができる。

音声分析には従来からよく挙げられる困難な問題が附随している。すなわち、音声信号そのものが過渡的性質を持つことをはじめ、音声には個人差、音声的環境による変化、抑揚による変化など、多くの変動が伴うことである。しかしこれらの変動も、音声分析、すなわち、物理的な音声信号から単音の音響的対応者を得ようとする試みの真の妨げにならないで、得られた音響的特性の統計的分布を附加することによって、困難ではあるが解決せられる問題である。音声分析の根本的な困難または限界は、音声現象の多元性そのものにあるように考えられる¹⁹⁾。音声は人間の大脳内の神経過程の段階においても、また社会的集団的な段階においても扱うことができ、それぞれの段階において内部斉合的な音声の分類が原理的には可能である。これらの分類の段階相互間での対応はいうまでもなく、多対多関係である。現在のテーマに最も関係の深い、音響学的、音声学的、知覚的の三段階においても例外ではあることはできない。音声分析は音響的段階と音声学的段階の対応を求めるものであるから、音声の音声学的分類単位と音響的分類単位の一対一対応は不可能であるといわなければならない。

知覚的段階において分類原理とされている所謂 *distinctive teature* も、音響的性質として音声信号の中に存在するものではない。また *Haskins* 研究所においては音声の知覚的手掛りとなる音響的特性がいくつか明らかにされたが、*Lieberman* ら²⁰⁾は刺戟としての音響的特性と音声の知覚の直接的結び付きを考えず、それらの間に発声器官の潜在的な自己受容的な運動感覚の媒介を考えている。すなわち、音響学的段階と音声学的段階の対応を単純化する役割を音声信号として与えられていない運動感覚に求めたのである。一方音声分析の問題は、最近、音声の機械的識別装置また逆に合成装置と関連して、しばしば研究されてきた。*Wiren* ら²¹⁾は音声信号だけを入力として働く、音声識別装置の可能な究極の分類対象を *phoneme* に対立するものとして“*phonoid*”と名付けている。装置の発達につれて *phonoid* と *phoneme* または *allophone* の関係はかなり整理され、単純化されるであろうが、多対多関係にあることにはかわりはない。音声識別装置が本来の目的である音声信号を入力として、音声記号さらには文字を出力として出すようになるためには、ぼう大な記憶装置と複雑な論理判定回路が必要となる。これらは多対多関係を一対一関係に修正するのに必要な音声信号に含まれていなかった情報および操作を装置に与えるものである。

分割という点から識別装置をみれば、それには入力である連続的音声信号を識別を行うべき単位に分割する機能をもった部分が含まれる。切片の大きさは方式によって異なるが、分節的切片を更に音響的に準定常的又は単純な構造をもった音響的切片に分割し、この切片について比較的少数の音響的次元で定量化した特徴の各切片についての総和として一つの単音を判定しようとするものがある²⁹⁾。もちろん音響的切片全体が分析の対象になるのではなく、sampling されるのであるが、sampling 場所、sampling 時間（走行分析の積分時間）の決定が重大な問題となる。また人間の音声認知過程との比較において興味ある問題である。さらに会話音声を対象とした識別装置では分割の問題は非常に重要になるがとゞではふれない。

音声合成は音声研究の大きな領域であるが、合成の方式は音声から抽出されたパラメーターの組合せ、またそれらを dynamic に制御するものと、実際の音声（通常一人の）から要素的切片を取り出して再結合するものに二分される。パラメーターの決定には何が音声にとって重要であるかについて、十分な知識が要求されるのに対し、後者では必要なパラメーターが要素的切片中に a priori に存在するという利点がある。そのかわり必要十分な要素的切片の数を最小にするにはどう切片化するかと問題となる。

Harris⁹⁾は異った音節間で子音の切片を交換し、聴取試験を行うことによって合成のための要素的切片（Harris のいう building block）を決定しようとした。彼の場合子音、母音の組合せ数も限定されており、また prosodic な条件は考慮されていないが、方法論的には興味深い。Lisker¹⁰⁾は tape-cutting and splicing によって、言語学者のいう要素に到達しようとする方法を“armchair operationism”と批判しているが、Harris の方法は音韻論のいう commutation⁹⁾を音響的次元で実行したものとみることができる。彼はさらにこの原理による合成装置を試作している¹⁰⁾。

合成のための切片化の方法はその数を少なくするためには、単音に対応する切片をとるようにした方がよい。しかし単音間の交互作用、transion という厄介な問題を考慮するならば、単音に固執するのは得策とはいえない。Peterson ら²⁹⁾は単一の articulatory sequence pair (pair の始め、または終りの部分が silent なものを含めて)をもって一つの要素的切片とした。同じ sequence pair でも prosodic な条件の異なるものはすべて独立の切片とし、これを dyad と呼んだ。かくて米語の idiolect を自然性を失わずに合成するのに必要な dyad の数は約8000~8500であると推定している²⁹⁾。

2. 実験目的

音声の時間的冗長性または時間的手掛りの問題の研究にあたって、周期的断続や時間的に均一な切片の再結合による方法は、単音の段階における問題の考察に不適当なことは既述のとおりである。この研究においてこの点を主題とするために、音声を単音に分割し、各々の単音のもつ継続時間を一定の割合に切断短縮する方法を用いた。かくて切断再結合された音声の知覚実験を通して、時間的冗長性の他に、知覚的手掛りの問題、切片的合成に対する可能性を考察しようとするものである。

