

単音節符号化装置とその会話音 分析への二,三の応用

坂井利之 堂下修司 橋本清

A Research Model of Japanese Monosyllable Recognition and Some Applications to the Conversational Speech Analysis;

Many attempts at the mechanical recognition of speech sound have been made and yet left many problems unsolved. Though the analysis method to be used is common to all the languages, the difficulty of performing recognition is largely dependent on the particular language. The Japanese language is considered to be easier in processing than the English; for example, rather fewer cardinal vowels, simple monosyllable structure, and one to one correspondence between these monosyllables and the Japanese alphabet.

In Japanese speech sound, the elemental units of articulation are monosyllables of one hundred odd, each of which consists of one consonant and one following vowel (some include semivowel, but no diphthong). The conversational speech sound is principally regarded as the successive utterances of these monosyllables with some modification by the influences from preceding and following sound.

From this point of view, we first tried to make a mechanical recognizer of Japanese monosyllables, and then extend it so as to be able to handle conversational speech sound by adding some functions to it.

In this paper, the outline of the research model of the phonetic typewriter, which operates for Japanese monosyllables, is described.

Speech sound has to be processed not only from the acoustical standpoint, but also from the linguistic one. According to the relations between them, many systems are obtained. For instance, the one which makes linguistic recognition directly from the data acoustically analyzed, the other which, after recognizing phonemes or monosyllables from the acoustical data, applies to them the linguistic informations on syntax and redundancies. The machine presented here takes the method of monosyllable discrimination at the acoustical level. The followings are

T. Sakai. 坂井利之 : P.H.D. Professor of Electrical Engineering Department, University of Kyoto
S. Doshita, 堂下修司 : Graduate School of Electrical Engineering Department, University of Kyoto
K. Hashimoto. 橋本清 : Electrical Communication Laboratory, Tokyo

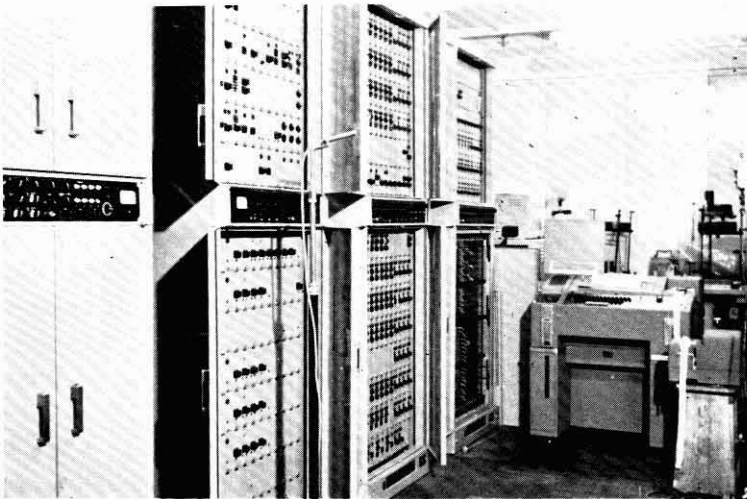
the outlines.

(1) This machine accepts as the input Japanese monosyllables separately articulated from each other. (2) Analysis is made by distinctive feature extraction and zero-crossing interval analysis which correspond to the manner of articulation and the place of articulation in utterance. From the result of the distinctive feature extractor, the rough classification of input monosyllables is carried out, such as pure vowel, unvoiced, voiced, nasal, stop and contracted. From the zero-crossing interval analyzer, channel outputs are obtained, which are classified by the zero-crossing time interval of each rectangular wave. (3) All the control signals are derived from the input speech sound wave itself. (4) Discrimination of vowel part and consonant part is made by separate circuits, using digital technique such as order pulse, AND gate, OR gate and binary circuit. For consonant part, each channel output of zero-crossing wave analyzer is divided with the threshold level which is previously decided by the statistical data. (5) All the obtained results are retained by register memories and gathered in the main matrix, where the final decision of input monosyllables is performed. (6) The results are indicated by the lamp indicator and also sent to high speed puncher or printer through code converter.

These operations are carried out in real time, and by detecting the end point of speech sound final output is generated.

In Fig. 1 the over-all block diagram of the phonetic typewriter is shown, and it is subdivided into the following parts:

(1) INPUT PART amplifies speech sound to a proper level and feeds it into each circuit of analyzing part. (2) ANALYZING PART performs phoneme classification and analyzes the vowel part and the consonant part. (3) JUDGING PART decides the phoneme of the vowel part and the consonant part, and then gets the final result of a whole monosyllable. (4) OUTPUT PART generates coded signals to control puncher and printer, and also resets the whole circuit in preparation for the next input.



Over-all view of the phonetic typewriter

Next, on the basis of the results of the above mentioned machine, an effort to translate conversational speech into machine codes is presented in the latter part of this paper. By adapting the machine to the continuous analysis and using displaying device, three dimensional pattern of distribution of the zero-crossing wave analysis in f_1 - f_2 domain vs time axis was visualized on oscilloscope. Fig. 8, Fig. 9, Fig. 10 are the examples of patterns of monosyllables, words which involve typical cases to be taken into account in conversational speech, spoken digits in Japanese and English, respectively.

Toshiyuki SAKAI, Shuji DOSHITA, Kiyoshi HASHIMOTO.

1. 緒 言

音声に含まれる情報は、(i) 個人差、感情等をあらわす自然性、(ii) 音韻情報、とになる。音声の伝達手段としての特徴は (i) にあるが、言語機能の点から見ると、基本的なものは (ii) に含まれている。音声の占める帯域は数千サイクルであるが、音韻情報のみに注目して音声を文字ないし符号で表現すれば、約 100 サイクルで、音声の符号化伝送が可能となり、通信の分野等において極めて有効な手段となる。

人間の情報伝達手段として音声は、意志伝達の速さ、エネルギーの経済性等において、最も一般的で不可欠なものであるが、もう一つの意志伝達手段である文字との間には、直接的な関連はなく、両者の変換には必ず人間の仲介を必要とする。これを機械により直接に音声から文字（又はそれに対応する符号）への変換を試みるのが、音声タイプライターであり、又逆の変換が自動発声装置である。話された言語を逐語的に文字に変換しても、完全な文章を形成せず、音声タイプライター出力を文字で表現する場合、困難な問題であるが、ある種の目的においては十分であり、又、狭帯域通信等で、音声として再現して聞く場合には、その様な問題はなくなるわけである。

以上の目的より、我々は先に試作した装置¹⁾の結果から、単音節符号化装置を試作し、この装置を利用して会話音声の分析への応用を試みたので、それらの概要を報告する。

2. 符号化装置の基本的条件

音声を扱うとき、音声的な特性の外、言語的な特性も重要である。従って、音声の最終的な識別は音響的な面の分析や識別のみでなく、言語情報をも含めて考えなければならない。この二つの面の組合せにおいて、音響分析の結果を直接に言語認識に用いる場合や、

音声の段階で先ず音韻として認識しそれに言語知識を用いる場合等考えられる。我々は、先ず音韻の識別を目的とし、音声的な面から分析及び弁別を試みた。その概要は次のようである。

- 1) 入力日本語単音節である。
- 2) 分析には、発声における調音形式 (manner of articulation) 及び調音位置 (place of articulation) に対応した専用の抽出回路を設ける。
- 3) 子音部、母音部は別々に分析して、最後に総合判定マトリクスで結果を総合する。このとき両分析結果は時間的に前後して得られるので、同時に判定するため、記憶保持回路を用いる。
- 4) あらゆる制御指令は入力となる被分析波より得る。
- 5) アナログ・デジタル変換の閾値、マトリクス回路等は、変更容易である。
- 6) 出力は、テレプリンタにより、カナ文字で得ることも、又それに対応する符号として、パンチする事も出来る。

日本語会話音声は比較的単純であり、その構成要素は百音節といわれるものであり、会話音声はこれらの連結したものに、若干の音結合による影響を考慮すればよい。従って日本語音声の基本的な性質はこの単音節に含まれているものと考えられ、又、装置の簡単のためにも入力として単音節のみを考える。

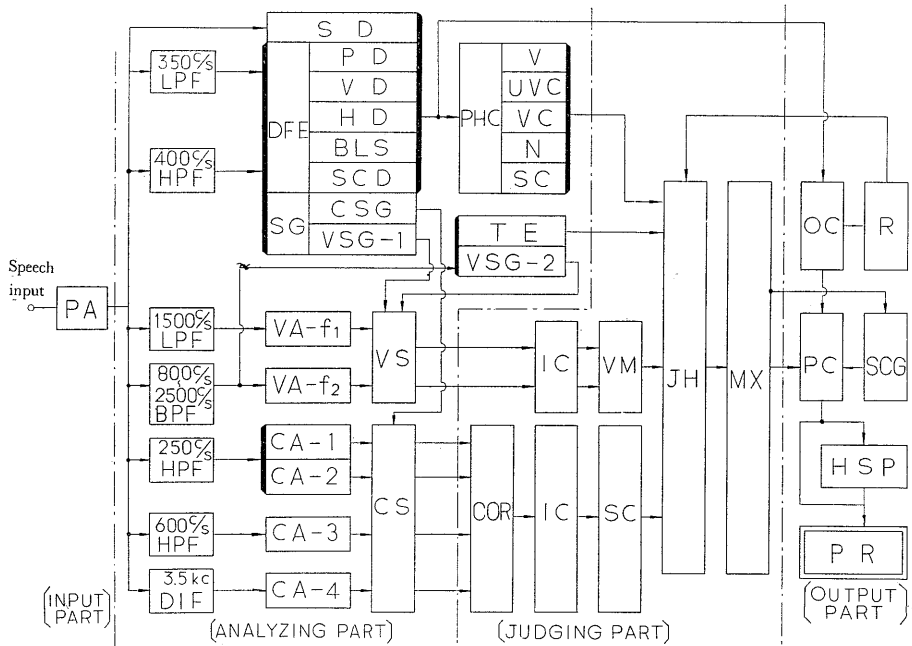
第1図は本装置のブロック図であり、次のような各部にわけられる。

- 1) 入力部 2) 分析部 3) 判定部 4) 出力部。

入力部では分析部の各回路に適当なレベルの信号を分配する。分析部では音韻種別の判定及び、母音、子音の分析を行ない、その結果から判定部で、子音及び母音の判定及び、単音節としての総合判定をする。出力部ではパンチャ、プリンタを動作させると共に、回路のリセットを行なう。

