

日本語母音における音継続時間と 明瞭度の関係について

太田文彦, 柳原尚明, 細田岩雄

The Intelligibility of Short Japanese Vowels as a Function of Duration

It is necessary to apply the speech sound continuously for a certain duration to identify it correctly. In this paper, studies on the intelligibility of Japanese vowels as a function of duration are described.

Procedure: A female voice (pitch ca. 280 c/s) was utilized, and the steady portion of the vowel wave was multiplied by the modulating rectangular pulse to produce the short sound wave. Duration of the rectangular pulse was varied from 5 msec to 100 msec. These samples were edited in random order and recorded to make a test tape.

Six trained listeners were selected in this experiment. On each stimulus presentation the listeners made two responses; (1) they attempted to identify the stimulus sound by writing down one of the five vowels; and (2) they assigned a rating of 4, 3, 2, 1, or 0 to indicate how confident they were that their attempted identification was correct.

Results: The rating score assigned by each listener was added up except a incorrect identification. The quotient that the total rating score was divided by total number of response was named "mean rating score" (MRS).

MRS rised with linearity against the logarithmic scale of vowel duration from 5 msec to 50 msec (Fig. 2 and 3). This linear regression was significant.

The regression coefficients of each of five vowels were not uniform; [I] was the greatest, and [E] [A] [O] and [U] followed in order.

The shortest duration necessitated to identify each vowel correctly was estimated from this regression line.

The shortest duration to be able to identify [I] was estimated at 19.6 msec;

[E] [A] and [O] was estimated 21.8 msec, 34.9 msec and 35.7 msec respectively. [U] was not identified even at 50 msec duration.

The soundspectrograms (Sonagram) were shown in Fig. 4-7. The longer the duration of vowel is, the more the figure clears.

Table 2 shows the confusion matrix for vowels. [U] was apt to be confused, and [A] [U] and [O] was confused each other. It is distinct with the similarity of the soundspectrogram among these three vowels.

Fumihiko OHTA, Naoaki YANAGIHARA and Iwao HOSODA

緒 言

音声の研究には種々の方法があるが、聴覚的立場からの研究は音声の本質から見て最も基本となるべきものである。聴覚的立場からの観察法としては明瞭度検査法がその主たる方法である。

我々は日本語母音の音声学的研究の一環として、母音認識に要する最短時間を測定せんとし、今回はその定常部につき検討を加えた。この問題に関する国内の知見は2, 3を数えるに過ぎず、しかもその判定に際しては研究者自身の主観的判断によりなされたものが多く、その成績は客観性に乏しい。

以上の見地より、我々は明瞭度検査法に rating response の概念を導入し、短時間母音の聞こえ方を客観的に把握すべく本実験を計画した。

実 験 方 法

1. 実験材料として女性アナウンサーの pitch 約280 c/s の5母音を用いる。
2. 実験装置の block diagram は Fig. 1に示す。5母音をそれぞれ endless tape と

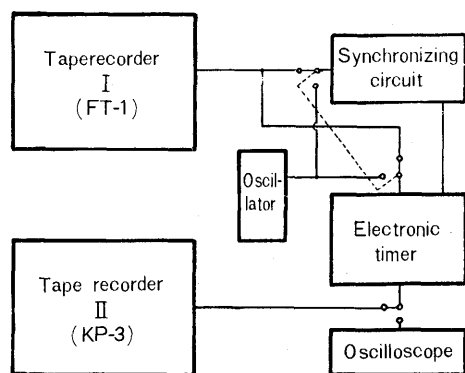


Fig. 1 Block diagram

して Taperecorder FT-1 (Sony) により再生し、その出力を synchronizing circuit および electronic timer の input に導く、timer については前著「聴覚における諸時間閾値に関する実験的研究」に詳述しているので省略する。synchronizing circuit の出力は negative trigger であって、その leading edge は入力波の負へ移る0点と時間的にほぼ一致している、そ

の negative trigger を timer の multivibrator の plate へ接続し、この trigger により multivibrator を働かせ波形の一定点から任意時間継続する矩形波を作り、それにより母音定常部を切りとり、任意時間の母音の sample を得ることが出来る。

Table 1. Results of rating responses.

