This is to certify that the

thesis entitled

COMPUTER VOICE IDENTIFICATION METHOD
BY USING INTENSITY DEVIATION SPECTRA
AND FUNDAMENTAL FREQUENCY CONTOUR

presented by

Hirotaka Nakasone

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Audiology and Speech Sciences

Major professor

Date November 8, 1983

COMPUTER VOICE IDENTIFICATION METHOD BY USING INTENSITY DEVIATION SPECTRA AND FUNDAMENTAL FREQUENCY CONTOUR

By

Hirotaka Nakasone

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

1984

ABSTRACT

COMPUTER VOICE IDENTIFICATION METHOD BY USING INTENSITY DEVIATION SPECTRA AND FUNDAMENTAL FREQUENCY CONTOUR

by

Hirotaka Nakasone

The major purpose of this study was to investigate the effectiveness of several speech parameters in eliminating the influence of the transmission and recording channels of unknown response characteristics. These speech parameters were tested for their effectiveness in text-independent voice identification by computer.

Text-independent speech samples were recorded from 10 male speakers randomly selected from a population of the native speakers of Midwest American-English dialect, each serving as unknown speaker and also known speaker. Recording was made simultaneously by three different transmission and recording devices. From these speech data, the following parameters were extracted to represent the unknown and the known speakers: 1) intensity deviation spectrum (IDS), 2) a set of fundamental frequency related measurements (FFC), 3) long-term averaged spectrum (LTAS), and 4) choral spectrum (SPT). The principal algorithm utilized for the measurement of the parameters 1), 3), and 4) was the Fast Fourier Transform (FFT); and for parameter 2), the interactive peak picking technique was employed. All speech parameters were

subjected to pre-processing in order to select optimum features from each parameter by using the hierarchical clustering, F-ratio, and data standardization. Distance between the speakers was measured by Euclidian distace. The decision rules employed were the nearest-neighbor rule and the minimum set distance rule. The rate of correct identification served as the criterion to determine the effectiveness of each parameter in elimination of the influence of the response characteristics.

From the results of this study, the following general conclusions were drawn: Both IDS and FFC were found to be effective in eliminating the influence of the transmission and/or recording channels, but their correct identification rates were only moderate (50-60%.) The composite parameter of FFC and IDS was found to be effective in eliminating the influence of the response characteristics although, the correct identification rate was not improved, *i.e.*, it was as good as each component of that composite parameter. The composite of FFC and LTAS was found to be the most effective parameter in eliminating the influence of the transmission system by achieving the highest possible correct identification rate (100%.)

Dedicated to
My mother, Haru Nakasone

dis

mos

mem

hel

tov

Dr

St

ma

pr

De

av

Ac

s

S

J

s

ACKNOWLEDGMENTS

I am deeply indebted to Dr. Oscar Tosi, my professor and dissertation director, whose guidance, encouragement, efforts, and most of all patience, made the completion of this study possible.

I would like to extend my sincere appreciation to Committee members, Dr. Leo V. Deal and Dr. Paul A. Cooke, for their helpful suggestions, corrections, and constructive criticisms toward the completion of this study. I also would like to thank Dr. Jose-Luis Menaldi of the Department of Mathematics, Wayne State University, who also served as a Committee member, for his mathematical assistance and critical review of the computer programms developed for this study.

Special acknowledgments are due Dr. Richard C. Dubes, Department of Computer Science, who kindly made the graphic plotter available, and Dr. Ernest J. Moore, Chairman of the Department of Audiology and Speech Sciences, for his interest and valuable suggestions during the final oral defense meeting.

My warmest gratitude is expressed to Ms. Nancy Brown of the State Bar of Michigan who proofread my draft, Dr. Christina Jackson-Menaldi for her companionship and her enthusiasm about this study, and Dr. Paul N. Deputy of Idaho State University, who first introduced me to this exciting field of Speech Science during the early stages of my graduate program.

Special recognition is noted to a number of individuals who provided their voice samples for this study. I am very grateful

for thei

Fina

who gave

in typin

final ve

for their cooperation.

Finally, but certainly not the least, I thank my wife, Aiko, who gave me her unconditional support and most qualified assistance in typing and retyping the numerous rough drafts, as well as the final version.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

Tables

LIST OF TABLES

LIST OF FIGURES

Mo

genera

include

methods

(short

recorde

speaker

upon

charact

voices

judgmen

produce

examine

the u

the spe

object

decisi

examine

subject

progra

identi

CHAPTER I

INTRODUCTION

Methods of voice identification can be classified into two general groups: subjective and objective. The subjective methods include aural and spectrographic examination; the objective methods are usually performed by using a computer. Aural methods (short term and long term memory) are performed by listening to the recorded voices of an unknown and a known or by remembering the speaker-dependent features of a voice. These are primarily based upon perceptual extraction of the speaker-dependent speech characteristics. The final decision regarding the identity of voices is made by the human examiner based upon his subjective judgment. With the spectrographic methods, the sound spectrograms produced from speech samples under study are examined. The examiner compares the acoustical characteristics of the known and he unknown voices displayed on the three dimensional plottings of he spectrograms (frequency, intensity, and time). In spite of the bjective means of displaying the speech parameters, the final ecision still belongs to the subjective judgment of the human xaminer. Hence, both aural and spectrographic methods are called ibjective methods (Tosi, 1979). When a computer is properly ogrammed with a set of algorithms, the results concerning entification of the voices are reproducible -- when similar types of data

expected

commonly

recogni

The

name

determi

samples

speakers

voice i

Acc
sam
be
ope
tes
sam
whe
In
sam
unk
sam
the
sev
unk
sam

In

included

equival

'examin

'voice i

specifi

of an u

speaker

data are submitted to the same procedure, the same output is cted. Hence, computer method is considered to be objective, only referred to as automatic or semi-automatic speaker gnition.

The term 'voice identification' has been applied as a generic which encompasses various aspects in the process of mining the identity of an unknown speaker, given his/her voice es and voice exemplars collected from one or more known ers. To be more specific, Tosi (1979) classifies tasks of identification as follows:

> According to the composition of unknown and known voice samples, tests of voice identification or elimination can be classified into three groups: discrimination tests, open tests, and closed tests. In the discrimination tests the examiner is provided with one unknown voice sample and one known voice sample. He has to decide whether or not both samples belong to the same talker.... In the open tests the examiner is given one unknown voice sample and several known samles. He is told that the unknown sample may or may not be found among the known samples.... In the closed tests of voice identification the examiner is also given one unknown voice sample and several known voice samples but he is told that the unknown voice sample is also included in the known voice samples....(pp. 4-5).

n this study, since in all tests the unknown voice was always ed within the known voices, the task is considered to be lent to the 'closed test' quoted above, excepting that the ner' is being replaced by a 'computer'. Hence, the term identification' as used in this study covers only this c task which is described as follows: Given voice samples nknown and a group of knowns, the task is to select a whose voice sample is the closest to that of the unknown.

be b

measu

speed

stor

are

such

featu

the

emplo

'non

obta

wher

inte

(day

of v

than

conc

spea

the

comm

samp

comp

The

The standard procedure of voice identification by computer can
oroadly divided into three major stages: data collection,
urement, and identification processing. In the first stage,
ch samples from a given group of speakers are recorded and
ed. In the second stage, speech parameters (characteristics)
measured. This stage includes a series of pre-processings,
as filtering, deletion of pauses and gaps, extraction of
ires, and statistical processing for feature optimization. In
hird stage, the identification operation is performed by
ying appropriate decision rules or criteria.

Speech samples collected can be either 'contemporary' or
ontemporary' (Tosi, 1979). Contemporary speech samples are
ned from each speaker during the same recording session,
as noncontemporary speech samples are recorded over some time
vals depending upon the scope of the researcher's interest
, weeks, months, or years). It has been noted that the task
ce identification is easier with the contemporary samples
ith the noncontemporary samples.

nother aspect involved during the data collection stage is
ned with the type of the phonetic content spoken by the
rs: 'text-dependence' vs. 'text-independence'. When all
peech samples of the speakers are the same in context, it is
ly referred to as 'text-dependence'. In this case the speech
s of the speakers under an identification process are
ed word by word, phrase by phrase, or sentence by sentence.
jor advantage of text-dependent speech samples is that they

can be u

of this a

many r

commerci

In

speakers

text-ind

duration

who use

(1974)

sentence

them in

of text-

identifi

reported

seconds

rate. I

differen

when the

influenc

dominate

duration

identifi

'text-in

consiste

speakers

be utilized in every method of voice identification. Because
his advantage, this type of text has been rigorously studied by
researchers, but mainly directed toward industrial or
ercial applications.

In 'text-independence', all the speech samples spoken by the
kers are different in phonetic content. The duration of the
-independent materials must be relatively long, and the minimum
tion appears to vary in length depending upon the researchers
use the term 'text-independence' somewhat differently. Atal
4) generated the text-independent speech sample from a single
ence by cutting it into 40 equal segments and later recombining
in random order. He reported that the minimum of two seconds
ext-independent speech sample resulted in a high correct
ification rate. Bunge (1978) and Furui *et al.* (1972)
ted the minimum duration of a close agreement of 11 and 10
ds is required for a sufficiently high correct identification
In Bunge's study, 41 male and 9 female speakers produced 50
rent texts, each text lasting 11 seconds. It was found that
the text length was decreased to below 11 seconds, the
ence of the text became increasingly obvious and finally
ated the speakers identity. He concluded that an 11-second
ion was a limit for text-independence for a high correct
ification rate. Markel and Davis (1979) defined the term
-indepenent' a little more stringently. Their speech data
sted of the extemporaneous speech material from 17 male
ers, each speaker recorded in five interview sessions at one

week ir

from li

speech

39-seco

high c

scope c

speech

languag

Each s

from bo

by one

text-ir

promisi

Sp

acousti

speech

which a

element

air pr

stream

or turb

and plo

taking

vocal t

contair

speaker

ek intervals. They attempted to make the speech samples free om linguistic constraint and further free from the manner of eech production. With this type of text-independent materials, a -second text length (containing only voiced frames) resulted in gh correct identification. Tosi *et al.* (1979) extended the ope of text-independence to different languages. In their study eech samples were obtained from 20 speakers who could speak three guages (Piamontes, Italian, and French) with equal fluency. h speaker was recorded while reading a 10-minute long passage m books and newspapers in three sessions, each session separated one week, and concluded that automatic voice identification with t-independent speech materials of different languages is nising.

Speech samples so collected are then processed for extracting ustic speech parameters to represent the speakers. Acoustic ech parameters are the measurements derived from speech signals h are considered to consist of three major elements. The first ent is the energy source coming out of the lungs as a stream of pressure. The second element is a modulation of this air am into vibratory motions (for voicing) set by the vocal cords urbulence of air in a constriction of the vocal tract (friction plosive). The third element involves resonance phenomena ng place as the modulated air pressure traverses through the l tract (pharynx, oral and nasal cavities). In each element ained is information of the linguistic contents as well as the ker characteristics.

intens
fundam
rise
consi
partic
output
charac
S
identi
combir
short-
freque
basic
furthe
the
identi
set
speake
T
a spe
the
speake
variat
variab
variab

The first element carries variation of overall speech sity as a function of time. The second element determines the nental frequency and its harmonics of voiced phonemes giving to perceptual pitch of the speaker. The third element is dered to be the most important because resonance gives a cular shape or envelope to the spectrum of the speech sound which includes both a phonetic content and the individual teristics of each speaker (Tosi, 1979).

peech characteristics (parameters) used in voice fication by a computer are generally extracted from one or a ation of these acoustic elements. Often used parameters are term spectra, long-term spectra, formants, fundamental ncy, and other variations statistically derived from these parameters. Usually, a certain number of features are selected from the parameter and these features represent speaker in a multidimensional space. In general, the ication process is based on the distance measured between a features of the unknown speaker and that of the known s.

e basic implicit assumption for voice identification is that ker can be distinguished by his/her speech signals and that riation of speech characteristics within an individual differs from that of the other speakers. The former on is commonly referred to as the 'intra-speaker lity' and the latter is referred to as the 'inter-speaker lity'. The sources of the intra-speaker variability can be attri

produ

physi

over

inter

confi

and i

Stater

proce

metho

indiv

and/o

the

alter

impers

conte

study

frequ

speak

This

ident

long

the

ributed to different emotional status, various manners of speech duction demanded by different circumstances, and small siological changes in articulatory apparatus of the same person an interval of time. On the other hand, the sources of the er-speaker variability are the different vocal tract igurations, physiological characteristics of the vocal cords, idiosyncratic speaking habits of different speakers, etc.

ement of the Problem

There are several sources which could interact with the ess of voice identification by a computer or by any other od. These are distortion of speech characteristics of an vidual due to the unknown response curve of the transmission or recording devices, various types of noise which deteriorate intelligibility of speech samples and intentional self ration of voice either to disguise the identity or to sonate another person. In addition, differences in phonetic nt and duration of the speech utterances of the speaker under interact with the procedures of voice identification. On ent occasions, the phonetic content spoken by the unknown er can be different from the one spoken by the known speaker. condition calls for so-called 'text-independent' voice ification, which usually requires speech samples of relatively duration obtained from each speaker.

This study focuses on the problem caused by the influence of transmission and recording devices by using the text-inc

Te
variabl
junctio
conditi
real li
the unc
call, wh
moment
for lat
respons
distort

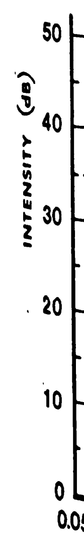
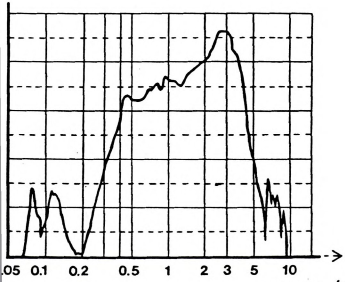


Figure
incl
atta
1979

-independent and contemporary phonetic materials.

Telephone transmission contains in itself several sources of ables such as the carbon microphone, number of connecting tions, line distance, and carrier systems. Under this ition, the response characteristics of the telephone line, in a life setting, cannot be determined. This is largely due to uncertainty of the telephone line involved in each telephone which is random, according to the existing traffic at the t of the call. Ordinarily the speech signals must be stored ater processing in a recording medium which also has its own nse characteristics. An example of this type of combined rtion in response characteristics is given in Figure 1.



e 1. Response curve of a transmitting and recording system, cluding a commercial telephone line and a magnetic pick-up ached at the receiver end of the line (taken from Tosi, '9).

transm

charac

sampl

conver

respon

of the

and r

only c

to pr

distor

resear

propos

.
c
b
r
t
t

To dat

O

freque

parame

and N

parame

freque

1977;

implem

In many cases, speech samples of the unknown speaker ansmitted and recorded through the system which has the aracteristics as shown in Figure 1 are compared with speech nples of the known speakers transmitted and recorded through a ventional microphone-to-tape recorder which has relatively flat ponse characteristics (linear). In such a case, the magnitude the distortion present in two entirely different transmission recording systems may be even more serious than the case where y one type of the transmission system is being used. It seems preclude any reasonable effort to eliminate this type of tortion made up of the unknown sources of variables. One earcher (Tosi, 1979) being pessimistic about this phenomenon, posed a very interesting idea as a possible solution:

> ....elimination of perturbing telephone influence could consist of including a 'standard' burst sound at the beginning of every telephone communication. Because the real spectrum of such a 'standard' sound would be known, the transfer function (response characteristics) of the telephone line could then be easily computed. (p. 55)

ate, unfortunately, this idea has not yet been realized.

Other alternative methods to eliminate the influence of the uency distortion are 'normalization' procedures on the speech meters extracted (Atal, 1978; Furui, 1981a; Bunge, 1978; Tosi Nakasone, 1980) and selection of the appropriate speech neters which are considered to be inherently resistive to the uency response characteristics (Atal, 1972; Markel *et al.* Hunt *et al.*, 1977). Since these alternative methods are mented in this study, details will be discussed in a later cha

Pur

eff

inf

und

tes

eve

(te

rat

coo

cha

the

all

con

and

pre

min

bet

voi

rec

tha

col

ter.

ose of the Study

The major purpose of this study was to investigate the ctiveness of several speech parameters in eliminating the uence of the transmission and recording devices of the ined response characteristics. These speech parameters were ed in text-independent voice identification by a computer.

It is known that a speaker can be recognized by his/her voice when the content of the text spoken is different -independent). Among many conditions necessary for a high of success, two factors are essential: 1) The speakers are rative (no disguise/mimicry, small variation in voice cteristics from one recording session to another, etc) and 2) ame transmission recording device is used for acquisition of ne unknown and known voices.

n the present study, one of these two factors was under l, *i.e.*, all the speakers were cooperative, rendering clean elatively uniform texts. Each speaker read an excerpt ibed for him and recorded in a single recording session, thus zing possible variation in his speech due to a time interval n sessions. The other factor was intentionally varied, *i.e.*, samples were collected through different transmission and ing devices. An annoying outcome of the latter procedure is wo identical speech samples of one speaker (one sample being ted by one transmission and recording device, another sample

bein
resu
iden
tran
reco
prob
tran

this
char
appl
spec
fund
spee
appl
3).
succ
proc
two
11),
unkr
char
comp
(de
revi
func

ing collected by another transmission and recording device) may sult in different spectral shapes and consequently yield a false entification. This type of problem can be easily solved if the ansfer functions of the transmission systems involved in the cording are well-defined. But the prime obstacle in solving this oblem is that in most cases the true transfer function of the nsmission recording channels is not known.

The present study focused on alternative approaches to solve s problem of eliminating the influence of the undefined response racteristics upon the speech samples. These approaches were 1) lication of the speech parameter IDS (intensity deviation ctra), 2) application of the speech parameter FFC (several damental frequency related measurements), 3) application of the ech parameter LTAS (long-term averaged spectrum), and 4) lication of the composite parameter of 1) and 2) and of 2) and

The IDS is a spectrum statistically derived from a set of cessive short-term spectra. The prototype of computational cedure for IDS prameter was introduced by Bunge (1978). IDS has properties: By definition (details to be discussed in Chapter it cancels the influence of the transmission systems of the own response characteristics, and it represents dynamically ging spectral structures of the speech signals. The FFC is osed of a set of fundamental frequency related measurements ined in this study) derived from a pitch contour. Literature ewed in the field of voice identification indicates that the amental frequency is a sufficiently effective speech parameter in di

respon

record

I

parame

pre-pr

optimi

statis

proced

PDP 11

Scienc

T

parame

1.

2.

3.

distinguishing speakers and is relatively insensitive to the onse characteristics of the transmission line used for rding the speech samples.

In order to enhance the 'elimination' effects, all speech neters (except choral spectra) were subjected to a series of processing procedures, such as pause elimination, feature nization by the hierarchical clustering technique and F-ratio stics, and standardization of features. Most of the dures were carried out by Fortran softwares implemented on a 1/40 minicomputer at the Department of Audiology and Speech ces, Michigan State University.

The following assumptions on the performance of each speech eter were set up:

The IDS is sufficiently effective in eliminating the influence of the frequency distortion from the voice samples of the unknown and the known speaker recorded through different transmission and recording systems, and it is equally effective for the voice samples recorded through the same transmission and recording system.

The FFC is sufficiently effective in eliminating the influence of the frequency distortion from the voice samples of the unknown and the known speaker recorded through different transmission and recording systems, and it is equally effective for the voice samples all recorded through the same transmission and recording system.

The LTAS is highly effective in eliminating the influence of the frequency distortion from the voice samples of the unknown and the known all recorded through the same transmission and recording system, but the effectiveness is decreased when the voice samples of the unknown is recorded through one transmission and recording system and that of the known through another system.

Choral spectrum is assumed to be as effective as the LTAS for the same conditions described in 3.

The composite parameter of FFC and IDS increases the effectiveness level in eliminating the influence of the frequency distortion from the voice samples of the unknown and known recorded through the different transmission and recording system.

The composite parameter of FFC and LTAS increases the effectiveness level in eliminating the influence of the frequency distortion from the voice samples of the unknown and the known recorded through the different transmission and recording systems.

o test the above assumptions, a total of 24 voice fication operations were conducted in different designs ing to various combinations of the speech parameters and the ssion systems used for recording the unknown and the known s. Each operation yielded the results in terms of the rate rect identification, which served as the measure of the veness of each parameter tested.

Signi

alone

perso

chang

ident

much

the o

searc

distu

there

direc

other

resea

trans

text-

large

repor

optim

have

study

text-

the p

and

speak

ignificance of the Study

It is known that a person can be recognized by his/her voice one when heard live or over the telephone line, provided that the rson is somebody the listener is familiar with. Despite some ange in the perceptual quality of the voice, the judgment on the entity of a speaker does not seem to be critically influenced too ch by the transmission line. This is the underlying fact that e objective of this study rests on. This study was designed to arch and study several speech characteristics which are not sturbed by transmission and/or recording devices. It is hoped, erefore, that the results from this study will contribute rectly or indirectly to understand more about human speech sound ner than the linguistic message it carries.

Another justification for this study is the scarcity of earch reports dealing with the problem of the influence of the nsmission and/or recording devices coupled with the t-independent speech materials for voice identification. A ge body of research reports is available, though most of these orts are based on the text-dependent speech data recorded under imaly controlled conditions. To date, only a few researchers e been concerned with these two problems together. One such dy conducted by Hunt *et al.* (1977) consisted of the -independent materials spoken by 13 speakers transmitted over FM radio broadcast. However, they used only one transmitting receiving system of high quality for recording all the kers. Therefore, even though excellent identification rates

were

transm

into a

text-

the u

and

throu

Admit

there

Limit

are a

was

three

softw

utili

infer

text)

condi

intra

be ge

reported in their study, the adverse influence of the mission and recording media does not seem to have been taken account.

The problem addressed in the present study involved independent voice identification tasks by using the voice of nknown speakers distorted by the telephone transmission system the relatively clean voices of the known speakers recorded gh a conventional microphone-to-tape recorder system. tedly, although this problem is very difficult to fully solve, is a legitimate need for investigation.

ation of the Study

This study was exploratory in nature. Therefore, the results oplicable only within the following limitations.

irst, a nominal size of 10 speakers were employed. This size nontrivial considering that actual speaker size was tripled by simultaneous transmission systems and that only Fortran res were available (no real-time hardware processor was ed). Because of this small speaker size, no statistical nce was attempted for generalization of the results.

econd, all the speakers were recorded only once (contemporary while reading in more or less the same style. These two ons obviously contributed in producing the unusually small peaker variability. Thus, the results from this study can ralized only to these types of speech data.

for

cri

dema

in

for

ide

was

thi

exp

mod

app

<u>Bac</u>

of

Whe

dur

con

con

rev

con

pre

cho

(ID

Third, though a mini computer was applied as the major tool data processing including decision algorithms, at a few cal points arbitrary intervention by the experimenter was ded. For this reason, the design of the voice identification is study was not meant to be a completely 'objective' method.

Finally, the term 'voice identification' operationally defined the purpose of this study was 'closed' type, *i.e.*, in each ification process, the speech sample of the unknown speakers always included among those of the known speakers. Presumably type of identification procedure can be considered as ratory, or linient, in terms of its credibility; hence, the tested in this study, as it is, has little immediate cability to real environment.

ound in Selecting Speech Parameters

Speech samples in this study were not only under the influence known response characteristics but also text-independent. ext-independent voice samples are involved, a relatively long on of speech is required in order to homogenize the phonetic t of the different texts from all speakers. Within this aint, several different types of speech parameters were ed from the literature in this field. Speech parameters ered *primae facie* were cepstral coefficients, linear ive coefficients (LPC), long-term averaged spectrum (LTAS), spectrum, variance spectrum, intensity deviation spectrum and fundamental frequency contour (FFC).

its

to th

trans

cepst

to be

The L

respo

remai

spect

sampl

trans

to ta

hiera

witho

LTAS,

the f

and

In sp

spect

ident

same

LTAS,

this

Of these seven parameters, the cepstrum was discarded despite
well recognized useful property -- it is relatively resistant
ne influence of the frequency response characteristics of the
smission system. It was found to be impractical to apply the
tral analysis for the amount of ' text-independent' speech data
e processed by using only Fortran softwares on a mini computer.
PC was also discarded because of its susceptibility to the
onse characteristics of the transmission system used. The
ning five parameters, LTAS, choral spectrum, variance
rum, IDS and FFC were tested in the pilot study. Speech
es were collected from five male speakers through two
mission systems, telephone line and a conventional microphone
pe recorder. The identification process was performed by the
rchical clustering technique with complete-link method,
ut feature optimization procedure.

From the results of the pilot study, it became apparent that
variance spectrum, and choral spectrum were very sensitive to
requency response characteristics of the transmission system
hat IDS and FFC were relatively insensitive to the influence.
te of their susceptibility to the influence, LTAS and choral
a were found to be very effective speech parameters for voice
fication provided all voice samples were recorded through the
transmission and recording channels. Consequently, IDS, FFC,
nd choral spectrum were adopted as speech parameters for
udy.

Lit

rep

div

and

In

pro

the

stu

pur

sev

arr

of

fil

et

cas

rature Review

A review of the related literature revealed that research rts on voice identification by a computer contain vast rsification in the methodology employed depending upon the type number of speakers, phonetic materials, and the tasks involved. ontrast, there are very few reports primarily dealing with the lem of the transmission and/or recording media. It was, efore, felt that organized presentation of these diversified ies in methodology was neither practical nor essential for the ose and the scope of this study. Hence, this section presents al studies dealing with automatic voice identification, nged by the type of speech parameters employed.

A short-term spectrum is generated from a very brief portion speech signal. It can be produced by the use of a bank of rs simulated on a digital computer (Prusansky, 1963; Bricker l., 1971) or by using the Fast Fourier Transform. In either the principal expression used is the Fourier transform:

$$F(\omega) = \int_0^{NT} f(t)e^{-j\omega t}dt$$

(for continuous signal)

or

wher

spac

conv

a s

appr

know

spee

spec

shor

Pruz

shor

17-c

use

time

of

were

in

two

the

that

$$F(\omega) = \sum_{n=0}^{N-1} f(nT)e^{-j\omega nT} \cdot T \qquad \text{(for discrete signal)}$$

ere N = number of samples (discrete) points of a time function, aced apart by a sampling period of T second. As used nventionally, a large number of short-term spectra are created in succession as a function of time. Hence, this spectrum is ropriately called 'temporal acoustic spectra' (Tosi, 1979) and wn to carry "....nearly all of the important information in ech...." (Atal, 1978). The duration of a single short-term ctrum is typically in the order of 20-30 milliseconds.

Some researchers investigated the effectivenesss of the rt-term spectra for a text-dependent speaker recognition. zansky (1963) generated the short-term spectrum by passing the t phrase of four words taken from each of 10 speakers through a hannel filter bank, frequency ranging from 100-7000 Hz. By the of the product-moment correlation coefficients computed on the -aligned spectra, she reported the correct identification rate 89% out of a total of 393 trials tested. The same speech data later tested by Bricker *et al.* (1971) with some modification procedure, resulting in a higher recognition rate of 97%. The studies cited were performed on speech data recorded through same transmission and recording equipment, though it was noted the spectra were strongly influenced by the frequency cha

has

and

imp

alt

tem

spe

spe

197

al.

thi

fil

whe

con

tex

tex

res

197

rec

'ch

as

(Ta

of

ch

acteristics of the recording devices. The short-term spectrum
peen often applied to the text-dependent data for which tedious
complex procedures for time alignment become extremely
tant during the recognition procedure.

Later long-term averaged spectrum was considered as an
native to compensate for the tedious procedures of the
ral alignment of the set of the text-dependent short-term
ra. Typically, long-term averaged spectrum is obtained from a
h signal of a relatively long duration (1-2 minutes in Tosi
70 seconds in Markel *et al.* 1977; 10 seconds in Furui *et*
1972, 11 seconds in Bunge 1978). When a computer is used,
spectrum can be easily produced either by using a bank of
rs, or by FFT algorithm. A unique property of this spectrum,
taken from a sufficiently long speech, is that phonetic
nts of different utterances can be balanced, thus enabling
ndependent voice identification. Its potential use for
ndependent voice identification has been recognized by many
chers (Tosi *et al.*, 1979; Bunge, 1978; Majewski and Hollien,
provided there is no influence of the transmission and/or
ing media.

he variation of the long-term averaged spectrum is called
spectrum' developed by Tosi (1979). He defines the spectra
long-term Fourier transforms of temporal choral speech
zy, 1958), which is produced from a temporal rearrangement
speaker's normal speech. The major difference between
spectrum and long-term averaged spectrum is seen in the

com

fas

con

cho

rec

tec

dep

of

How

lon

be

spe

197

198

me

the

fu

pr

al

co

th

st

sp

re

hi

nputational economy: The former is said to be generated much ster than the other by a factor of about 20. Tosi *et al.* (1979) nducted a text-independent voice identification by applying the oral spectra. Speech samples of different languages were orded from 20 speakers. By using the hierarchical clustering hnique, they reported the identification error rate of 5 to 30% ending upon the method used and suggested the promising utlity the choral spectra for text-independent voice identification. ever, inasmuch as the spectrum bears the same property as the g-term averaged spectrum, the choral spectrum is also known to susceptible to the influence of the transmission media.

Linear predictive coefficients (LPC) has been studied by many ech researchers for automatic speaker recognition (Atal, 1974, 3; Markel *et al.*, 1977; Markel and Davis, 1979; He and Dubes, 2.) The LPC is usually derived by using the autocorrelation hod from speech signals, revealing the spectral properties of speech as a function of time. LPC can represent the amental frequency and its harmonics when the order of the ictor coefficients is relatively high (40 coefficients) and can represent formants when the order of the predictor is low (12 ficients), but it is susceptible to the frequency response of recording apparatus and the transmission systems. Atal (1978) ied the effectiveness of LPC (12 coefficients) for automatic ker identification by using 10 female speakers. All speakers rded six repetitions of the same short sentence, using a -quality microphone on two occasions at 27-day intervals. Each

utt
pre
dim
non
(19
(of

ide
rec
utt
int
div
spe
ser
fir
seg
was
met
fea
pit
pit
al
cor
ana
ide
cor

erance was divided into 40 segments; and from each segment, 12 dictor coefficients were extracted to form a vector of 40 ensions. The identification decision was based on the -Euclidian distance measure defined by Shafer and Rabiner 75.) The correct rate of identification was found to be 63.8% 60 total judgments.)