扱う量がアナログ・デジタルにわたるため、トランジスターを用いて回路を構成し、実験に便利なように、シートに組まれている。

入力音声のレベルは発声時に出来るだけ揃えるようにし、各単音節の母音部において10 db 以内の変動におさえる。機械の動作・制御・指令は全て入力音声から取り出し、他の部分へ送り、母音及び子音サンプル区間、及び出力回路を制御する。音声は発声過程において、調音形式と調音位置によって特徴づけられる。実際には前後の音韻間の相互影響による二次的な変化も大きな要素となってくるが、基本的には上の二つの面が一次的特徴



PA	Preamplifier	前置増中回路	SC	Stop Consonant	破裂音
LPF	Low Pass Filter	低域フィルタ	TE	Transient Extractor	拗音分離回路
HPF	High Pass Filter	高域フィルタ	VA	Vowel Analyzer	母音分析回路
BPF	Band Pass Filter	帯域フィルタ	VS	Vowel Sampling Gate	母音サンプルゲート
DIF	Differenciater	微分回路	CA	Consonant Analyzer	子音分析回路
DFE	Distinctive Feature Extractor	弁別要素抽出回路	CS	Consonant Sampling Gate	子音サンプルゲート
SD	Speech Detector	音声区間抽出回路	IC	Integrating Counter	蓄積計数回路
PD	Pitch Detector	ピッチ抽出回路	VM	Vowel Judging Matrix	母音判定マトリクス
VD	Vowel Detector	母音区間抽出回路	COR	Consonant Channel OR Circuit	子音チャンネル再編成
HD	High Frequency Detector	高周波成分抽出回路	SC	Schmidt Circuit	2値判定回路
BLS	Buzzless Speech Detector	ブズレス音声区間抽出回路	JH	Judgment Holding Register	判定保持回路
SCD	Stop Consonant Detector	破裂音抽出回路	MX	Main Judging Matrix	総合判定マトリクス
SG	Sampling Interval Generator	サンプル信号発生回路	OC	Output Control	出力制御回路
CSG	Consonant S. G.	子音サンプル信号発生回路	R	Reset Circuit	リセット回路
VSG	Vowel S. G.	母音サンプル信号発生回路	PC	Puncher Coder	コード変換回路
PHC	Phoneme Classifier	音韻大別回路	SCG	Special Code Generator	特殊符号発生回路
V, UVC } see Table 1		第1表参照	HSP	High Speed Puncher	高速度パンチャ
VC, N }			PR	Printer	プリンタ

第1図 装置のブロック図

Fig. 1. Block diagram of a research model of the phonetic typewriter

であると思われる。そこで我々は、第1表に従ってこの両面から音声を扱い、音韻大別回

第1表 日本語音韻表 (右欄は第1図 PHC による分類を示す)

Table 1. Classification of Japanese phonemes (The right column shows the Classification in PHC of Fig. 1)

音 韻 Phoneme	有 声 音 Voiced	母 音 Vowel ; a , i , u , e , o 	母 (V 音)
		子 音 Consonant	} 一般有声音 (VC)
摩擦音 Fricative ; z , ʒ , (dz , dʒ) (破擦音を含む Including affricate)			
破裂音 Stop ; b , d , g , r (弾音を含む Including flapped)			
鼻 音 Nasal ; m , n 	鼻 子 音 (N)		
		半 母 音 Semi-vowel ; w , j 	半 母 音 (T)
	無 声 子 音 Unvoiced	摩擦音 Fricative ; s , ʃ , h (気音を含む Including aspirate)	} 無 声 子 音 (UVC)
		破 擦 音 Affricate ; ts , tʃ	
		破 裂 音 Stop ; p , t , k	

路において調音形式に対応する特徴的な性質から音声をいくつかのグループに大別し、更に調音位置に対しては、零交叉波間隔による分析を行なう方法を取り、又一部の音韻に対しては時間的な変化をも考慮した。

子音及び母音の区間の区切りのはっきりした定義は困難であるが、雑音的、過渡的な子音部分と、準定常的、周期的な母音部分とは存在し、この基準に従って、母音子音の区間を分離して扱う。その必要は、両者が性質を異にするものである事、及び、母音の強大なエネルギーによわい子音部がマスクされるのを防ぐためである。母音の判定は第1ホルマント (f_1)、及び第2ホルマント (f_2) に相当するピークを求め、 f_1-f_2 図上に於て行なう。子音部分は主に零交叉波分布を求め、各音韻に適した分類特性を選び、その結果を2値化して行なう。これらの結果は保持され、総合判定マトリクスにおいて組合わされ、音声終端信号で出力回路に出される。従って、子音は母音に先行するが子音の分析結果に母音の結果を組合わせることにより子音に対する母音の影響が考慮されている。

3. 弁別要素抽出回路及び音韻大別回路

零交叉波ではレベルに関する量は全く除かれているので、レベル情報は、これらの弁別要素抽出回路によって検知し、音韻大別を行ない、又、分析系統に対して制御信号を送る。

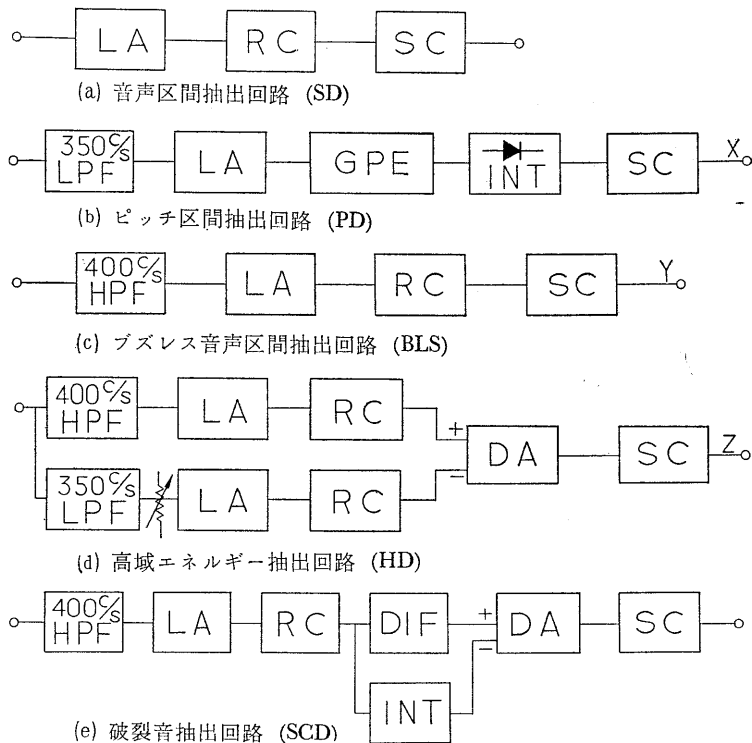
(1) 弁別要素抽出回路

日本語単音節の構成を弁別の面から大きく

先行音部+子音部+過渡部+母音部

にわけて扱う。入力音声はこれらのうちいくつかは組合さって構成されているので、入力にどれが含まれているかを検出し、純母音、無声子音、一般有声子音、鼻音、拗音に大別し、更に子音部の立ち上りの波形の鋭さより破裂音を検出する。

先ず音声を350c/s LPF 及び400c/s HPF によって大きくわけ第2図のように回路を構成



LA ; Linear Amplifier	増 巾 器	INT ; Integration	積 分
RC ; Rectifier	整 流 器	DIF ; Differentiation	微 分
SC ; Schmidt Circuit	シュミット	DA ; Difference Amp.	差 動 回 路
GPE ; Grüentz Type	ピッチ抽出	LPF ; Low Pass Filter	低域フィルタ
Pitch Extractor	回 路	HPF ; High Pass Filter	高域フィルタ

第2図 弁別要素抽出回路ブロック図

Fig. 2. Block diagram of the Distinctive Feature Extractor

する。同図 (a) は音声区間を求める回路で、音声を直接増巾整流した後、夫々約0.1ms, 10ms の充電、放電時定数で包絡線を求め、2値化する。この出力は特に弁別に寄与しないが、出力制御、回路のリセット等、装置全体としての制御を行なう。同図 (b) は 350c/s

LPF を通した後, Grüntz 型のピッチ抽出器を構成している. 得られたピッチ周期のパルスは, 蓄積回路に通し, ピッチの区間を取り出している. この回路は有声音の区間でのみ働くように設定されるが, 母音初期や無声破裂音等で正しく有声音区間を分離することは困難である (この出力を X で表わす). (c) 図では, 400c/s 以上の成分を取り出し 2 値化した出力 (Y で表わす) を得, 無声子音, 母音, 鼻音性先行者の区間で動作する. 又, 破裂性の音の早い立ち上りにも応答するように, 整流の充電時定数は十分小さく選ばれている. (d) 図では, 400c/s 以上の成分が, 350c/s 以下の成分に比しある値より大きいとき "1", 小さいとき "0" という出力を生じる (Z で表わす). (e) 図は破裂音検出回路で, 400c/s 以上の成分をとり出し, 成分が全くないときに, 急激な立ち上り波形が生じた場合にそれを検出する. 又破裂音でも |p|, |t| と |k| では性質を異にするため独立の回路を設ける. これらの出力は最終判定まで保持され, 無声子音の弁別に用いられる.