Duration of sound time (msec)	U			O			A			E			I								
	rating score			Confusion	rating score		Confusion	rating score		Confusion	rating score		Confusion	rating score		Confusion					
	4	3	2	1	0	4	3	2	1	0	4	3	2	1	0	4	3	2	1	0	
5	1	1	1	1	8	3 o, o, a, 3 o, a, a, 4 o, o, a, a, 4 o, o, a, a, 4 o, a, a, a, 3 a, e, i, 4 o, a, a, i, 3 a, a, e, 4 o, o, a, a, 4 o, o, a, a, 4 e, e, i, i, 4 a, a, a, e,	2	1	5	3	1	u,	1	3	5	2	1	o,	1	6	3
10	2	2	3	2	2		3	2	1	4	1	u,	2	u, o,	2	3	5	4	3	2	2
15	3	2	2	2	2		4	5	2	4	1	e,	3	u, u, o,	3	4	4	4	2	3	4
20	3	4	1	1	4		3	2	3	1	1	2	u, u,	1	u,	3	4	3	2	o, o,	6
25	3	1	3	1	4		3	5	3	3	4	1	3	5	4	6	5	1	1	o,	10
30	3	3	2	2	4		4	6	1	1	u,	2	5	5	1	5	5	1	1	o,	7
40	4	3	2	2	4		3	7	2	2	u, u,	5	7	10	2	5	5	2	1	o,	12
50	6	2	4	4	3		4	7	3	3	5	7	2	u, u,	5	7	10	2	1	o,	12
60	4	4	1	1	3		3	5	6	1	e,	6	5	1	e,	8	4	10	2	u,	10
70	7	1	4	4	4		4	8	1	2	1	a,	4	6	2	u, e,	7	4	1	u,	11
80	5	3	4	4	4	4	4	3	3	2	a, e,	5	5	2	u, u,	10	1	1	9	3	
90	8	4	4	4	4	4	4	3	4	1	e,	7	4	1	10	1	1	1	12	12	
100	5	2	1	4	4	4	3	5	2	2	e, e,	7	4	1	u,	7	2	2	1	12	

timer の出力は Brown 管 oscilloscope で観察するとともに、Tape-recorder KP-3 (東通工) に導き録音する。

3. 音時間の長さを 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, および 100 msec の 13 種とし、5 母音につき得た 65 個の sample を random に並べかえたものを test 用 tape とする。test 用 tape における各 signal 間の間隔は 7 sec とする。

使用 tape は Synchro-tape (TDK), tape speed 19 cm/sec とする。

4. 明瞭度検査は京都大学音声科学研究室に於て行い同研究部員 6 名よりなる crew に聴取させる。応答は forced choice とし、その response に対する rating response を行なわせる。rating の与え方は明瞭にある母音にきこえるものを 4 とし、あいまいなものを 2, 全く判定不能のものを 0 とし、程度に応じて、

3あるいは1の rating を与えるよう指示する。

実験成績

1. rating の与え方: 6名の crew に2回ずつ聴取させた12の response の成績を Table 1 に示す。4あるいは3の rating を与えられた数は、各母音とも音時間の増すにつれて増加しているが、[O], [E] では70~80 msec 以上でかえって rating 2が増してくる。各母音とも 50msec 前後で rating 4 あるいは3が多い。

Table 2 Confusion Matrix

		Response					Total
		U	O	A	E	I	
Signal	U	×	13	22	5	4	44
	O	8	×	2	6		16
	A	10	2	×	2		14
	E	1	5		×		6
	I	3	1			×	4
	Total	22	21	24	13	4	84