He and Dubes (1982) presented a paper on speaker tification by using LPC and pitch contour. Speech samples were rded in a sound booth by eight Chinese male speakers, each ring 15 repetitions of a short sentence in the Chinese language a microphone attached to a tape recorder. Each utterance was ded into 4-second epochs, resulting in five speech data per ker (each datum, thus containing either two complete spoken ences or one complete and partially complete sentence.) Then, ly, each datum was partitioned into 40 segments. From each nt, 12 predictor coefficients were computed. A pitch contour also prepared from each datum by two different ds: cepstrum method and peak detecting technique. Five res measured on a pitch contour were maximum, minimum, average period, the maximal slope, and the larger one of the two period values determining the largest slope. Subsequently, data were subjected to feature optimization procedures sting of: The hierarchical clustering technique, discriminant sis, and F-ratio as discussed in Chapter II. For fication decision operation, the Euclidian distances were ted between the test pattern and reference patterns and the dec

rul

pi

to

dis

96

di

th

at

95

di

th

sp

19

fu

do

th

B

r

c

r

r

t

t

ecision criterion was based upon the nearest-neighbor decision ule. The results from their study were as follows: 81.9% by itch contour when all speech data were included, but the rate rose o 96.4% when the data containing partially complete sentences were iscarded; 75.6% by LPC for the entire data, but increased to .4% when the data containing partially complete sentences were scarded. A combination of features from the pitch contour and e LPC was also tested with all speech data included (in an tempt to compensate for varying text contents). This resulted in .8% correct identification. Although the authors of this study d not specify, the speech data were rather text-dependent even if e proper alignment of each phonetic unit was not attempted.

The cepstrum of a speech signal is defined as the power ectrum of the logarithm power spectrum of the signal (Noll, 67). This method was introduced as a means to separate ndamental frequency from the speech signal in the frequency main. Much attention has been given to the cepstral method in field of automatic speaker recognition (Atal, 1978, Luck, 1969; nge, 1978; Tosi, 1979; Furui, 1981b; He and Dubes, 1982). The son for its popular application appears to be twofold: The strum is mathematically well defined, *i.e.*, renders itself for iable algorithmic implementation to a computer, and it is atively resistive to the frequency response characteristics of transmission system as well as the recording devices.

Furui (1981a) published a comprehensive study on the hniques for automatic speaker verification based on the cepstrum coe

(te

con

mal

Sev

fro

the

seg

el

se

wa

ut

pr

Fu

ce

ch

co

ut

co

s

"

t

c

i

c

ficients computed on a fixed, sentence-long utterance t-dependent) recorded over the conventional telephone ection. A total of 50 utterances from each of 20 speakers (10 s, 10 females) were recorded over the period of two months. ral kinds of utterance sets were prepared, all band limited 100 Hz to 3.0 KHz. Cepstral coefficients were derived from predictor coefficients (LPC) obtained from the same speech ent. After several pre-processings -- such as pause ination, time registration, normalization, and optimum feature ction -- were applied, the results of verification error rate less than 1% even if the test utterance and the reference rance were subjected to different transmission conditions (but, ably, within the same telephone connection). The key factor, i commented, would be in the normalization procedure of the trum coefficients to remove the distortions of the response acteristics introduced by the transmission system. The ficients were averaged over the duration of the entire ance, and the average values were subtracted from the cepstrum ficients of every frame.

Pitch contour is a plotting of the time-varying pitch of the h signal. Or, synonymously expressed, pitch contour is he glottal frequency - characteristics melody curve of the r." (Tosi, 1979). There are two properties of pitch ur: First, it is sufficiently speaker-dependent; second, it sistive to the distortion introduced by the frequency response cteristics of the transmission and recording devices. Many res

al

spe

hi

ch

th

id

10

se

up

re

(9

pl

ap

v

T

e

(

L

s

m

e

f

a

earchers (Atal, 1972; Markel, Oshika, and Gray, 1977; Hunt *et* , 1977; He and Dubes, 1982) explored pitch contour as one of the ech parameters for speaker recognition, and confirmed it to be a hly (or, at least sufficiently) reliable speaker-dependent racteristic.

Atal (1972) studied the average pitch and the measurements of temporal variation of pitch for automatic speaker ntification. Speech data were text-dependent and collected from male speakers, each producing 6 repetitions of the short ence, and each sentence lasting 1.8 to 2.8 seconds depending the different rate of utterance by individual speakers. He rted that the measurement of the average pitch was far better correct identification) than that of the temporal variation of h.

Markel *et al.* (1977) also investigated pitch contour but ied it to text-independent speech materials obtained from a small group of speakers with homogeneous pitch distribution. used F-ratio (analysis of variance) as the measure of the ctiveness of the average pitch and the standard deviation puted on pitch contour) as a function of the number of frames, from Lv = 10 to Lv = 1000, which correspond to about 70 nds). It was concluded that the average pitch is significantly effective than the standard deviation of pitch and that the nated standard deviation about the average pitch was reduced about 18 Hz (for Lv= 10) to about 6 Hz (for Lv= 1000). They ed two other parameters,the spectral-related feature obtained

by

of

eff

spe

ide

pa

ea

co

fr

me

fr

fu

pr

pr

Fi

a

cc

cc

ac

sp

cc

o

f

t

d

LPC and gain variation, to the same speech data for the purpose comparing the three speech parameters. Ranking of the ectiveness in discriminating speakers was in the order of ectral feature, pitch contour, and gain variation.

Hunt *et al.* (1977) conducted text-independent voice entification by using pitch contour (and including other speech ameters). They used a group of 12 professional meteorologists, ch reading two sets of different text transmitted over the munication channel. Seven different kinds of fundamental equency related measures were derived from a pitch contour, viz, n fundamental frequency, group mean of the mean fundamental quency and its rate of change, and proportion of time that damental frequency is rising or falling. The pitch contour was pared by the use of a hardware implemented real time cepstral cessor. Identification performance was tested in two ways: st, texts of the unknown and the known speakers were arranged in oncontemporary manner, resulting in 89% (133 out of 149 samples) rect identification; and second, texts were arranged in a temporary manner, resulting in 100% correct identification. In ition to pitch contour, they included two other parameters, ctral related and gain related. When all three parmeters were pared in terms of identification performance, ranking was in r of spectral parameters, followed by pitch contour, then lly, gain related parameters. The study by Hunt *et al.* showed voice identification can be done even if the speech data of erent content were transmitted over the communication channel;

h

t

a

s

owever, the degree and nature of the distortion of the ransmitting system were not specified clearly. It appears that ll speech data in their study were transmitted and recorded by the ame system.

Defi

defi

Cate

Cros

Feat

FFC

IDS

Lin

inition of the Terminology

Key terminologies used throughout this study are operationally ined as follows.

egory: The term refers to a set of patterns (or samples) and is used synonymously to a speaker in this study.

ss-transmission: This term refers to a voice identification procedure in which the speech data from the unknown speaker and the known speaker are prepared through two different transmission systems.

ture: A feature refers to an individual measurement component within a speech parameter. The number of features determines the dimensionality of a parameter. For instance, the first frequency component in the IDS, or the average fundamental frequency in the FFC, is called a feature.

: The FFC is a speech parameter which consists of a set of fundamental frequency related measurements computed on a pitch contour of a running speech sample.

(intensity deviation spectrum): The IDS is a speech parameter derived from a set of successive short-term spectra. The IDS reflects, by definition, temporal variations of the spectra of speech sound.

ar system: This system refers to a path in which the speech sound is transmitted by a microphone and recorded onto an audio tape by using a tape recorder. In this system the microphone and the tape recorder are characterized as having the relatively flat response curve (Linear) covering the

LTAS

Norm

Pat

Sho

Spe

SP

speech frequency range. The term 'Linear speech' or 'Linear -' in this study refers to the speech sound, or processed speech data made available by using this system.

S (Long-term averaged spectrum): The LTAS is a speech parameter computed by superposing and averaging of n short-term spectra. Each one of these spectra is originated by successive segments from the speech samples of about 11 seconds utilized in this study. The LTAS reflects static spectral feature of speech sound.

al system: It refers to a path in which the speech is transmitted by a microphone and recorded onto a magnetic tape by a tape recorder. The system is assumed to have undefined response characteristics. The term 'normal speech' or 'normal -' in this study denotes the speech sound, or processed speech data made available by this system.

ern: It is composed of the set of features chosen from a single or more of the speech parameters. A pattern is equivalent to a speech sample and is the basic data set to represent the voice characteristics of the speaker.

t-term spectrum: This spectrum is generated from a short segment (in this study 25.6 msec) of each processed speech sample by using Fast Fourier Transform (FFT).

ch parameter: The term refers to the measurement(s) derived from acoustic speech signal. The individual feature is extracted from a speech parameter.

(choral spectrum): The SPT is a speech parameter produced

Telep

Text

Tran

Voi

from choral speech by processing it through FFT. An elaborated definition and algorithm to generate this spectrum is presented by Tosi (1979). In this study, choral speech is obtained by superimposing 0.4096 second long segments of on-going speech.

ephone system: It refers to a path in which the speech is transmitted via the telephone transmitter, received at the remote end of the local line telephone set, and recorded onto an audio tape recorder. The term 'telephone speech' or 'telephone -' in the text refers to the speech sound or processed speech data made available by this system.

:-independence: This term refers to the type of phonetic materials used as the speech data for voice identification. Text-independent voice identification uses different texts from the unknown and the known speakers. Counterpart of this term is the 'text-dependence'.

smission system: Restricted to this study, 'transmission system' refers to a system of the devices used for transmitting the speech sound, such as a microphone, telephone transmitter and its attachment, and so forth. The term 'system' is used to refer to a path of the speech signal from the speaker's mouth to the sound storage device, an audio tape recorder.

e Identification: It is defined as a process of selecting a speaker (from a group of the known speakers) whose voice sample is the closest to that of the unknown.

thin-transmission: This term refers to a voice identification procedure in which all speech data from the unknown and the known speakers are prepared by the same transmission system.

Organization of the Study

This study is divided into four chapters.

Chapter I presented a general introduction to voice identification, the statement of the problem, the purpose, significance and limitation of this study, a review of the literature, and a list of operational definitions of the terminologies.

Chapter II is devoted to the description in detail of experimental procedures: Recording of speakers, digitization, pause elimination, generation of speech parameters, feature optimization and standardization, and identification operations.

Chapter III presents the results from the identification operations.

Chapter IV concludes this study by presenting discussions, conclusions, and implications for further research.

T

of p

materi

operat

record

proced

speech

standa

sectic

voice

list c

List c

For re

CHAPTER II

EXPERIMENTAL PROCEDURE

This chapter is organized into three sections: 1) recording phonetic materials, 2) pre-processing of these phonetic terials, and 3) experimental procedure for voice identification erations. In the first section, recording and arrangement of the corded speech data are discussed. In the second section, ocedures and algorithms for pause elimination, extraction of the eech parameters, optimization of the features, and andardization of the features are discussed. In the third ction, distance measurements, decision processes, and designs of ice identification operations are covered. The following is a st of equipment and softwares used throughout this study.

<u>t of Equipment</u>

recording speech data:

Condenser microphone, Bruel & Kjaer, type 4132
Cathode follower, Bruel & Kjaer, type 2619
Microphone amplifier, Bruel & Kjaer, type 2603
Dynamic microphone, Ampex, model 2001
Local line telephone sets
Open-reel tape recorder, Teac, model A-7010
Open-reel tape recorder, Sony, model TC-106A
Cassette tape recorder, Marantz, Superscope, model C-202-LP
Open-reel tapes, Scotch, low noise, 1.5 mil, 1200 ft
Cassette tapes, 3M, low noise, 30-minute (one side)

For

Spe

6-m

ran

Mic

fre

rea

a

as

no

re

op

ma

tr

tw

processing speech data:

PDP 11/40 mini computer, 64 k(byte) memory, with 2 disk drives
16-bit A/D and D/A converters, 3 Rivers Computer Corp.
RK05 disks, 2.4 Mega bytes
CRT monitor
Light pen connected to the CRT monitor
Deckwriter II, Digital Equipment Corp.
Open-reel tape recorder, Ampex, model 4000G
Fortran software (see Appendix H)

RECORDING OF PHONETIC MATERIALS

kers and Phonetic Materials

The phonetic materials used in this study consisted of nute long speech samples. The subjects were 10 male speakers omly selected from a population of native speakers of estern American-English dialect, ages ranging from 20 - 35, from defective or pathological voice conditions. All speakers different excerpts from a nontechnical book (Appendix A shows mple excerpt) at a 'normal' reading speed. Each speaker was d to rehearse by reading aloud a brief paragraph (one which was going to be included as speech data for him) while all the rding equipment and the telephone line were checked for proper ation. During recording, each speaker was instructed to tain approximately the same distance from his mouth to the smitter of the telephone set (about 3-5 cm) and to the other microphones (about 15 cm). No additional instructions as to

the

Reco

spea

syst

to

tran

reco

micr

char

tele

tran

ins

tel

Spe

att

the

Sup

cha

tra

in

mod

cal

mic

he manner in which the speaker should read the excerpt were given.

ecording Setting

Figure 2 illustrates the simultaneous recording of each peaker through three different transmission and recording ystems: 1) through a telephone line with the remote end connected a tape recorder by an inductive pick up ('telephone ansmission'); 2) through a conventional microphone-to-tape corder ('normal transmission'); and 3) through a crophone-to-tape recorder of an almost linear frequency response aracteristics. Hereafter, these three systems are referred to as lephone transmission, normal transmission, and linear ansmission system, respectively.

For telephone transmission, the telephone set was placed side of a sound booth and dialed up to the other end of the local lephone system (campus line at Michigan State University.) eech signals were drawn by the use of an inductive coil directly tached around the receiver of the telephone set, and connected to microphone input of a cassette tape recorder (Marantz, perscope model C-202LP). No care was taken to check the response racteristics of this telephone transmission system. For normal nsmission, a dynamic microphone (Ampex, model 2001) was placed a sound booth and connected to an open-reel tape recorder (Sony, el TC-106A) outside the booth. No care was taken for ibrating the frequency response characteristics of this rophone and tape recorder either. Thus, a normal transmission

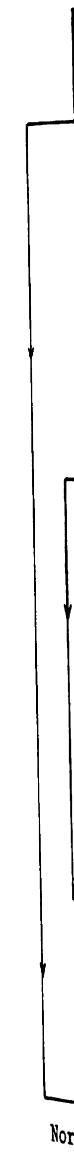
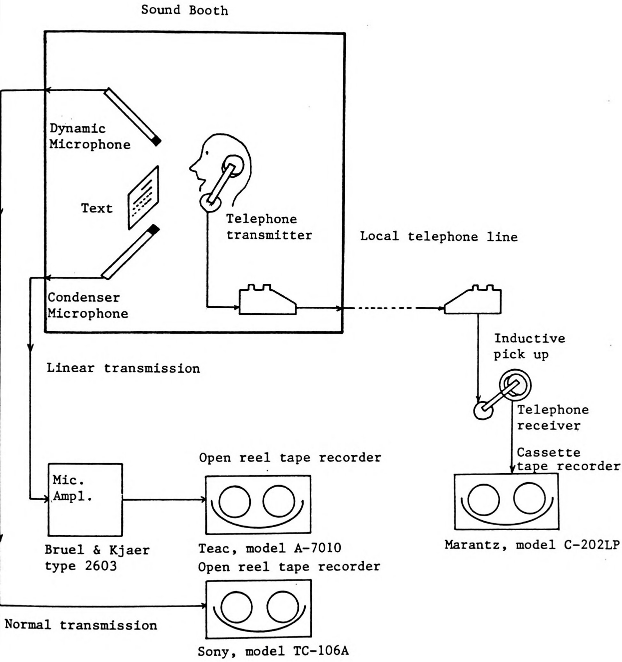


Figure 2.
trans



e 2. A diagram showing equipment used for three simultaneous
ansmission and recording systems.

system

For 1

type 4

type 2

2603)

model

were

charac

these

becau

trans

long

Arran

known

as ar

a kn

were

iden

illu

Figu

tran

The

tran

system was assumed to have the undefined response characteristics. For linear transmission, a condenser microphone (Bruel & Kjaer, type 4132) was coupled with the cathode follower (Bruel & Kjaer, type 2619) connected to a microphone amplifier (Bruel & Kjaer, type 2603) outside the booth, then to an open-reel tape recorder (Teac, model A-7010). This condenser microphone and Teac tape recorder were calibrated for their linearity of the response characteristics. Plottings of the response characteristics of these two devices are given in Appendix B.

Although each speaker read only one 6-minute long text, because of the above described simultaneous recordings by three transmission systems, each speaker produced a total of 18-minute long speech data.

Arrangement of Speech Data

In this study, all speakers served both as unknown and as the known persons. This required that the speech sample of a speaker as an unknown differed in context from that of the same speaker as a known. Therefore, the speech data stored in audio tape recorders were properly arranged to enable 'text-independence' from three identical texts simultaneously produced by a speaker. Figure 3 illustrates this procedure for proper arrangement. As shown in Figure 3, a 6-minute long speech of each speaker for each transmission was partitioned into three 2-minute long portions. The initial 2-minute portion was then segmented from telephone transmission (indicated by ①), the medial portion was segmented

Figu

Text B by
Speaker 2

Text A by
Speaker 1

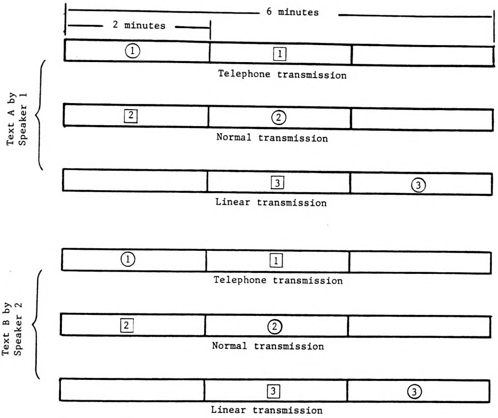


Figure 3. Sample arrangements of text-independent phonetic materials from two speakers. For the cross-transmission voice identificaiton operation: The segmented portion 1 in a circle in telephone transmission was used to prepare the data base for the speaker as an unknown while portions 1 and 2 in cirles in normal transmission and linear transmission were used to prepare the data base for the same speaker as the knowns. For the within-transmission voice identification: For each transmission system, the segmented portion in a circle was used to prepare the data base as an unknown and the portion in a square, as a known speaker.

from nc

was se

arrange

cross-

of por

as us

The la

in a

data.

conten

system

speech

l

Digit

stored

recor

analo

(PDP

sampl

in a

om normal transmission (indicated by ②), and the final portion, ıs segmented from linear transmission (indicated by ③). This rrangement was conducted to prepare input speech data for ross-transmission voice identification . In addition, another set f portions was necessary to prepare 'text-independent' speech data s used in within-transmission voice identification operations. he latter set of partitionings are indicated by the number encased n a square seen in Figure 3. Unmarked portions were not used as ata. Consequently, two 2-minute portions of different phonetic ontent were segmented as raw speech data from each transmission ystem and represented a speaker. In total, 60 2-minute long peech data resulted from the above arrangement:

10 (speakers) x 3 (transmissions) x 2 (portions) = 60.

PRE-PROCESSING OF SPEECH DATA

igitization

During the digitization process, each analog speech sample ored in the original magnetic tape was played back on the same cording equipment which was used to record it. A 16-bit alog-to-digital converter (ADC) interfaced with a mini computer P 11/40) digitized a 2-minute long speech one at a time, at a pling rate of 10000/sec. Then the digitized speech was stored a disk (2.4 Mega bytes) for the subsequent pause elimination proced

(=1 wo

the AD

F

of no

digit

varia

Pause

autom

elimi

alter

part

extr

each

abou

spec

elim

sta

def

def

pau

by

procedure. Each sampled point (digitized) was quantized by 2 bytes (=1 word) resulting in dynamic range of about 90 dB as specified by the ADC.

Frequency transfer function of the ADC indicated some amount of nonlinearity. However, this nonlinearity (distortion) in the digitization process was not considered as the source of the variables because it was constant for all input speech data.

Pause Elimination

The silent portions and pauses were detected and deleted automatically from the speech samples. The objective of pause elimination was to reduce the amount of speech data without altering speech characteristics. Also, this procedure was particulary important for properly computing IDS parameter whose extraction was based on a set of successive short-term spectra, each spectrum being transformed from a brief speech segment of about 25 msec long. It was deemed essential that no short-term spectrum was resulted from the 'silent' portions or 'pauses'.

In many studies on automatic speaker identification, elimination of the silent portions and pauses is included as a standard procedure, though no author has provided a specific definition of the pauses. In this study, pauses were clearly defined by applying the quantitative pausometric definition of pauses proposed by Tosi (1974). The entire process was implemented by Fortran software on the PDP 11/40.

T

that

determ

determ

is de

only

the v

resid

that

and

phone

pause

be sc

respe

the

repr

and

spee

a sp

fixe

and

each

elim

samp

This process was performed in the time domain in such a way
when the signal falls within two pre-set parameters (one to
rmine the amplitude threshold, 'Ap', and the other, to
rmine the time threshold, 'Tp') that portion of the speech wave
etermined as a pause. The result was a concatenated speech of
'signals' from which all pauses were eliminated. Initially,
values for Tp and Ap parameters were sought by listening to the
dual pauses (deleted and concatenated for audio playback so
the most unvoiced and weak consonants, such as /f/, /θ/,/h/,
brief portions preceeding to and following after plosive
emes such as /p/, /t/, and /k/ were detected and deleted as
es. Consequently, typical values for Tp and Ap were found to
omewhere between 15 to 20 milliseconds, and 0.90 to 0.99,
ectively.

Figure 4 illustrates how pauses were defined and deleted from
input speech wave. Figure 5 shows spectrographic
esentations of the original input speech containing all pauses
the resulting output speech without pauses.

The goal of pause elimination was to obtain a pause-free
h sample of about 55 to 60 seconds long from each portion for
aker per transmission. The lower boundary of 55 seconds was
so that five 11-second epochs were secured for each speaker,
e upper boundary was set to provide a margin of one second to
epoch. Summary of the results from the above pause
ation procedure is given in Appendix C. Subsequently, this
was subdivided into five 11- second long epochs, each epoch

Fi

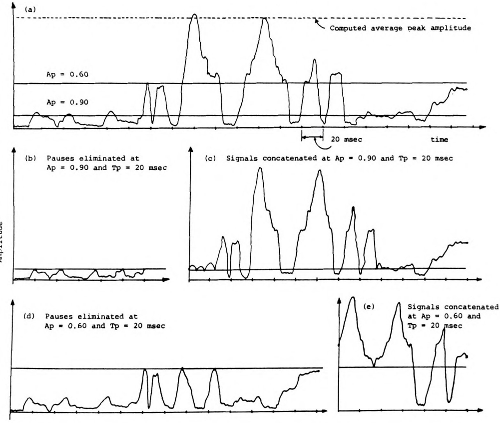


Figure 4. A graphic illustration of pause elimination procedure.
(a) Input speech signal of about 2 second long is subjected to the pause elimination procedure with a Tp = 20 msec and two different Ap values. A dotted horizontal line is the computed average peak amplitude determining the amplitude dynamic range. The Ap is the proportion of this dynamic range.
(b) Pauses are eliminated at Ap = 0.90 and Tp = 20 msec. (c) Signals are detected from (a) at Ap = 0.90 and Tp = 20 msec and concatenated. (d) Pauses are eliminated at Ap = 0.60 and Tp = 20 msec.
(e) Signals are detected from (a) at Ap = 0.60 and Tp = 20 msec and concatenated.

(a) Input speech

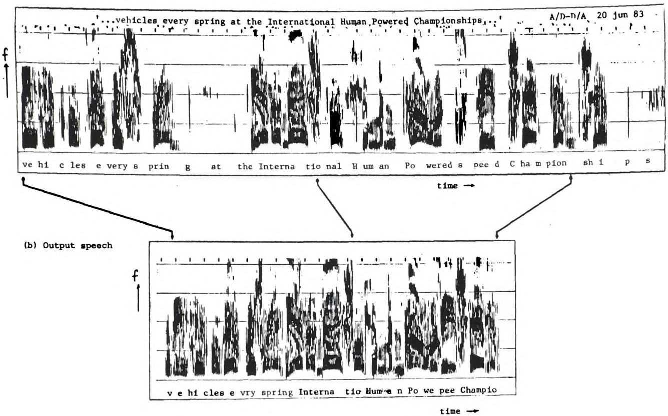


Figure 5. Two spectrograms showing examples of an input speech and a resulting output speech after pause elimination. (a) Input speech of about 4 second duration including pauses. (b) Output speech (pauses eliminated at Ap = 0.90 and Tp = 20 msec) resulting into about 2 second duration.

to be u

process

analog

tape re

and ch

Extrac

1. I

A

an 11

(one p

were

by us

was

cover

about

where

o be used for generating speech parameters, IDS and LTAS. All the rocessed speech segments, epoch by epoch, were also stored as nalog speech (by playing back through the DAC) onto audio tapes by ape recorder (Ampex, model 4000 G) for later use to generate FFC nd choral spectra.

<u>traction of Speech Parameters</u>

IDS (intensity deviation spectrum)

An IDS was generated from a set of short-term spectra. From 11-second epoch of the processed speech, a set of 440 portions ne portion = 25.6 millisecond, or window of 256 sampled points) re transformed to the corresponding number of short-term spectra using the FFT (Fast Fourier Trasform). Each short-term spectrum s then represented by the intensity at 128 discrete frequencies, vering the frequency range from 0 to 5000 Hz, with an interval of out 39 Hz. The IDS was computed by the following expression:

$$P_{ik} = \frac{1}{\bar{S}_{ik}} \sum_{j=1}^{J} \left| S_{ijk} - \bar{S}_{ik} \right|$$

ere; P_{ik} = intensity of the ith frequency of the kth IDS,

$\bar{S}_{ik}$ = average intensity of the ith frequency over J short-term spectra from the kth segment,

S_{ijk} = intensity of the ith frequency of the jth short-term spectrum from the kth segment.

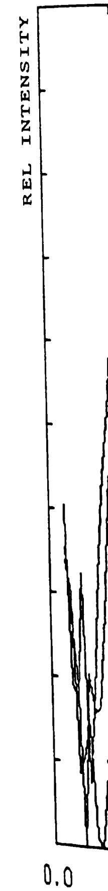
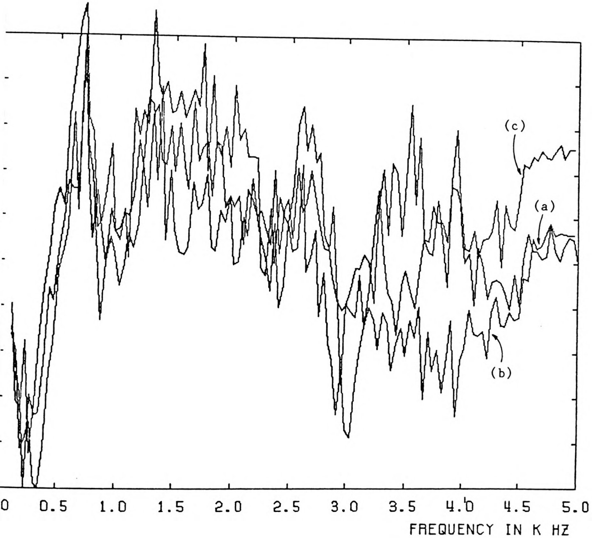


Figure
text
(a)



gure 6. Computer plottings of three IDS's generated from the
text-independent speech data of speaker 1:
(a) by telephone, (b) by normal, and (c) by linear system.

Es

express

these

numbers

frequen

at th

differ

3) Re

spectr

C

for e

assign

were g

Figur

Speec

trans

2.

set

Which

Essentially, the above expression to compute the IDS can be pressed by 1) Adding the intensities within each frequency of ese 440 short-term spectra and dividing the sum by the total mbers of spectra, thus obtaining the average intensity for that equency; 2) Subtracting the average intensity from each intensity that particular frequency, and taking the sum of all absolute fferences, then dividing this sum by the same average intensity; Repeating above steps for all ordinates over 440 short-term ectra.

Consequently, an IDS was represented by 128 intensities -- one each frequency available in the spectrum. Five IDS's were signed to each speaker per transmission. In total, 300 IDS's re generated:

5(IDS) x 10(speaker) x 3(type of transm.) x

2(cross- or within- transm.) = 300

ure 6 shows three computer plottings of IDS's generated from ech samples of a speaker recorded via three different nsmission systems.

LTAS (Long-term averaged spectrum)

An LTAS was computed by averaging the ordinate values across a of successive 440 short-term spectra, the same set of spectra ch generated the IDS. It was computed by:

$$L_{ik} = \frac{1}{J}\sum_{j=1}^{J} S_{ijk}$$

where; L

L

S

s

Total nu

three co

speaker

3. FF

An

segment

been pr

speech

10000/s

for mea

A

S

peak

review

and Du

requir

should

The o

avera

were

ere; L_{ik} =average intensity of the ith frequency for the kth LTAS,

S_{ijk} = intensity of the ith frequency of the jth short-term spectrum for the kth segment.

tal numbers of LTAS's generated was also 300. Figure 7 shows ree computer plottings of LTAS generated from speech samples of a eaker recorded via the three different transmission systems.

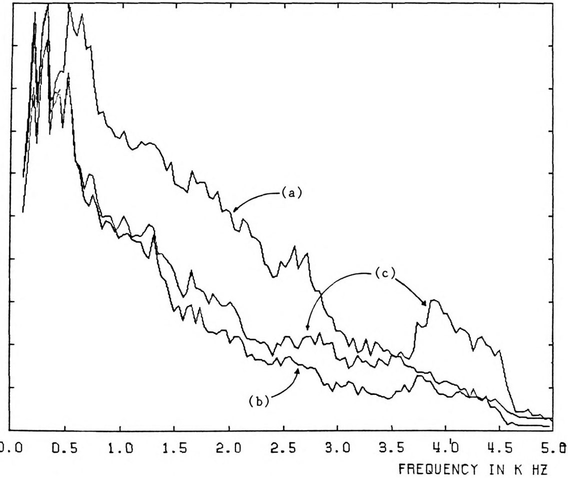
FFC (Fundamental frequency related measurements)

An FFC was prepared from the first five seconds of each gment of the pause-deleted speech sample which has previously en processed and stored in the audio tape. Once again, this eech segment was digitized by the ADC at a sampling rate of 000/second, one segment at a time. Then it was further processed measuring fundamental frequencies (Fo) as described below.