3. 実験方法

3.1 実験資料の音声

実験に使用したのは一名の成人男子（NHK アナウンサー）が単独に発声した日本語単音節から抽出した40個の音節である。[p, t, k, b, d, g, s, h, dz, d₃, r, w, j, m, n]の子音と原則的には母音系列の両端および中央の[a, i, w]または[a, e, o]の三母音と組合されたものである（第3図参照）。録音された音節のテープ上での正確な位置の決定にはソナグラフを使用した。すなわち15°/secで録音されたテープをソナグラフの磁気円板（この回転速度は15°/secである）の上に貼りつけて直接ソナグラムをとる（第1図のソナグラムはそのうちのいくつかを示している）。描かれたソナグラムの任意の点は、ソナグラフの記録針と磁気ヘッドの位置からテープ上にマークすることができる。このようにするとソナグラフの磁気ヘッドの出力は通常の場合より低下するのでトランジスターヘッドアンプを附加して補償した。

3.2 分割・再結合の原則

母音のフォルマントがソナグラムで定常状態に入った位置を子音と母音の境界と操作的に定義する⁴⁾。子音が摩擦音の場合、この分割操作は比較的簡単であるが、transitionがゆるやかな有声子音の場合にはかなり困難である。第1図に示すように、音声信号の始まりの線b（有声子音ではその前のbuzz barの始めの線aではなく）と、終りの線cおよび子音と母音の境界線oを引く。bo, ocをそれぞれ二等分する線m, nにおいてテープを垂直に切断して、主として子音の移行部分と母音の前半を含む切片を切り取り、子音部

註 4) Pitch 成分によつて子音（特に無声）と母音に分離する一般的方法よりは子音時間が長くなる。

分に注目してこれを立上り音節と呼ぶ。移行音節を除かれた残りの切片を線 m, n が重なるように接合して一つの音節を合成し、立上り音節と呼ぶ。このようにして一つの音節から、継続時間の等しい二つの刺激音節が作られた。おのおの40箇の刺激音節を立上り群、移行群ごとに、刺激間隔4秒でランダムに配列し、4秒の休止間隔の中にナンバーコールを入れて、編集再録音し、聴取試験用のテープを作った。

3.3 聴取試験

テストテープを通常の室内でスピーカーにより再生した。編集、再生に使用した録音器は、ソニー製 CP-1, KP-3 の各型及び東京電気音響の TD-102 型である。被験者は心理学専攻の学生20人で数人づつ集団的に聴取りを行った。被験者達は音声学的訓練の経験はなく、聴取りにあたっては日本語音に最も近似した音を仮名書きするよう教示した。

4. 結果と考察

4.1 分割音声の了解度

4.1.1 実験方法の節で述べたように、分割再結合された音節の了解度について子音別に第1表に示した。音節の了解度は、音節に含まれている母音、子音両者がともに元のものに認知された場合を正反応とし、その全反応数に対する百分率で表わされている。一音節の全反応数は20人×3母音であるが、子音によっては60にならないものがあり、その子音はかっこで囲んである。第1表によって、立上り群、移行群を比較すると、全子音音節平均でそれぞれ68%, 12%で、この差は勿論有意である（以下検定はすべて χ^2 によった²⁹⁾）。子音別にみると [b], [dz], [d₃], [w] 以外では両群間の差はすべて有意で立上り群が高く、例外は [r], [n] 音節だけである。

4.1.2 音節中の母音だけの了解度を同じように子音別に示したのが第2表である。各母音が平均されているが、両群ともに約90%の高い了解度で、各子音別にみてもいずれも

Table I: Syllable intelligibility percentages in comparison for the two kinds of segmented syllables, "initial" and "transitional". Those statistically significant percentages ($P < 0.05$) which prevail in each pair of syllables is presented in bold faced type. Most of the percentages are based on 60 responses (20 listeners × 3 vowels), but those of (dz), (d₃) and (w) are based on 40, 20 and 20 responses respectively.

	p	t	k	b	d	g	s	h	(dz)	(d ₃)	r	(w)	j	m	n	Total
Initial	70	63	66	53	75	35	76	76	33	40	13	40	35	70	15	52
Transitional	30	10	12	38	2	2	3	38	43	25	35	55	2	40	45	24

Table II : Vowel intelligibility percentages. Conditions are the same as in Table I, except that correct identifications of CV syllables are scored for vowels. Note that there is no significant difference ($P>0.05$) between the two kinds of syllables.

	p	t	k	b	d	g	s	h	(dz)	(dʒ)	r	(w)	j	m	n	Total
Initial	100	92	100	75	96	98	96	95	80	95	92	65	63	88	78	88
Transitional	100	77	93	96	91	100	93	95	93	100	98	100	81	100	90	93

Table III : Confusion matrix for vowels. The horizontal rows show frequencies of various vowel judgements made for each of the stimulus vowels. Results for “initial” and “transitional” syllables are mixed in this matrix by combining their respective figures in each cell.