2. 異聴: confusion matrix を Table 2に示す。出異聴は [U] [O] [A] [E] [I]の順に少なくなり、[U]では終始同程度の出異聴が見られる。また入異聴は [A] [O] [U]に多く [E] [I] に少ない。[I] は出入異聴とも著しく少なく入異聴はすべて [U] からである。

3. mean rating score (MRS) の推移: 異聴された場合は rating score を0として与えられた rating の合計を求め平均したものを

mean rating score (MRS) と仮称する。

これを横軸に音継続時間の対数をと縦軸に MRS をとって plot すると (Fig. 2.3)

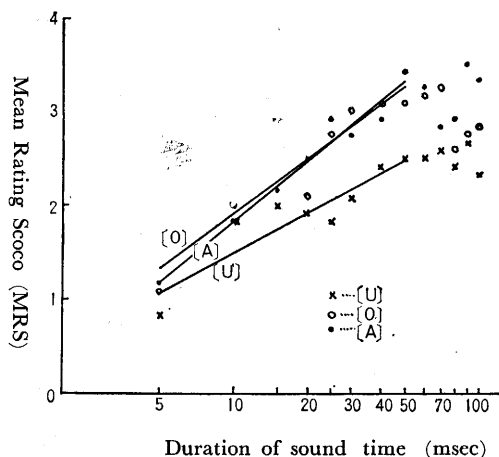


Fig. 2: Relation of MRS to sound duration of vowels [U] [O] and [A].

The abscissa shows the logarithmic scale of sound duration. The ordinate indicates the mean rating score (MRS) obtained for six listeners. The solid lines show the regression line for [U] [O] and [A].

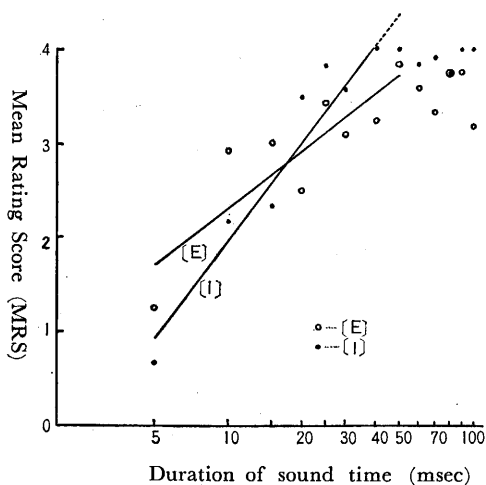


Fig. 3 shows the data of vowels [E] and [I]; same scale as Fig. 2.

各母音とも40~50msec までは上昇するが、50m sec 以上では横這いあるいはむしろ低下する傾向が見られる。

5~50m sec の MRS 上昇はほぼ直線的であるのでこの回帰直線を求めると、

$$[U] \quad y = 1.41 \log x + 0.10 \quad F(0.01) = 30.8^{**}$$

$$[O] \quad y = 1.94 \log x - 0.01 \quad F(0.01) = 23.2^{**}$$

$$[A] \quad y = 2.41 \log x - 0.30 \quad F(0.01) = 177.5^{**}$$

$$[E] \quad y = 2.03 \log x + 0.28 \quad F(0.01) = 17.4^{**}$$

$$[I] \quad y = 3.46 \log x - 1.47 \quad F(0.01) = 68.7^{**}$$

となり、何れも1%の危険率で有意である。

[I] は38.1 msec で MRS 4 に達するが他の4母音は 50 msec でも MRS 4 に達しない。

5母音全体の平均では

$$y = 2.20 \log x - 0.28 \quad F(0.01) = 106.9^{**}$$

となる。

4. Sonagram : (Fig. 4-7) 音時間10msec では formant 帯域は不分明であるが、

5母音間に少しずつ pattern の差を認めることができる。

20 msec ではその差はさらに明らかとなり、30msec, 50msec と音時間の長くなるにつれて著明となってくる。しかし [U] [O] [A] については、その pattern はよく似ており、特に低音域の pattern が不鮮明である。