A. Detection of Fo

Several techniques of the computer implementation of direct k detection, which estimate Fo's from the digitized speech were iewed from the existing literature (Gold and Robiner, 1969; He Dubes, 1982.) The review indicated that this type of technique uires frequent heuristic adjustments when the incorrect Fo's uld occur, tending to result in a grossly smoothed Fo contour. one developed by He and Dubes was tested for computing the rage Fo from a half second long speech signal, and the results e compared to the ones estimated by laboratory equipment (sound

Figure
te
(a



igure 7. Computer plottings of three LTAS's generated from the text-independent speech data of speaker 1: (a) by telephone, (b) by normal, and (c) by linear system.

spectro

excelle

indicat

informa

problem

Fo

applyir

a new p

study.

experin

technic

procedu

C

domain

Figure

segmen

while

method

of abc

wave

line b

fundar

entire

delete

segme

were

ctrograph, visipitch, and oscilloscope). The match was ellent. Nevertheless, this preliminary experimentation cated that a direct peak picking technique may lose significant ormation of Fo variations. Markel (1977) also attests to this lem.

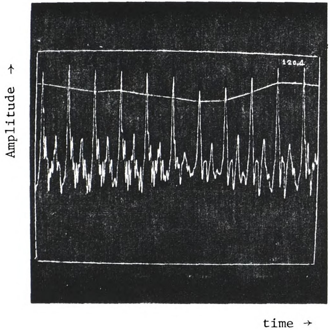
For the reasons listed above and inasmuch as the aim of ying the FFC was to represent fine glottal dynamic variations, w peak detecting technique was devised specifically for this y. This technique demands interactive participation of the rimenter; hence, it is called 'interactive peak detecting nique.' Discussion of this technique and measurement edures for the FFC follows.

Cycle-to-cycle Fo's were measured directly from the time in speech by the use of an interactive peak picking method. re 8 shows a photograph of the actually displayed speech ent on the CRT (display screen operated via the PDP 11/40) e Fo detection was in progress. Major steps involved in this od were as follows: 1) displaying of the digitized speech wave bout 100 milliseconds on the CRT, 2) visually inspecting the to determine recurrent wave patterns, 3) drawing a flexible by a light pen capturing those recurrent peaks associated with amental frequencies, and 4) repeating the above steps until an e 5-second segment was exhausted. Inevitably, because of ed pauses, there were discontinuity points within the speech nt under process. These discontinuities in the displayed wave carefully avoided so that they would not be falsely considered

(a)

(b)

(a)



(b)

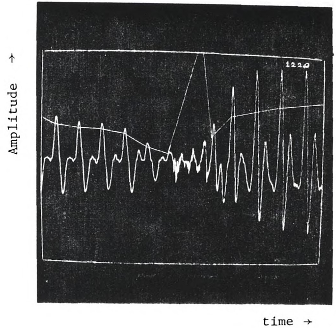


Figure 8. Photographs of the CRT displaying the interactive peak detecting procedures. Each speech segment shown is about 102.4 msec (1024 sampled points) long. A flexible horizontal line traversing the recurrent peaks is a trace entered by a light pen operated by the experimenter. The number appearing on the upper right corner is the updated fundamental frequency measured from the right most peak interval(period). (a) The segment with no gap present within the frame displayed. (b) The segment with a gap, or unvoiced phoneme, present which is skipped (jumped) by light pen maneuver.

as pitch

pen mane

display

automat

An

pitch c

two end

partici

periods

reliab

B

N

Fo con

below.

number

of the

i.

ii. (

ɔitch periods. This manipulation was carried out by the light maneuver. However, the continuity from one frame of the ɔlayed signal to the subsequent frame was maintained ɔmatically by a software.

An output of this interactive peak detecting technique was :h contour containing successive pitch periods and amplitudes of ends of the pitch periods. Although it required a careful icipation of the experimenter to correctly target pitch ods, this method was found to be quite simple, quick and able.

B. Extraction of FFC

Nine features (measurements) were computed for the FFC on each ontour created earlier. Computational procedures are described w. For all procedures; Fo= fundamental frequency, N = total ers of Fo's (in each Fo contour), and Ao = relative amplitude he peak (of Fo).

$\bar{F}_o$: The average Fo computed on a Fo contour.

$$\bar{F}_o = \frac{1}{N} \sum_{n=1}^{N} F_o n$$

σF_o : The standard deviation of Fo.

$$\sigma F_o = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (F_o n - \bar{F}_o)^2}$$

iii. $\bar{\Delta}F_0$

iv. $\bar{\Delta}F_0/$

v. σA_0

vi. $\bar{\Delta}A_0$

vii. F_0

viii. F_0

$\bar{\Delta}F_o$: The average temporal variation of Fo in successive cycles.

$$\bar{\Delta}F_o = \frac{1}{N-1} \sum_{n=1}^{N-1} |F_o n+1 - F_o n|$$

$\bar{\Delta}F_o/\bar{F}_o$: The ratio of the temporal variation of Fo to the average Fo.

σA_o : Standard deviation of cycle-to-cycle peak amplitudes

$$\sigma A_o = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (A_o n - \bar{A}_o)^2}$$

$\bar{\Delta}A_o$: The average temporal variation of peak amplitudes

$$\bar{\Delta}A_o = \frac{1}{N-1} \sum_{n=2}^{N-1} |A_o n+1 - A_o n|$$

$F_o(max)$: The maximun Fo in a pitch contour.

$$F_o(max) = Max[F_o 1,\ F_o 2, \cdots,\ F_o N]$$

$F_o(min)$: The minimum Fo in a pitch contour.

$$F_o(min) = Min[F_o 1,\ F_o 2, \cdots,\ F_o N]$$

ix. F

Figure

frequen

above.

4. S

S

elsewh

spectr

a ch

speech

durat

stack

speec

gener

(1-12

inter

the

para

show

$F_o(rng)$: The range of the Fo. The Fo(rng) is simply computed by

$$F_o(rng) = F_o(max) - F_o(min)$$

re 9 illustrates three basic measurements (periods, fundamental uency, and amplitude) required for computations described e.

SPT (choral spectrum)

Since the detailed procedure to create an SPT is given where (Tosi, 1979), here only a brief description of the trum is presented. A choral spectrum is a Fourier Transform of horal speech. A choral speech is generated by segmenting a ch wave of a certain duration into n portions of the same tion t (in this case 0.4096 second, or 4096 sampled points) and king entire n portions, one on top of another, resulting in one ch segment of length t. Choral spectrum used in this study was rated by using FFT resulting in 1-byte integer intensity values 20 dB) at 2048 discrete frequency components with about 2.44 Hz rval. For this study, all choral spectra were generated from previously processed speech segments which were used for other meters. A total of 300 choral speech resulted. Figure 10 s a computer plotting of a sample choral spectrum.


Amplitude

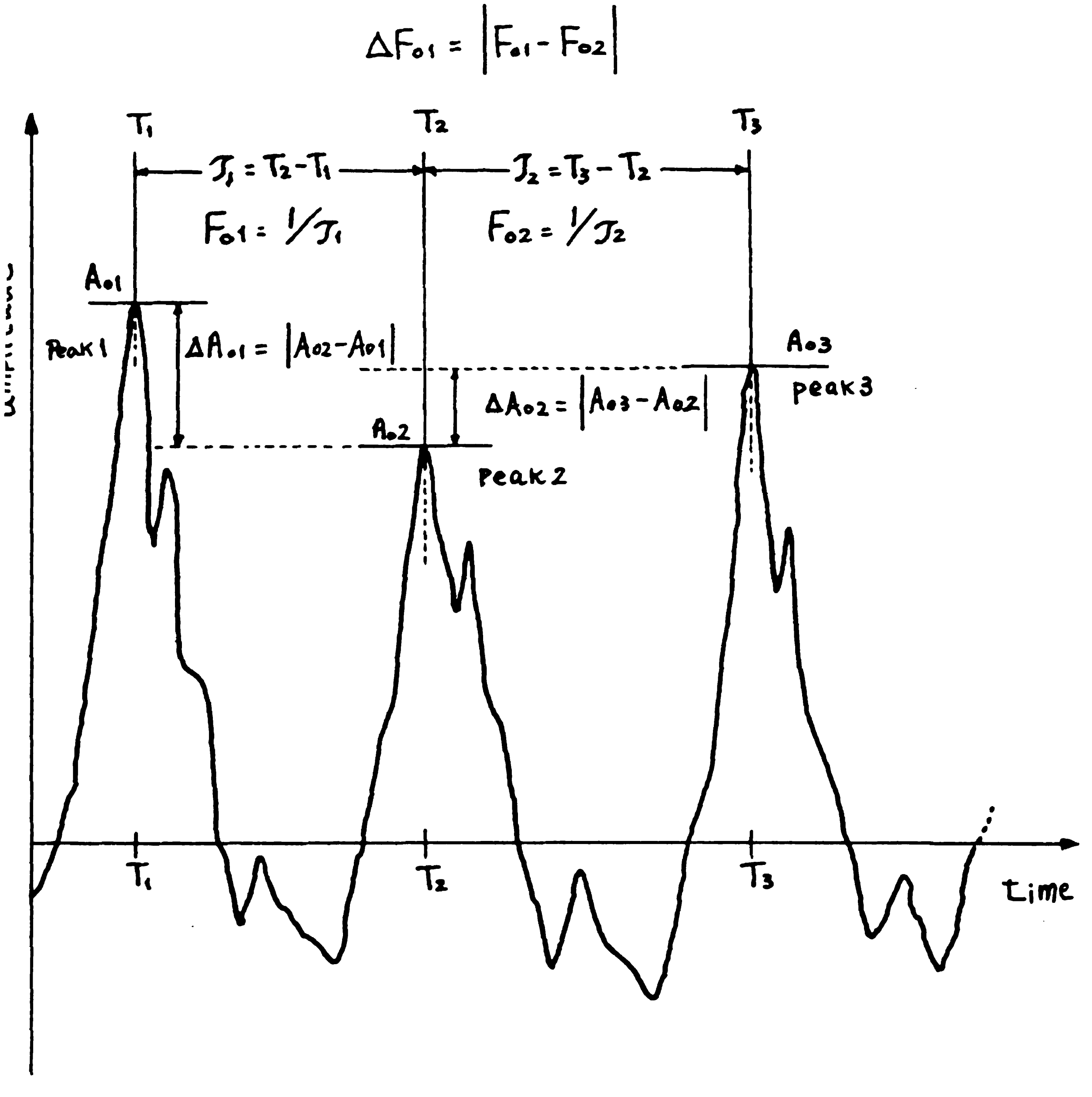


Figure 9. Illustration of Fo, ΔFo, and ΔAo by using three consecutive peaks in a simplified short speech segment. In this figure, Fo = fundamental frequency, $\mathcal{J}$i= ith pitch period, and Aoi = amplitude of the ith peak.

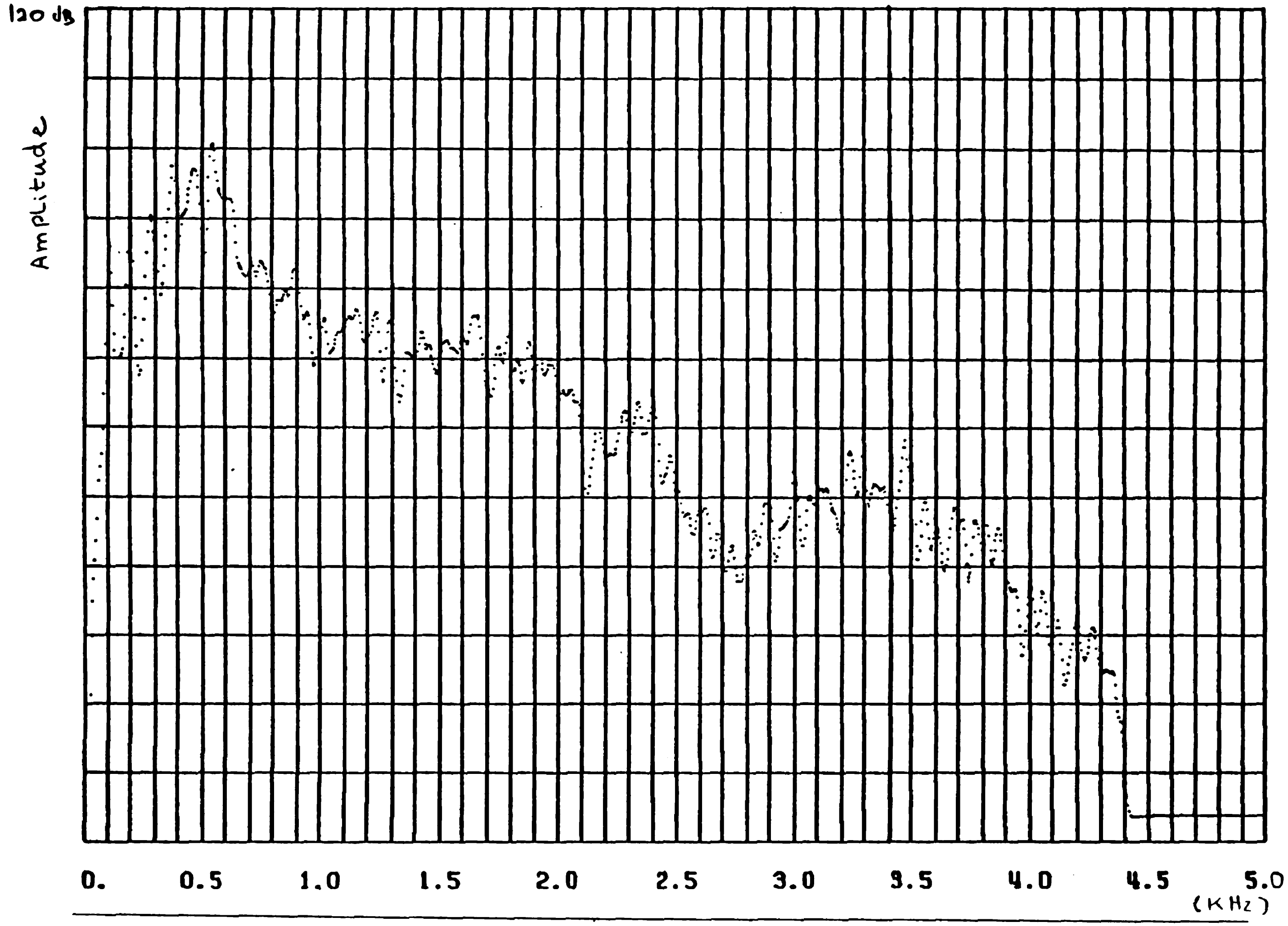


Figure 10. Computer plotting of a sample choral spectrum based on a 11-second long compressed speech taken from speaker 1: The duration of choral speech from which this choral spectrum was produced by FFT was 0.4096 second (or 4096 sampled points used as window size by FFT).

Optimizat

The

of featu

features

between t

has been

identific

relative

in a ge

the mean

increase

number,

number

sizes of

(1978)

should b

Ta

optimiz

contain

of fea

for the

hierarc

F-ratio

hierarc

two-st

used.

Optimization of Features

The objective of feature optimization was to reduce the number of features for computational simplicity by selecting only those features which were determined to be effective in discriminating between the speakers. The importance of the optimization procedure has been recognized in many studies on automatic voice identification, especially when the number of speaker is very small relative to the number of features utilized. Hughes (1968) showed in a general statistical model that for a fixed number of samples, the mean identification accuracy increased when the dimensionality increased until an optimum value was reached. Beyond the optimum number, the accuracy decreased linearly. He suggested the optimum number of features to be 5, 10, 20, and 100 or greater for sample sizes of 20, 100, 500, and larger, respectively. Jain and Dubes (1978) suggested as a rule of thumb that the number of features should be at least five for each sample.

Taking these notions suggested above into consideration, the optimization procedure was applied to the original set of features contained in each speech parameter used in this study. The number of features was 128 for both IDS and LTAS, nine for FFC, and 2048 for the choral spectrum. Optimization was carried out by using the hierarchical clustering technique with complete-link method, and/or F-ratio statistics. For the parameters, IDS and LTAS, both the hierarchical technique and the F-ratio statistics were applied in a two-step sequence, whereas for the parameter FFC, only F-ratio was used.

The

tool ir

speech

word

identif

and th

(1973).

particu

cluster

and Du

pattern

themse

statis

resear

as a p

voice

parame

1.

were

of co

compu

compo

high

by 1

The clustering technique has been employed as an effective tool in many scientific endevors including the field of general speech research. This technique has been applied to the study of word recognition (Rabiner *et al.*, 1977) and also to voice identification by computer (Tosi *et al.*, 1979). Elaborated review and theory of the clustering technique is provided by Anderberg (1973). The use of this technique in this study was in only one particular way to illustrate the diverse applicability of the clustering technique. This particular usage was suggested by Jain and Dubes (1978) as a means to reduce the number of features in a pattern by finding those features which are highly correlated among themselves. This technique was later coupled with F-ratio statistics, which also has been often implemented by some researchers (Paul *et al.*, 1975; Markel and Davis, 1979; Atal, 1978) as a part of the procedure for selecting effective features for voice identification.

Feature optimization procedures applied to the speech arameters used in this study are discussed next.

. IDS and LTAS

First, the original number of 128 features for IDS and LTAS ere reduced to 100. This reduction was imposed by the limitation f computer memory capacity available on PDP to carry on further omputational procedures. Sacrificing the lowest three frequency omponents (78 Hz and lower) and highest 25 components (4017 Hz and igher up to 5000 Hz), the new IDS and LTAS were then represented y 100 features of frequency ranging from 117 to 3978 Hz, with the

interval
carried
complete-
Hier
hierarchi
features
Spearman-
all pai
speaker)
correlat
hierarch
prepared
three s
transmis
From e
compone
horizon
six den
F-
statist
the de
each c

terval of 39 Hz. Optimization procedures were sequentially rried out by the hierarcical clustering technique with nplete-link method, and then by F-ratio statistics.

Hierarchical Clustering: The objective of using the erarchical clustering technique was to determine subsets of atures which were highly correlated among themselves. The earman-product moment correlation coefficient was calculated for l pairs of features over a set of 50 patterns (5 for each eaker). Then, a similarity matrix was prepared from these rrelation coefficients and submitted as input data to the erarchical clustering technique. Six similarity matrices were epared from six sets of 50 patterns (three sets for the IDS, and ree sets for the LTAS, both based upon three different ansmission systems), resulting in six complete-link dendrograms. om each dendrogram, 10 clusters (groupings by frequency nponents) were systematically chosen by means of placing a izontal line at the appropriate proximity level. The resulting dendrograms are presented in Appendix D.

F-ratio statistics: The objective of application of F-ratio tistics was to pick the best feature from each cluster formed in dendrogram. The F-ratio was computed for every features within h cluster. This F-ratio was expressed as:

$$F = \frac{\text{Between (inter-) speaker variance (considering each speaker as a group)}}{\text{Within (intra-) speaker variance (considering 5 samples per speaker)}}$$

Basica

the d

F-rati

and a

best f

T

of th

featu

dendr

featu

indic

whil

the

woul

type

was

fol

fea

a.

b.

c.

Basically, the larger the F value of the feature is, the greater the discriminating power as indicated by that feature. All F-ratio's so computed within each cluster were then rank-ordered and a feature which yielded the largest F-ratio was chosen as the best feature in that cluster.

Table 1(a) shows that three different transmission data bases of the IDS parameter resulted in varying compositions of the features and their F-ratios whithin each cluster formed by the dendrogram. For example, in the case of telephone IDS, two features, *viz.*, 117 Hz and 195 Hz, were grouped in one cluster indicating a relatively high correlation between the two features, while the same two features were grouped in different clusters in the case of normal IDS and of linear IDS.

Since it was necessary condition that each resulting pattern would be composed of the same set of features regardless of the type of transmission systems, some degree of arbitrary interaction was introduced by the experimenter. By inspecting Table 1(a), the following strategic steps were taken to determine 10 optimum features for IDS.

a. The minimum F-ratio denoted by Fm was chosen as a cut-off criterion which was arbitrarily set to Fm = 5.0.

b. From each cluster of features in the telephone column, a feature with the largest F value (and equal to or greater than Fm) was chosen.

c. F values of the same feature in other two transmission data bases (normal and linear) were checked if they exceeded or

Table 1(

cl. 1

cl. 2

cl. 3

cl. 4

cl. 5

cl. 6

cl. 7

le 1(a). Results of the feature clusterings and the corresponding F-ratio's of IDS in three transmission systems.

Telephone transmission

Cluster	Feature	Frequency	F-ratio
	61	2461	3.293
	59	2383	2.780
	60	2422	1.849
	65	2617	3.369
	64	2578	5.121
	63	2539	2.961
	62	2500	2.231
	53	2148	2.921
	49	1992	3.212
	52	2109	3.242
	58	2344	3.767
	54	2187	3.056
	51	2070	3.815
	50	2031	6.411
	56	2266	3.613
	55	2226	4.742
	9	430	8.020
	46	1875	3.351
	7	351	6.984
	4	234	4.408
	43	1758	2.581
	45	1836	1.490
	42	1719	3.559
	48	1953	2.755
	47	1914	6.170
	2	156	11.321
	1	117	11.018
	6	312	6.974
	3	195	30.276
	10	469	3.411
	8	391	4.934
	5	273	13.982
	23	976	2.670
	20	859	3.976
	27	1133	4.741
	24	1016	10.585
	57	2305	2.838
	98	3906	18.708
	97	3867	22.920
	100	3984	14.274
	99	3945	17.697
	14	625	10.975
	21	898	8.353
	11	508	7.399
	12	547	2.399
	96	3828	14.268
	22	937	8.211
	19	820	6.266
	18	781	6.994
	17	742	12.342
	16	703	12.674
	15	664	11.810
	39	1601	2.666
	13	586	5.546
	38	1562	4.660
	35	1445	4.358
	33	1367	6.723
	93	3711	17.425
	90	3594	20.345
	92	3672	17.379
	95	3789	22.447
	94	3750	23.281
	91	3633	21.033
	86	3437	26.927
	85	3398	34.200
	83	3320	23.286
	88	3516	39.217
	84	3359	34.283
	87	3476	31.000
	89	3555	21.620
	⋮		

Normal transmission

Cluster	Feature	Frequency	F-ratio
cl. 1	45	1836	3.288
	44	1797	2.070
	46	1875	2.268
	43	1758	1.290
	66	2656	2.109
	64	2578	2.367
	68	2734	4.379
	67	2695	4.844
	65	2617	5.109
cl. 2	61	2461	5.266
	58	2344	4.103
	62	2500	1.932
	59	2383	2.347
	63	2539	3.615
cl. 3	22	937	5.290
	21	898	5.250
	12	547	6.083
	2	156	20.341
	1	117	23.746
	60	2422	5.333
	5	273	19.856
	33	1367	3.360
	14	625	6.213
	13	586	12.211
	11	508	13.913
	34	1406	8.474
	95	3789	10.824
	93	3711	22.716
	96	3828	13.245
	94	3750	12.615
	98	3906	12.290
	97	3867	14.553
	99	3945	9.521
	100	3984	9.638
	10	469	3.699
cl. 4	54	2187	2.846
	51	2070	11.504
	53	2148	2.800
	49	1992	3.475
	76	3047	4.404
	55	2226	3.547
cl. 5	39	1601	4.215
	38	1562	4.213
	6	312	8.012
	3	195	61.350
	17	742	6.774
cl. 6	57	2305	1.426
	50	2031	2.276
	25	1055	1.756
cl. 7	30	1250	5.421
	27	1133	4.669
	24	1016	4.148
	26	1094	5.000
	29	1211	7.751
	28	1172	10.974
	32	1321	9.278
	31	1289	4.883
cl. 8	69	2773	1.679
	42	1719	0.909
	73	2930	5.055
	71	2851	2.389
	72	2891	3.949
	70	2812	2.496
cl. 9	36	1484	12.580
	35	1445	5.071
	37	1523	7.574
	23	976	3.416
	75	3008	5.343
⋮			

Linear transmission

Cluster	Feature	Frequency	F-ratio
cl. 1	29	1211	6.235
	28	1172	7.687
	26	1094	4.688
	27	1133	11.984
	25	1055	3.455
	49	1992	4.106
	48	1953	2.629
	51	2070	3.194
	52	2109	1.546
cl. 2	10	469	5.267
	9	430	10.258
	1	117	31.983
	44	1797	13.130
	43	1758	4.475
	45	1836	7.967
	5	273	16.527
	2	156	18.285
	13	586	10.644
	40	1641	4.126
	39	1601	4.517
	42	1719	2.639
	41	1680	2.675
cl. 3	50	2031	1.731
	46	1875	8.842
	47	1914	3.855
	75	3008	7.173
	74	2969	4.399
	73	2930	4.134
	78	3125	8.638
	77	3086	10.922
	79	3164	7.935
	76	3047	7.315
cl. 4	83	3320	7.073
	82	3281	9.127
	81	3242	11.393
	80	3203	9.333
	84	3359	4.984
	89	3555	15.484
	88	3516	14.932
	87	3476	6.577
	86	3437	8.393
	85	3398	8.692
	4	234	14.374
	23	976	7.859
	22	937	5.213
	7	351	9.498
cl. 5	55	2226	8.202
	11	508	12.706
	57	2305	8.896
	56	2266	7.368
	8	391	2.353
	54	2187	4.741
	53	2148	5.740
	61	2461	2.961
	59	2383	4.693
	58	2344	6.804
	60	2422	4.327
cl. 6	65	2617	4.966
	62	2500	7.408
	64	2578	6.832
	63	2539	7.289
	68	2734	3.514
	66	2656	4.289
	69	2773	2.462
	67	2695	4.210
cl. 7	16	703	14.795
	15	664	13.952
	20	859	9.604
	19	820	6.673
	17	742	6.945
⋮			

(continued)

Tabl

Fe

cl. 8

cl. 9

cl. 10

* Onl
sto
fea
col
wh
di
th
ci
by
by

Table 1(a) continued.

Telephone transmission

Feature	Frequency	F-ratio
41	1680	3.163
34	1406	7.282
44	1796	4.797
37	1523	2.436
36	1484	3.145
32	1328	9.552
29	1211	4.438
26	1094	3.381
74	2969	7.161
31	1289	6.680
40	1641	2.176
25	1055	2.753
82	3281	15.643
81	3242	12.489
78	3125	11.438
80	3203	9.770
77	3086	9.556
79	3164	4.496
75	3008	7.932
76	3047	5.472
30	1250	9.789
28	1172	3.033
72	2891	6.292
71	2851	6.043
70	2812	3.670
73	2930	5.819
67	2695	6.294
66	2656	5.058
69	2773	2.388
68	2734	3.374

Normal transmission

	Feature	Frequency	F-ratio
cl. 9	9	430	6.985
	41	1680	1.852
	15	664	7.295
	40	1641	4.523
	8	391	2.244
	52	2109	4.648
	56	2266	3.183
	74	2969	4.660
cl. 10	78	3125	7.069
	77	3086	9.605
	83	3320	9.691
	79	3164	7.089
	80	3203	10.972
	82	3281	12.201
	81	3242	11.386
	84	3359	11.308
	4	234	8.257
	91	3633	1.629
	89	3555	15.626
	92	3672	4.086
	86	3437	5.115
	85	3398	1.730
	90	3594	5.862
	87	3476	4.933
	88	3516	9.714
	7	351	8.883
	47	1914	2.014
	18	781	3.582
	20	859	4.739
	19	820	4.236
	48	1953	4.863
	16	703	7.826

Linear transmission

	Feature	Frequency	F-ratio
cl. 7	71	2851	4.784
	6	312	4.839
	3	195	28.833
	72	2891	2.370
cl. 8	70	2812	2.999
	14	625	6.974
	32	1328	5.046
	31	1289	7.034
	33	1367	4.157
	30	1256	5.544
cl. 9	36	1484	7.317
	35	1445	7.749
	34	1406	10.611
	100	3984	15.634
	98	3906	18.884
	99	3945	14.675
	37	1523	8.333
	28	1562	3.828
cl. 10	24	1016	5.196
	18	781	7.164
	21	898	11.390
	97	3867	21.679
	94	3756	32.102
	96	3828	28.935
	95	3789	21.514
	93	3711	20.590
	92	3672	22.930
	91	3633	16.660
	90	3594	17.117
	12	547	4.634

Only features (frequency components) of IDS considered in this stdudy were the ones circled. Full line circles indicate those features selected as optimum from the corresponding transmission column. Dotted line circles also indicate optimum features, but which were resulted from the interaction in the selection strategy discussed in the text. Note that in every transmission column there are 10 common features as indicated by full and dotted line circles. Clusterings of 100 features into 10 subsets (as indicated by cl. 1, cl.2, etc) in each column in the above table were produced by dendrograms in Appendix D.

d.

the

deter

stra

feat

resu

corr

in

whi

ste

bas

sum

pr

de

LT

equal to Fm = 5.0.

Then, if both F values in two trasmission data bases met the criterion of Fm > 5.0, that feature was selected as an optimum feature from the cluster considered. Otherwise, another feature of the next largest F value from the same cluster was processed for the same sequence until a feature was found which met the criterion in 3, or all features were exhaused.

The above steps were repeated for the remaining clusters of e telephone transmission data base. Optimum features so far termined are marked by the solid circles in Table 1(a). The same rategy was applied to further determine the remaining 5 optimum atures basing on the other two transmission data bases. The sults are also marked by full line circles entered in the rresponding columns of Table 1(a). Dotted line circles appearing each transmission column of the table are also optimum features ich were resulted from the interaction of the aforementioned eps.

Consequently, 10 mutually common optimum features of the IDS sed on three transmission data bases were selected. These are mmarized in Table 2.

Another set of 10 optimum features of the LTAS were also epared by referring to Table 1(b) according to the similar steps scribed above. The results of 10 optimum features determined for AS are summarized in Table 3.

Table 1(

cl. 1	cl. 2	cl. 3	cl. 4	cl. 5	cl. 6	cl. 7

ble 1(b). Results of the feature clusterings and the corresponding F-ratio's of LTAS in three transmission systems.