(2) 音韻大別回路

我々の装置では原則として音韻大別と分析とを別々に扱い, 両者の相互関係は考慮してい

入力音声 Input monosyllable	母音 V	無声子音 UVC+V	有聲子音 VC+V	鼻音 N+V
X				
Y				
Z				
母音 (XZ) Vowel				
無声子音 (XY) Unvoiced				
先行音 (XȲ) Buzz				
☒ 鼻音 (XYȲ) Nasal				
判定 Decision of PHC	右以外の場合 Nodetection of consonant	$X\bar{Y} > 10\text{ms}$	$X\bar{Y} > 30\text{ms}$	$X\bar{Y}\bar{Z} > 30\text{ms}$

註 Vowel Part 母音部 Noise Part 雑音部
Notes. Buzz Part 先行音部 Nasal Part 鼻子音部

第3図 音韻大別回路動作説明図

Fig. 3. Operation of the phoneme classifier

ない。このような簡単化された方法で音韻大別を完全に行なう事は困難であり、分析結果をも併せて音韻大別をする必要がある。

音韻大別は第1表の右欄に示したように行なう。弁別要素抽出回路で得た出力X, Y, Zの各種音韻入力に対する関係は第3図に示される。先ず区間が選ばれる。即ち、母音区間はXZ, 無声子音区間は $\bar{X}Y$, 先行音区間は $X\bar{Y}$, 鼻子音区間は $XY\bar{Z}$ として得る。音韻大別はこの各々の区間出力が第3図に示された期間以上つづいた場合にそれぞれの音韻グループとして別けられる。即ち、「無声子音」は $\bar{X}Y$ が10ms以上、「有聲子音」は $X\bar{Y}$ が30ms以上、「鼻音」は $XY\bar{Z}$ が30msつづいた場合に判定が行なわれ、その結果は保持され、以後生ずる他の判定は採用されない。上の様な状態が無く、XZなる状態が生じた場合には「純母音」と判定される。

これらの回路による方法の誤判定について考える。弁別要素抽出は、高域、低域の2個のフィルターのみでは不十分であり、又回路は絶対レベルで動作するために、入力音声レベルをある一定範囲内に保つ必要があり、この制限は特にピッチ区間抽出回路で問題となる。又、母音の発声初期でははっきりしたピッチ性をもたない場合も多く、一方|P|, |t|等ではピッチ類似の成分をもつため、この回路の動作開始特性は安定したものでなく、「無声子音」と「母音」との間の誤判定を生ずる（これに対しては分析の面からの結果も考慮する予定である）。

一般有聲音に関しては先行音の存在を仮定している。先行音は、有聲音の弁別に本質的なものでないが、ほとんどの場合伴って生じ、又、合成の面からも有聲の知覚に有効な事が報告されている²⁾。先行音が弱く、十分に検出し得ない場合には母音に誤判定する。一般有聲子音の先行音は原則として鼻に抜けないが単音節では、鼻音性の先行音を伴う場合も比較的多い。しかしこのときには、子音部の前に先行音の減衰及び急激な雑音成分の発生があり、これを用いて「有聲子音」を「鼻音」と区別する。

上の回路とは別に拗音の分離は第2ホルマンントの大きな動きを検出して行なう³⁾。後述する第2ホルマンント分析システムを用いて、帯域フィルターで800c/s~2500c/sの成分を取り出し、零交叉波に変換してから、ホルマンント周波数を電圧に直し、2400c/s附近から、後続母音の第2ホルマンントへと収斂していく場合に拗音と判定し、結果は保持される。母音もこのような時間的な関係において判定するのが好ましいが、装置が大規模となるため、時間的な関係は拗音、母音の分離のみに用いている。

4. 零交叉波分析

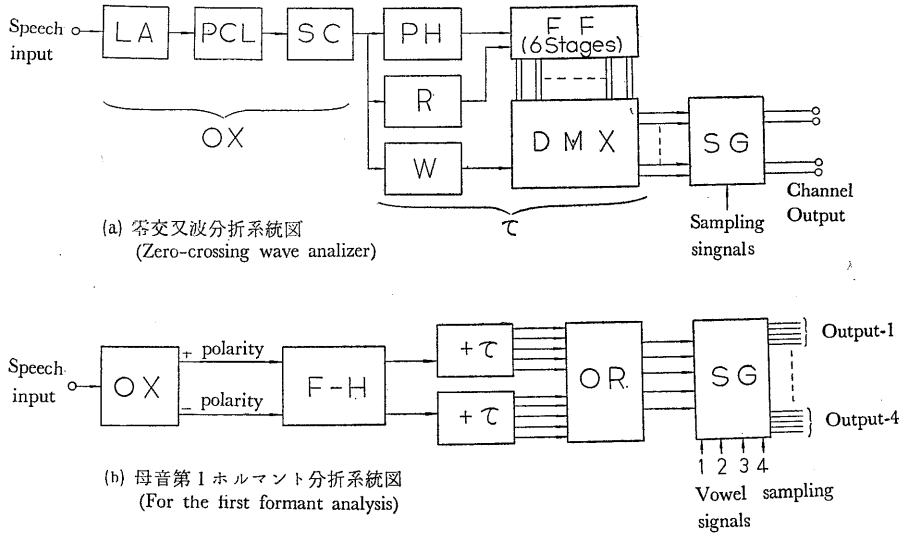
音声の分析の方法として、フィルター、ソナグラム等によるスペクトル分析が多く用いられ、多くのデータが発表されている。一方、音声を infinite clipping により振巾を2値化した零交叉波についても、良好な明瞭度を有することが英語について Licklider 等により報告されており、日本語についても確かめられている。

零交叉波における情報は零レベルを切る点の時系列にあるが、基本的な性質は相隣る零交叉時間の間隔にあるものと考えられ、我々もこれについて分析を行なう。零交叉波に変換する場合変換方法、及び原波形のスペクトルとの関係が問題となる。変換方法としては P. Marcov 等⁹⁾ の考えに従って一度 SSB に変調してから Clipping を行なった後に、復調する方法が考えられるが、直接に時間間隔を分析する場合には、音声原波形のまま Clipping する方が適していると思われる。スペクトル分布と零交叉波との関係の一般的な扱いは困難で、零交叉間隔分類統計による分布とスペクトル分布の対応はあまり良くない。母音においてはいくつかのホルマントを、零交叉波に対応させる事はむづかしく、又子音のように過渡的な波形に対してははっきりした特徴を示さない。日本語について零交叉時間間隔 τ による分析は、坂井・井上等¹⁰⁾ により行なわれ、 τ をいくつかのチャンネルに分類して測定し、子音母音各々について統計を行なっている。

我々はこれらの結果にもとづいて次のようにして扱う。音声は、子音部、母音部によって性質が異なるために分離して分析し、又、零交叉波分析の結果は、変換前に通す周波数特性によって大きく影響をうけるので、最も良い結果が得られるように予め音声にフィルター操作をし、母音、子音各部の分析に適する様にいくつかの分析系統を並列に用いる。

零交叉波は雑音の影響をうけやすい。そのため、零交叉波出力から、音声入力より得たサンプル信号により必要な部分のみを取り出し、又定量的な分析が出来るようにしている。

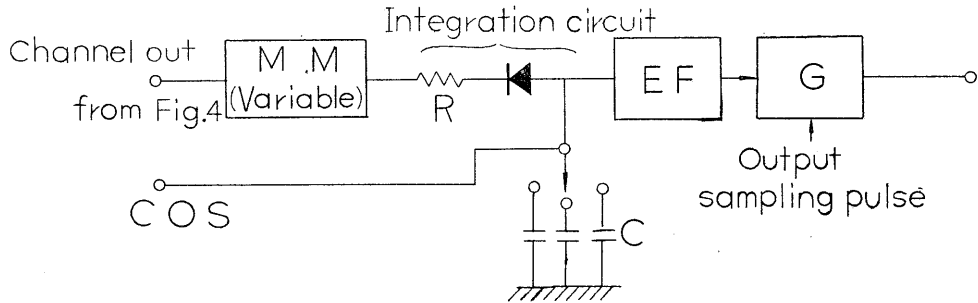
零交叉波分析システムのブロック図を第4図に示す。予め適当なフィルターを通してから零交叉波変換器に入れ、直線増巾器で数百ボルトの電圧を発生し、これを一挙に約0.5ボルトにダイオードで対称クリッピングを行なう。その後整形し、全体として40db以上の変換レンズをもつ。この矩形波でパルスハートレを動作させて、20.7KCで量子化し、それを6段のフリップ・フロップ・カウンターで数え、分類回路に入れる。チャンネル分類特性は量子化周波数により制限をうけるが、ダイオード・マトリクスを変えることにより容易に変更出来る。その後制御指令部よりのサンプル信号により、必要な部分のみを取り出