		Response						
		a	i	u	e	o	Omission	Total
Stimulus	a	527		14	1	14	4	560
	i	1	274	41	2	1	1	320
	u		1	386	1	10	2	400
	e			5	6	149		160
	o				39		120	160
Total		528	280	486	153	145	8	1600

有意差が認められない。母音別の反応数は第3表に confusion matrix の形で示す。これによると [u], [a], [e], [i], [o] の順に低くなっている。[i]と[o]の間には有意差があり、もちろんこれらと他の [u], [a], [e]の間にも有意差がある。しかし [u], [a], [e]の三者間には有意差はない。

4.1.3 母音の了解度が非常に高いから、第1表の音節了解度は子音の了解度にかゝっているわけである。第4表にこれを示した。立上り群には第1表よりやゝ高くなったものがあるが、その傾向と有意性はまったく第1表と同じである。

Table IV : Consonant intelligibility percentages. Conditions are the same as in Table I except that this test of intelligibility is only concerned with the correct identification of consonants. There is almost no difference between Tables I and III.

	p	t	k	b	d	g	s	h	(dz)	(dʒ)	r	(w)	j	m	n	Total
Initial	87	63	66	66	75	35	60	80	33	40	13	40	72	80	23	57
Transitional	30	10	12	38	2	3	12	38	43	25	35	55	2	40	45	24

4.2 考察

4.1の結果の考察に先立って実験に使用された音節の継続時間を検討しておく必要がある。3.2で述べたソナグラムの区分の長さによって継続時間を求め、第5表に示す。母音

Table V : Durations of the sounds in milli seconds. These measurements are based on the lengths of the Sonagram patterns. Buzzbars in the patterns of voiced consonants are omitted from the measurements.

Sounds	p	t	k	b	d	g	s	h	(dz)	(dʒ)
Durations	36	29	63	42	55	54	112	84	81	90
N. of sample	3	3	3	3	3	3	3	3	2	1

r	(w)	j	m	n	a	i	u	e	o
59	99	131	47	47	126	140	141	147	140
3	1	3	3	3	14	8	9	4	5

の継続時間が長く、破裂音がその他の子音より短いのは従来よくいわれていることである。さらに継続時間が破裂音、摩擦音、破擦音内での弁別の特徴になると考えられている⁷⁾。また、機械的弁別装置において、[t]と[k]の弁別に継続時間の差が利用されている¹⁴⁾。分割によって継続時間が半分にされても、もとの時間の長いもの程了解度が高いのではないかと一応考えられる。子音について、継続時間と了解度の間の列位相関を求めてみると、立上り群、移行群でそれぞれ $\rho = -0.25$, $\rho = 0.09$ となりこの考えは否定される。母音が子音に較べて了解度が高く、両群で差のないのは、いうまでもなくその周期的性質、したがってその継続時間を通じて均一なことによるのであって、継続時間の長いためではない。母音の継続時間短縮によって起る了解度の変化について、Stevens²⁵⁾は合成母音の時間を20~500msに変化したところ、/i, I/, /u, U/ はほとんど影響されないのに対して、/ɛ, æ/, /ʌ, a/ は影響を受けると報告している。本実験では [i], [o] の了解度が低下しており、合成音声との違い、国語の差など、今後検討されねばならない。

次に先行声帯音の問題であるが、スペクトログラムにおいて、buzz bar と呼ばれるこの部分は、子音部分の大きな部分を占める。しかし発声の仕方によって、その長さは大きく変化し、高い周波数成分が欠けているので、子音の識別に有力な手掛りでないことは容易に考えられるが、有声子音の有声性の手掛りをなすとの意見もある¹⁹⁾。本実験では子音部分の分割にあたって、情報密度の比較的均一な部分を二等分するために、buzz bar を子音継続時間に入れなかった。したがって、立上り音節は buzz bar だけ時間的にも余分に含んでいるが、果して、弁別に役立たぬのであるか、この点を検討するために予備実験を行った。

有声子音の buzz bar だけを切除して聴取試験を行ったが、その結果被験者30名で、平均了解度86%、しかもその異聴には対応無声子音への異聴の傾向は明らかでなかった。こ

の点佐藤³⁰⁾も同様な結果をのべている。Haskins 研究所の Pattern Playback による合成音声と実際の音声とでは buzz bar に対する知覚的手掛りの重みは異なるのではないか。

本実験ではテープを垂直に切断したので、この効果について考えてみよう。テープを垂直に接続しても空隙を作らぬように注意すればクリック音を殆んど聞えないし、ソナグラムにもパルスは見られない。次に断続によるスペクトラムの変化であるが、最も小さい切片でも約15msであるから、音声の重要な周波数領域での妨害は僅かであろう。しかし音声の包絡線の立上り時間そのものが子音の知覚的手掛りをなすと考えられる点もあるので¹⁰⁾、切断の影響はこの点で特に考慮されねばならないであろう。

Ahmed ら³¹⁾は CVC の構成をもつ単音節の、音声の立上りの時刻を基準とし、それに続く任意の時間だけ音声を通じたり、また切ったりする装置を用い、知覚実験を行った。その結果、子音の継続時間のうち母音に近い部分の方が、立上りの部分よりも、子音の弁別に必要な情報をより多く含んでいること、さらに、この傾向は破裂音、摩擦音ほど大きいことなど、本実験と全く反対の結論を出している。この喰い違いの原因としては、Gate 回路によって子音の立上り部分がとり出されている場合、後続の母音を取り除かれているのに対して、本実験でいう移行部分が聞かれている場合には、それに続く母音とともに音節として刺激が与えられ、音節の了解度から子音の部分の了解度が求められた。このような刺激提示が立上りの部分の了解度を著しく低下させた原因と考えられる。さらにこの実験では被験者が二人で、各子音を40回反復しているなど知覚実験の方法にかなり問題がある。