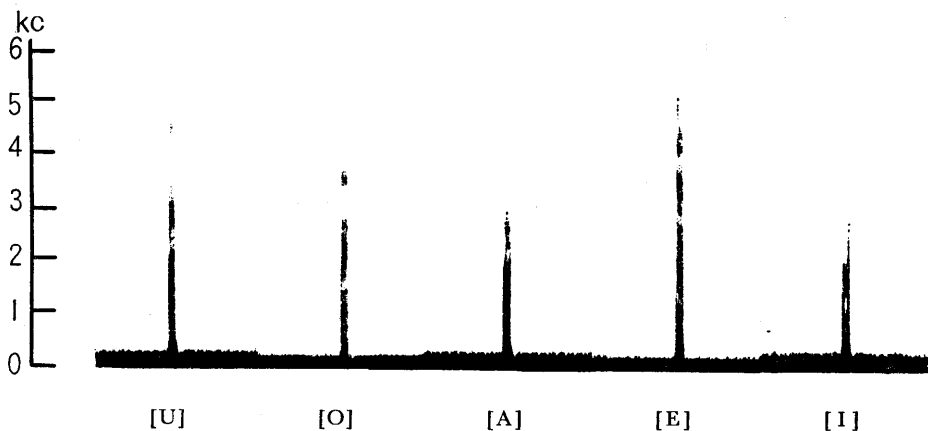


Fig. 4

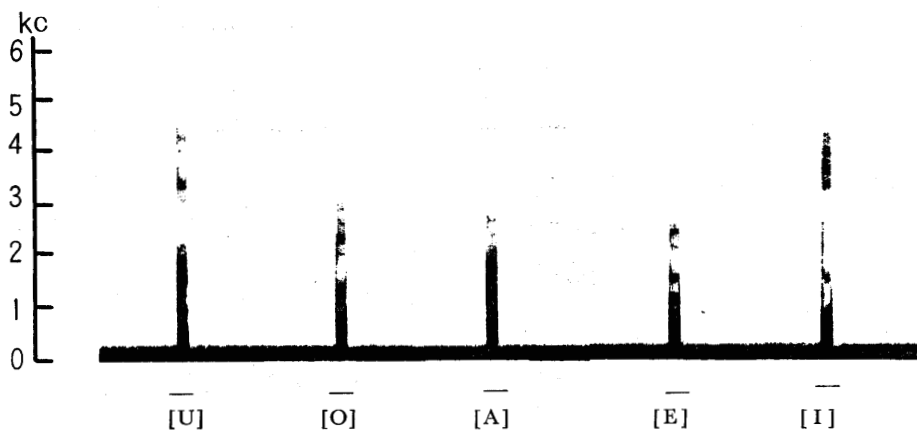


Fig. 5

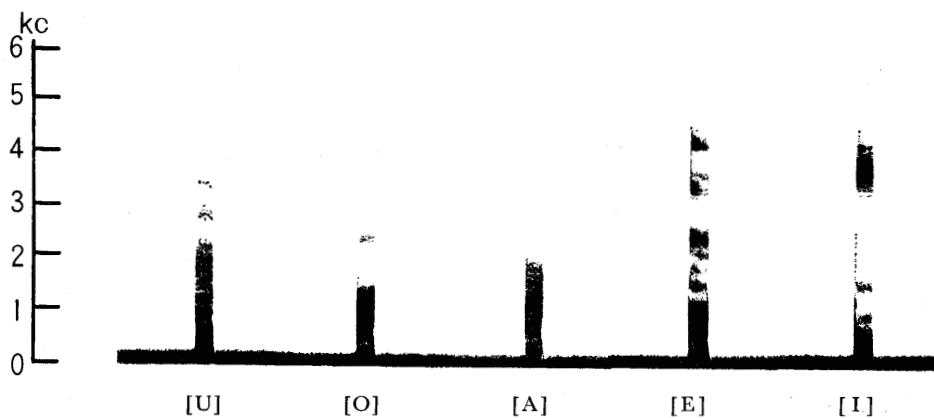


Fig. 6

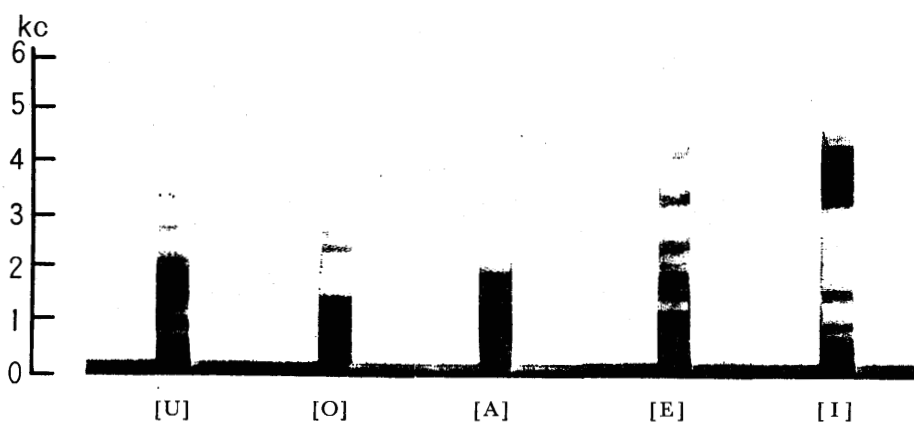


Fig. 7

Fig. 4, 5, 6 and 7 illustrate the sound-spectrogram (Sonagram) of short Japanese five vowels. Duration of these short vowels is 10 msec in Fig. 4, 20msec in Fig. 5, 30msec in Fig. 6 and 50 msec in Fig. 7 respectively.

綜括ならびに考按

1. 短時間音の物理的特性について

純音を短時間の矩形波で切断すると、その物理的特性は原音の周波数成分のみより成るものではなくなる。その最大 energy は発振周波数のところにあるが、その両側にも対称的に energy の分布が見られ、その上 energy は $f_0 \pm n/d$ (f_0 : 発振周波数, n : 倍音の数, d : 音時間) のところで0となる。すなわち中心帯の幅は $2/d$ である。それ故音時間が2倍になれば中心帯の幅は $1/2$ になるが、中心帯の占める energy の比率は大となる。そして音時間の増加により音は click より純音に近づく、一方音の始めと終りでは energy の大部分は中心帯よりもむしろ側波帯にある。