Telephone transmission

Cluster	Feature	Frequency	F-ratio
cl. 1	4	234	50.969
	2	156	36.392
	1	117	131.098
	9	430	15.068
	3	195	17.619
cl. 2	35	1445	15.200
	34	1406	20.897
	36	1484	11.153
	38	1562	13.054
	37	1523	8.227
	42	1719	14.422
	41	1680	17.839
	40	1641	9.571
	39	1601	8.292
cl. 3	28	1172	3.210
	27	1133	7.858
	33	1367	18.652
	31	1289	18.481
	32	1328	12.612
	30	1250	11.374
	29	1211	6.040
	8	391	16.120
	7	351	13.136
	5	273	20.889
cl. 4	66	2656	11.325
	65	2617	11.406
	64	2578	9.614
	62	2500	8.869
	61	2461	6.112
	63	2539	7.125
	47	1914	7.965
	46	1875	8.026
	48	1953	3.168
	26	1094	5.584
	25	1055	2.932
	24	1016	6.775
	23	976	3.942
	22	937	4.256
cl. 5	47	2969	10.370
	73	2930	7.450
	72	2891	8.366
	71	2851	15.235
	68	2734	10.239
	67	2695	12.312
	70	2812	8.807
	69	2773	11.396
cl. 6	60	2422	5.075
	59	2383	6.835
	58	2344	7.724
	57	2305	5.486
	56	2266	5.213
	55	2226	4.228
	54	2187	3.161
	50	2031	5.383
	49	1992	5.790
	53	2148	4.565
	52	2109	4.422
	51	2070	5.076
	44	1797	7.488
	43	2758	8.521
	45	1836	4.647
	6	312	4.559
cl. 7	82	3281	16.610
	81	3242	17.938
	80	3203	9.887
	78	3125	17.225
	77	3086	10.802
	79	3164	14.610
	76	3047	14.853
	75	3008	10.028

Normal transmission

Cluster	Feature	Frequency	F-ratio
cl. 1	20	859	12.959
	19	820	18.354
	70	2812	8.907
	69	2773	10.409
	72	2891	8.715
	71	2851	8.362
	17	742	12.901
	16	703	16.073
	18	781	17.325
cl. 2	95	3789	13.799
	94	3750	22.689
	98	3906	12.673
	97	3867	14.588
	99	3945	16.670
	96	3828	15.930
	100	3984	16.525
	3	195	50.299
	12	547	21.086
	11	508	48.115
	10	469	26.683
cl. 3	86	3437	8.971
	85	3398	8.768
	88	3516	8.735
	87	3476	12.500
	93	3711	23.368
	92	3672	9.701
	91	3633	18.468
	90	3594	12.744
	89	3555	11.660
	22	937	6.458
	21	898	6.847
cl. 4	74	2969	17.181
	73	2930	10.026
cl. 5	60	2422	32.843
	59	2383	21.795
	62	2500	13.660
	61	2461	19.744
	34	1406	39.174
	33	1367	15.605
	36	1484	18.857
	35	1445	16.923
	84	3359	17.061
	83	3320	19.710
	82	3281	53.856
	81	3242	46.599
	80	3203	43.491
	79	3164	35.587
	78	3125	71.138
	77	3086	35.397
	76	3047	18.257
	75	3008	16.621
cl. 6	4	234	75.829
	2	156	187.373
	1	117	147.548
	39	1601	15.202
	38	1562	15.491
	37	1523	21.668
	13	586	22.938
cl. 7	58	2344	22.609
	57	2305	18.117
	56	2266	14.462
	55	2226	13.313
	47	1914	6.148
	46	1875	9.658
	45	1836	11.533
	44	1797	7.431
	43	1758	4.060
	42	1719	5.809
	41	1680	8.388
	40	1641	9.641

Linear transmission

Cluster	Feature	Frequency	F-ratio
cl. 1	87	3476	7.318
	86	3437	8.495
	88	3516	8.072
	84	3359	10.846
	83	3320	12.140
	82	3281	21.026
	81	3242	22.170
	85	3398	10.840
	80	3203	56.845
	79	3164	41.886
	76	3047	39.477
	75	3008	40.212
	78	3125	44.471
	77	3086	31.598
	74	2969	25.188
	73	2930	20.035
	72	2891	24.349
cl. 2	48	1953	6.470
	47	1914	9.887
	49	1992	10.494
	51	2070	9.243
	50	2031	9.969
	53	2148	9.939
	52	2109	11.612
cl. 3	42	1719	8.757
	41	1680	17.588
	57	2305	38.191
	56	2266	41.343
	55	2226	17.351
	54	2187	10.135
	38	1562	16.481
	37	1523	19.336
	40	1641	15.954
	39	1601	9.169
	35	1445	15.913
	34	1406	47.204
	36	1484	21.670
	62	2500	25.483
	61	2461	48.127
	59	2383	29.503
	58	2344	35.593
	60	2422	33.442
	13	586	17.404
cl. 4	71	2851	23.296
	70	2812	20.467
	69	2773	11.245
	68	2734	9.152
	67	2695	6.890
	66	2656	12.437
	65	2617	10.618
	64	2578	16.509
	63	2539	15.591
cl. 5	.95	3789	15.214
	94	3750	16.246
	93	3711	23.821
	99	3945	17.820
	98	3906	17.961
	97	3867	33.640
	96	3828	26.384
	100	3984	7.843
	3	195	18.299
cl. 6	23	976	12.142
	22	937	11.380
	24	1016	12.915
	19	820	9.535
	16	703	25.364
	18	781	23.249
	17	742	18.841
	21	898	12.353
	20	859	7.273

(continued)

* cl. 10 cl. 9 cl. 8

Table 1(b) continued.

Telephone transmission

Cluster	Feature	Frequency	F-ratio
	100	3984	11.982
	99	3945	17.497
	96	3828	24.616
cl. 8	84	3359	37.874
	83	3320	16.394
	87	3476	32,655
	98	3906	10.631
	97	3867	21.443
	95	3789	11.462
	94	3750	13.388
	93	3711	13.085
	92	3672	14.301
	90	3594	10.374
	91	3633	12.009
	89	3555	13.085
	86	3437	14.807
	88	3516	18.315
	85	3398	20.899
cl. 9	13	586	7.356
	12	547	12.854
	14	625	8.019
cl. 10	16	703	13.959
	15	664	11.573
	18	781	18.674
	17	742	36.773
	19	820	12.322
	21	898	7.869
	20	859	5.798
	11	508	9.390
	10	469	24.984

Normal transmission

Cluster	Feature	Frequency	F-ratio
cl. 7	54	2187	11.382
	53	2148	10.170
	52	2109	16.789
	51	2070	12.407
	50	2031	5.674
	49	1992	11.061
	48	1953	7.628
cl. 8	26	1094	9.599
	25	1055	7.505
	8	391	17.578
	29	1211	19.242
	28	1172	22.832
	27	1133	17.779
	32	1328	27.179
	31	1289	19.942
	30	1250	17.065
	24	1016	10.195
cl. 9	15	664	17.506
	7	351	31.731
	14	625	13.378
	23	976	5/905
	5	273	83.671
cl. 10	64	2578	13.558
	63	2539	14.048
	67	2695	15.116
	66	2656	17.382
	68	2734	11.231
	65	2617	29.236
	9	430	20.091
	6	312	12.267

Linear transmission

Cluster	Feature	Frequency	F-ratio
cl. 7	92	3672	16.137
	91	3633	18.267
	90	3594	18.290
	89	3555	8.917
	6	312	8.771
	10	469	22.361
	9	430	16.746
cl. 8	12	547	19.424
	11	508	39.773
	4	234	82.301
	2	156	143.825
	1	117	191.461
cl. 9	8	391	18.092
	5	273	37.258
	15	664	39.859
	14	625	23.351
	32	1328	19.659
	31	1289	18.872
	33	1367	26.845
	7	351	23.878
cl. 10	46	1875	14.662
	45	1836	14.907
	44	1797	24.200
	43	1797	17.924
	26	1094	11.755
	25	1055	7.656
	27	1133	19.880
	29	1211	18.722
	28	1172	14.288
	30	1250	18.073

* Only features (frequency components) of LTAS considered in this study were the ones circled. Full line circles indicate those features selected as optimum from the corresponding transmission column. Dotted line circles also indicate optimum features, but which were resulted from the interaction in the selection strategy discussed in the text. Note that in every transmission column there are 10 common features as indicated by full and dotted line circles combined. Clusterings of 100 features into 10 subsets (as indicated by cl. 1, cl. 2, etc.) in each column in the above table were produced by dendrograms in Appendix D.

Table

Feature (#)
1
2
3
4
5
6
7
8
9
10

* F-rat

Df of

able 2. Optimum features selected for IDS parameter (as denoted by circles in Table 1(a)).

ture (#)	Frequency (Hz)	F-ratio* in		
		Telephone transmission	Normal transmission	Linear transmission
1	117	11.02	23.75	31.99
2	195	30.28	61.35	28.83
3	430	9.02	6.96	10.26
4	508	7.40	13.91	12.71
5	1406	7.28	8.47	10.61
6	3086	9.56	9.61	10.99
7	3281	15.64	12.20	9.13
8	3555	21.62	15.63	15.48
9	3750	23.28	12.62	32.10
0	3867	22.92	14.55	21.68

-ratio is statistically significant (p=0.05) for F > 2.12 with
f of numerator = 9, and Df of denominator = 40.

Table

Featur (#)
1
2
3
4
5
6
7
8
9
10

* F-r
Df

Table 3. Optimum features selected for LTAS parameter (as denoted by circles in Table 1(b)).

		F-ratio* in		
ature #)	Frequency (Hz)	Telephone transmission	Normal transmission	Linear transmission
1	117	131.10	147.55	191.46
2	195	17.62	50.30	28.36
3	547	12.85	21.09	18.84
4	742	36.73	21.90	18.84
5	1406	20.90	39.17	47.20
6	2617	11.41	29.24	10.62
7	2851	15.24	8.36	23.30
8	3203	9.89	43.49	56.85
9	3359	37.87	17.06	10.85
10	3867	21.44	14.59	33.64

F-ratio is statistically significant ($p=0.05$) for $F > 2.12$ with Df of numerator = 9, and Df of denominator = 40.

frequ

compo

great

extre

resul

syste

extre

norma

p=0.0

value

was

param

chara

and

featu

inte

2.

larg

atte

effe

feat

valu

thre

The general trend as shown in Table 2 was that the lowest two frequency components, 117 Hz and 195 Hz, and highest four components, 3281 Hz, 3555 Hz, 3750 Hz, and 3860 Hz had relatively greater F values than those of the frequencies between the two extremes. Unlike the case of IDS, most features of the LTAS resulted in somewhat similar F values across the three transmission systems excepting for one, 117 Hz. This feature yielded the extremely high values of 131.10, 147.55, and 191.46 for telephone, normal and linear transmission LTAS parameter, respectively.

In all cases, F-ratio was statistically significant (at p=0.05) for F>2.12. All features (for IDS and LTAS) had greater F values than this critical value of 2.12. One interesting outcome was the composition of the optimum features in the IDS and LTAS parameters in spite of the presumably different speech characteristics that they carried: Four features (117, 195, 1406 and 3867 Hz) were shared by both parameters. The remaining six features were distributed somewhat differently along the frequency interval.

2. FFC

Since the number of features included in an FFC was not very large, feature optimization for the FFC speech parameter was attempted by the use of F-ratio alone. To study the relative effectiveness in discriminating between different speakers, all features in an FFC were subjected to F-ratio statistics. Each F value was computed over 50 patterns (5 for each speaker) and for three different transmission systems.

com
in
in
tel
dev
in
tel

rep
(19
in

de
(t
ha
sy
σ_A
F=
3.

a'
co
op

Table 4 is a list of F-ratios of nine features of the FFC computed from speech data in three transmission systems. As shown in Table 4 $\bar{F}o$ (mean fundamental frequency) had the largest F-ratio in all transmission systems (F=429.705, 147.915, 313.346, for telephone, normal, and linear respectively). σFo (standard deviation of Fo) resulted in much smaller F-ratio than that of $\bar{F}o$ in all transmission systems (F =55.378, 28.488, and 4.976, for telephone, normal, and linear, respectively).

These results of $\bar{F}o$ and σFo in this study comply with the ones reported by Markel *et al.*, (1977), Hunt *et al.*, (1977), and Atal (1972) in view of the relative effectiveness of these two features in discriminating speakers.

Less conspicuous features were found to be σAo (standard deviation of amplitudes of successive peaks of Fo) and $\bar{\Delta}$Ao (temporal variation of amplitudes of successive peaks of Fo) which had the smallest F-ratio among the rest, in all transmission systems. Especially, in linear transmission, these two features, σAo and $\bar{\Delta}$Ao yielded F-ratios smaller than the critical value of F=2.12.

3. SPT

No feature optimization procedure was taken for the SPT, *i.e.*, all SPT's retained the original dimensionality of 2048 frequency components and entered as they were to voice identification operations.

Table

Featur name
$\bar{F}_0$
σF_0
$\bar{\Delta}F_0$
$\bar{\Delta}F_0$
F_0(ma
F_0(mi
F_0(rn
$\bar{\Delta}A_0$
σA_0

* F-
a
on
sa
(p
de

** $\bar{F}$o
σF
$\bar{\Delta}$F
$\bar{\Delta}$F
Fo
Fo
Fo
$\bar{\Delta}$
σ

Table 4. Nine features and F-ratios of the FFC.

Feature** name	F-ratio* in Telephone transmission	Normal transmission	Linear transmission
$\bar{F}_0$	429.705	147.915	313.346
σF_0	55.378	28.488	40.976
$\bar{\Delta} F_0$	34.616	37.447	27.443
$\bar{\Delta} F_0$ / $\bar{F}_0$	5.274	5.404	6.499
F_0(max)	41.384	51.388	9.240
F_0(min)	7.289	11.570	4.440
F_0(rng)	18.457	17.272	7.107
$\bar{\Delta} A_0$	5.274	5.320	1.152
σA_0	2.486	4.374	1.370

* F-ratio of each feature for each transmission was computed on a data set of 50 samples. Between speaker variance was based on 10 speakers and the within speaker variance was based on 5 samples for each speaker. F-ratio was statistically significant ($p = 0.05$) for $F > 2.12$ with Df of numerator = 9, and Df of denominator = 40.

** $\bar{F}o$ = mean fundamental frequency (Fo).

σFo = standard deviation of Fo.

$\bar{\Delta}$Fo = average temporal variation of Fo.

$\bar{\Delta}$Fo / $\bar{F}$o = ratio of the average variation of Fo to the mean of Fo.

Fo(max) = maximum (highest) Fo.

Fo(min) = minimum (lowest) Fo.

Fo(rng) = range of Fo.

$\bar{\Delta}$Ao = average temporal variation of peak amplitude of Fo.

σAo = standard deviation peak amplitude of Fo.

Orga

by

tran

in

voi

unk

whi

lin

rep

in

kno

Her

fro

me

De

Di

pa

kn

v

a

VOICE IDENTIFICATION EXPERIMENTS

Organization of Experiment

A total of 24 voice identification experiments were conducted by different combinations of the speech parameters and types of transmission systems. In addition, all the parameters were tested in the cross-transmission as well as in the within-transmission voice identification experiments. In the cross-transmission, all unknown speakers' voices were based upon the telephone system, while all known speakers' were based upon either the normal or linear transmission system. It was assumed that all speakers were represented by the biased response characteristics. In contrast, in the within-transmission experiment, both the unknown's and the known's voices were based upon the same transmission system. Hence, in the latter case, all the speakers were represented free from the influence of the transmission system.

Two major steps involved in each experiment were 1) measurement of distance and 2) application of the decision rules. Description of these two steps follows.

Distance Measurement

As a measure of proximity, or separation, between a pair of patterns, one belonging to the unknown and another belonging to the known, Euclidian distance was applied. Euclidian distance is a vectorial summation of the differences between a pair of features available in the patterns. This implies that if the values assigne

pattern

For thi

all fe

standa

standa

Z_i

where

featur

trans

these

dista

wher

assigned to the features are not distributed homogeneously across a pattern, the distance measure may introduce a highly biased result. For this reason, prior to the computation of Euclidian distance, all features in the parameter — IDS, LTAS, and FFC — were standardized by Z-transformation. Each feature in a pattern was standardized by transforming into a Z-score as described below.

$$Z_{ij} = \frac{P_{ij} - \bar{P}_i}{\sigma P_i}$$

for i= 1, 2, ..., I
(number of features)
j= 1, 2, ..., J
(number of patterns)

where $\bar{p}_i$ and σp_i are the mean and the standard deviation of the ith feature computed over J (=50 in this study) patterns. Z_{ij} is a transformed score for the ith feature of the jth pattern. Then, these standardized Z values were used in the subsequent Euclidian distance measurement.

Euclidian distance was calculated by the following expression:

$$D_{ij} = \sqrt{\sum_{k=1}^{K}(Z_{ik} - Z_{jk})^2}$$

where; D_{ij} = Euclidian distance between the ith pattern and the jth pattern,

Z_{ik} = kth feature of the ith pattern,

Z_{jk} = kth feature of the jth pattern,

K = total number of features within a pattern.

Deci

mini

iden

1.

unk

dia

Up

we

pa

o

s

t

e

Decision rules

Two decision rules, the nearest-neighbor decision rule and the minimum set distance rule were applied concurrently for all voice identification experiments.

1. The nearest-neighbor decision rule:

This decision rule assigned one of the known speakers to the unknown by the following sequence. Figure 11 is a simplified diagram to illustrate the following sequence.

a. Designating one of the n patterns belonging to the unknown as the test pattern, and all other patterns belonging to the knowns as reference patterns.

b. Computing the Euclidian distance between this test pattern and all other reference patterns.

c. Assigning the test pattern to the known one of whose reference patterns is the closest (the nearest) to the test pattern: One decision has been rendered.

d. Repeating a through c until all patterns of the unknown are processed as test patterns.

Up to this point, n identification decisions (n=5, in this study) were reached, *i.e.*, the Euclidian distances from each of the n test patterns available for an unknown to all other reference patterns of the known speakers (the number of reference patterns in this study was 50). Then the entire sequence was repeated to process the remaining unknown speakers.

A total of 50 decisions for each voice identification experiment were yielded by the nearest-neighbor decision rule.

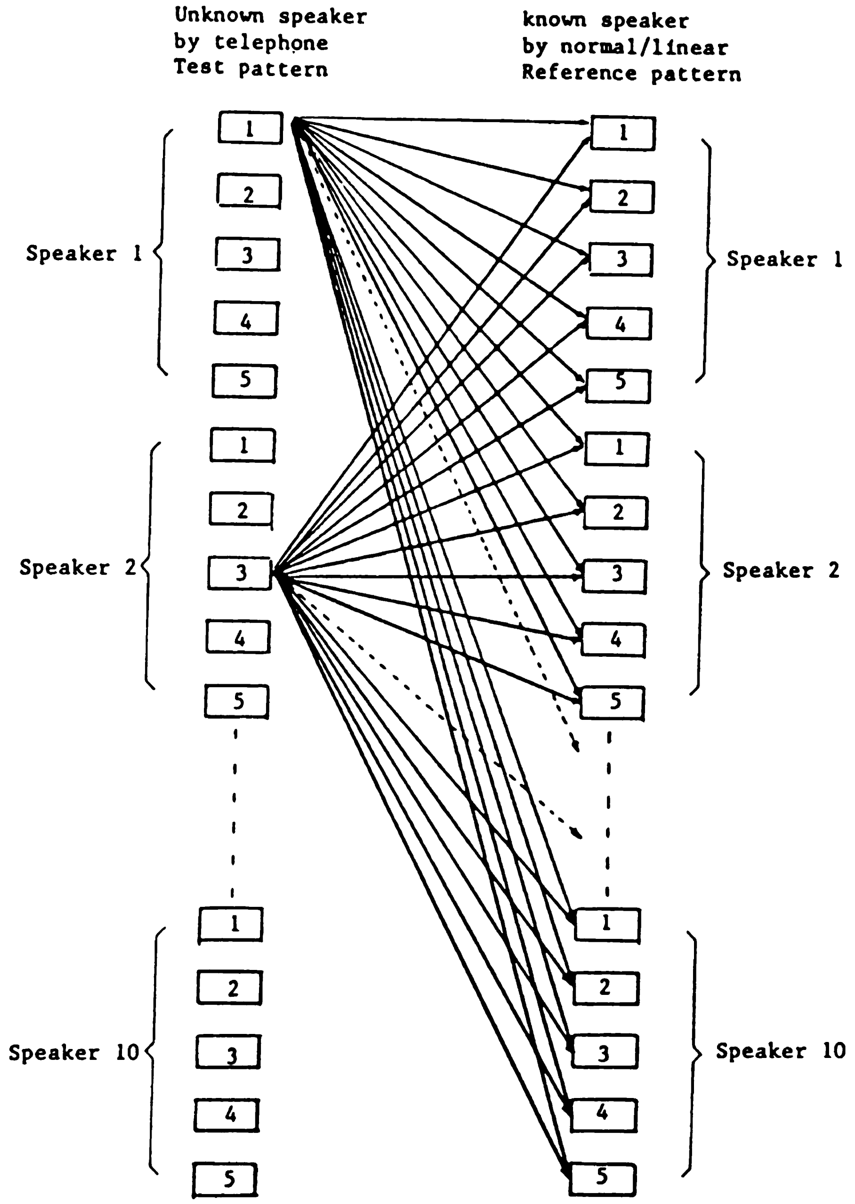


Figure 11. A diagram illustrating the nearest-neighbor decision rule. All the speakers are treated as the unknowns basing on the telephone transmission as well as the knowns basing on the linear/normal transmission system. An arrow indicates Euclidian distance between the test pattern of the unknown to all reference patterns of the knowns. Note that Euclidian distance is not computed among the unknowns nor among the knowns, and that the length of an arrow is not proportional to the actual Euclidian distance.

2. Th

Ur

require

set di

consist

compute

illustr

a

b

c

d

The a

speake

experi

T

to v

combi

2. The minimum set distance rule:

Unlike the former decision rule, this set distance rule requires *a priori* category (speaker) information in determining the set distance between two speakers under process. A set is consisted of n patterns assigned to each speaker. Major steps to compute a set distance in this study is discussed next. Figure 12 illustrates these steps by using n=3 for simplicity.

a. The Euclidian distance from the set of unknown patterns to each set of known patterns are computed.

b. Then, the maximum distance from each unknown pattern to each known pattern within a category is chosen.

c. From this set of maximum Euclidian distances the minimum is chosen to represent the Euclidian distance between the unknown speaker to every known speaker. These sets of distances are ranked from the shortest to the longest distance.

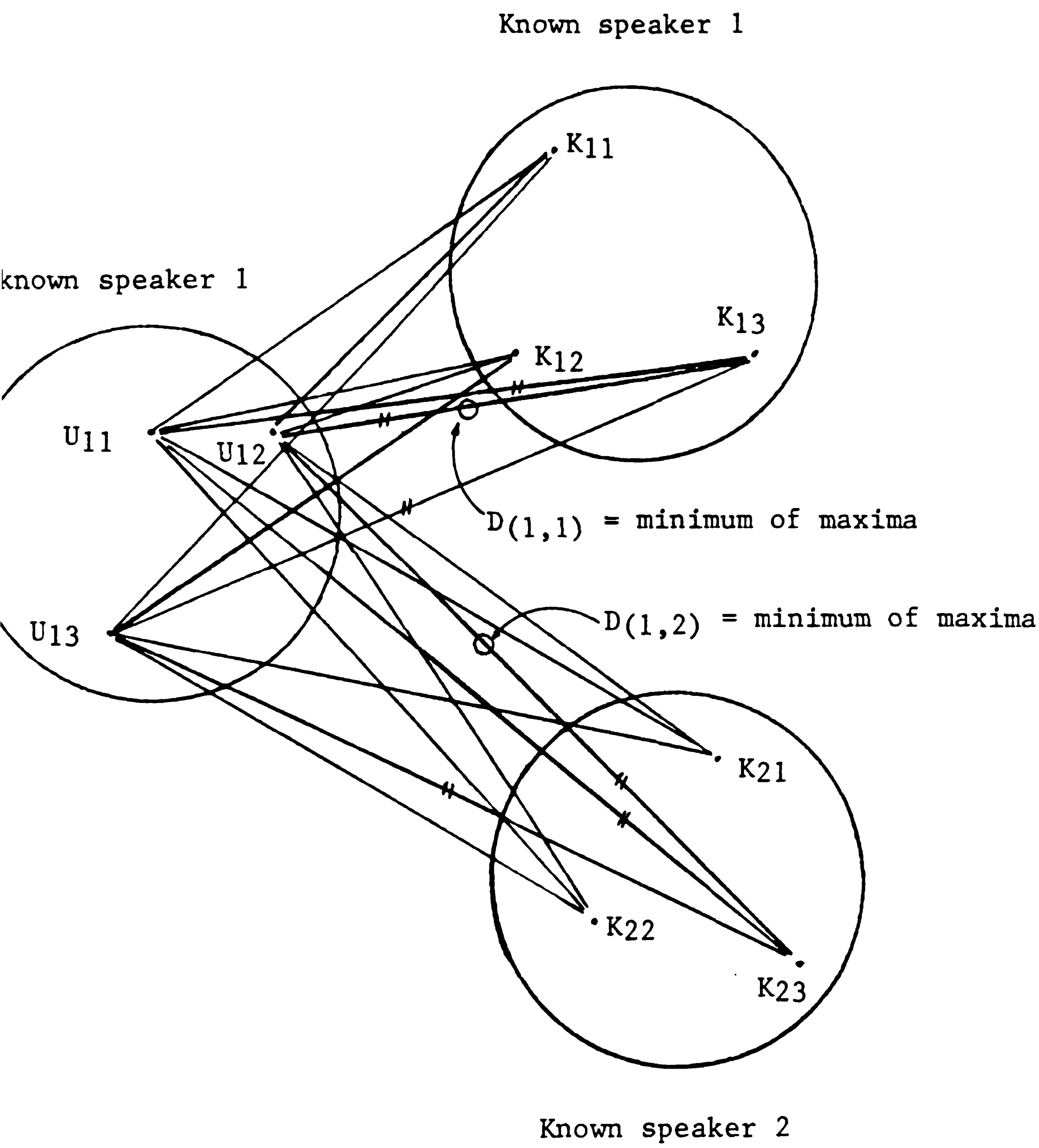
d. Finally, the known speaker whose set distance to the unknown is the shortest is assigned to the unknown.

The above procedures were then repeated for the remaining unknown speakers. A total of 10 decisions for each voice identification experiment were yielded by the minimum set distance decision rule.

Two decision rules described above were concurrently applied to voice identification operations conducted under various combinations of the speech parameters and transmission systems.

Unknown s

Figure
Thr
rep
the
pa
unl
le
A
fr
th
In
kn
ar
ex



igure 12. A diagram showing an example of the minimum set distance rule. Three speakers are shown, one as an unknown and two as knowns, each represented by three patterns. In this diagram, the symbol U_{11} denotes the first pattern of the unknown speaker 1, and a symbol K_{11}, the first pattern of the known speaker 1. A line drawn between a pattern of the unknown and that of the known indicates the Euclidian distance. The length of the line is proportional to the Euclidian distance computed. A '//' indicates the maximum distance among Euclidian distances computed from a pattern of the unknown to all patterns of the known. A 'O' is the minimum distance of the maxima between the unknown and the known. In this example, the minimum set distance between the unknown and the known speaker 1 is designated as D(1,1) and that between the unknown and the known speaker 2, as D(1,2). Since D(1,1) < D(1,2) in this example, the unknown is identified with the known speaker 1.

ider
spec
sys
wer
spe
spe
and
as
the
Cha

id
te
ei
sp
id
hi
co
SP
L

CHAPTER III

RESULTS

This Chapter focuses on the results of the voice identification operations which were conducted by using different speech parameters tested under various combinations of transmission systems (telephone, normal, and linear). Speech parameters tested were IDS (intensity deviation spectra), LTAS (long-term averaged spectra), FFC (fundamental frequency contour), SPT (choral spectra), and two composite parameters of IDS and FFC and of LTAS and FFC. In each identification operation, 10 male speakers served as both unknown and known speakers. Speech data obtained from all the speakers were 'text-independent' as described in the previous Chapter.

Table 5 summarizes the results of the cross-transmission voice identification (all the unknown speakers were recorded through a telephone system and all the known speakers were recorded through either normal or linear systems). The relative effectiveness of speech parameters are depicted in Table 5 in terms of the correct identification rates. It is clearly seen from this table that the highest correct identification rate of 100 % was achieved by the composite parameter of LTAS and FFC, and the lowest rate of 20 % by SPT. The identification rates of the remaining parameters, IDS, LTAS, and FFC (each tested as a single parameter) and the composite

Table

*

Table 5. Summary of the results of the cross-transmission voice identification operations.

Type of parameter and transmission	Rate of correct identification (%)*
IDS	
Telephone vs. Normal	70
Telephone vs. Linear	60
LTAS	
Telephone vs. Normal	70
Telephone vs. Linear	70
FFC	
Telephone vs. Normal	50
Telephone vs. Linear	40
SPT	
Telephone vs. Normal	20
Telephone vs. Linear	20
IDS + FFC	
Telephone vs. Normal	60
Telephone vs. Linear	60
LTAS + FFC	
Telephone vs. Normal	100
Telephone vs. Linear	100

* By the minimum set distance rule.

of IDS

Ta

the in

the in

Further

presen

as se

Theref

identi

F

the r

the fa

the c

opera

obtai

the

known

teste

(wher

this

this

IDS.

iden

over

the

of IDS and FFC fell in the intermediate range of 40 to 70 %.

Table 6 provides more comprehensive results in order to enable the interpretation of the elimination effect of each parameter upon the influence of the response curve of the transmission systems. Further detailed identification results for all the operations are presented in Appendix G. Two independently applied decision rules, as seen in Table 6, resulted in close conformity to one another. Therefore, the following discussion of the results is based on the identification rates yielded only by the minimum set distance rule.