LA Linear Amplifier	直線増巾器	DMX Diode Matrix	分類マトリクス
PCL Peak Clipper	ピーク・クリッパ	SG Sampling Gate	サンプル回路
SC Schmidt Circuit	シュミット回路	F-H Full-Half Cycle	全、半サイクル
OX Zero-crossing Wave Converter	零交叉波変換器	Selector	切換え回路
		τ Time Interval Classifier	時間々隔分類回路
PH Pulsed Hartley	量子化回路	OR Channel OR Gate	チャンネル毎の正負OR回路
R Reset Circuit	リセット回路		
W Writing Circuit	出力パルス回路		
FF Flip Flop Counter	2進カウンタ(6段)		

第4図 零交叉波分析回路ブロック図

Fig. 4. Block diagram of zero-crossing wave analyzing circuit



MM Monostable Multi	モノスティブルマルチ	EF Emitter Follower	エミッタフォロウ
COS Clamping Off Signal	グランプオフ信号	G Gate	出力ゲート

第5図 蓄積計数回路

Fig. 5. Schematic circuit of integrating counter

し、第5図のような蓄積計数回路に入る。ここでサンプルされた分類出力は一定巾のパル

スにし、時定数 $C \cdot R$ でもって積分され、パルス本数に比例した電圧を得る。クランプはサンプルの開始と共にはずされ(クランプ・オフ)、サンプル区間終了と共にアナログ電圧を判定回路に送り、コンデンサーを再びクランプする。

5. 母音分析・判定回路

母音分析は零交叉波分析により近似的に第1ホルマント (f_1)、第2ホルマント (f_2) を求める方式である。音声原波形は零交叉波変換の前に両ホルマント帯を分離するようにし、 f_1 に対しては1500c/sの低域フィルター、 f_2 に対しては800c/s~2500c/sの帯域フィルターを通した後に微分を行なう。両ホルマントを上記の固定フィルターで完全に分離する事は出来ないが、日本語の5個の母音弁別の面から考えて、十分である。即ち f_1 について |ウ|、|オ|、|ア| の f_2 が混入するが、それらは特に大きい影響を及ぼさない。しかし f_2 の系統のフィルターは男声に対して選ばれたものであり、ピッチの高い女声に対しては |エ| の第1ホルマントが混入し、又、|イ|、|エ| の第2ホルマントは f_2 帯域に落ちない(ピッチ周波数に従って、 f_2 帯域を移動することを考慮している)。

f_1 分析系統は第4図b)に示されるように、切換え回路により、零交叉間隔を1周期について測定する場合(全サイクル)と、正又は負の極性の区間について測定する場合(半サイクル)とにわけられ、どちらかをスイッチで切りかえて用いる。又、半サイクルに対しては、正、負、別々の回路で分類し、その後各チャンネル分類出力毎にORしてまとめる。母音サンプルは独立に4個設けてある。第1サンプルとして、母音区間始端から10ms遅らせて40msをサンプルする。その後拗音分離回路からの再サンプル指令によって再サンプルが40ms行なわれるが、これは f_2 の移動の検出によるものであり、 f_2 よりも f_1 が大きく上昇傾向を示すような有声子音に伴なわれた母音では誤判定を生ずる事がある。

母音分析系統のチャンネル分類特性を第2表に示す。 f_1 は5チャンネルにわけ、 f_2 は、先ず8チャンネルにわけ、それをまとめて3チャンネルとして用いる。各出力パ

第2表 母音分析系統チャンネル分類特性及び判定マトリクス

Table 2. Frequency characteristic of the channels of the vowel analyzing part and its decision matrix in f_1 - f_2 domain

		第2ホルマント 2nd formant			
		分類特性 (c/s)	2500—1700	1700—1130	1130—720
第1ホルマント 1st formant	1440—790	No output	ア (a)	ア (a)	
	790—680	エ (e)	ア (a)	オ (o)	
	680—470	エ (e)	オ (o)	オ (o)	
	470—400	エ (e)	ウ (u)	オ (o)	
	400—170	イ (i)	ウ (u)	ウ (u)	

ルスは、蓄積計数回路で、各チャンネル中心周波数に逆比例した重みづけをして、アナログ電圧に変換され、サンプル区間終了と共に取り出す。 f_1 , f_2 各系統における零交叉間隔分布は、ほぼホルマントの分布に近い形を示すので、 f_1 , f_2 の各々につき、多くの電圧比較器により最大電圧を示すチャンネルを選び出力“1”を出す。この“1”の位置により、第2表に従って母音を判定する。全平面は $5 \times 3 = 15$ のメッシュにわけられ、これを5つの音韻にまとめている。又最上左端のメッシュは、雑音等により母音サンプルが誤って生じた場合、分布が正常でなければ、この部分に判定され、又判定がリセットされた状態でもこの部分に指示される。この回路による母音判定結果は、子拗音をのぞき男性に対し約94%である。誤りの方向は、 f_1 において |イ||エ|間、|ウ||オ|間、 f_2 において |イ||ウ|間等が主なものである。

6. 子音分析判定回路⁹⁾

子音部の弁別は、音韻大別回路により、無声子音、有声子音、鼻音のグループ分けと、破裂音の検出を行ない、各グループに属する音の間の弁別を子音分析系統により行なう。子音の分析及び判定は、エネルギーもよわく、過渡的、短時間のものが多いので、母音に比して困難である。又、子音は母音の影響をうける。従って次の様にして分析を行なう。

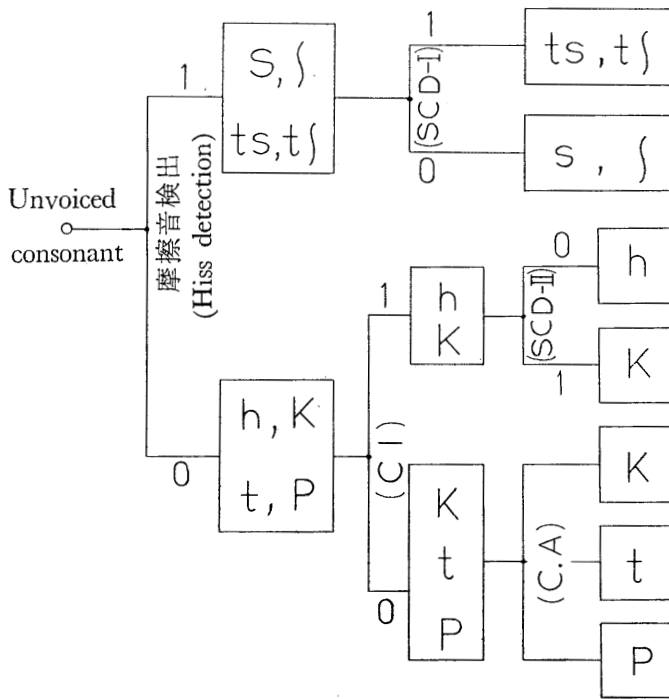
- 1) 子音の性質は本質的には雑音的な子音部にあり、先ずこの部分について分析を行なう。ついで、母音へのわたりの部分にも重要なものについて母音分析系統で扱う。
- 2) 無声子音、有声子音、鼻音に対し独立の分析系統を設ける。
- 3) 子音の区間をサンプルして取り出す。
- 4) 子音に対する母音の影響を考慮して判定する⁷⁾。

子音分析系統ブロック図は第1図にあり、第4図、第5図のような回路で構成される。3系統の分析系統をもち、各々250c/s HPF, 600c/s HPF, 微分を行なった後、零交叉波に変換する。分類回路においては、先ず16チャンネルに分類するが、次にダイオード・マトリクスにより、いくつかのチャンネルをORでまとめて、より広い巾をもつ新しいチャンネルに変換する(新チャンネル編成)。これは、子音の分布は、鋭いピークよりも、広い分布をし、又、各音韻の判定に必要とするチャンネル分類特性が異なるためである。この新チャンネルについて、蓄積計数を行ない(電圧変換比はパルス巾、及びコンデンサーの切換えにより弁別に適した様に選ぶ)、出力を一定レベルでもって Schmitt 回路により2値化し、判定回路への論理変数とする。

子音サンプルは、ブズレス音声区間 (Y) が1になるときに始まり、その後最小10ms, 最大40ms, の範囲で母音区間の始るまで行なう。無声子音の場合、母音始端はピッチ抽出回路により得られるが、有声音に関しては、明確な点は捉えられない。又鼻子音の部分は介別に有効な結果を与えないので、子音系統では扱わない。