複合音に対してはこの関係は複雑ではあるが、切断による周波数成分の変化により、短時間母音は原母音と異なった周波数成分を示すことは当然考えられる。これを Sonagram により観察すると、時間が長くなるにつれてその pattern がはっきりしてくる経過は明らかである。たとえば [I] の Sonagram を時間を追って詳しく観察してみると Fig. 4 (10 msec) では始めと終りに click と考えられる縦の線があり、中央部のみが [I] の pattern らしきものを呈している。Fig. 5 (20 msec) では両端の黒線はやや細くなり、黒化度を見てもその部の energy の減少を示している。中央部の模様は [I] の pattern に近くなっている。Fig. 6 (30 msec) では前端の click はやや強いが中央部の pattern はますます明瞭となってくる。Fig. 7 (50 msec) でもほぼ同様であるが pattern の縞模様が 30 msec よりもさらに鮮明になっている。他の4母音についてもほぼ同様である。すなわち 10 msec では前後端の click の勢力が大きい点、各母音の pattern がまだはっきりしない点等より当然異聴のおこることが予想される。しかし pattern そのものも 50 msec でもまだ低音域の黒化度が強く energy 量の大きいことを示しており、切断による周波数成分の変化があると考えねばならない。[U] [O] [A] 間の異聴の多い点は Sonagram から観察されるこれらの点とも一致している。