立上り音節群は一応切片による新しい音節の合成であると考えられる。しかもこの群が切片の不完全さにもかかわらずかなりの了解度を保ったことは、切片的合成に有望な見通しを与えるものといえよう。合成といっても、もともと一つの sequence に含まれていた切片による合成であるから、異った発話から抽出した切片によって、連続音声を合成する際に困難な問題となる基本 pitch、抑揚、単音相互間の交互作用などは事前に避けられているわけである。しかし、日本語は単独の単音節をならべても、カナ電報のように、かなりの情報伝達ができるから、この点切片的合成にとって有利な国語であると考えられる。

4.2 分割音声の異聴

音声の聴取試験においてみられる異聴は、音声の音響的構造、認知の手掛りを知る上に重要な示唆を与えるものである。本節ではソナグラフによる分析、分割がもたらした音声構造上の変化をもとにして、本実験に見られた異聴の分析を試みる。第6表は聴取試験の

Table VI : Confusion matrix comparing consonants in the two kinds of segmented syllables. Horizontal rows show the frequencies of various consonant judgements made for each of the stimulus consonants. For each cell, upper left-hand figures refer to "initial syllable" responses and lower right-hand figures to "transitional syllable" responses. Bold face figures in the diagonal show frequencies of correct judgements.

		Response																			Total
		p	t	k	b	d	g	s	h	(dz)	(dʒ)	r	(w)	j	m	n	F	ts	tʃ	Omissions	
Stimulus	p	52	2		2	11		1	3		1					2			1	25	60
	t	16	38	1		2										2			1	9	60
	k	6	22	40	1	2	1	4	1	1		2	2	1	1				3	17	60
	b	9	1		40	8	23		1	3		2	1		4				2	18	60
	d	8	1		8	45					1	5	1				1			18	60
	g	9	2	11	6	7	9	1	1	1		2	2	2		2			5	29	60
	s		4	21	1				46	7	3	2					1	2	31	1	60
	h	6	5	6	21		1	1		48	23								4	4	60
	(dz)	1	1			12	14	1	3		13	17			3	5		5			40
	(dʒ)		1		8			1	1			8	5							1	20
	r				26	15	11	2		1	1				18	6	4			4	60
	(w)	1	1	2	1	4						2	8		4	1	4		2	20	60
	j	3	12	1	11	6		1		2		10		43	13	2	7		1	6	60
	m		6			23									48	11	2			1	60
	n	1	6		2	12	1				2	5		6	42	24	14	27		5	60
	*Error	49	13	15	54	15	13	4	6	3	18				81	23	5	7	1	21	
	Total	84	46	39	114	34	7	2	21	5	40				10	10	23	32		140	

* Error totals are the sums of numbers in the column without the number on the diagonal.

子音部分だけの結果を confusion matrix の形で示したもので、横の行は刺激を、縦の列は反応された単音を表わす。横の行を辿ることによって、該当刺激音が如何なる単音に聞かれたかを知ることができる。各 cell 内の左上の数字は立上り群、右下の数は移行群の反応の頻数を表わす。該当子音を含む母音の結果が加算されている。

4.2.1 [p, b, k, g (a, i, u)], [t, d (a, e, o)]

破裂音の子音継続時間は短い、非常に短い期間の burst (破裂による雑音) と続いて起る transition によって特徴づけられる。有声破裂音の場合は burst に続いて transition が起るが、無声音の場合は [k], [p], [t] の順に短くなる aspiration の後に母音が始まる。しかし aspiration も雑音であるから、後続母音のフォルマント領域が強調されているのがソナグラムに見られる。これも一種の transition とみるならば、立上り群には burst と transition の一部の、二重の手掛りが与えられている。移行群では逆にこれらのものが失われるだけでなく、切断によって一種の破裂が附加されることになり、ここに両群の了解度の差をもたらした原因が考えられる。立上り群にみられる高い了解度および移

移行群にみられる子音脱落の増加は破裂音にとって burst の重要性を思わせる。佐藤²⁰⁾は無声破裂音の burst だけを残して, aspiration の全部(時間的には子音部分の大部分)を除去してもよく了解されると報告している。Lieberman¹⁹⁾は破裂音の合成には burst と transition の間に, 即ち aspiration にあたる部分に silent interval が必要であると述べ, この点佐藤と見解を異にしている。

移行群についてみると無声 [k], [t] いずれも [p] と最も多く聞かれている。これは分割による人工的な破裂が [p] に類似するものと考えられるが, [p] の場合には脱落が最も多く, 問題の複雑さを思わせる。有声の場合は脱落が非常に増加するが, 同様に [b] への confusion の傾向がうかがわれる。transition の始まる点の周波数的位置 (locus) が破裂音内の弁別の手掛りになり¹⁹⁾, それも後続母音によって影響されるが, [b], [d], [g] の順に複雑になるといわれている¹⁹⁾。Hall ら¹³⁾は, burst および transition だけを gate 回路で抽出し, 分析すると同時に, 聴取試験も行った。彼らは final stop を問題にしたのであるが, /l/ からの transition が最も聞きやすく, /k/, /g/ の認知が困難であると報じている。これらの諸条件が重なって, 移行群の [b] の優位, および [d], [g] の反応数低下をもたらしたようである。