(1) 無声子音：

無声子音の判定は第6図のように構成して行なう。音韻大別からは「無声子音」「破裂



CI Consonant Interval 子音継続時間 CA Consonant Analysis 子音分析

第6図 無声子音弁別系統

Fig. 6. Discrimination method of unvoiced consonants

I 及び II」を得、又、子音継続時間（閾値30ms）も使い、摩擦音の検出及び分析は零交叉波分析(上述の子音系統)を用いる。検出は3500c/s以上のチャンネルに、定常的な雑音成分を検出し、その後2値化して、判定するが|k(i)|と|t|の間で誤判定を生じやすい。|k| |t| |P|は後続母音の子音に及ぼす影響を考慮して、後続母音別にカナ文字式に判定する。

無声子音特に破裂音は特徴的な分布を示さず、又発声の方法によって著るしい変化をうけるために弁別能力は良くなく、今の所70%位である。

(2) 有声音音：

有声子音の特徴は雑音成分のみでなく、特に破裂音においては、母音へのわたりの部分も重要であるが、我々は先ず、無声子音の場合と同じ様にして、雑音成分（子音部）の分析を試みた。無声子音に比し雑音成分は弱く、短かく、又、ピッチ成分が存在するため、まだまだ良いスコアを得ていない。又、わたりの部分については、別に考慮する（第9章参照）。

7. 総合判定マトリクス

以上のべた判定や分析の結果は全て保持され、総合判定マトリクスに入る。即ち、音韻大別の結果（母音、無声子音、有声子音、破裂音、拗音、鼻音）、母音判定の結果（ア、イ、ウ、エ、オのうちの1つ）、及び子音の新チャンネルについて2値化された判定値である。これら入力のうち適当なものをダイオードAND回路で組合わせて、百音節のうち1つを出力として表示する。子音の判定回路が後続母音により異なる事は前述の通りである。なお母音、子音の結果の保持回路は、サンプル区間終了と共に以前の結果をリセットし、ついで新たに書き込むようにする。このため、判定保持回路には常に何んらかの結果が保持されている。

8. 出力回路

総合判定マトリクスの結果は、音声区間抽出回路出力の終端でパンチ又はタイプされる。このため雑音により二重の出力を出すことをさけるため、一定時間以上音声エネルギーを検出した場合のみ出力指令を出す。パンチは8単位で、タイプはテレプリンタによりカナ文字で行なう。百単音節の外、スペース記号、長母音記号、判定結果が出ないで出力指令のみ出たときの「？」の記号等も組み込まれている。

9. 会話音分析への応用例

以上のべて来た弁別装置は単音節に対して動作するように設計したものである。しかし実用化のためには、会話音声を取扱う様に考慮しなければならない。このため、装置を会話音声弁別に適する様に変更し、先ず始めに、会話音声に対する分析データを得た。

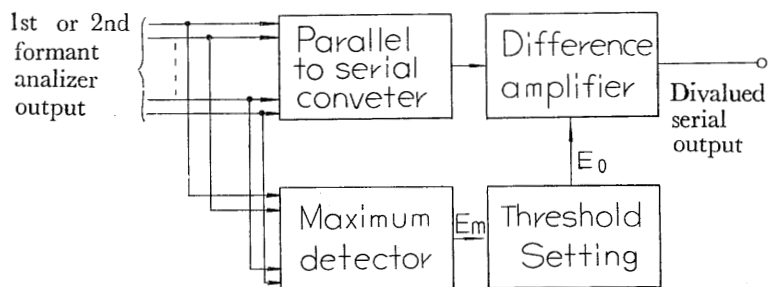
日本語は、基本的には「子音+母音」の構成をもつ単音節から構成されるが、会話音声を分析、弁別していく上には単なる単音節のつながりといった面以外に次の点も考慮しなければならない。（i）判定の最小単位の選択；文章、単語等も考えられるが、構成要素の

数などの点から、音韻単位の弁別が良いと思われる。(ii) 上記の判定単位に対して、連続した会話音声から音と音の区分点を求め、各単位を分離しなければならない。(iii) 単音節をもとにして会話音を扱う場合、前後の音による音結合の影響を考慮しなければならない。(iv) ピッチ変化、レベルの上昇、下降の傾向、1つの音単位の継続時間、スペース等も重要な性質である。

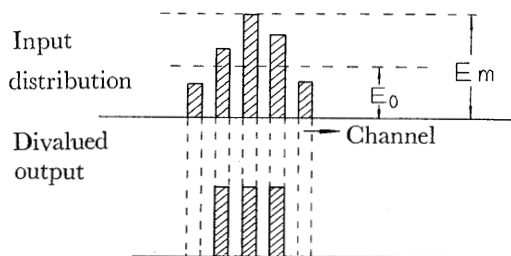
我々は、単音節弁別を基礎として、それに会話音声弁別に必要な上の諸事項を考慮していく形をとる事により会話音声を扱う事にした。

(1) 零交叉波分布模様直視装置

以上の様な会話音分析には、音の特性を時間的な変化の形でとらえていかなければならない。又この方法は単音節の扱いにも有効である。これらのうち最も重要なホルマントの特性を知るため、単音節弁別装置の母音分析用第1、第2両ホルマント分析システムを利用し、これに零交叉波分布模様直視装置を附加して、両ホルマントの時間的な変化を捉えた。第4図の母音第1、第2ホルマント分類出力を各々5及び9チャンネルに第3表の様にわける。その後第5図のような蓄積計数回路(2系列)により、10ms 毎に交互に連続してサンブ



(a) ブロック図 Block diagram



(b) 入力-出力関係

第7図 零交叉波分布直視装置2値化回路

Fig. 7. Circuit block diagram of divaluing the zero-crossing interval distribution and its operation

ルし、各サンプル周期毎に一組の零交叉波分布を得る。ついで、各ホルマント別々に分布の最大値 E_{max} を求め、 $E_0 = E_{max}/\alpha$ ($\alpha \geq 1$) を閾値としてもとの分布を2値化すると共に、並列チャンネル出力を直列信号に変換する。(第7図)

以下に示すデータは、ブラウン管上に得られた図形を写真記録したものであり、輝点は2値化出力の“1”をあらわし、横軸は時間(1つの点が1サンプルの結果)、縦軸はチャンネル(周波数)を表わす。2値化で $\alpha = 1$ とするとピーク検出となるが、ここでは $\alpha = 2$ 、即ち閾値 $1/2$ を採用した。

(2) 分析データ例

第8図、第9図、第10図にデータ例を示す。又、そのチャンネル割当ては、第3表のように下から第1ホルマント、弁別要素抽出回路出力、第2ホルマントの順である。いづれも試料は男声である。図において8-1~8-5は日本語5母音の例で、イ、エでは安定したホルマント分布を示すのに対し、ア、オ、ウの第2ホルマントでは広がった分布をする。これは各ホルマントが完全には分離されていない事を示すものと考えられる。又有声子音から母音部へのわたりの部分の変化は8-6~8-8に示される。このように後続母音の第2ホルマントの低い場合には、子音の調音点により特徴づけられるホルマントの変化が見られる。8-9~8-12は半母音の例である。

会話音声の構成は無数に作りうるので、個々に分析するのは良い方法ではない。そこで会話音声分析に考慮すべきいくつかの項目を選び出して、それについて分析をした(第9図)。

考慮した項目と、それに対するデータ例は次のようである。[]は発音を示し、又()は実際には発音されない音である事を示す。

- i 連続母音; 9-1 恋 [koi] 9-2 声 [koe] 9-3 塩 [jio]

第3表 零交叉波分布模様のチャンネル割りあて (第8, 9, 10図について)
Table 3. Channel location of the zero-crossing interval pattern of Fig. 8, Fig. 9, and Fig. 10

	Channel number	Frequency Characteristics (c/s)	Channel width (c/s)
2 nd formant	9	3500 ~ 2500	1000
	8	2500 ~ 2020	480
	7	2020 ~ 1680	340
	6	1680 ~ 1440	240
	5	1440 ~ 1250	190
	4	1250 ~ 1120	130
	3	1120 ~ 920	200
	2	920 ~ 780	140
	1	780 ~ 720	40
Output of DFE	F ₂ D	2 nd formant detector	
	HD	} See Fig. 1	
	BLS		
	PD		
1 st formant	5	1450 ~ 920	530
	4	920 ~ 690	230
	3	690 ~ 470	220
	2	470 ~ 340	130
	1	340 ~ 160	180

- ii 拗音と連続母音の関係; 9-4 仕上げ [ʃiage] 9-5 医者 [ija]
9-6 視野 [ʃija]
- iii 促音; 9-7 一茶 [i'sa] 9-8 学校 [ga'ko:]
- iv 母音消失; 9-9 司会 [(i)kai] 9-10 イチ [itʃ(i)] 9-11 イチ [itʃi]
- v 撥音; 9-12 人員 [dʒiNiN] 9-13 辞任 [dʒiniN]
9-14 信任 [ʃiNniN]

このような例について、分布の変化する点などから各音の区切りが求められれば、単音節におけるのと同様の分析方法により、会話音声を扱う事が出来る。

一般的な言語を対象とする外に、数字音声のように、語彙の限られた言葉を対象とすることも実用上意義がある。この目的のために、日本語及び英語の発声された数字に対する分析を行なった。結果の一例を第10図に示す。数字は1つずつ区切って、

1 [ichi], 2 [ni], 3 [saN], 4 [joN], 5 [go], 6 [roku], 7 [nana],
8 [hachi], 9 [kju], 0 [rei] のように発声されている。