2. 音時間と明瞭度の関係について

すでに述べたごとく音時間 50 msec までは MRS は時間の対数に比例して上昇するが 50 msec 以上では MRS はほとんど上昇しない。Fig. 2, 3 に示す直線を延長しても

plot された MRS 値は直線より離れ, 水平あるいはむしろ低下の傾向を示す。そこで 50 ~100 msec の値について回帰直線を求めてみると, 回帰係数の値は, [U] -0.198 , [O] -1.500 , [A] 0.016 , [E] -1.250 , [I] 0.090 となり, [A] [I] はほぼ水平, [U] [O] [E] は低下の傾向を示している。しかしこれらの回帰直線は [A] については 5% の危険率で有意であるが, 他の 4 母音については有意性をみとめない。したがって 50 msec 以上における上述の傾向は偶然の所産ではないとは言えない。

次に 5~50 msec で得た直線を MRS 4 まで延長外挿してみると, [I] はすでに 38.1 msec で MRS 4 に達しているので除外すると, [U] 583.4 msec, [O] 116.7 msec, [A] 102.2 msec, [E] 68.0 msec となり, 5 母音全体では 88.2 msec となるから, もし MRS が 50 msec 以上でも直線的に上昇するとすれば, 88.2 msec ですでに全 response が rating 4 を示していなければならない。しかるに [U] [O] [A] の値はそれより遙かに大きく, 88.2 msec では MRS は 4 以下であり矛盾する。すなわち 50 msec 以上までの直線を外挿することはできないと考えられる。これは [U] [O] [A] に異聴の多いことと考え合せて, MRS を求める際の異聴の取扱いに問題が残されていると考えられる。そこで異聴を除外し正答のみについての rating score の合計を正答数で除してみると (Table 3)

Table 3

		U	O	A	E	I
Duration sound time (msec)	5	0.83	1.18	1.17	1.36	0.80
	10	2.44	2.18	2.20	2.92	2.60
	15	2.67	3.18	2.49	3.00	2.34
	20	2.87	2.50	2.72	3.00	3.50
	25	2.74	3.00	2.92	3.42	3.83
	30	3.12	3.27	2.75	3.36	3.58
	40	3.22	3.08	3.18	3.25	4.00
	50	3.75	3.70	3.42	3.83	4.00
	60	3.34	3.46	3.54	3.66	3.83
	70	3.87	3.54	3.40	3.64	3.92
	80	3.62	3.10	3.50	3.75	3.75
	90	4.00	3.00	3.50	3.75	4.00
	100	3.50	3.00	3.64	3.45	4.00

Table 3 shows the quotient that the total rating score was divided by the number of correct identification.

These values are larger than MRS in general.

上昇率は変るがやはり 50 msec の値が最高値に近い値を示している。すなわち MRS は 50 msec 附近までは音時間の対数に比例して上昇するが, それ以上は時間を増しても良くなないと考えられる。50 msec における MRS 値は [I] を除いては 4 以下である。これを回帰直線より推定すると, [U] 2.48 ± 0.25 , [O] 3.29 ± 0.41 , [A] 3.34 ± 0.16 , [E] 3.73 ± 0.49 となり, [E] 以外はその上限も 4 に達しない。しかし 5 母音の弁別がつく程度のものには少なくとも rating 3 が与えられているので, MRS が 3 に達する音時間を求めてみると, [O] 35.7 msec, [A] 34.9 msec, [E] 21.8 msec, [I] 19.6 msec で, [U] は 50 msec でも MRS 3 に達しない。すな