First, IDS produced the elimination effect on the influence of the response curve. This interpretation is clearly supported by the fact that similar identification rates were obtained in both the cross-transmission and the within-transmission identification operations. This effect can also be seen by comparing the rates obtained by IDS (60 -70 %) with the ones obtained by SPT (20 %).

Second, LTAS was found to be susceptible to the influence of the type of transmission systems used. This susceptibility is known by the decrease in the identification rates from 100 % (when tested under the within-transmission operations) to 60 to 70 % (when tested under the cross-transmission operations). In spite of this susceptibility of the LTAS to the type of transmission system, this parameter yielded about the same identification rate as the IDS. The reason that both LTAS and IDS resulted in the same identification rate could be in the way that LTAS was extracted by overlapping the set of short-term spectra. Practically, LTAS was the same as the denominator used in the expression to compute the

Table

Desig

Cross
trans

With
tran

Desi
Cros
tran

with
tran

Desi

Cro
tra

Des
Cro
tra

Wit
tra

*
**

Table 6. Summary of the results of 24 voice identification operations

	Type of transmission and speech parameter used for: unknown speakers		known speakers		Identification rate (in %): rule 1*	rule 2**
Design 1						
	Telephone	(IDS) vs.	Normal	(IDS)	52	70
Cross-	Telephone	(IDS) vs.	Linear	(IDS)	56	60
transmission	Telephone	(LTAS) vs.	Normal	(LTAS)	70	70
	Telephone	(LTAS) vs.	Linear	(LTAS)	66	70
	Telephone	(IDS) vs.	Telephone	(IDS)	58	60
	Normal	(IDS) vs.	Normal	(IDS)	64	70
Within-	Linear	(IDS) vs.	Linear	(IDS)	70	70
transmission	Telephone	(LTAS) vs.	Telephone	(LTAS)	100	100
	Normal	(LTAS) vs.	Normal	(LTAS)	98	100
	Linear	(LTAS) vs.	Linear	(LTAS)	98	100
Design 2						
Cross-	Telephone	(FFC) vs.	Normal	(FFC)	52	50
transmission	Telephone	(FFC) vs.	Linear	(FFC)	58	40
within-	Telephone	(FFC) vs.	Telephone	(FFC)	48	40
transmission	Normal	(FFC) vs.	Normal	(FFC)	48	40
	Linear	(FFC) vs.	Linear	(FFC)	56	50
Design 3						
	Telephone	(IDS+FFC) vs.	Normal	(IDS+FFC)	62	60
Cross-	Telephone	(IDS+FFC) vs.	Linear	(IDS+FFC)	56	60
transmission	Telephone	(LTAS+FFC) vs.	Normal	(LTAS+FFC)	92	100
	Telephone	(LTAS+FFC) vs.	Linear	(LTAS+FFC)	94	100
Design 4						
Cross-	Telephone	(SPT) vs.	Normal	(SPT)	10	20
transmission	Telephone	(SPT) vs.	Linear	(SPT)	20	20
Within-	Telephone	(SPT) vs.	Telephone	(SPT)	92	80
transmission	Normal	(SPT vs.	Normal	(SPT)	94	90
	Linear	(SPT) vs.	Linear	(SPT)	88	80

* The nearest-neighbor decision rule.
** The minimum set distance decision rule.

IDS

of

corr

were

were

bei

suf

App

spe

fre

9

6

hor

sp

fe

di

co

th

va

g

i

o

s

IDS parameter.

Third, FFC was also shown to be quite free from the influence of the frequency response curve -- FFC resulted in very similar correct identification rates no matter what transmission systems were used for both unknown and known speakers. However, the rates were only moderate at 40 to 50 %. This implies that FFC, although being free from the influence of the response curve, may not be a sufficiently effective speech parameter for voice identification.

With reference to the FFC features, the raw data presented in Appendix E were inspected. It revealed that certain groups of speakers shared extremely similar $\bar{F}o$'s (average fundamental frequency) and other features. For example, speakers 1, 3, 5, and 9 had among themselves almost interchangeably close Fo's. Speakers 6 and 10 formed another group with very close $\bar{F}o$'s. This homogeneity of the distribution of $\bar{F}o$'s within the certain group of speakers appears to be rather contradictory to the fact that this feature, $\bar{F}o$, resulted in the highest F-ratio (indicating good discriminating power) in all the transmission data bases. Such a contradiction, however, may not be surprising considering the fact that F-ratio only reflected (as applied in this study) the variation of the feature values among the speakers as a whole group, instead of the variation between all possible parings of the individual speakers. Apparently, interpretation of the face value of the F-ratio as the measure of discriminating power for the speakers must be made with caution.

ident

were

rate

spea

noti

corr

that

cha

(as

the

com

dis

hi

op

pr

as

ch

pr

a

c

r

s

Fourth, SPT came out as predicted. High correct identification rates (80 to 90 %) were produced when the voices were recorded only by one type of the transmission system, but the rates decreased (20 %) when voices of the unknown and the known speakers were recorded through different transmission systems.

Fifth, the composite of IDS and FFC did not show any noticeable improvement in terms of elimination effect and the correct identification rates. This was probably due to the fact that both parameters included the same type of speech characteristics. In view of the results that these two parameters (as tested separately) were relatively free from the influence of the transmission system, the identification rate of 60 % by a combination of the features from IDS and FFC was a rather disappointing outcome.

Finally, the composite of LTAS and FFC showed the unexpectedly high correct identification rate of 100 % in two cross-transmission operations (telephone vs. normal and telephone vs.linear). The probable reason for this high identification rate can be expressed as follows. LTAS and FFC carried different types of speech characteristics working in a complementary fashion, *i.e.*, LTAS provided the static spectral features, thus reflecting more or less average vocal tract shape during speech production, while FFC contained the fundamental frequency related features, thus reflecting information about the glottal dynamics in on-going speech.

T

the s

projec

1 un

is de

from

appro

known

two-

of

from

i i

spea

is i

repr

spe

13(

ide

spe

the

dis

co

wi

sp

The following Figure 13(a-e), 14(a-e), and 15(a-e) show, for the sake of illustration, two-dimensional projections (nonlinear projection algorithm by Sammon, 1969) each projection consisting of 1 unknown and 5 known speakers. Briefly, the Sammon's projection is described to perform a point mapping of N L-dimensional vectors from the L-space to a lower-dimensional space to preserve approximate data structure. In this study, N=6 (1 unknown and 5 known speakers) and L=5 (5 samples/speaker) was plotted into a two-dimensional space. In each projection, five patterns (samples) of the unknown speaker are denoted by ui (i=unknown speaker index from 1 to 5), and a center of the dispersion of the unknown speaker i is indicated by Uci. Known speakers are simply denoted by the speaker index and the center of the dispersion of known speaker i is indicated by Kci.

In Figure 13(a-e) all the speakers (unknown and known) are represented by telephone IDS data base. It is shown that unknown speaker 1 is closest to known speaker 1 (correct identification, 13(a) unknown speaker 4, closest to known speaker 4 (correct identification, 13(d)), but unknown speaker 5 as closest to unknown speaker 3 (incorrect identification, 13(c). As clearly shown in these projections (Figure 13(a-e), relatively tight spatial distribution of 5 speakers could be accounted for rather modium correct identification rate achieved by IDS as tested under all the within- and cross-transmission operations.

In Figure 14(a-e), each projection shows 1 unknown and 5 known speakers, all the speakers represented by telephone LTAS. In

Fi

U
K
U
Fi

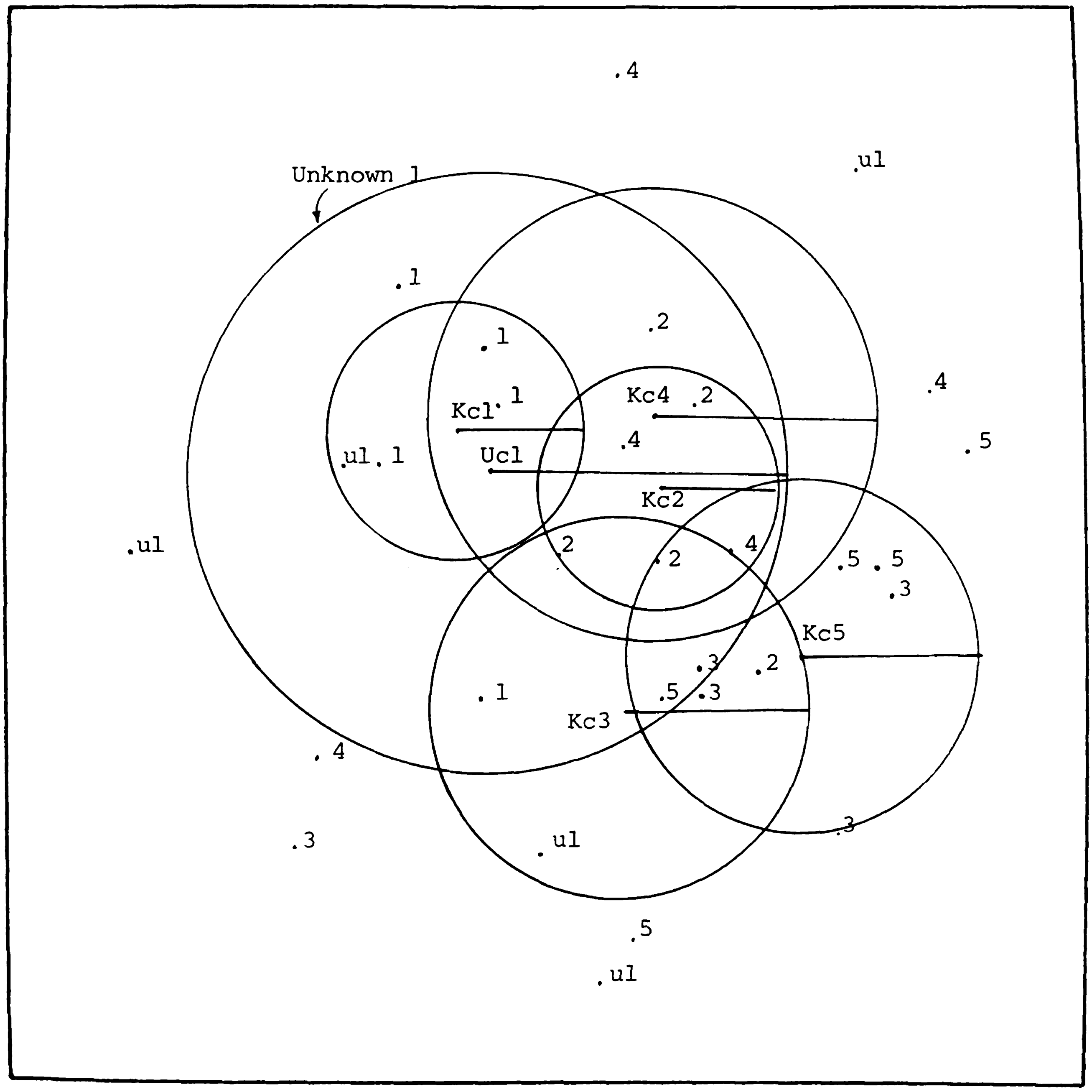


Figure 13(a). 5 known speakers (telephone IDS) vs. unknown speaker 1 (telephone IDS):

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 1 = ul
Kcj = The center of dispersion of Samples of the jth known speaker
Uc1 = The center of dispersion of samples of unknown speaker 1

Figure 13 (a-e). Sammon's projections of 5 known speakers and 5 unknown speakers, all represented by telephone IDS parameter: (a) 5 knowns vs. unknown 1; (b) 5 knowns vs. unknown 2; (c) 5 knowns bs. unknown 3; (d) 5 knowns vs. unknown 4; and (e) 5 knowns vs. unknown 5.

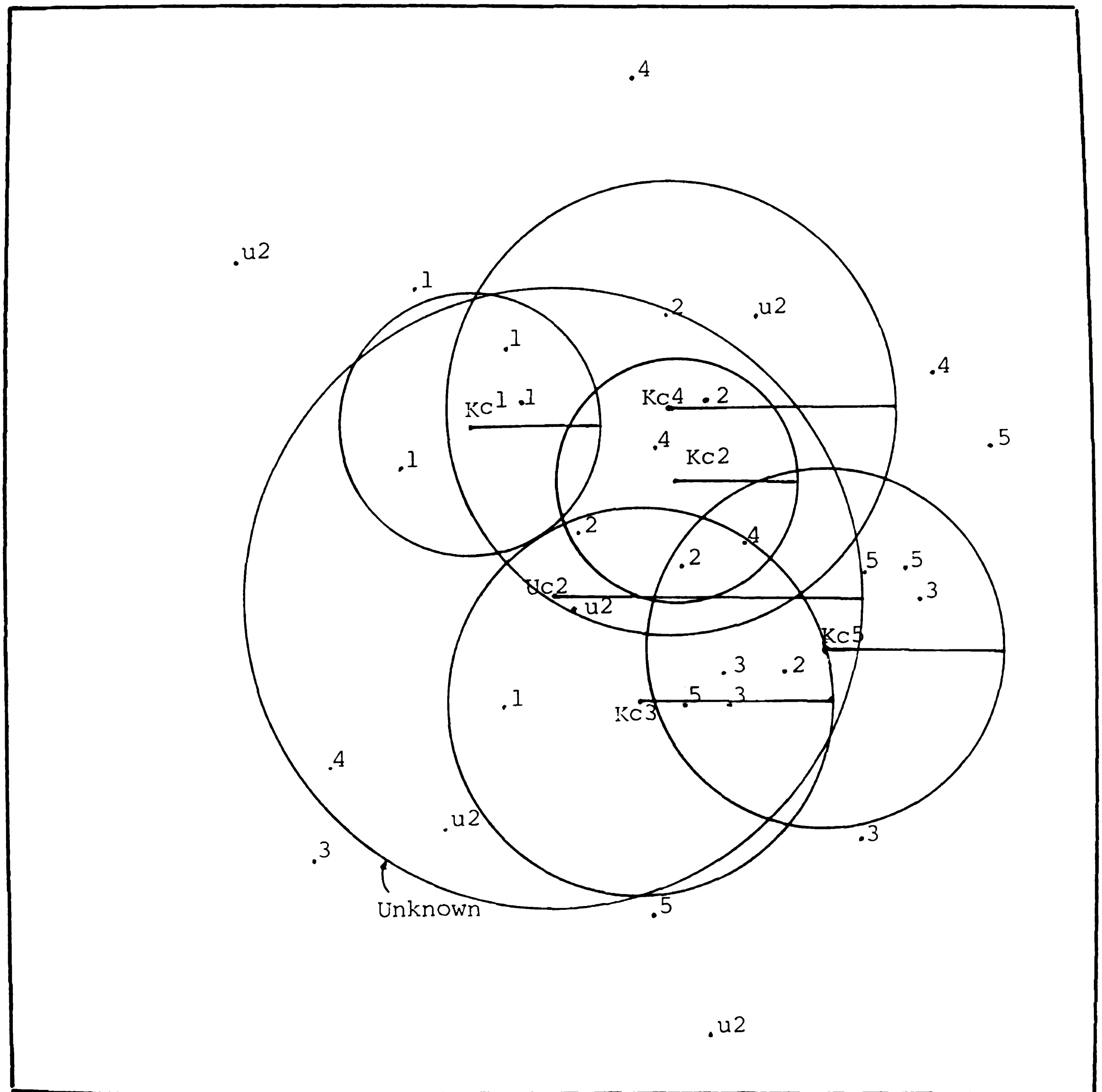


Figure 13 (b). 5 known speakers (telephone IDS) vs. unknown 2 (telephone IDS):

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 2 = u2
Kcj = The center of dispersion of samples of the jth known speaker.
Uc2 = The center of dispersion of samples of unknown speaker 2.

Fig

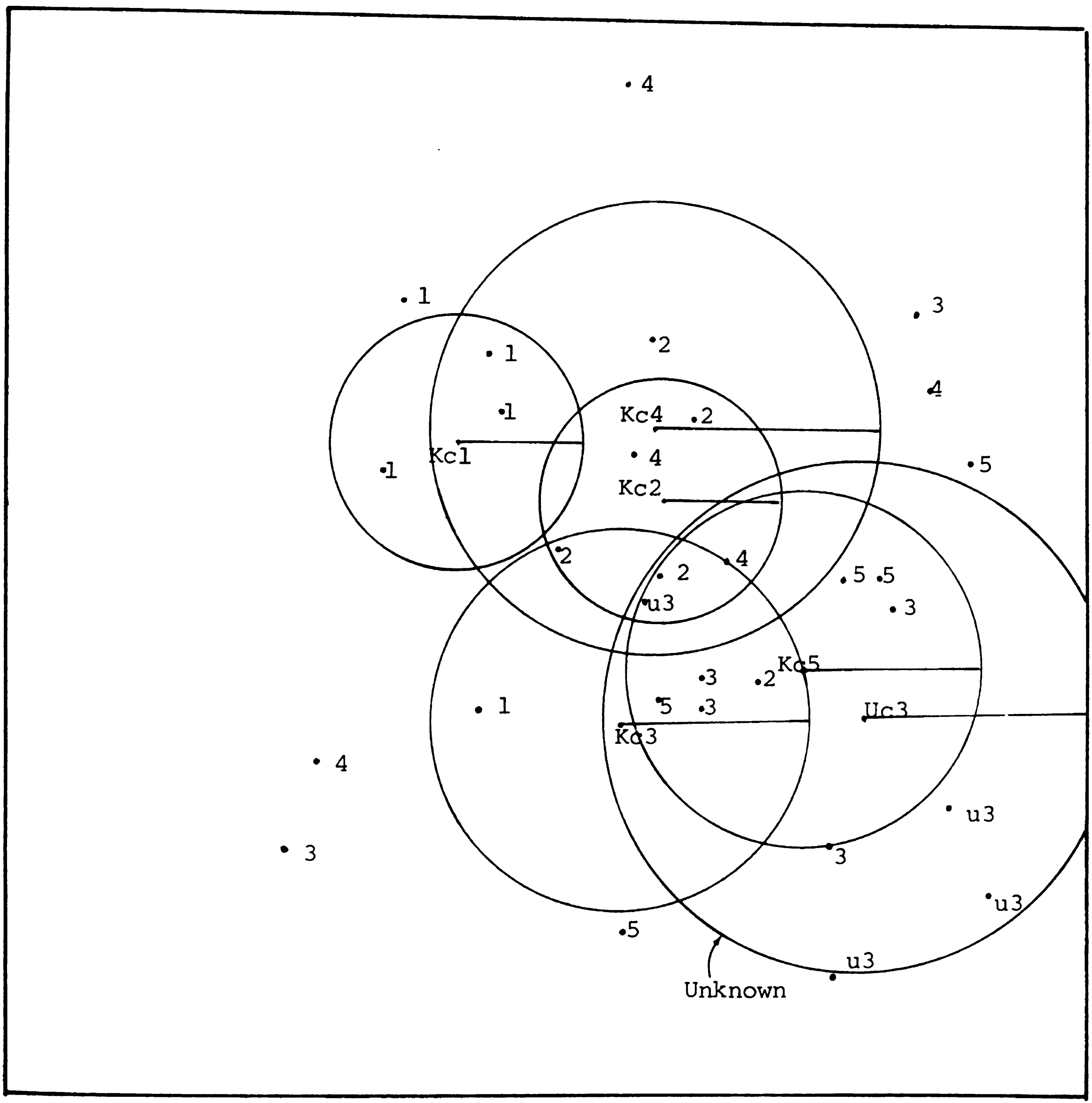


Figure 13 (c). 5 Known speakers (telephone IDS) vs. unknown 3 (telephone IDS:

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 3= u3
Kcj = The center of dispersion of Samples of the jth known speaker.
Uc3 = The center of dispersion of samples of unknown speaker 3.

Fig

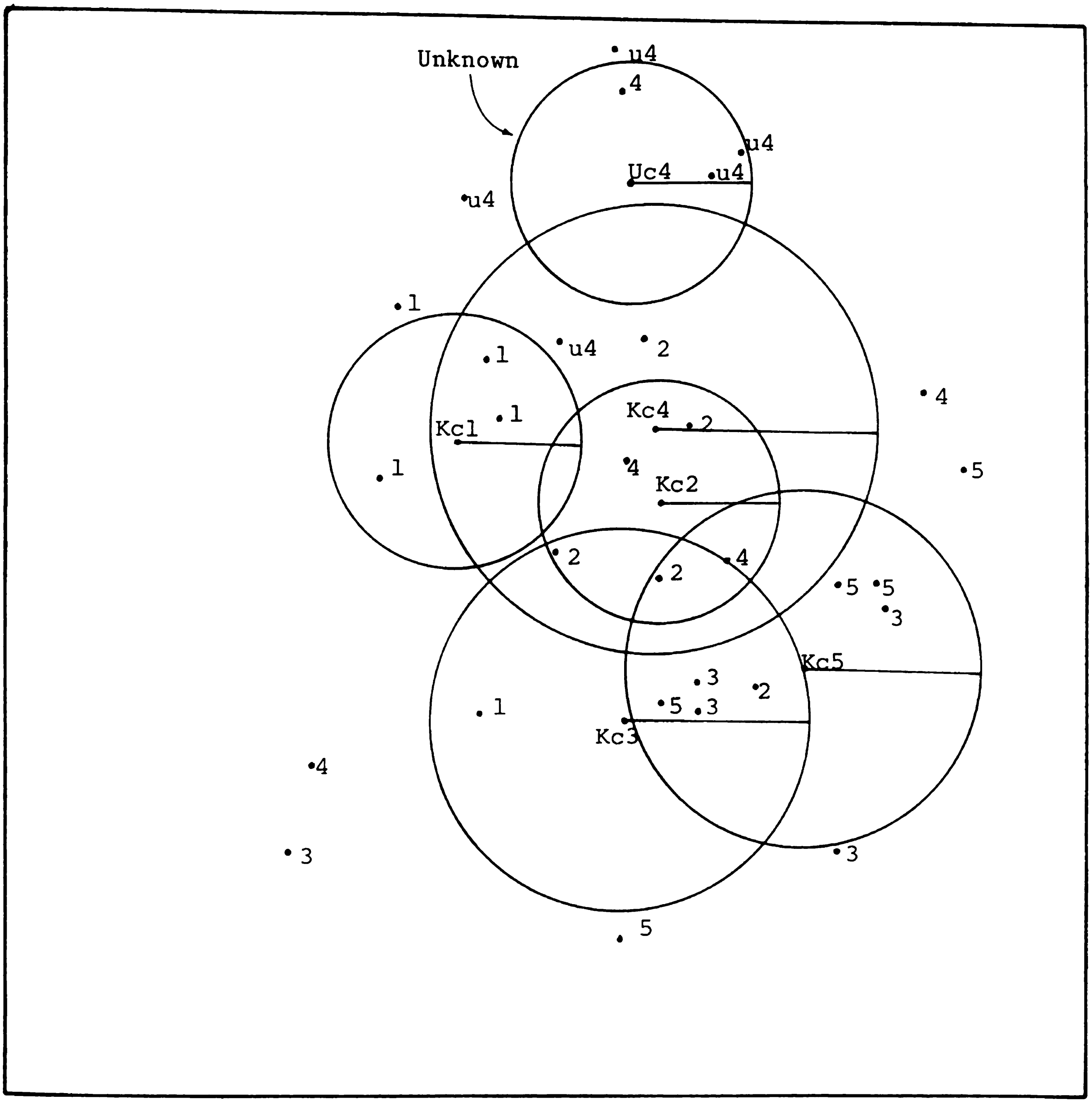


Figure 13(d). 5 known speakers (telephone IDS) vs. unknown speaker 4 (telephone IDS):
Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 4= u4
Kcj = The center of dispersion of samples of the jth known speaker.
Uc4 = The center of dispersion of samples of unknown

Fig

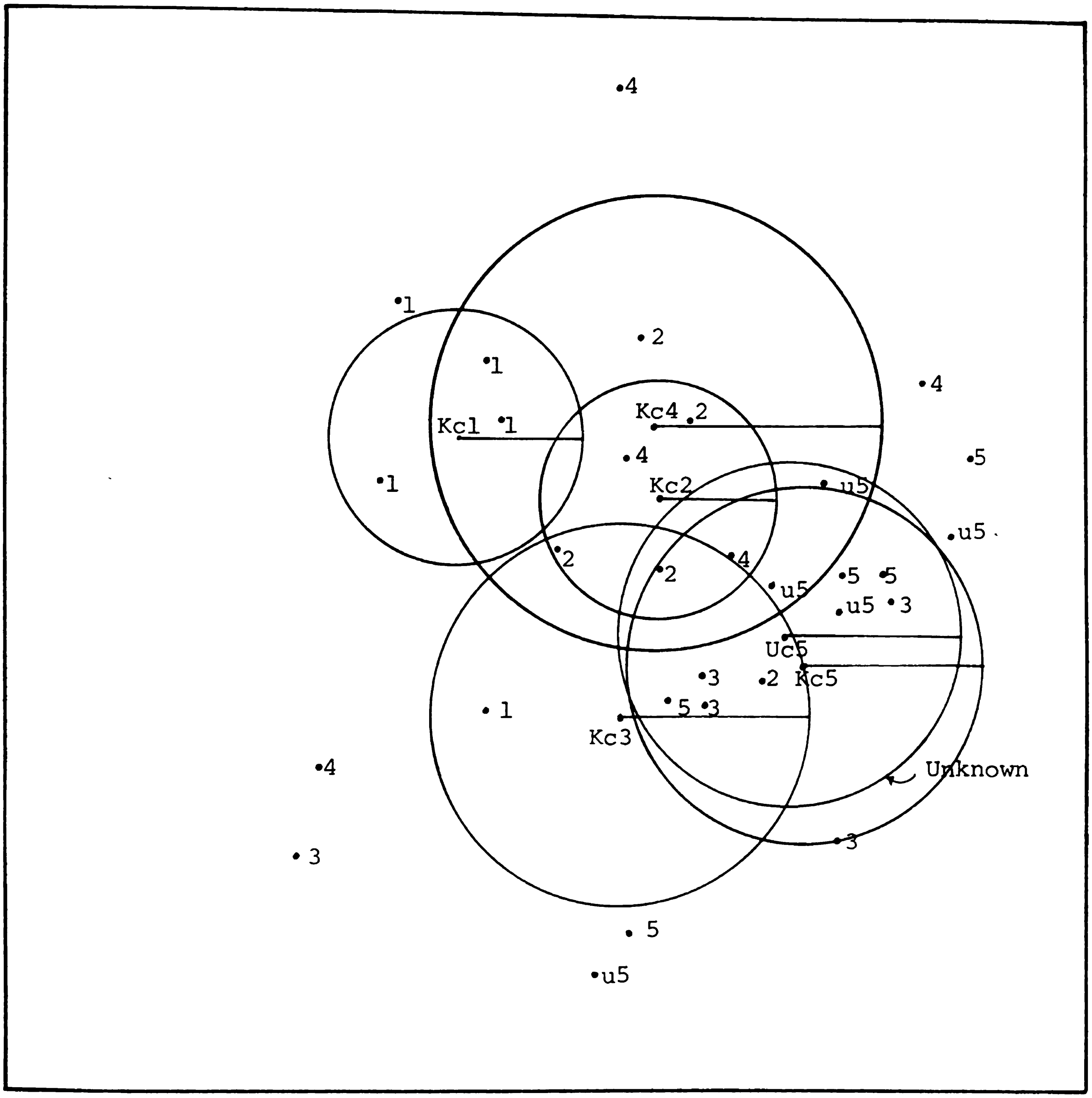


Figure 13(e). 5 known speakers (telephone IDS) vs. unknown speaker 5 (telephone IDS).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 5= u5
Kcj = The center of dispersion of samples of the jth known speaker.
Uc5 = The center of dispersion of samples of unknown speaker 5.

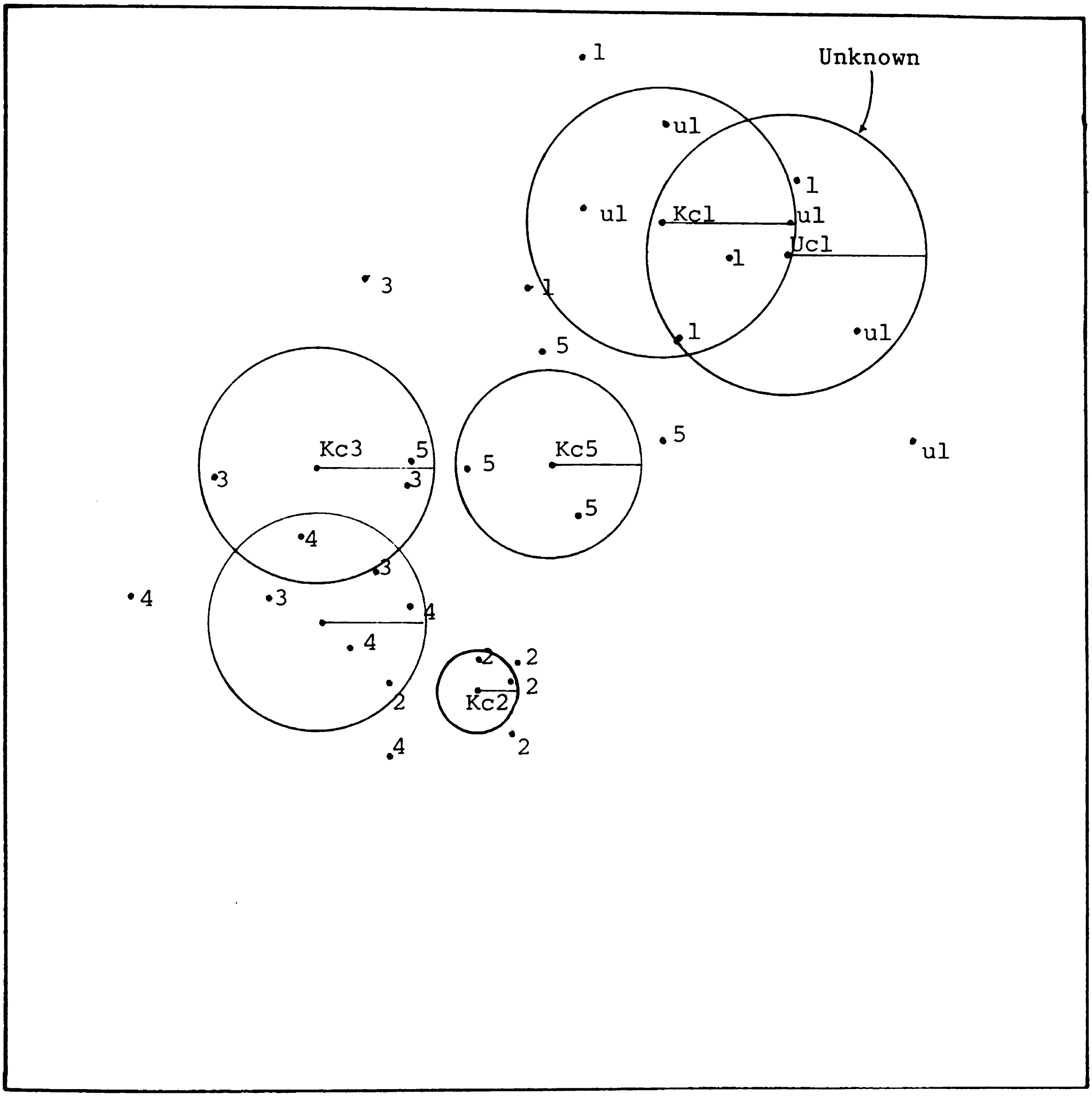


Figure 14(a). 5 known speakers (telephone LTAS) vx. unknown speaker 1 (telephone LTAS):

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 1 = ul
Kcj = The center of dispersion of samples of the jth known speaker
Uc1 = The center of dispersion of samples of unknown speaker 1

Figure 14 (a-e). Sammon's projections of 5 known speakers and 5 unknown speakers, all represented by the telephone LTAS parameter:
(a) 5 knowns vs. unknown 1; (b) 5 knowns vs. unknown 2;
(c) 5 knowns vs. unknown 3; (d) 5 knowns vs. unknown 4;
(e) 5 knowns vs. unknown 5.