又、英語のサンプルはリングフォン米語コースより取ったものである。

(3) フィルターによる分析例

零交叉波の場合と同様の方法でフィルターを用いて、周波数、時間、振巾の三次元表示を試み、上記の零交叉波分析の結果と比較した。高域強調回路で、日本語会話音声の長時間スペクトル分布がほぼ一樣になる様に補償した後、帯域巾100c/sのフィルター24個で0~2.4KCを順次カバーする。分析は主として、ホルマント区間を対象とする。結果の例を第11図に示す。輝度はスペクトル振巾をあらわし、縦軸にチャンネル、横軸に時間(5ms毎のサンプル)をとつてある。この結果は、各フィルタ出力をそのまま表示したものであり、冗長なものを全て含んでいる。これに対し、零交叉波では、レベルに関する情報及び反共振等の微細な構造に欠けている上に、2値化により簡単化しているので、両者を直接比較する事は出来ないが、ほぼ近い傾向を示している。以上のようなデータをもとにして、会話音声の弁別について、研究をすすめている。

10. 結 言

日本語単音節について働作する弁別装置についてのべた。また、日本語全音のうちには確実なパラメータを把握するに至っていないものや、又データより得た結果を、装置の関係で金物に置きかえられないものもある。このような問題点を解決すると共に、この装置

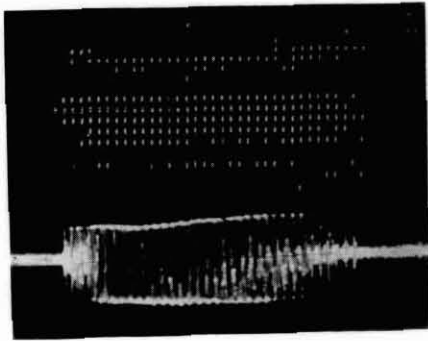
を利用して得られたデータより、会話音声を扱う装置を設計している。

謝 辞

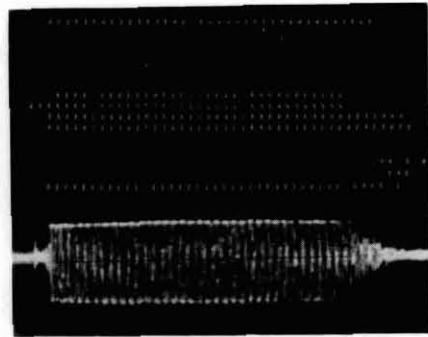
末筆ながら、日頃御指導頂く京都大学電子工学教室、前田憲一教授に感謝し、併せて、この研究は文部省の科学研究費によるものであることを記して謝意を表します。又、試作装置に関しては、日本電気株式会社との協同研究によるもので、黒川武夫伝送工業部長、仲丸由正元研究室長、友成治夫課長、永田邦一、関本忠弘、津崎賢郎、金子尚志、加藤康雄、近藤寛の諸氏に、深く感謝します。

文 献

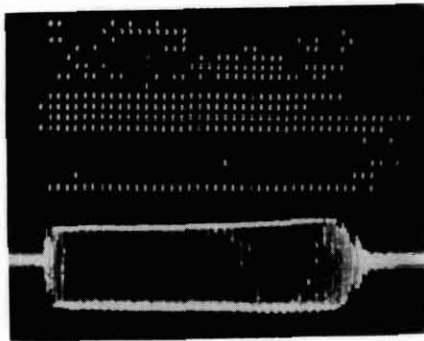
- (1) 前田憲一、坂井利之、他；「音声タイプの基本設計」電気通信学会，オートマトンと自動制御研究専門委員会資料，1960年1月。
- (2) たとえば、中田和男；通信学会誌，44，pp221～227 (1961)。
- (3) 前田，坂井，堂下；電気通信学会全国大会，No. 34，昭34。
- (4) P. Marcov, J. Daguet : Information Theory 3rd London Symposium pp231～244 (1955)。
- (5) T. Sakai, S. Inoue : JASA 32 pp441～450 (1960)，その他通信学会大会等における発表。
- (6) 前田，坂井，倉下；通信学会全国大会，No. 33，昭34。
- (7) 前田，坂井，白井；電気四学会連合大会，No. 707，昭34。
- (8) 前田，坂井，白井；通信学会全国大会，No. 35，昭34。



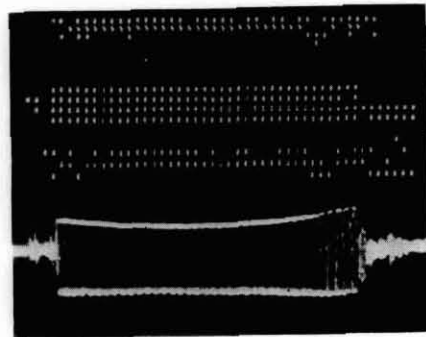
8~1 ア [a]



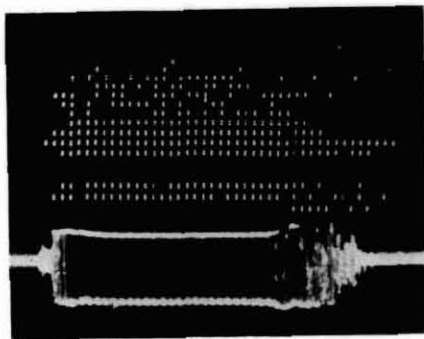
8~2 イ [i]



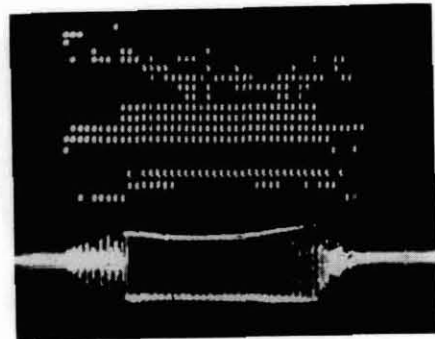
8~3 ウ [u]



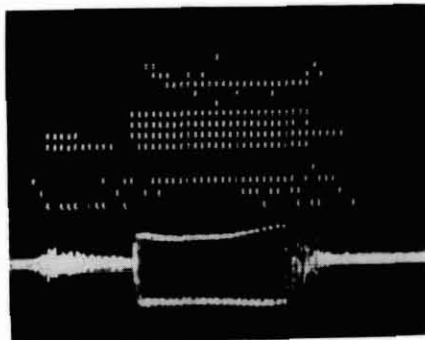
8~4 エ [e]



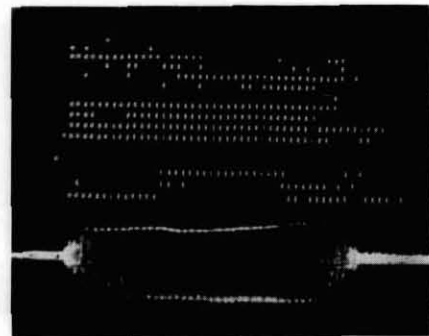
8~5 オ [o]



8~6 ロ [ro]



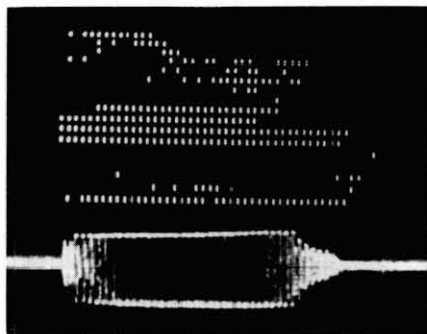
8~7 ド [do]



8~8 ノ [no]



8~9 ヤ [ja]



8~10 ュ [ju]



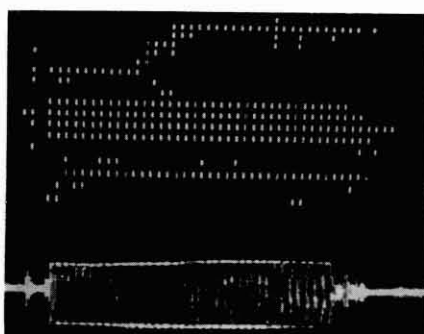
8~11 ュ [jo]



8~12 ヲ [wa]