4.2.2 [s (a, e, u), h (a, e, o), dz (a, u), d₃ (i)]

摩擦, 破擦者は狭さくによる雑音が特徴で, その継続時間も長く, その間比較的一様で母音の影響も少ない。したがって狭さく雑音のスペクトラム, 特にスペクトラムの周波数下限, 雑音強度, 立上りの包絡線の形状など, かなり単純な次元においてこのクラス内の弁別特徴が探索されてきた。

第6表によれば脱落が少なくなり, 立上り群の了解度の高いのはこのクラスの単純構造的性から理解されるので, 移行群の異聴の分析に重点を置く。[s] の移行群では [t] と [ts] が全反応の97%を占めている。これは分割による人工的な立上りが [t] に聞かれるのであるから, [t] の burst に続く aspiration が知覚的手掛りをなさないことがこゝでも確かめられる。逆に [s] は摩擦雑音の周波数構造だけでなく, 立上りの時間的性質に依存することを示すものである。同様の結果は釜本¹³⁾も報告している。しかし, Lotz 等²⁰⁾は /s/ の摩擦部分を除去すると, アメリカ人は /b, d, g/ を聞くことから, 英語の有声, 無声破裂の差異の一つである aspiration が, それらの知覚的弁別特徴であるといっている。さらに, 同じ音が aspiration を伴わぬ無声破裂音をもった国語人にとって, 無声破裂音と聞かれたと報じている。Gerstman⁷⁾は [j] を短かくすると [k], [t] への異聴を認め, これら三者

の間の調音点がほぼ同じであることを挙げて、運動知覚的因子による解釈をしている。Denes⁹⁾は final の [s], [z] についてそれらの継続時間を変えることによって互換性が得られること、すなわち [s] を縮めることによって [z] に聞こえると報告している。1/2短縮の本実験ではこれに類似の結果はみられない。

[h] では後続母音の構えがとられたまゝ、摩擦雑音が発せられるので、当然雑音成分のうち、母音のフォルマント領域が強調せられる。しかしこの種の母音の影響は transition と違って、その継続時間を通じて変化しない。したがって同一音節内で再結合する本実験では、母音の影響はほとんど問題にならない。立上り群の了解度が高いのはこのように理解せられるが、移行群では [k] への異聴が目立つ。この21箇の異聴を母音別にみると、17は [ho] の場合である。これは [o] の第一、第二フォルマントにあたる低音部の強調された雑音に、分割による破裂性が加えられたためとみられる。Liberman 等¹⁵⁾は個々の無声子音の特徴的な burst の location が後続母音によって、変化することを述べている。それによると、burst が後続母音の第二フォルマント又はその上にある場合、[k] に聞こえ、かつ、後続母音が /o/ /o/ /u/ のときこの傾向が著しいと報告しているが、[h(o)]→[k] の異聴はこれと符合する。

[dz] は両群の了解度に有意差のないのは既述の通りであるが、両群とも [d] への異聴が著しい。[ザ], [ズ] は [dz] と表わされる破裂音であるからと考えられるが、移行群では [dz] の摩擦部分が主であるから、前述の [s]→[t] の異聴との対応が考えられる。[d₃] では例数が少ないが [r] への異聴が著しい。

4.2.3 [r (a, i, u)], [w (a)], [j (a, u, o)]

このクラスの音声に関する研究は他の音に較べて少ない。Liberman 等^{8,16)}によれば、transition の時間が長いこと、locus はある期間停留し、破裂音とは逆に、locus と transition の start の間に空隙がないこと、および、第三フォルマントの transition も手掛りの役割を果していると述べている。[r] についてみれば立上り群では [b], [m] への異聴が多い。母音別にみると [ra] の場合70%が [m] で、[ri], [ru] になると、60%, 65%が [b] である。ソナグラムを見ると、[r(a)] は locus への停留が数十msに及び、ゆるやかな transition が続くのに対して、[r(i, u)] では停留がほとんど見られず、有声破裂音に似たパターンを示す。[r(a)] はむしろ鼻音に近い。したがって立上り群において、locus が主として働いて [m] が聞かれ、[r(i, u)] の場合は transition の後半が失われたために、transition の時間の短縮が破裂音への異聴をもたらしたと考えられる。移行群でも同様に

破裂音への異聴が多い。

4.2.3 [w (a)], [j (a, u, o)]

[w], [j] については Liberman 等¹⁰⁾は transition の周波数範囲を一定に保ちながら, transition の時間を長くすることによって, 上向き transition では [be]→[we]→[ue], 下向きでは [ge]→[je]→[ie] と認知が変化することを見出している。日本語ではこれに近似する音節もなく, [wa] だけでは比較できない。[j] は実験に用いた単音中, 最も急激な transition の見られるものであるにもかかわらず, 立上り群の了解度が高い。移行群では [t, b, r, n] 等へ異聴が分布し一定の傾向がみられない。