わち [I] [E] は 20 msec 前後, [O] [A] は 35 msec で弁別可能となる。

村上が [A] [E] [I] は 15~20 msec で充分であるが, [O] [U] はこの時間内に判断不能な例があると述べている成績ともほぼ一致する. しかし服部は tape を切断して 5 母音とも 10 msec まで短縮しても正しく判断出来ると述べている. また田口は [O] 音は波数 2 個で [O] と明瞭に聞こえる述べ, 彼の行った純音の実験成績と一致すると述べている. 純音につき著者の一人太田が以前に実験した結果では, 閾値上 50db における音高知覚時間閾値は 250c/s で 16.68msec, 1000c/s で 11.63msec, 4000c/s で 8.16 msec, 8000c/s で 6.36msec であり, 高周波数音ほど短縮しているが波 2 個で pitch が判るという成績は得られなかった. しかし我々の純音の成績からみても 250c/s で 16.68msec という値は [I] [E] の値と近く, 母音認識に要する時間は大体その pitch 附近の純音の高さを認識するに要する時間に近いと見ることが出来る. ただ, 田口, 服部の値との差は, 我々の用いたのが母音定常部である点と, 明瞭度検査の方法とにその原因があるのではないかと考える.

3. 異聴について

Table 2 に見られる通り, 出異聴は [I] [E] [A] [O] [U] の順に多くなり, 入異聴は [A] [O] [U] の三つがほぼ同数で他の二者に比し遙かに多い. Table 1 から判るように [U] は音時間に関係なく出異聴が多く, その大部は [A] または [O] と聞かれている. このことは [U] という母音が切断という歪に対して抵抗性がなく [U] としての pattern を失い易いのではないかと考えられる.

ここで異聴現象について少し考えてみたい. 我々がある音声を聞いてそれを正しく判断するために, 我々のもつ音韻体系の中にその音声に対する手がかり (cue) を持っている. しかしその cue は個人によっても少しずつ広がり異なるかも知れない. 万人が正しく聞く音声というものは各人の cue の一致している部分にその音声の pattern があるからである. 歪語音を用いると各人のもっている cue は変らないが, 歪のために音声の pattern は変り, 聞こえたときに他の母音の cue の範囲内に入るか或はどの cue にも入らず外へ出てしまうかであろう. しかしその場合でも聴取者は 5 母音の何れかに聞こうと努力するために, 個人によっていくらか異なる cue の中へそれをとり込もうとするため異聴がおこってくる. どの cue へとり込もうとするかが入異聴から推定され, そこから 5 母音のもつ cue の広さが推測されうる.

こういう考えから見ると母音定常波を矩形波で切断するという歪は [U] の pattern を乱しやすく, 一方他の母音の pattern の乱れは音韻体系のなかにある [U] [A] [O] の cue の中にとりこまれやすいと考えられる.

結 論

我々は日本語5母音の定常波を矩形波で切断し, その継続時間と明瞭度の関係につき実験を行ない, 次の結論を得た.

1. 母音明瞭度の MRS は音時間 50 msec までは音時間の対数に比例して上昇するが 50 msec 以上ではこの関係は認められない.

2. MRS の上昇率は5母音間に差があり, [I] が最も急峻で, [E] [A] [O] の順に緩やかとなる. [U] は最も緩やかな上昇を示し, 50 msec でも MRS が3に達せず異聴されることが多い.

3. Sonagram の pattern は音時間が長くなるに従って明瞭となるが, [U] [O] [A] では低音域の energy が強く, 類似した pattern を示す.

擧筆にあたり恩師後藤光治教授の御指導御校閲を深謝します.

本論文の要旨は第6回耳鼻咽喉科臨床会 (37.11.10) において口演した.

主 要 文 献

- 1) Stevens, S.S. : Handbook of Experimental Psychology (1960)
- 2) Clarke, F.R. : J.A.S.A. 32 : 35 (1960)
- 3) Guilford, J.P. (秋重訳) : 精神測定法 (1959)
- 4) 村上 : 日耳鼻 57 : 416 (昭29)
- 5) 中島 : 聴感と音声 (昭35)
- 6) 太田 : 耳鼻臨床 50 : 659 (昭32)
- 7) 服部 : 耳鼻臨床 49 : 886 (昭31)
- 8) 竹内 : Studia phonologica I : 70 (1961)