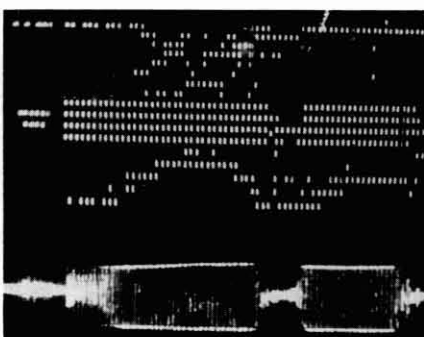
9~1



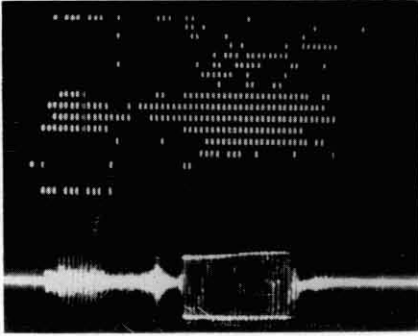
9~2



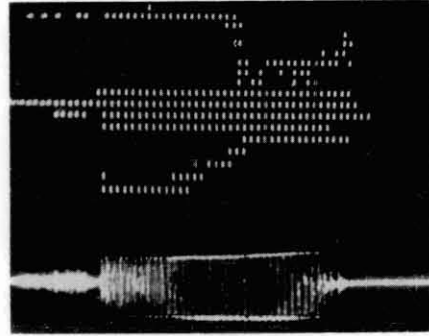
9~3



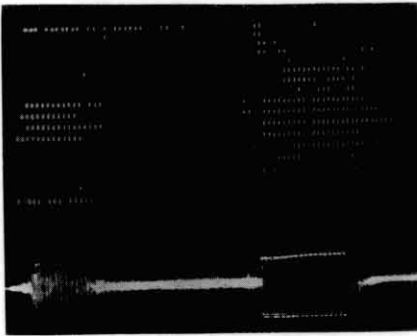
9~4



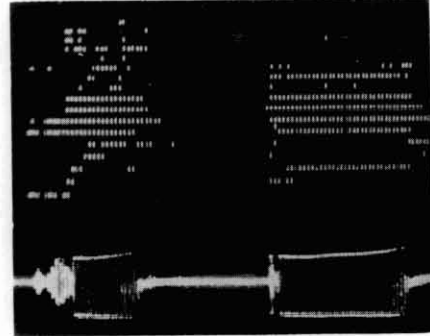
9~5



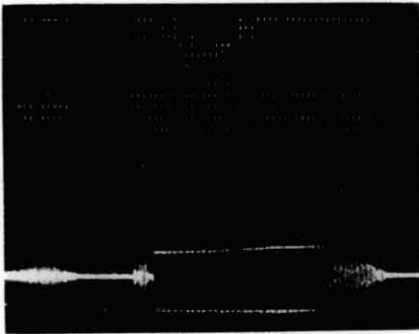
9~6



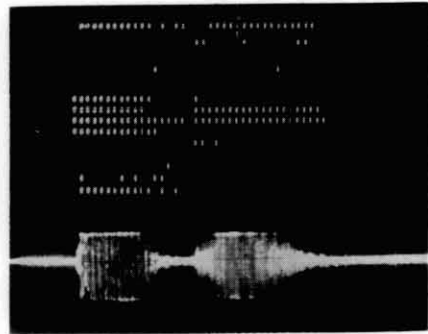
9~7



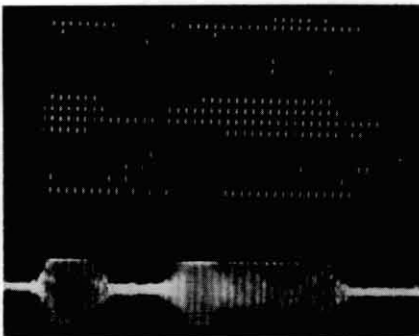
9~8



9~9



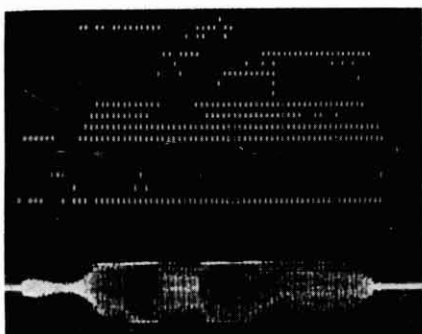
9~10



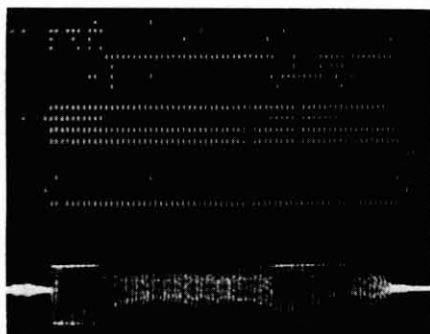
9~11



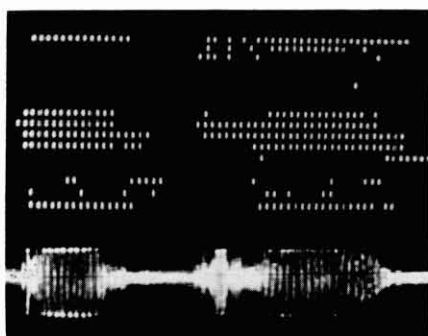
9~12



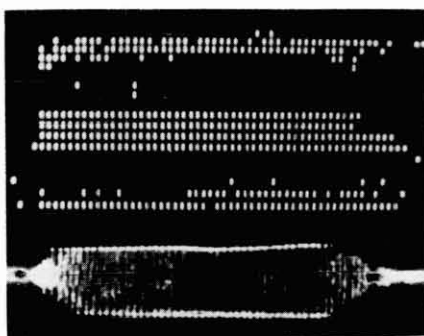
9~13



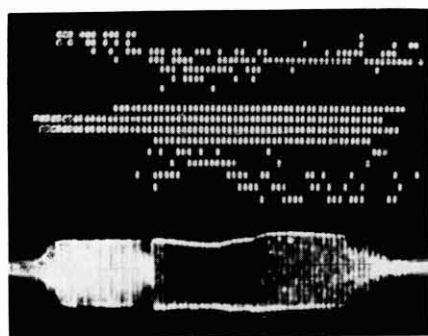
9~14



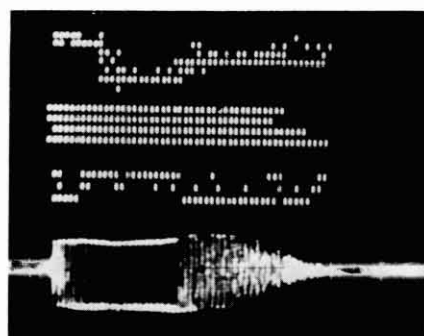
0~1 イ (1)



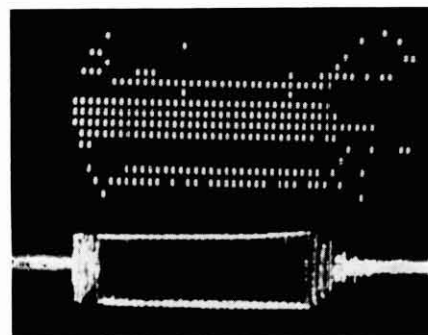
10~2 エ (2)



10~3 ア (3)



10~4 ウ (4)



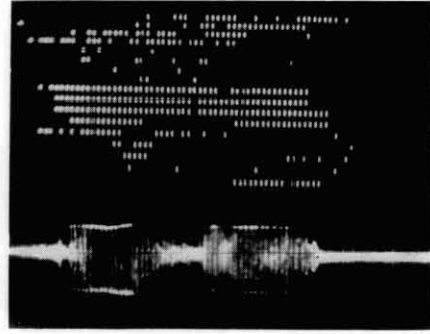
10~5 オ (5)



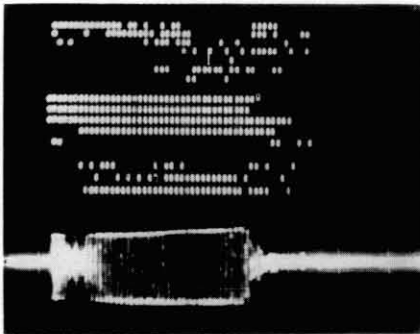
10~6 カ (6)



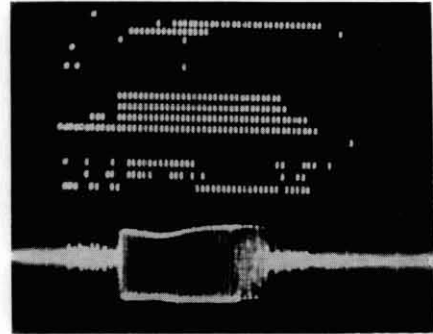
10~7 ナナ (7)



10~8 ハチ (8)



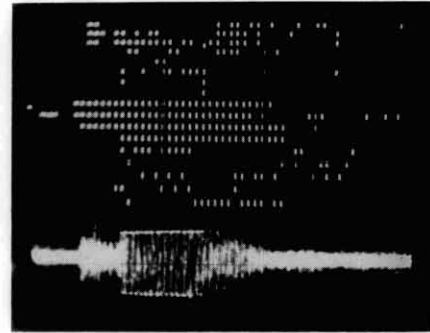
10~9 キユウ (9)



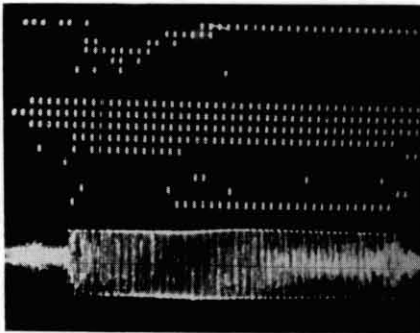
10~10 レイ (0)