4.2.4 [m (a, i, u)], [n (a, i, u)]

鼻音は鼻腔共鳴によって, 他のクラスから特徴づけられ, クラス内では調音点の差異が transition に表われていると云われている^{16,19)}。

[m], [n] を通じて第6表に著しいことは, 立上り群で, [m], [n] とともに [m] に聞かれるのが多く, [n] がこれにつぐ。すなわち [m], [n] の差は少ないといえる。移行群では, 正反応に続いて [b] が多くなっている。Malecot²¹⁾は実際に発声された鼻音音節 [m, n, ŋ (æ)] のテープの母音部と子音部を切り離し, 組合せを変えたところ, 最も多く [m] が聞かれたと報告している。このことから, 彼は鼻腔共鳴⁵⁾は鼻音の中立的な class marker でなく, やゝ調音点の手掛りになっていると述べている。これは本実験で立上り群の [m] 反応の増大に符合する。

Haskins 研究所^{10,20)}では [m], [b], [p] の間に, また [n], [d], [t] の間に transition の共通性があり, これを調音点の類似性によるとしているが, 本実験において [m] の移行群での異聴が [b], [p] に集中しているのも同様に考えられる。[n] の場合には, [d], [t] よりも, かえって [b], [p] が多く, この考え方はあてはまらない。

5. 要 約

日本語単音節40箇 ([p, t, k, b, d, g, s, h, dz, dʒ, r, j, w, m, n] を含み, これらと 3乃至1箇の母音の組合されたもの) を対象として, 各単音節の録音テープに含まれている子音および母音を, おのおの継続時間の1/2の点で分割して, 二組の子音および母音の切片にする。子音の立上り部分を含む切片と, これに後続する母音の後半の切片を接合し

註 5) 後続母音が [æ] のときには transition がほとんど起らない。

て新しい音節を合成して、立上り音節と呼び、立上り音節の切除された残余の子音と母音の連続した切片を、移行音節と名付ける。もとの自然発声された音節から派生した、これら両群の人工音節について、了解度試験を行った結果次のことが明らかにされた。

1. 分割された母音は、立上り、移行両群とも高い了解度を保ち（約90%）、両群の間に有意差はない。

2. 分割された子音の了解度は、母音に較べると低いが、立上り群は移行群より了解度が高く、大部分の子音について、両群の差は有意である。

3. 子音間の了解度を比較すると、子音の継続時間と了解度の間には相関がない。

その他、分割された子音の知覚の confusion matrix と、ソナグラムによる分析にもとづいて、個々の子音の知覚的弁別特性の問題が検討せられた。

文 献

- 1) Ahmend, R., R. Fatehchand : Effect of sample duration on the articulation of sounds in normal and clipped speech. *J. Acoust. Soc. Amer.*, 1959, 31, 1022~1029.
- 2) Cooper, F. S., P. C. Delattre, A. M. Liberman, J. M. Borst & L. J. Gerstman : Some experiments on the perception of synthetic speech sounds. *J. Acoust. Soc. Amer.*, 1952, 24, 597~606.
- 3) Denes, P. : Effect of duration on the perception of voicing. *J. Acoust. Soc. Amer.*, 1957, 29, 761~764.
- 4) Fairbanks, G. : Experimental study of time compression. *J. Acoust. Soc. Amer.*, 1956, 28, 591.
- 5) Fisher-Jørgensen, E. : The phonetic basis for identification of phonemic elements. *J. Acoust. Soc. Amer.*, 1952, 24, 611~617.
- 6) Garvey, W. D. : The intelligibility of speeded speech. *J. exptl. Psychol.*, 1953, 45, 102~108.
- 7) Gerstman, L. J. : Noise duration as a cue for distinguishing among fricatives, affricative and stop consonants. *J. Acoust. Soc. Amer.*, 1956, 26, 160.
- 8) Gerstman, L. J., L. Lisker, J. D. O'connor, A. M. Liberman, P. C. Delattre & F. S. Cooper : Initial and intervocalic cues for the perception of liquids and semivowels. *J. Acoust. Soc. Amer.*, 1957, 29, 778.
- 9) Harris, C. M. : A study of the building blocks in speech. *J. Acoust. Soc. Amer.*, 1953, 25, 962~969.
- 10) Harris, C. M. : A speech synthesizer. *J. Acoust. Soc. Amer.*, 1953, 25, 970~975.
- 11) Halle, M., G. W. Hughes & J. P. A. Radley : Acoustic properties of stop consonants. *J. Acoust. Soc. Amer.* 1957, 29, 107~116.

Figure 1 : Broad band sound spectrogram showing the segmentation of syllables. In each spectrogram, cutting lines for the magnetic tape are indicated as dotted lines 'm' and 'n', which respectively bisect the recorded consonant segment 'bo' and the vowel segment 'oc'. The initial half of the consonant segment 'bm' was recombined with the latter half of the vowel segment 'nc' to make a stimulus syllable named "initial syllable". The residual portion 'mn' is then another stimulus syllable named "transitional syllable".