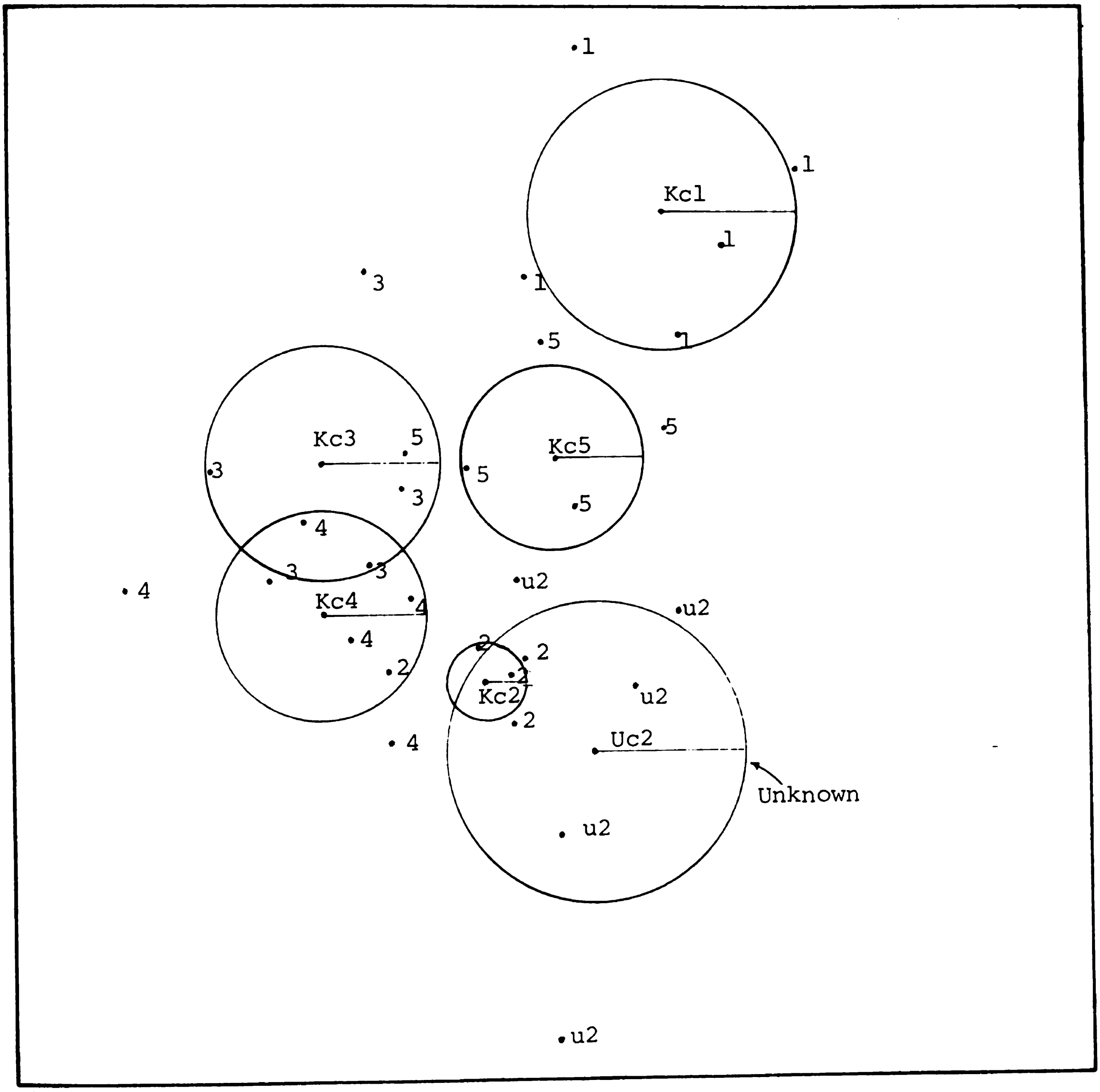


Figure 14(b). 5 known speakers (telephone LTAS) vs. unknown speaker 2 (telephone LTAS).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 2 = u2
Kcj = The center of dispersion of samples of the jth known speaker
Uc2 = The center of dispersion of samples of the unknown speaker 2

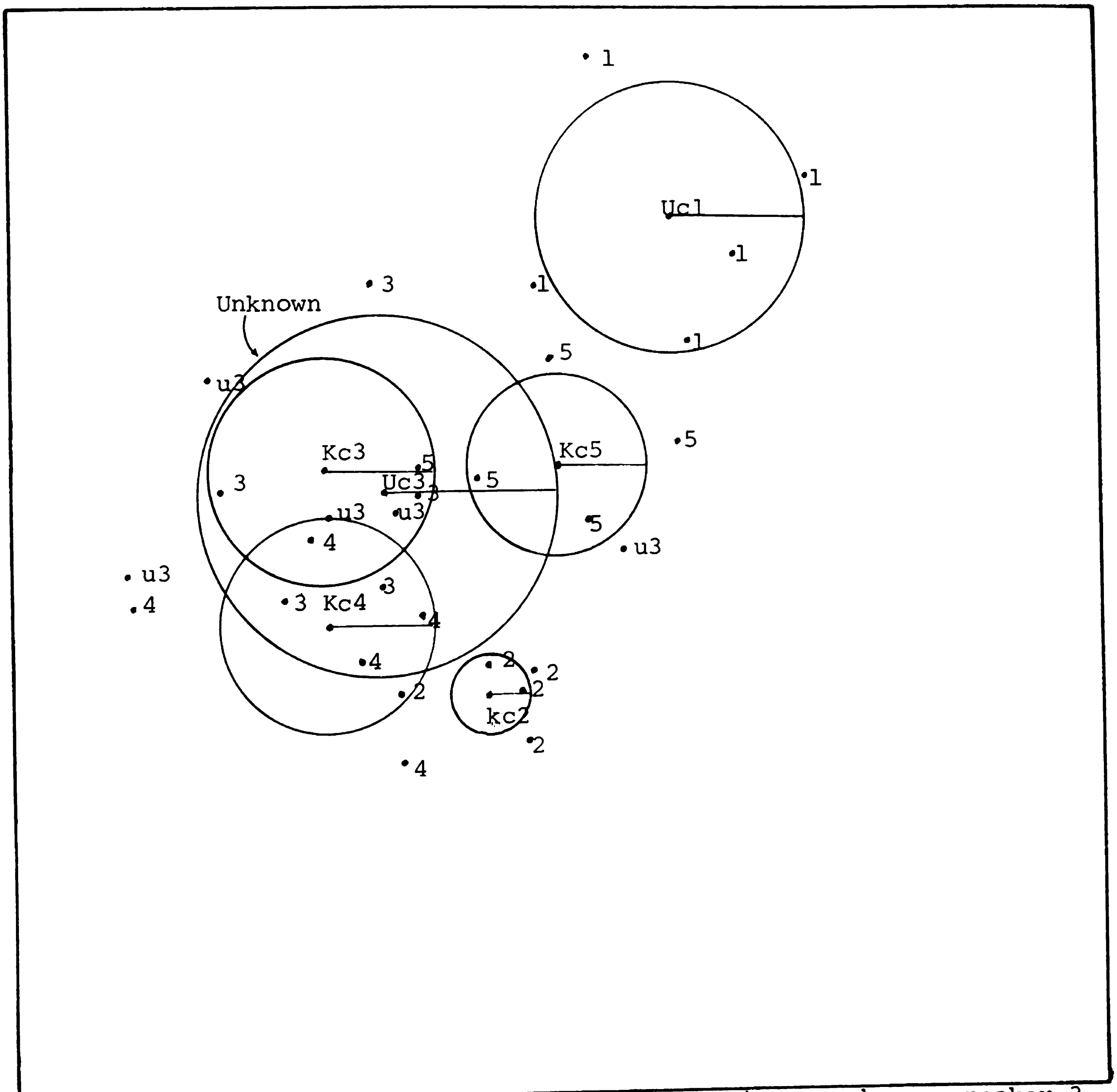


Figure 14 (c). 5 known speakers (telephone LTAS) vs. unknown speaker 3 (telephone LTAS).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 3 = u3
Kcj = The center of dispersion of samples of the jth known speaker
Uc3 = The center of dispersion of samples of the unknown speaker 3.

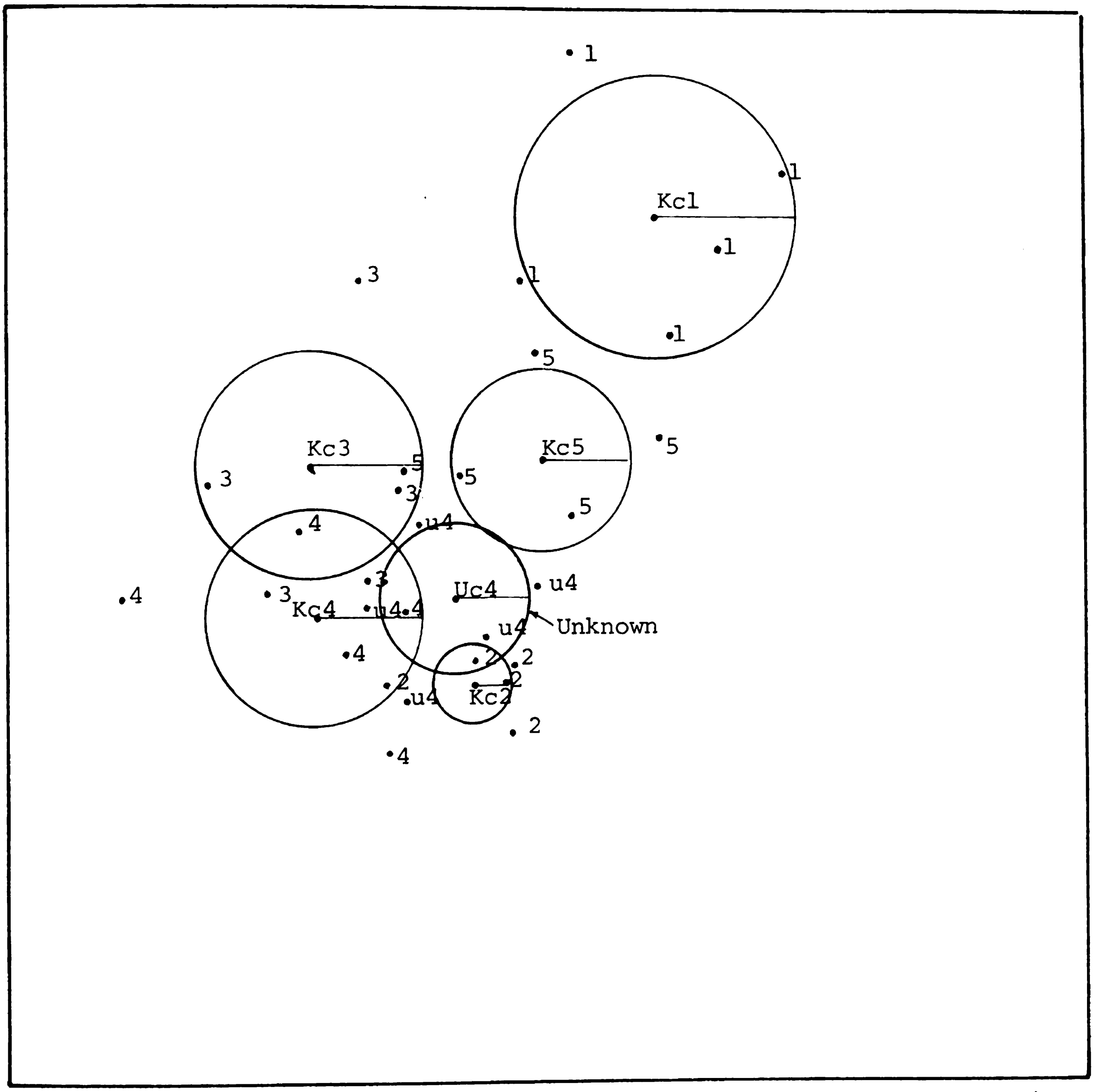


Figure 14(d). 5 known speakers (telephone LTAS) vs. unknown speaker 4 (telephone LTAS).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 4 = u4
Kcj = The center of dispersion of samples of the jth known speaker.
Uc4 = The center of dispersion of samples of the unknown speaker 4.

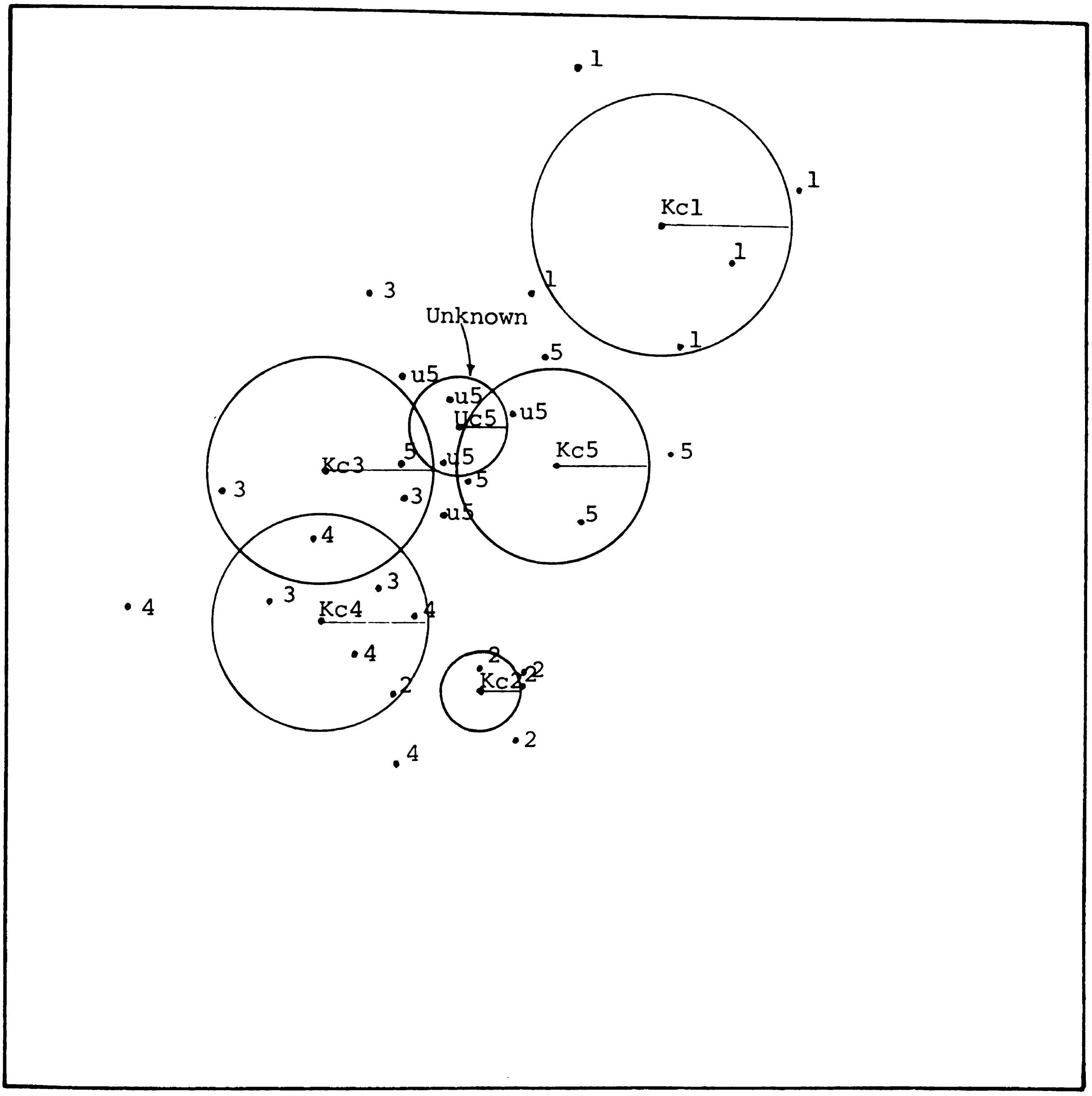


Figure 14(e). 5 known speakers (telephone LTAS) vs. unknown speaker 5 (telephone LTAS).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 5 = u5
Kcj = The center of dispersion of samples of the jth known speaker.
Uc5 = The center of dispersion of samples of the unknown speaker 5.

contrast to the spatial dispersion of 5 known speakers by IDS given in Figure 13(a-e), here, all the speakers (both unknown and known) are more clearly separated.

In Figure 15(a-e), each projection shows 1 unknown and 5 known speakers, all speakers being represented by the composite of LTAS and FFC parameters, and the unknown speaker recorded through a normal transmission system. These projections (a-e) indicate that all 5 unknown speakers were correctly identified and that all the known speakers were shown to have the relatively small intra-speaker variation.

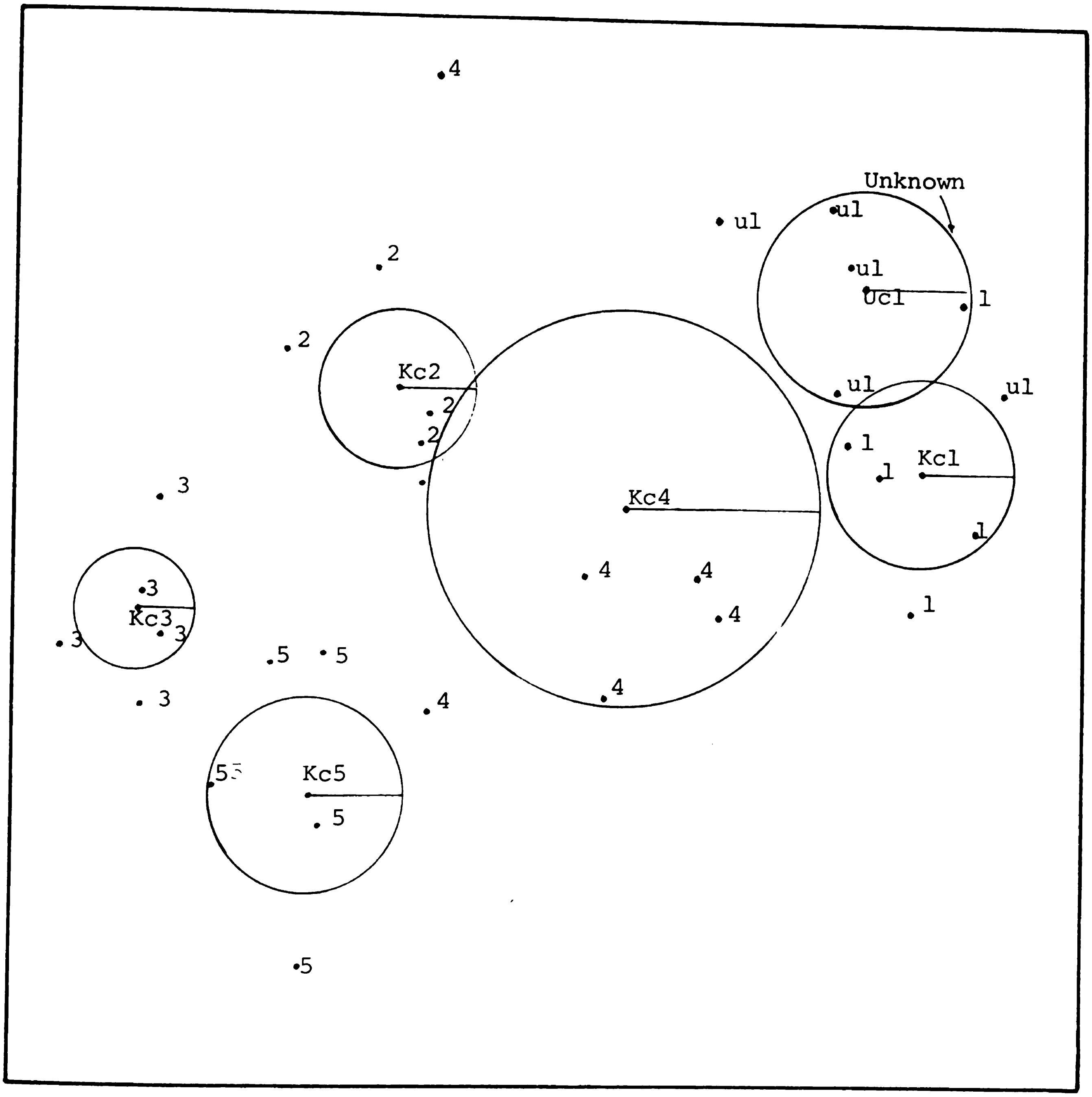


Figure 15 (a). 5 known speakers (composite of FFC and LTAS by normal transmission vs. unknown speaker 1 (composite of FFC and LTAS by telephone transmission).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 1 = ul
Kcj = The center of dispersion of samples of the jth known speaker.
Ucl = The center of dispersion of samples of the unknown speaker 1.

Figure 15 (a-e). Sammon's projections of 5 known speakers and 5 unknown speakers - knowns represented by the composite parameter of FFC and LTAS by normal transmission system and unknowns represented by the same composite parameter by telephone transmission system.
5 known speakers vs. (a) unknown speaker 1; (b) unknown speaker 2; (c) unknown speaker 3; (d) unknown speaker 4; (e) unknown speaker 5.

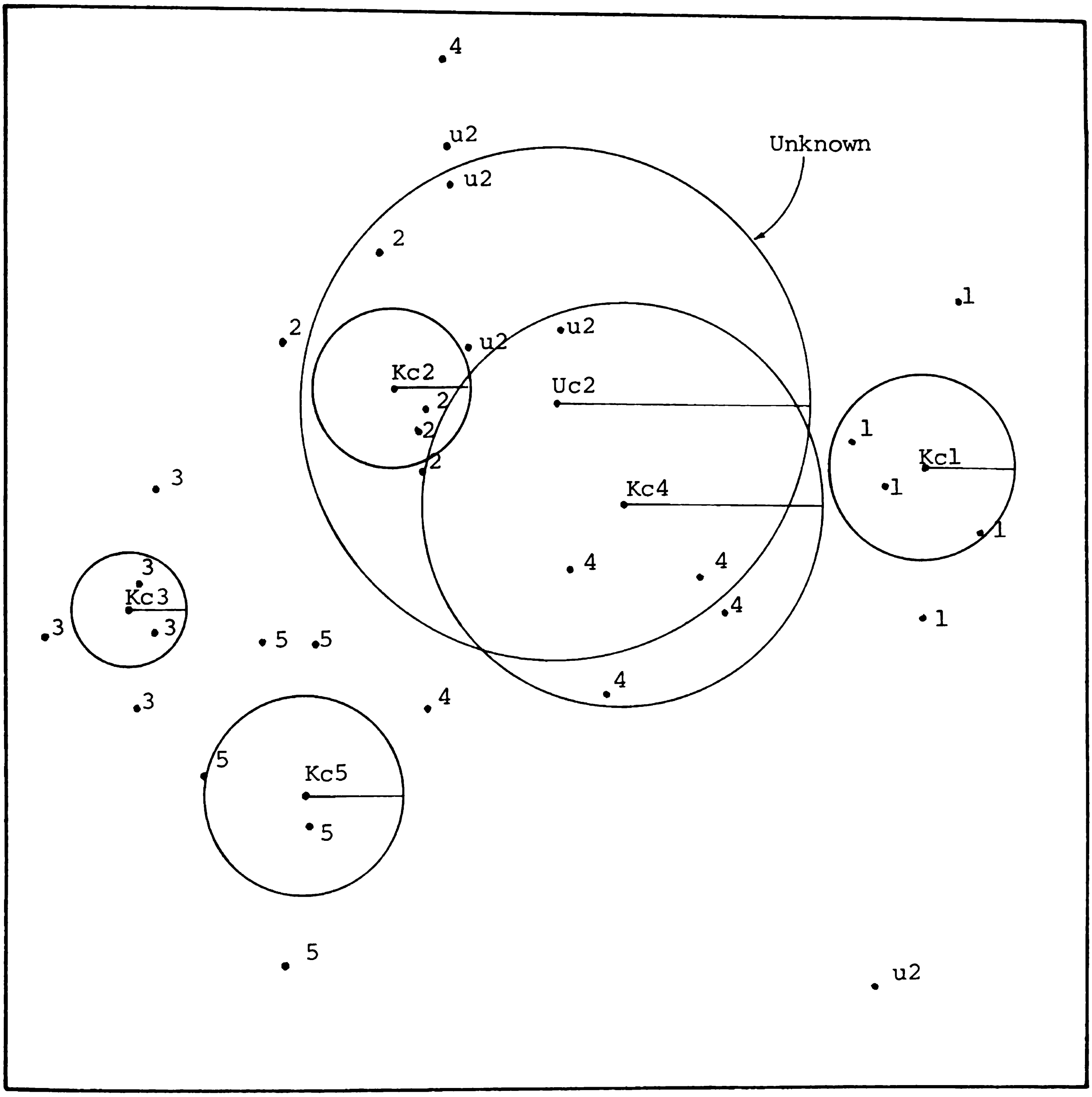


Figure 15 (b). 5 known speakers (composite of FFC and LTAS by normal transmission) vs. unknown speaker 2 (composite of FFC and LTAS by telephone transmission).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 2 = u2
Kcj = The center of dispersion of samples of the jth known speaker.
Uc2 = The center of dispersion of samples of the unknown speaker 2.

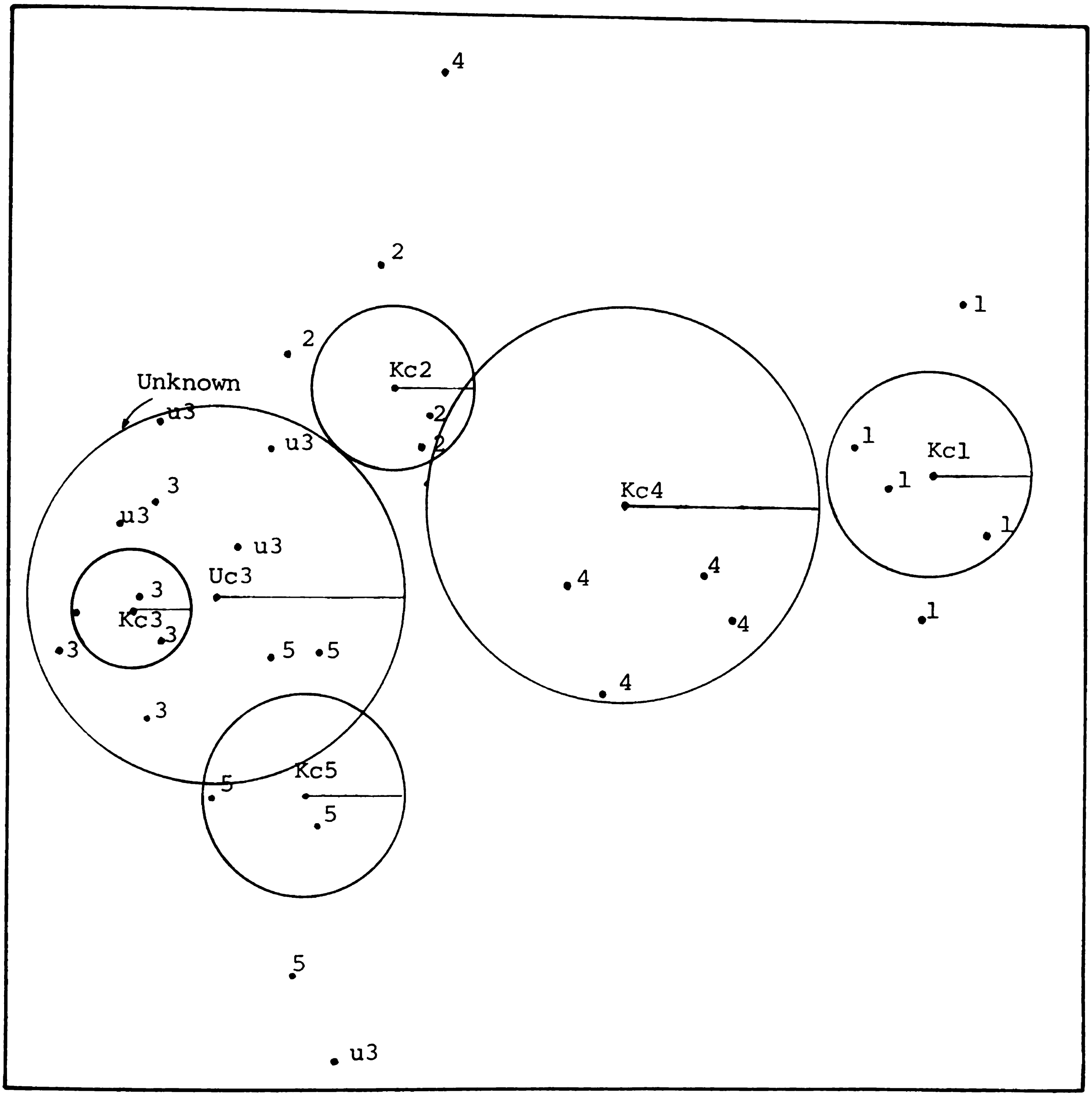


Figure 15(c). 5 known speakers (composite of FFC and LTAS by normal transmission) vs. unknown speaker 3 (composite of FFC and LTAS by telephone transmission).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 3 = u3
Kcj = The center of dispersion of samples of the jth known speaker.
Uc3 = The center of dispersion of samples of the unknown speaker 3.