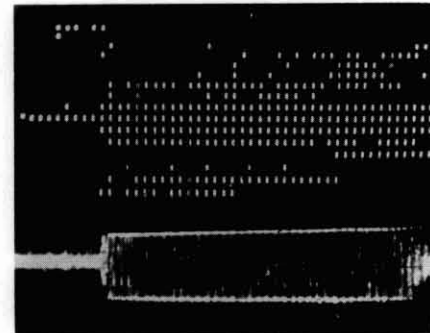
10~11 one



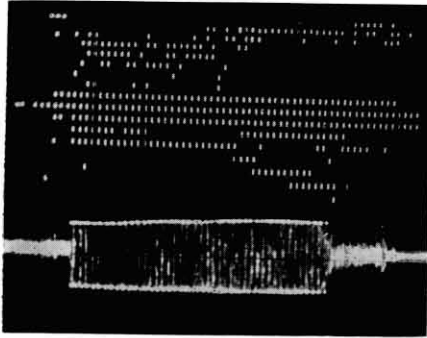
10~12 two



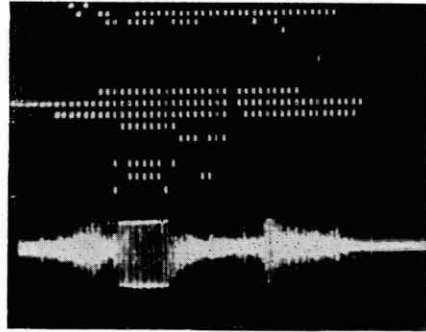
10~13 three



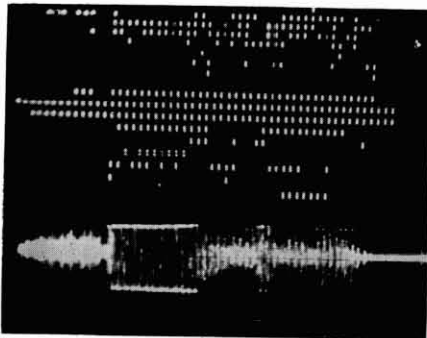
10~14 four



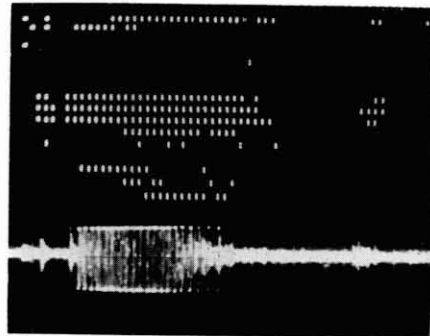
10~15 five



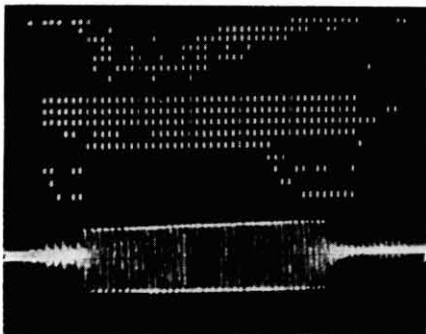
10~16 six



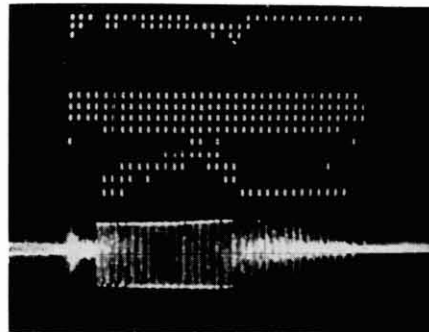
10~17 seven



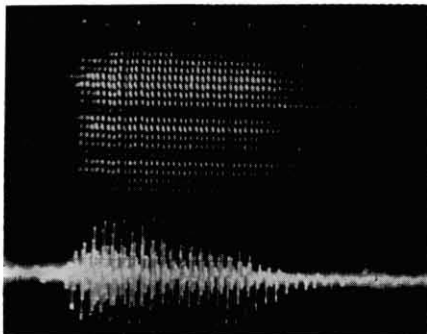
10~18 eight



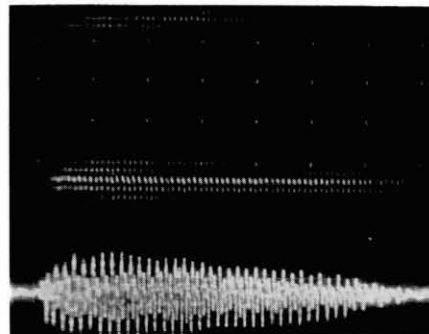
10~19 nine



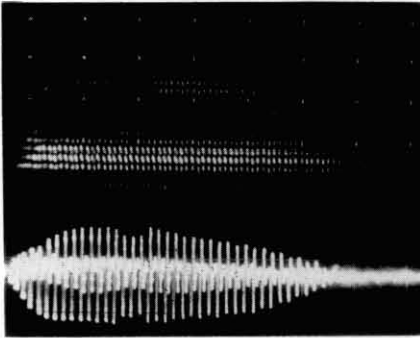
10~20 ten



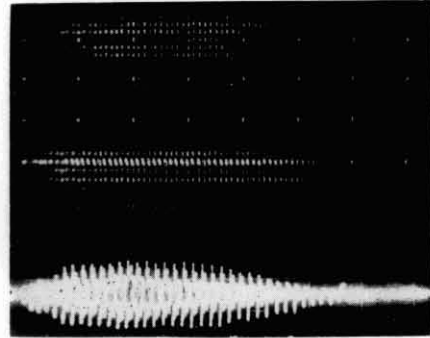
11~1 ア [a]



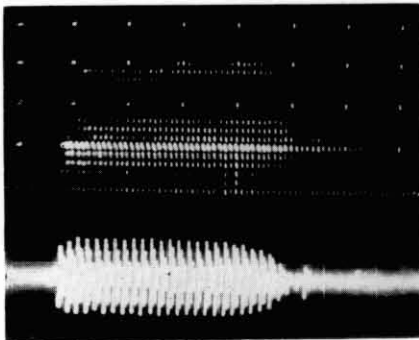
11~2 イ [i]



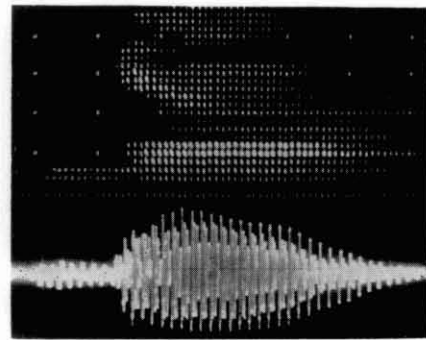
11~3 ヴ [u]



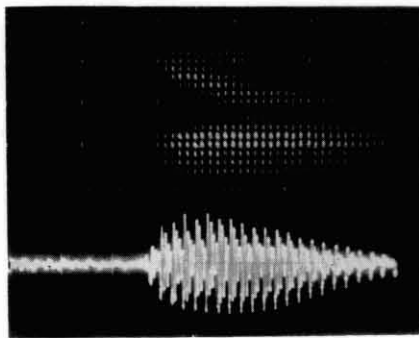
11~4 エ [e]



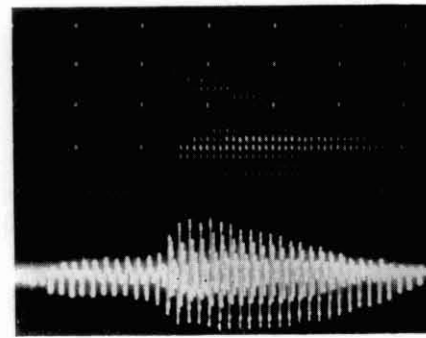
11~5 オ [o]



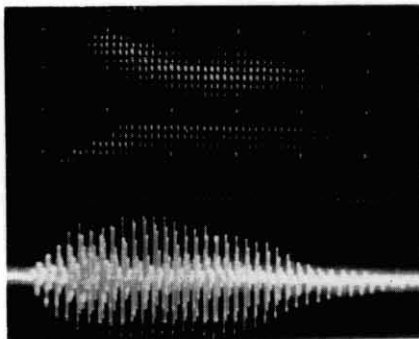
11~6 ロ [ro]



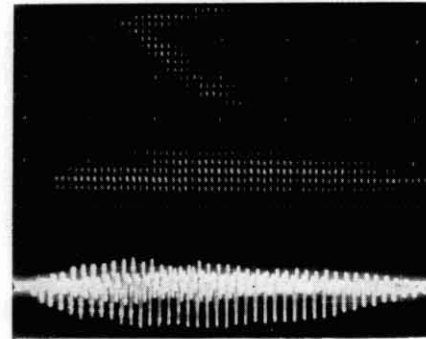
11~7 ド [do]



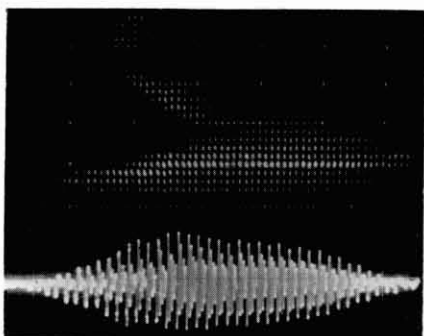
11~8 ノ [no]



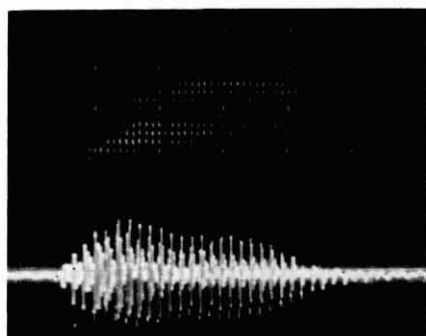
11~9 ヤ [ja]



11~10 ュ [ju]



11~11 ㇰ [jo]



11~12 ㇱ [wa]