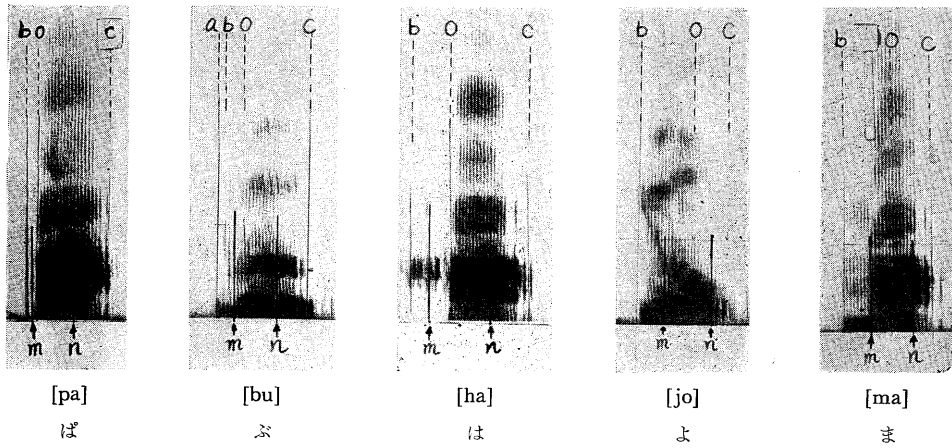


Figure 2 : Broad band spectrograms for the "initial syllables" and "transitional syllables". Each vertical pair of syllables was derived from one syllable indicated in the middle row.

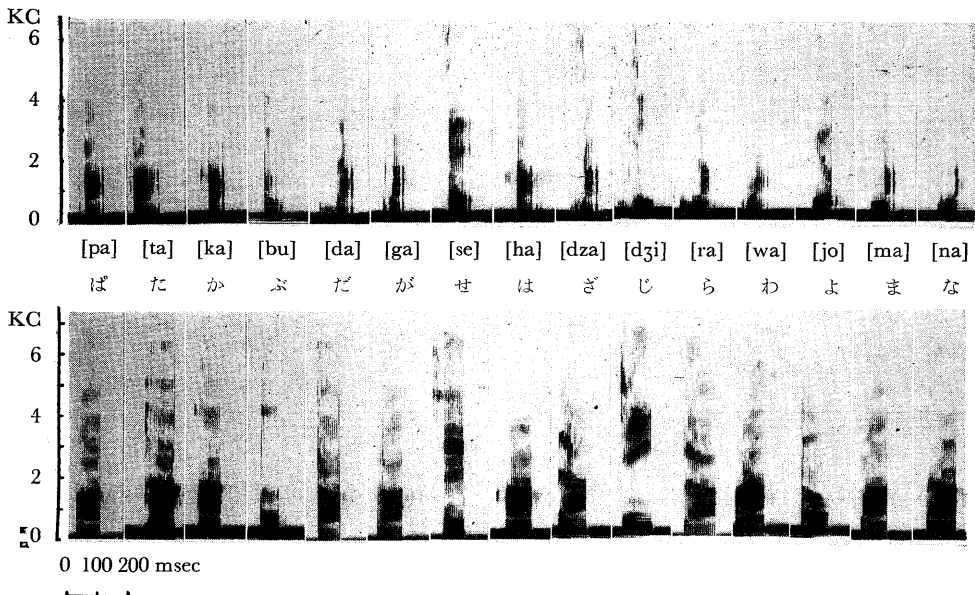
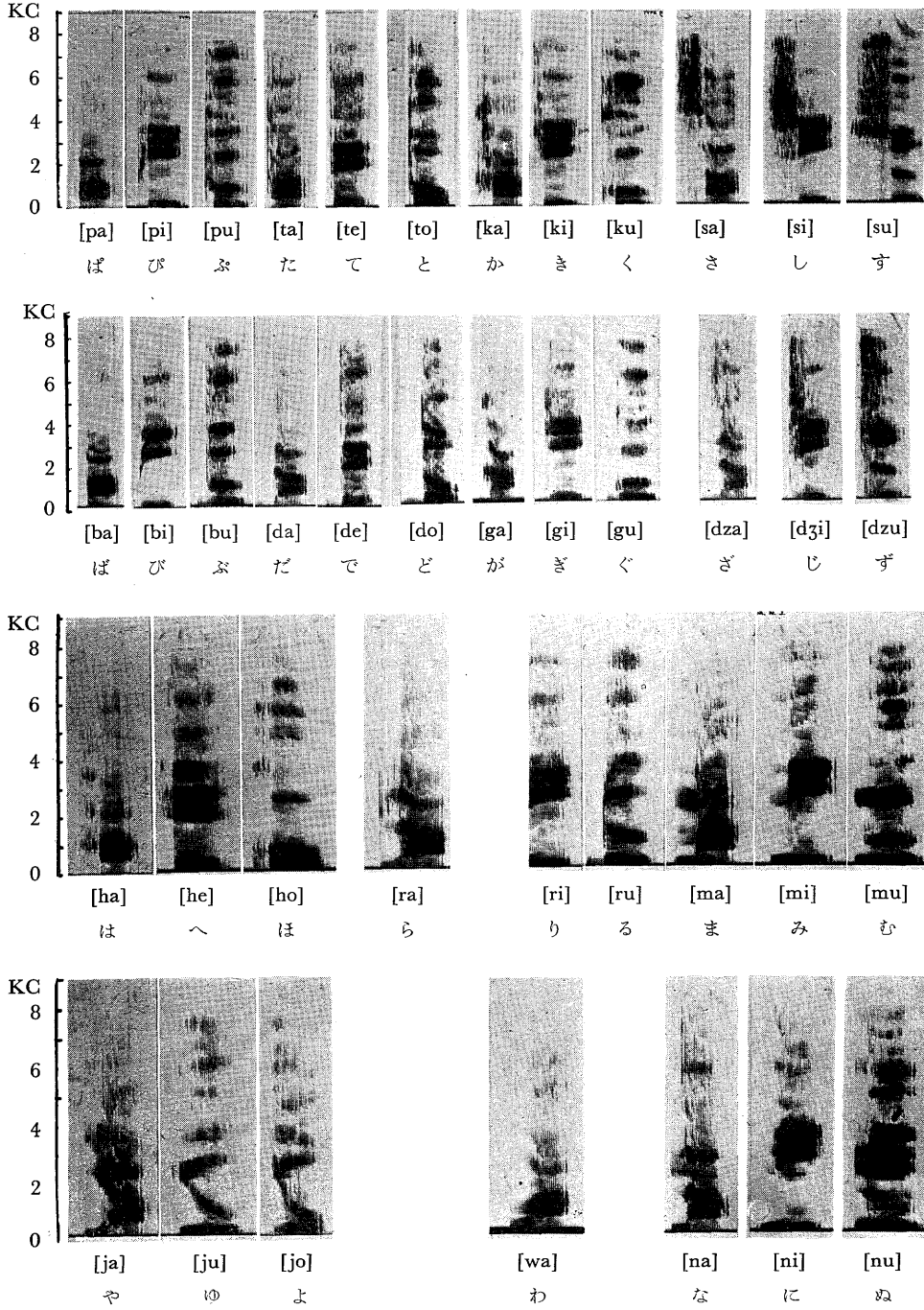