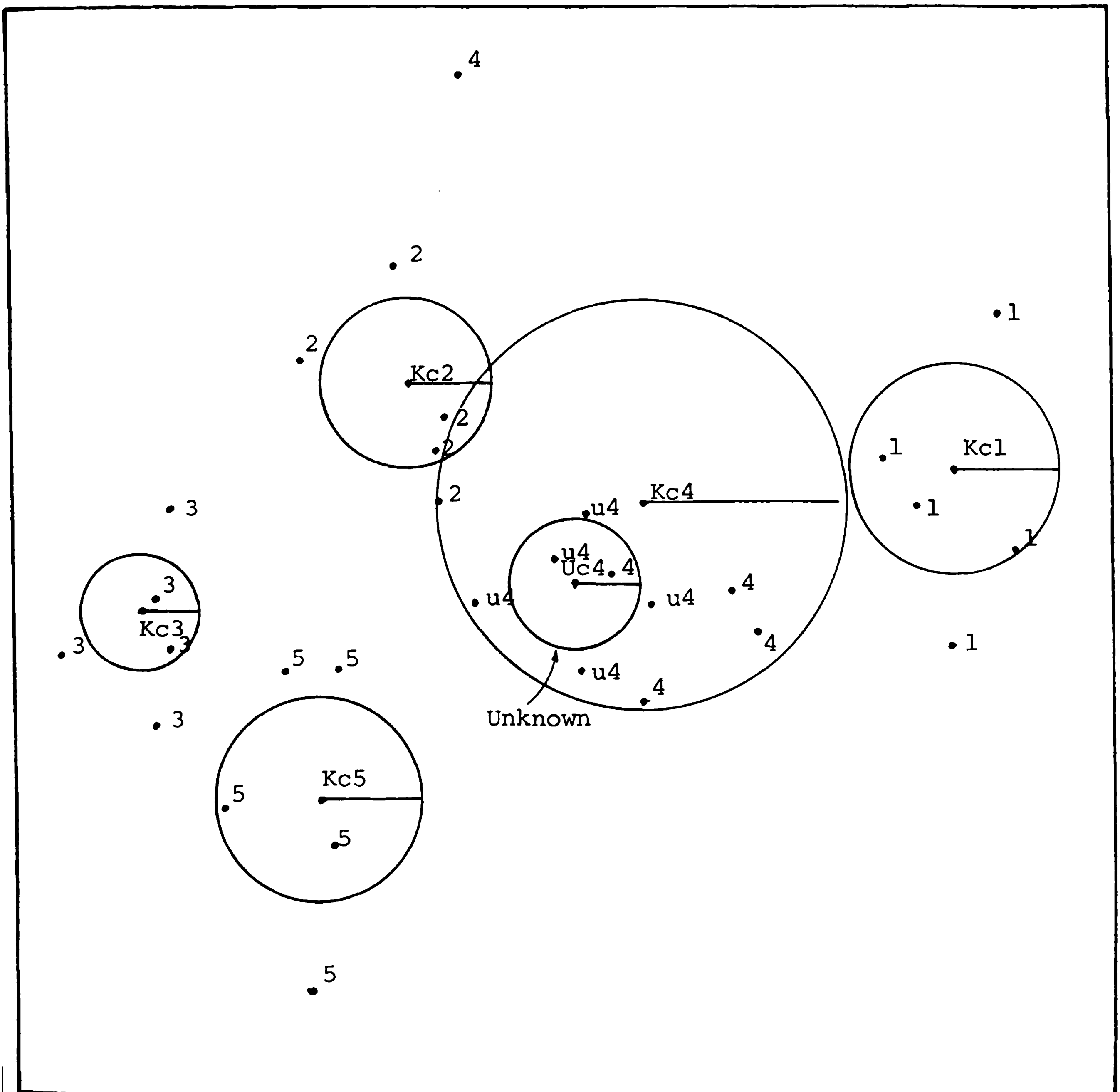


Figure 15 (d). 5 known speakers (composite of FFC and LTAS by normal transmission) vs. unknown speaker 4 (composite of FFC and LTAS by telephone transmission).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 4 =u4
Kcj = The center of dispersion of samples of the jth known speaker.
Uc4 = The center of dispersion of samples of the unknown speaker 4.

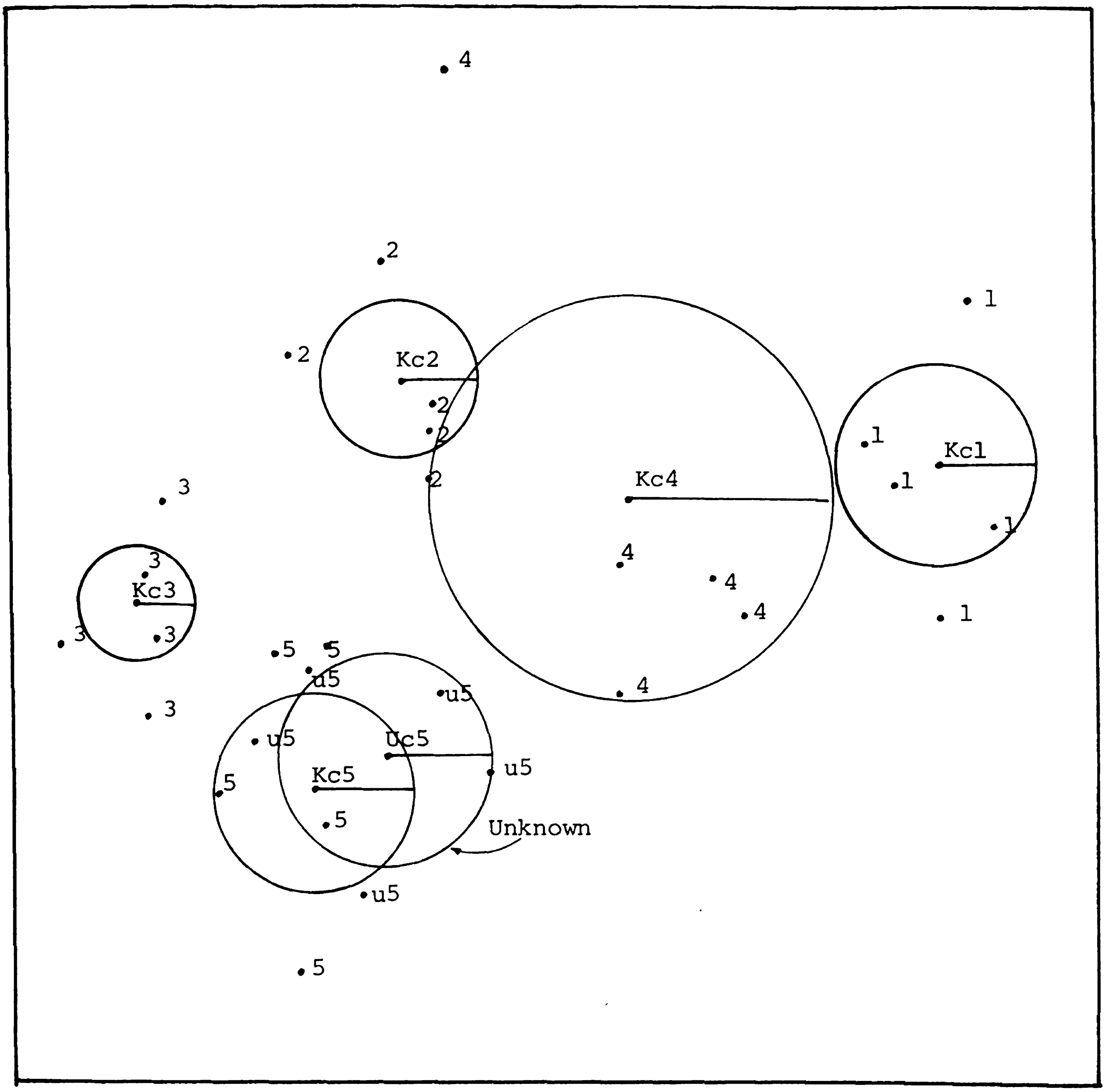


Figure 15(e). 5 known speakers (composite of FFC and LTAS by normal transmission) vs. unknown speaker 5 (composite of FFC and LTAS by telephone transmission).

Known speaker 1 = 1
Known speaker 2 = 2
Known speaker 3 = 3
Known speaker 4 = 4
Known speaker 5 = 5
Unknown speaker 5 =u5
Kcj = The center of dispersion of samples of the jth known speaker.
Uc5 = The center of dispersion of samples of the unknown speaker 5.

CHAPTER IV

DISCUSSIONS AND CONCLUSIONS

In the present study, the speech materials were produced from 10 male speakers, each speaker simultaneously recorded by three different transmission and recording systems. Four types of speech parameters were extracted from the text-independent materials to represent the speakers, as unknown and known persons. Then these parameters were studied in voice identification operations for their effectiveness in eliminating the influence of the response characteristics of the transmission and recording devices.

DISCUSSIONS

Analytical comparisons of the results obtained in this study to those reported in the literature do not appear feasible nor meaningful due to the large variation in the types of phonetic materials, size and type of the speaker population, methodology and procedure employed, *etc.* Nonetheless, in order to facilitate some reasonable interpretation of the results from this study, several factors other than the distortion due to the transmission system which could have been critically involved in the process are discussed below.

On Speaker Population

Typically the number of speakers included in studies of voice identification by computer is small. This fact is mainly due to the huge amount of information present in a brief segment of speech to be processed by the limited memory capacity and computational speed of the most computers. Clearly this constraint upon the size of speaker population makes it difficult to generalize many study results. Doddington (1974) presented a computer simulation on the expected error rate for speaker identification as a function of population size. It was shown that the overall probability of an incorrect decision is a monotonically increasing function of population size. Some examples of expected error rate (for optimum identification) with population size N were 0.01 for N=2, 0.025 for N=5, 0.08 for N=10, 0.18 for N=20, and 0.5 for N=100. According to Doddington's study, then, the error rate of 0.08 or conversely the correct identification rate of 0.92 (92%) can be interpreted as optimum, or sufficient for voice identification by computer; but it seems to be far from the ultimate goal for practical use.

In addition to the number of speakers employed, the homogeneity of the speaker population is also known to considerably affect the identification rate. On the issue of 'homogeneity' of speaker population, Bogner (1981) presented a good example in explaining the varying results of identification rates reported in much of the literature:

> One possible 'explanation' is that this talker's speech is exceptionally similar to the average of the population of talkers. Let us denote talkers of this type as 'type-X.' Assuming that a proportion 0.1 of the talker population is of type-X, we find that the expected number of such talkers in a sample of 10 is 1, with a standard deviation of 1. Thus, it would not be surprising to find some sets of 10 talkers, with no type-X talkers, and some with 2 or 3, causing the resultant error rates to differ greatly, by factors of more than 3.

Raw data (from Appendix E) were inspected to check how the speaker population in this study can be explained in Bogner's view. The sample average of $\bar{F}o$, which yielded the largest F-ratio value, was calculated over five FFC's for all the speakers. Based on this feature alone, speakers 1, 3, 5, and 9 were found to be extremely similar, as their $\bar{F}o$'s were 109.8, 110.1, 111.1, and 111.7 Hz, respectively. Visual inspection of Appendix G (11-15) also indicated that these four speakers were frequently misidentified among themselves. Speakers 2 and 10 were seldom misidentified with any other speakers. Speaker 2 had the lowest $\bar{F}o$, whereas speaker 10 had the highest $\bar{F}o$.

Another interesting interpretation can be drawn from the results obtained by the use of the LTAS speech parameter. Judging from the 100% identification rate achieved under the within-transmission voice identification, it appears that the speakers in this study were less homogeneous when represented by their long-term spectral characteristics than when represented by glottal dynamic characteristics. In other words, given a group of speakers, the spectral features have greater discriminablity in distinguishing speakers than glottal features do. In general, this complies with the results of Markel *et al.* (1977) and Hunt *et al.*

(1977).

On Effects of Feature Optimization

The major topic of this study was the 'elimination' of the distortion of the frequency characteristics existing in the speech samples of the unknown speakers and the known speakers who were recorded through the different transmission systems. Two approaches were tried. The first approach was attempted by applying the IDS and the LTAS parameters whose features were optimized (reduction and selection). The second approach was carried out by means of selecting several time-varying fundamental frequency related features forming the FFC parameter whose features were also subjected to the optimization procedure (selection). Since the design of this study was not intended to study the effect of the optimization procedure *per se*, no substantiating grounds for interpreting this effect is given in this study. Despite this lack, indirect interpretation of this optimization effect was attempted by referring to the resulting correct identification rates as the measure of effectiveness. It became obvious that the optimization procedure induced the 'elimination' effects to different degrees according to each speech parameter in use.

First, the effect of the optimization upon the IDS was undeterminable simply because there were no contrasting results to be compared. Second, in view of the fact that all the original features of the FFC were retained for voice identification operation, the optimization procedure applied to the FFC was not in

effect at all for the 'elimination'. Contrary to this, when only five features with relatively high F-ratio values were used, the FFC yielded very poor identification rates. Third, in view of the fact that the LTAS resulted in the better identification rate than the choral spectrum did, the feature optimization procedure was probably in effect for 'elimination' of the distortion. An earlier pilot study carried out with the LTAS which included all the original 128 features (no optimization) also resulted in identification rate of only a chance level of 10-20%. A reasonable speculation may be that 10 features (frequency components in Table 3 chosen for the LTAS were those which represented speaker-dependent characteristics, leaving out other characteristics related to linguistic content and to the transmission and recording devices. This speculation, of course, is difficult to verify but appears to be worthy of further elaborate investigation.

On Composite Parameter

Besides the 'elimination' of the influence of the response characteristics, another issue of concern in this study was the discriminating power of the speech parameter used. The goal of this study was to investigate speech parameters which are relatively resistive to distortion and also highly effective in identifying the speaker. The results indicated that the composite of the LTAS and FFC was the best approach to such a goal. Probable effects of the composite are suspected to be twofold: That the

number of features were simply increased by a combination of 10 from the LTAS and 9 from the FFC resulting into a total of 19 features; or different types of speech characteristics were integrated into one parameter containing both static spectral information and dynamic glottal information.

With respect to the first notion, Hughes (1968) showed that the tendency of monotonically increasing the error rate of identification beyond the certain optimum number of features and suggested that the number of the optimum features was 5, 10, 20, 100 or greater, for the size of 20, 100, 500, and larger. A similar result was also reported by Cheung and Eisenstein (1978). Using 32 features (pitch, log energy, 10 partial correlation coefficients, 10 cepstral coefficients, normalized absolute prediction error energy, and 9 normalized autocorrelation coefficients) extracted from text-independent speech data, they plotted the identification error rate as a function of the number of features. They concluded that the identification error rate gradually decreases by increasing the number of features, but it starts to taper off with 6 features -- no further improvement is gained by increasing by more than 6. Many other researchers in automatic speaker recognition appear to select a set of a certain number of features according to prior knowledge about their effectiveness without giving specific attention to the optimum number of features.

Since not only the number but the type of speech parameter from which features are derived may interact with the results, both

theoretically and practically it does not seem feasible to establish the optimum number of features. For these reasons, the possible effect of the increased number of features in the composite parameter upon the improvement of the identification rate in this study remains unanswered.

The latter notion perhaps provides more reasonable interpretation of the possible effect for improving the identification rate. As discussed earlier, one particular subgroup of speakers was consistently misidentified (correct rate 50%) among themselves when represented by one parameter (FFC), but another subgroup of speakers was misidentified (correct rate 70%) when represented by the other parameter (LTAS). In one particular condition, telephone vs. linear cross-transmission voice identification operation, the correct rate of 100% (by minimum set distance rule) was achieved when the above two parameters were made into a single composite parameter.

Cheung and Eisenstein (1978) studied text-independent voice identification and showed that if the speakers were represented by the set of features from three different types of parameters including pitch, energy, and spectral information, the identification performance was much better than if the speakers were represented by the set of features from only one of these parameters. By using text-dependent speech samples, He and Dubes (1982) also concluded that the combined feature set from the LPC and pitch contour resulted in the similar trend. Although these two studies were not concerned with the influence of the response

characteristics, the results may indicate general efficacy of the composite parameter of different types.

A reason for the poor identification rate obtained by the composite of the IDS and the FFC parameters may largely be due to the similar type of speech characteristics contained in both parameters. Since IDS was derived by the time-normalization formula reflecting the intensity variation of the feature (frequency component) as a function of time -- hence partially dynamic in nature, and FFC was mainly consisted of the variation of the fundamental frequency related features extracted from the time domain speech -- also dynamic in nature, these two parameters can be considered to share partially common characteristics.

More convincing supportive evidence for the usage of composite parameter of different types might be seen in the other methods of voice identification, namely the spectrographic method and aural method. Typically in the spectrographic method, multiple speech characteristics (parameters) are concurrently examined in a paired set of spectrograms, one for the unknown speaker and the other for the known speaker. These include mean frequencies and bandwidths of vowel formants, gaps and type of vertical striation , slopes and transition of formants, duration of similar phonetic elements and plosive gaps, energy distortion of fricatives and plosives, and interformant acoustic density patterns (Tosi *et al.*, 1972). Perceptual speech characteristics predominantly used by the human examiner in the aural method are known to be clarity, roughness, magnitude, and animation (Voiers, 1964), pitch, intensity, quality,

and rate (Holmgren, 1967), or quality, rhythm, melody pattern, pitch, rate and respiratory group (Tosi, 1979). To date there is no evidence alluding to the superiority of the perceptual method over the computer method of identification, or vice versa. Nonetheless, in practice , it is a general notion that the computer method of the present state of art would often fail to correctly identify the speaker given the text-independent speech materials containing other undefined characteristics in addition to those of speaker dependent characteristics. Under the same circumstances, the human examiner can often recognize the speaker with relative ease. In this sense, the ultimate goal of the computer voice identification seems to be attained only by simulating yet unknown perceptual mechanisms of the trained examiner who can extract multiple speech parameters singly or in any combination, depending upon the type of speech data at hand.

To sum up, considering the results from this study and reports in the literature cited above, the main reason for the improved identification rate by the composite parameter seems to be attributed to the inclusion of two different types of speech parameters, one type carrying static spectral features, the other type carrying dynamic glottal features. Though much effort in the study of automatic voice identification has been focused on the search of the cardinal speech parameter(s), it is likely that the identification system demands several types of parameters to fully represent the speakers who are text-dependently or -independently, and/or with or without influence of the transmission system.

On Influence of Pause Elimination

During the recording session, the speaker was allowed to read the text material at his preferred reading rate. Consequently, the varying rate of each individual speaker was reflected in the resulting speech data in terms of the duration of the voiced frames or in terms of the overall articulatory patterns. For example, compressed speech (processed speech data from which pauses are deleted) of a speaker who read at a relatively faster rate might have contained more voiced frames per unit of time than that of a speaker who read at a slower rate. Inevitably, the rate of voiced frames included in the compressed speech appeared to have interacted with each speech parameter derived to the different degree and in different aspects.

One of the speech parameters, LTAS (long-term averaged spectrum) is believed to have been affected to the minimum degree by the varying rate of voiced frames. Because LTAS was a spectrum of the averaged intensities of frequency components, it reflected no dependency upon the time variation. Also, the duration of the total voiced frames in the compressed speech of all the speakers was assumed to be long enough to counterbalance the different phonetic contents for each speaker.

In contrast, because all the features in IDS and in FFC (except Fo) were computed as a function of time (IDS, computed from a successive temporary varying short-term spectra, and FFC, computed from time varying Fo contour), these two parameters are believed to have been under the influence of the rate of voiced

frames. The influence seems to be double faceted: One is that the rate of voiced frames successfully reflected speaker-dependent characteristics as it was intended; another is that it was involved as a confounding variable. The latter aspect of the influence certainly leaves some room for further investigation.

On Interaction by the Experimenter

Certainly any kind of interaction by the experimenter should be avoided, or at least minimized if a completely objective or automatic method of voice identification is desired. In this study there were two spots which demanded the intensive participation of the experimenter. One spot took place during feature optimization procedure, where some amount of experimenter's strategy was inevitable in choosing the 10 features for the IDS and LTAS parameters. This interaction during the feature selection appears to be a drawback in view of the repeatability of the procedure. Since it was found that the feature optimization was a crucial procedure in the 'elimination' of the influence of the response characteristics, the need for further study for interaction-free algorithms for this feature optimization scheme is obvious.

Another spot of interaction took place during the interactive peak detecting method applied for the measurement of 9 features of the FFC parameter. This method was implemented to acertain accurate measurements of fine temporal variations of the fundamental frequency. Since the compressed speech data included many discontinuous points between successive voiced frames (signals), the application of the interactive method was crucial in order to prevent these discontinuities from resulting in erroneous measurements. Fortunately, it became clear that as long as the experimenter is familiar with the wave pattern of speech sound depicting the recurrent peaks, this interactive method would result in the stable measurements from experiment to experiment. For this reason, the interaction involved in the measurement procedure for the FFC parameter does not appear to have contributed to the resulting identification rate as a confounding factor.

CONCLUSIONS

This study was exploratory in nature regarding the methodology applied and the types of problems dealt with. Despite the application of a computer as the major computing source, there were several stages where interactions by the author were required. Within such a limitation set forth, the following general conclusions were drawn from the results of this study.

1. Both IDS (intensity deviation spectra) and FFC (fundamental frequency contour) are effective in eliminating the influence of the response characteristics of the transmission and recording channels. But their correct identification rates were only moderate 50-60%.
2. LTAS (long-term averaged spectra) is susceptible to the

influence of the response characteristics, but even under that influence, the correct identification rate was 60-70%.

3. The composite parameter of IDS and FFC is effective in eliminating the influence of the response characteristics. However, the correct identification rate is not improved, *i.e.*, it is only as good as each component.
4. The composite parameter of LTAS and FFC is the most effective speech parameter in eliminating the influence of the response characteristics of the transmission systems. It achieved the highest possible correct identification rate of 100%.

IMPLICATIONS FOR FURTHER RESEARCH

The major findings of this study was that the speech parameter composed of the optimized features of LTAS and the derived features of FFC can successfully eliminate the influence of the biased frequency response characteristics of the transmission and recording devices. In relation to this finding, further immediate research topics focusing on the same problem investigated in this study are suggested below.

1. The methodology used in this study could be replicated by using one composite parameter of LTAS and FFC with the increased speaker population of 50 or more.

2. Choral spectra could be investigated for its feasibility for feature optimization procedure. It is clear that choral spectra and the LTAS possess equally useful property in distinguishing the speakers provided all the voices are collected by the same transmission systems. The major advantage of choral spectra over LTAS (long-term averaged spectra) is a drastic reduction in the amount of time required to generate it. If the feature optimization can be made feasible with choral spectra, then it can form a composite parameter with FFC.

3. Feature optimization procedures applied in this study could be investigated for its further improvement. Although it was found that this procedure played a nontrivial role toward the elimination effects, it required some amount of arbitrary interaction by the experimenter. Ideally, there should be no interaction taken by the experimenter during the feature optimization process. This aspect appears to be worthy of serious investigation to make the computer method voice identification more objective.

4. The measure of the intra-speaker variability could explicitly be taken into account to establish the basis for assigning the probability of errors of identification.

5. Speech parameters other than IDS, LTAS, choral spectra, and FFC could be investigated not as alternative parameters, but as possible candidates to be included in the composite parameter. One such candidate is the pause characteristics

(considered to be independent of the frequency response characteristics) in an on-going speech of the speakers.

REFERENCES

REFERENCES

Anderberg, M.R. Cluster Analysis for Application, Academic Press, New York, 1973.

Atal, B. S. 'Automatic recognition of speakers from their voices' in Automatic Speech & Speaker Recognition, N. Rex Dixon and Thomas B. Martin (ed.) IEEE Press, New York, 1978, pp.349-364.

Atal, B.S. 'Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification', J. Acoust. Soc. Amer., 1974, Vol. 55, pp. 1304-1312.

Atal, B.S. 'Automatic speaker recognition based on pitch contour', J. Acoust. Soc. Amer., 1972, Vol. 52, pp. 1687-1697.

Bogner, R.E. 'On talker verification via orthogonal parameters', IEEE Trans. Acoust., Speech, and Signal Processing, 1981, Vol. ASSP-29, No. 1, pp.1-12.

Bricker, P.D. *et al.* 'Statistical techniques for talker identification', Bell System Technical Journal, 1971, Vol. 50, pp.1427-1454.

Bunge, E. 'Autmatic speaker recognition system AUROS for security systems and forensic voices identification' in Automatic Speech & Speaker Recognition, N. Rex Dixon and Thomas B. Martin (ed.). IEEE Press, New York, 1978, pp. 414-420.

Cheung, R.S. and Eisenstein, B.A. 'Feature selection via dynamic programming for text-independent speaker identification'. IEEE Trans. Acoust., Speech, and Signal Processing, Oct.1978, Vol. ASSP-26, No.5, pp.397-403.

Das, S.K. and Mohn, W.S. 'A scheme for speech processing in automatic speaker verification', IEEE Trans. Audio Electro-Acoust., 1971, Vol. Au-19, pp.32-43.

Doddington, G.R. 'A method of speaker verification', Paper presented at The Eightieth Meeting of the Acoust. Soc. Amer. 1970, Nov. 3-8, Houston, Texas.

Doddington, G.R. 'Speaker verification - Final report', Rome Air Development Center, Griffiss AFB, N.Y., Tech. Rep. April, 1974, RADC 74-179.

Furui, S. 'Cepstrum analysis technique for automatic speaker verification' IEEE Trans. Acoust., Speech, and Signal Processing, April, 1981a, Vol. ASSP-2, No. 2. pp.254-272.

Furui, S. 'Comparison of speaker recognition methods using statistical features and dynamic features', IEEE Trans. Acoust., Speech, and Signal Processing, 1981b, Vol. ASSP-29, No. 3, pp.342-350.

Furui, S. Itakura, F., and Saito, S. 'Talker recognition by longtime averaged speech spectrum', Electronics and Communications in Japan, 1972, 55-A, pp.54-61.

Gold, B. and Rabiner, L. 'Parallel processing techniques for estimating pitch periods of speech in the time domain'. J. Acoust. Soc. Amer., 1969, Vol. 46, pp.442-448.

Hair, G.D. and Rekieta, T.W. 'Automatic speaker verification using phoneme spectra', J. Acoust. Soc. Amer., 1972, Vol. 51, p.131(a).

Hair, G.D. and Rekieta, T.W. 'Mimic resistance of speaker verification using phoneme spectra', J. Acoust. Soc. Amer., 1972, Vol. 51, p.131(a).

He, Q., and Dubes, R. 'An Experiment in Chinese speaker identification', presented at 1982 IEEE Int'l Conf. Trans. Acoust., Speech, and Signal Processing.

Holmgren, G. 'Physical and psychological correlates of speaker recognition', Journal of Speech and Hearing Research, 1967, Vol. 10, pp.57-66.

Hughes, G. F. 'On the mean accuracy of statistical pattern recognizers', IEEE Trans. on Information Theory, 1968, Vol. IT-14, pp.55-63.

Hunt, M.J., Yates, J.W., and Briddle, J.S. 'Automatic speaker recognition for use over communication channels', IEEE Int's Conf. Record on Acoust., Speech and Signal Process. May 9-11, 1977, pp.764-767.

Jain, A.K. and Dubes, R. 'Feature definition in pattern recognition with small sample size', Pattern Recognition, 1978, Vol.10, pp.85-97.

Luck, J.E. 'Automatic speaker verification using cepstral measurements', J. Acoust. Soc. Amer., 1969, Vol.46, pp.102-1032.

Majewski, Z. W., and Hollien H. 'Cross correlation of long-term speech spectra as a speaker identification technique', *Acustica*, 1975, Vol. 34, pp.20-24.

Markel, J.D. 'The SIFT algorithm for fundamental frequency estimation', *IEEE Trans., Audio, Electroacoust.*, 1977, Vol. AV-20, pp.367-377.

Markel, J.D., and Davis, S.B. 'Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base', *IEEE Trans. Acoust., Speech, and Signal Processing*, Feb., 1977, Vol. ASSP-27, No.1, pp.74-82.

Markel, J.D., Oshika, B.T., and Gray, A.H. 'Long-term feature averaging for speaker recognition', *IEEE Trans. Acoust., Speech, and Signal Processing*, 1977, Vol. ASSP-25, pp.330-337.

The National Research Council, *On the Theory and Practice of Voice Identification*, National Academy of Sciences, Washington D.C., 1979.

Noll, A.M. 'Cepstrum pitch determination', *J. Acoust. Soc. Amer.*, 1967, Vol.41, pp.293-309.

Paul, J.E., Rabinowitz, A.S., Riganati, J.P., Richardson, J.M. 'Development of analytical methods for a semi-automatic speaker identification system', *1975 Carnahan Conf. on Crime Countermeasures*, 1975, pp-52-64.

Pruzansky, S. 'Pattern-matching procedure for automatic talker recognition', *J. Acoust. Soc. Amer.*, 1963, Vol.35, pp.354-358.

Rabiner, L.R., Levinson, S.E., Rosenberg, A.E., and Wilson,J.G. 'Speaker-independent recognition of isolated words using clustering techniques', *IEEE Trans. Acoust., Speech, and Signal Processing*, Aug., 1977, Vol. ASSP-27, No.4, pp.336-349.

Sammon, J.W. Jr. 'A nonlinear mapping for data structure analysis', *IEEE Transactions on Computer*, May 1969, pp.401-409.