Figure 3 : Broad band sound spectrograms of Japanese syllables.



- 12) Harris, K. S. : Cues for the discrimination of American English fricatives in spoken syllables. *Lang. & Speech*, 1958, 1, 1~7.
- 13) 服部四郎 : 音声学, 1951, 岩波書店.
- 14) 井上誠一 : 日本語音声の性質とその弁別に関する研究, 京都大学工学研究科博士論文.
- 15) 釜本安敏 : 日本語音声の構音機構に関する電気音響器機による分析的研究, 耳鼻臨床, 1958, 51, 536~574.
- 16) Liberman, A. M. : Some results of research on speech perception. *J. Acoust. Soc. Amer.*, 1957, 29, 117~123.
- 17) Liberman, A.M., P.C. Delattre, L.I. Gerstman & F.S. Cooper : Tempo of frequency change as a cue for distinguishing classes of speech sounds. *J. exptl. psychol.*, 1956, 52, 127~137.
- 18) Liberman, A. M., P.C. Delattre & F. C. Cooper : The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, 1952, 65, 497~516.
- 19) Lisker, L. : Linguistic segments, acoustic segments, and synthetic speech. *Lang.*, 1957, 33, 370~374.
- 20) Lotz, J., A. S. Abramson, L.J. Gerstman, F. Ingeman & W.J. Nemser : Voicing cues : Perception of stop consonants after removal of a preceding fricative. *J. Acoust. Soc. Amer.*, 1959, 31, 1568~
- 21) Malecot, A. : Acoustic cues for nasal consonants. *Language*, 1956, 32, 274~284.
- 22) Miller, G. A., J. C. R. Licklider : The intelligibility of interrupted speech. *J. Acoust. Soc. Amer.*, 1950, 63, 176~185.
- 23) Peterson, G.E., S-Y. W. Wang & E. Silvertsen : Segmentation techniques in speech synthesis. *J. Acoust. Soc. Amer.*, 1958, 30, 739~742.
- 24) Potter, R. K., J. C. Steinberg : Toward the specification of speech. *J. Acoust. Soc. Amer.*, 1950, 22, 807~820.
- 25) 坂井利之, 堂下修司, 橋本清 : 日本語単音節の符号化, 電気通信学会, オートマトンと自動制御研究専門委員会資料, 1961.
- 26) 佐藤利男 : 有声無声破裂音の時間要素の差異について, 日本音響学会誌, 1958, 14, 117~122.
- 27) Steinberg, J.C. & N.R. French : Test of speech and music transmission and effect of distortion on speech and music. in H. Pender & K. Mc Ilwain (Eds.), *Electrical Engineer's Handbook* (3rd Ed.) New York : Wiley, 1936.
- 28) Stevens, K. N. : Effect of duration upon vowel identification. *J. Acoust. Soc. Amer.*, 1959, 31, 109.
- 29) Trites, D. K. : Graphic determination of significance of 2×2 contingency tables. *Psychol. Bull.*, 1957, 54, 140~144.
- 30) Wang, W. S-Y., G. E. Peterson : Segment inventory for speech synthesis. *J. Acoust. Soc. Amer.*, 1958, 30, 743~746.
- 31) Wiren, J., H. L. Staubs : Electronic binary selection system for phoneme classification. *J. Acoust. Soc. Amer.*, 1956, 28, 1082~1091.