Shafer, R. L., and Rabiner, L. 'Digital representation of speech signals', *Proceedings of IEEE* 1975, Vol. 63, pp.662-667.

Tarno'czy, T. 'Determination du spectre de la parole avec une methode nouvelle', *Acustica*, 1958, 8:392-395.

Tosi, O. *Voice Identification: Theory and Legal Applications*. University Press, Baltimore, 1979.

Tosi, O. 'Pausometry: Measurement of a low level of acoustic energy', in World Papers in Phonetics, Phonetics Society of Japan, The Phonetic Society of Japan, Tokyo, 1974, pp.129-144.

Tosi, O., and Nakasone, H. 'Cancellation of the telephone response curve', Paper presented at International Association for Identification, Aug., 1982, Rochester, New York.

Tosi, O., Pisani, R., Dubes, R., and Jain, A. 'An objective method of voice identification', Current Issues in the Phonetic Sciences, Harry & Patricia Hollien (ed.) In series of Current Issues in Linguistic Theory Vol. 9, in Amsterdam Studies in the Theory and Hearing of Linguistic Science IV, Amsterdam-John Benjamins B.V., 1979, pp.851-861.

Tosi, O., *et al.* 'Experiment on voice identification', J. Acoust. Soc. Amer., 1972, Vol. 51, pp.2030-2043.

Voiers, W. 'Perceptual bases of speaker identity', J. Acoust. Soc. Amer., 1964, Vol. 36, pp.1065-1073.

APPENDIX A

A SAMPLE TEXT EXCERPT

APPENDIX A

A SAMPLE TEXT EXCERPT

Halfway down the Canyon, you come to 400-million-year-old limestones. There are no reptiles to be found here, but there are the bones of strange armoured fish. An hour or so later – and a hundred million years earlier – the rocks contain no sign of backboned animals of any kind. There are a few shells and worms that have left behind a tracery of trails in what was the muddy sea floor. Three-quarters of the way down, you are still descending through layers of limestone, but now there is no sign of life whatever. By the late afternoon, you ride at last into the lower gorge where the Colorado River runs green between high rock walls. You are now a vertical mile below the rim and the rocks have been dated to the immense age of 2000 million years. Here you might hope to find evidence for the very beginnings of life. But there are no organic remains of any kind. The dark fine-grained rocks lie not in horizontal layers like all those above, but are twisted and buckled and riven with veins of pink granite.

Are signs of life absent because these rocks and the limestones directly above are so extremely ancient that all such traces have been crushed from them? Could it be that the first creatures to leave any sign of their existence were as complex as worms and molluscs? For many years these questions puzzled geologists. All over the world, rocks of this antiquity were carefully searched for organic remains. One or two odd shapes were found, but most authorities dismissed these as patterns produced by the physical processes of rock formation that had nothing whatever to do with living organisms. Then during the 1950s, the searchers began to use high-powered microscopes on some particularly enigmatic rocks.

A thousand miles northeast of the Grand Canyon, ancient rocks of about the same age as those beside the Colorado River outcrop on the shores of Lake Superior. Some of them contain seams of a fine-grained flint-like substance called chert. This was well known during the last century because the pioneers used it in their flintlock guns. Here and there, it contains strange white concentric rings a metre or so across. Were these merely eddies in the mud on the bottom of the primeval seas or could they have been formed by living organisms? No one could be sure and the shapes were given the non-committal name of stromatolite, a word derived from Greek meaning no more than 'stony carpet'. But when researchers cut sections of these rings, ground them down into slices so thin that they were translucent and examined them through the microscope, they found, preserved in the chert, the shapes of simple organisms, each no more than one or two hundredths of a millimetre across. Some resembled filaments of algae; others, while they were unmistakably organic, had no parallels with living organisms; and some looked to be identical with the simplest form of life existing today, bacteria.

It seemed almost impossible to many people that such tiny things as micro-organisms could have been fossilised at all. That relics of them should have survived for such a vast period of time seemed even more difficult to believe. The solution of silica which had saturated the dead organisms and solidified into chert was clearly as fine-grained and durable a preservative as exists. The discovery of the fossils in the Gunflint Chert stimulated further searches not only in North America but all over the world and other micro-fossils were found in cherts in Africa and Australia. Some of these, astonishingly, pre-dated the Gunflint specimens by a thousand million years. But if we want to consider how life arose, we have to look back a further thousand million years beyond even the earliest micro-fossils, to a time when the earth was completely lifeless and still cooling after its birth.

The planet then was radically different in almost every way from the one we live on today. The clouds of water vapour that had surrounded it had condensed to form seas, but they were still hot. We are not sure how the land masses lay, but they certainly bore no resemblance in either form or distribution to modern continents. Volcanoes were abundant, spewing ash and lava. The atmosphere was very thin and consisted of swirling clouds of hydrogen, carbon monoxide, ammonia and methane. There was little or no oxygen. This mixture allowed ultraviolet rays from the sun to bathe the earth's surface with an intensity that would be lethal to modern animal life. Electrical storms raged in the clouds, bombarding the land and the sea with lightning.

Laboratory experiments were made in the 1950s to discover what might happen to these particular chemical constituents under such conditions. Such gases, mixed with water vapour, were subjected to electrical discharge and ultraviolet light. After only a week of this treatment complex molecules were found to have formed in the mixture, including sugars, nucleic acids and amino acids, the building blocks of proteins. There seems no doubt that molecules such as these could have formed in the seas of the earth at the very beginning of its history.

As the millions of years passed, the concentrations of these substances increased and the molecules began to interact with one another to form even more complex compounds. It may even be that some ingredients were added from outer space, brought by meteorites. Eventually, among a great variety of substances, there appeared one that was to be crucial for the further development of life. It is called deoxyribonucleic acid, or DNA for short. Its structure endows it with two key properties. First, it can act as a blueprint for the manufacture of amino acids; and second, it has the capacity to replicate itself. With this substance, molecules had reached the threshold of something quite new, for these two characteristics of DNA are also those of living organisms such as bacteria. And bacteria, besides being the simplest form of life we know, are also among the oldest fossils we have discovered.

The ability of DNA to replicate itself is a consequence of its unique structure. It is shaped like two intertwined helices. During cell division, these unzip, splitting the molecule along its length into two separate helices. Each then acts as a template to which other simpler molecules become attached until each has once more become a double helix.

The simple molecules from which the DNA is mainly built are of only four kinds, but they are grouped in trios and arranged in a particular and significant order on the immensely long DNA molecule. This order specifies how the twenty or so different amino acids are arranged in a protein, how much is to be made, and when. A length of DNA bearing the information for an unbroken sequence of manufacture is called a gene.

Occasionally, the DNA copying process involved in reproduction may go wrong. A

(Source: Life on Earth by David Attenborough, pp. 18-19, 1979)

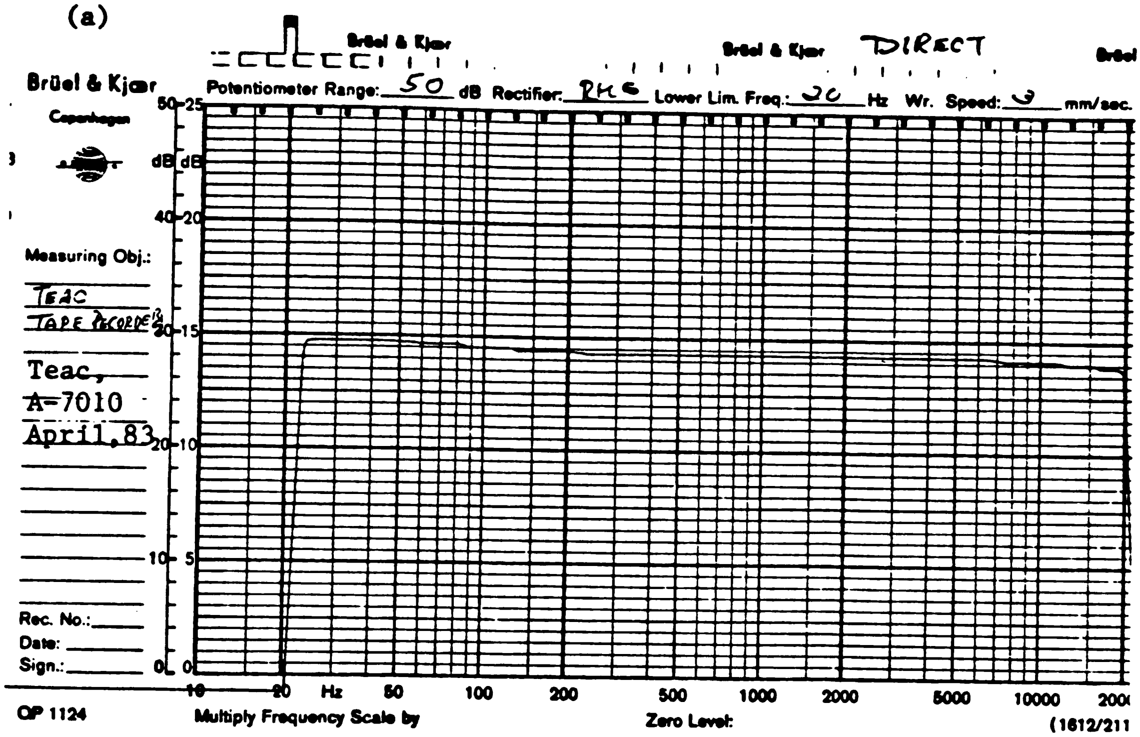
APPENDIX B

RESPONSE CHARACTERISTICS OF TEAC TAPE RECORDER AND BRUEL & KJAER MICROPHONE

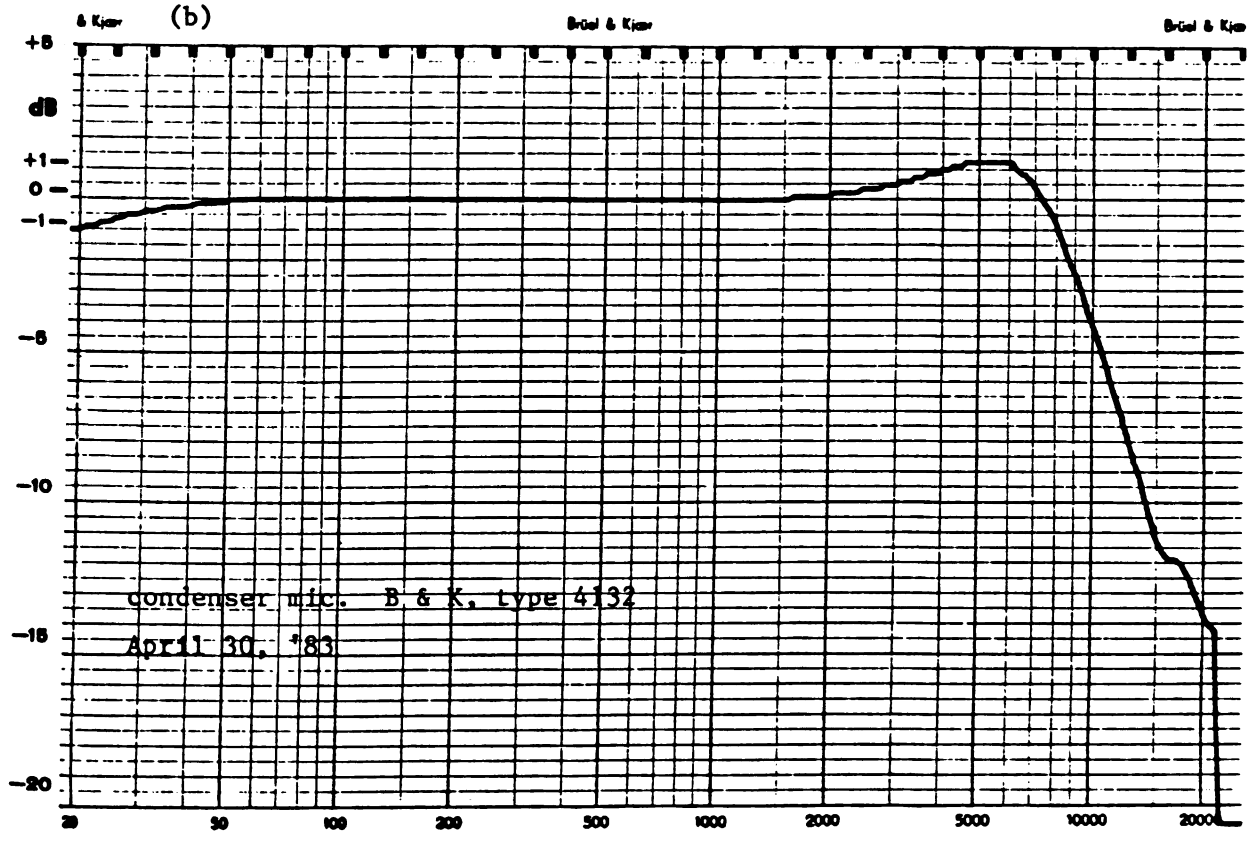
APPENDIX B

RESPONSE CHARACTERISTICS OF TEAC TAPE RECORDER
AND BRUEL & KJAER MICROPHONE

(a)



(b)



APPENDIX C

SUMMARY OF THE RESULTS FROM PAUSE ELIMINATION

APPENDIX C

SUMMARY OF THE RESULTS FROM PAUSE ELIMINATION

	Telephone transmission			Normal transmission			Linear transmission		
	Tp = 20 msec			Tp = 20 msec			Tp = 20 msec		
Speaker	Ap	output(sec)	ratio	Ap	output(sec)	ratio	Ap	output(sec)	ratio
1	.975	59.413	.495	.940	59.000	.492	.930	60.951	.508
2	.960	60.235	.502	.950	60.764	.506	.920	61.285	.511
3	.975	59.717	.497	.950	57.139	.476	.920	61.189	.510
4	.970	59.688	.497	.960	58.647	.489	.920	57.743	.481
5	.977	57.700	.481	.970	57.801	.482	.920	56.863	.474
6	.950	59.509	.496	.940	60.206	.502	.900	61.231	.510
7	.940	58.628	.489	.910	60.976	.508	.880	61.248	.510
8	.945	58.447	.487	.950	60.588	.505	.900	60.660	.506
9	.950	60.385	.503	.870	59.060	.492	.830	60.910	.508
10	.970	59.835	.499	.930	60.419	.503	.930	60.875	.507
mean	.960	59.356	.495	.937	59.460	.496	.905	60.296	.503
S.D.	.014	.802	.007	.027	1.262	.010	.029	1.521	.013

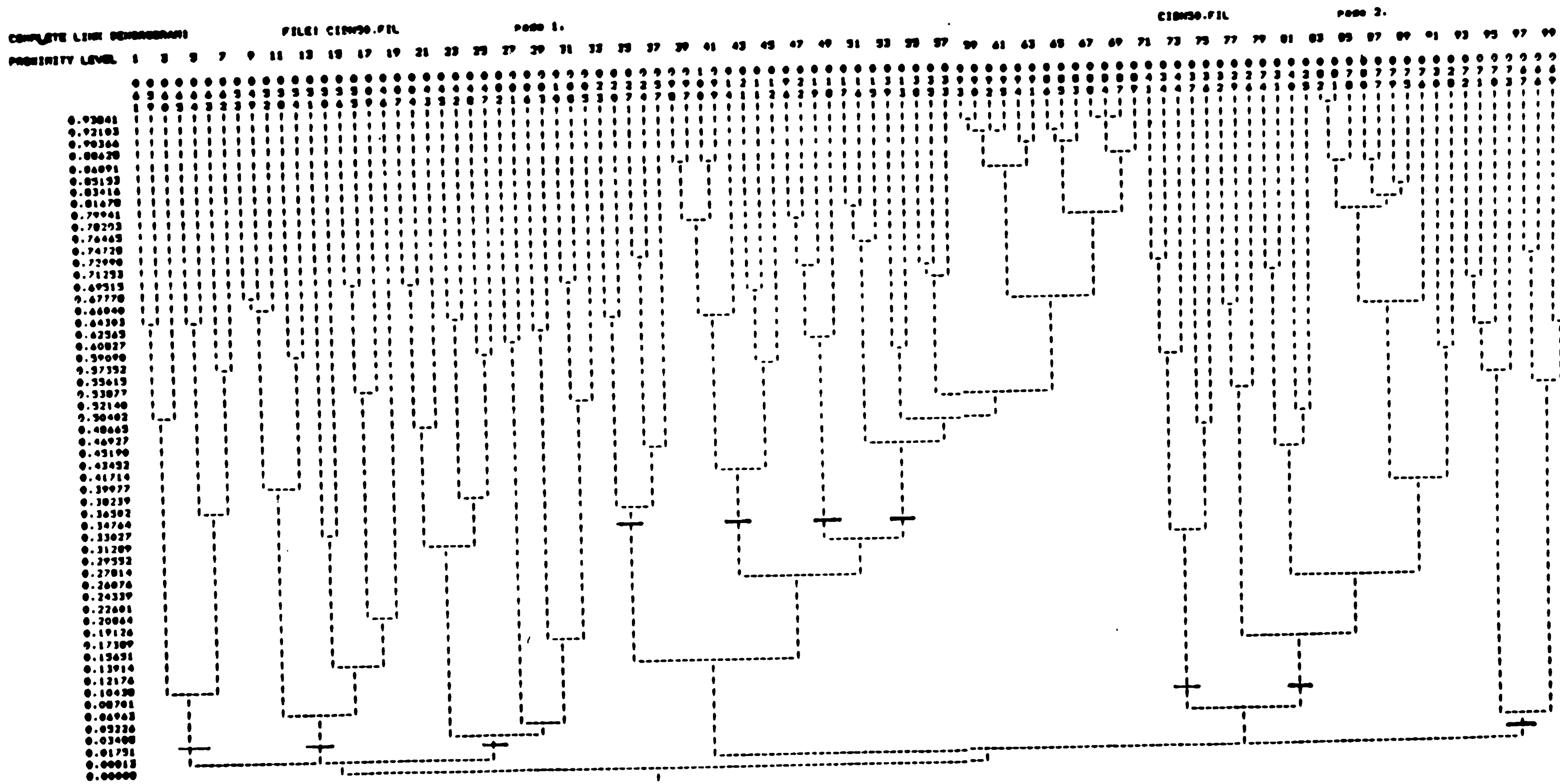
The duration of input speech signal was 2 minutes for all speakers. The ratio was computed by dividing output length (in sec) by input length (in sec). The Tp value was kept constant of 20 msec. The Ap value was made variable to obtain the output signal of the prescribed interval of duration.

APPENDIX D

COMPLETE-LINK DENDROGRAMS OF IDS AND LTAS FEATURES

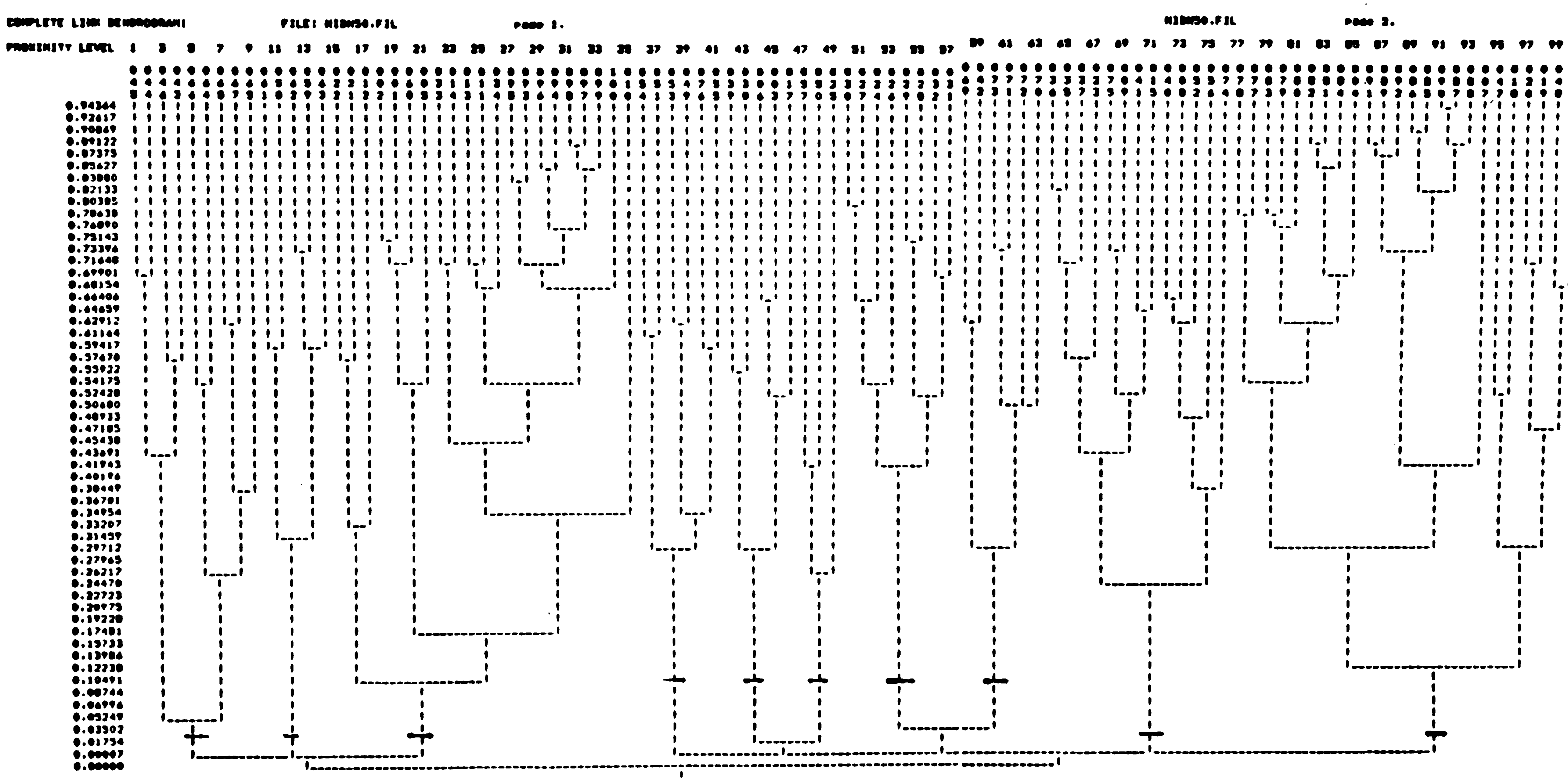
APPENDIX D(1)

DENDROGRAM OF 100 FEATURES BASED ON TELEPHONE IDS.



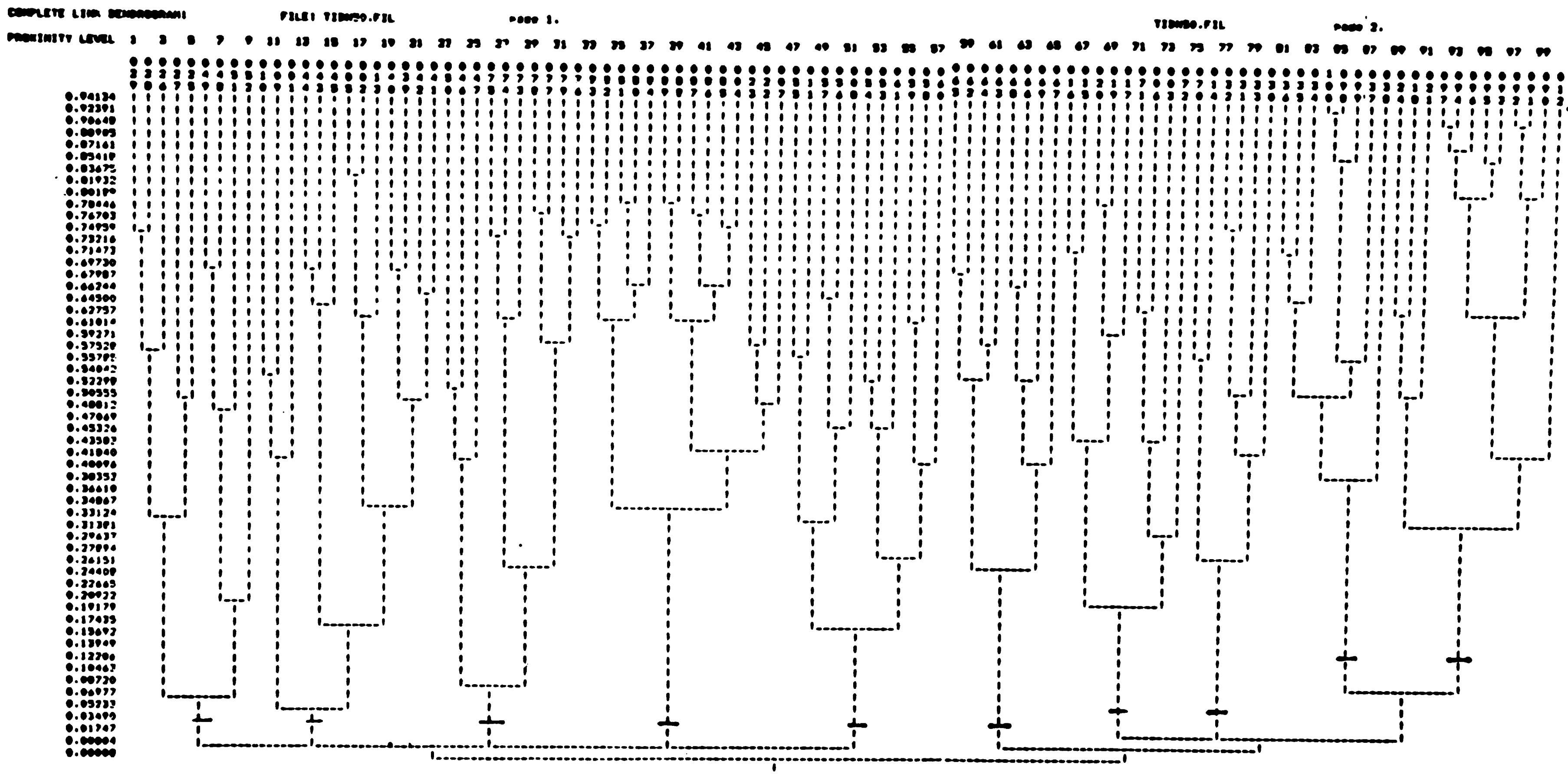
APPENDIX D(2)

DENDROGRAM OF 100 FEATURES BASED ON NORMAL IDS.



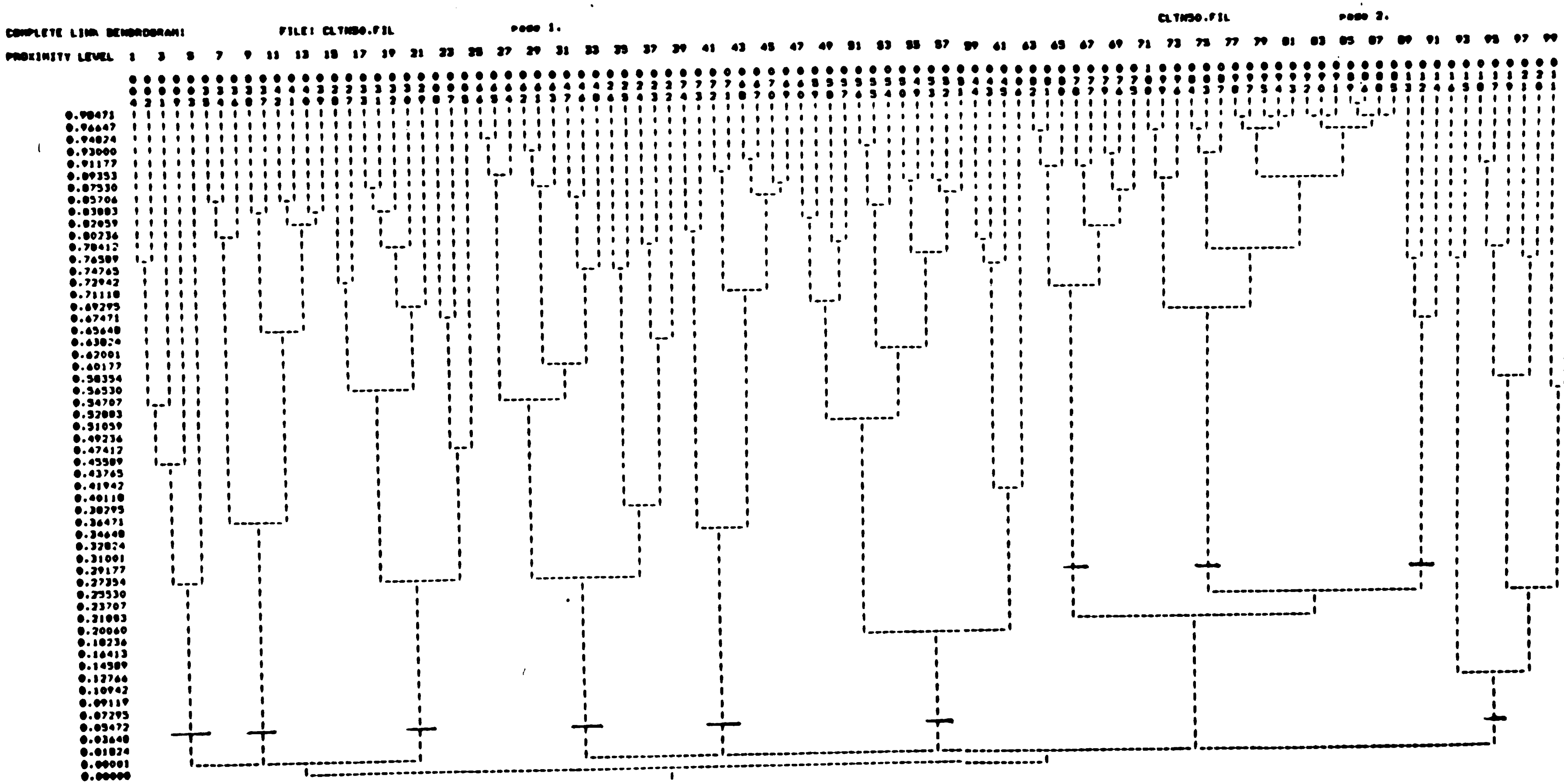
APPENDIX D (3)

DENDROGRAM OF 100 FEATURES BASED ON LINEAR IDS.



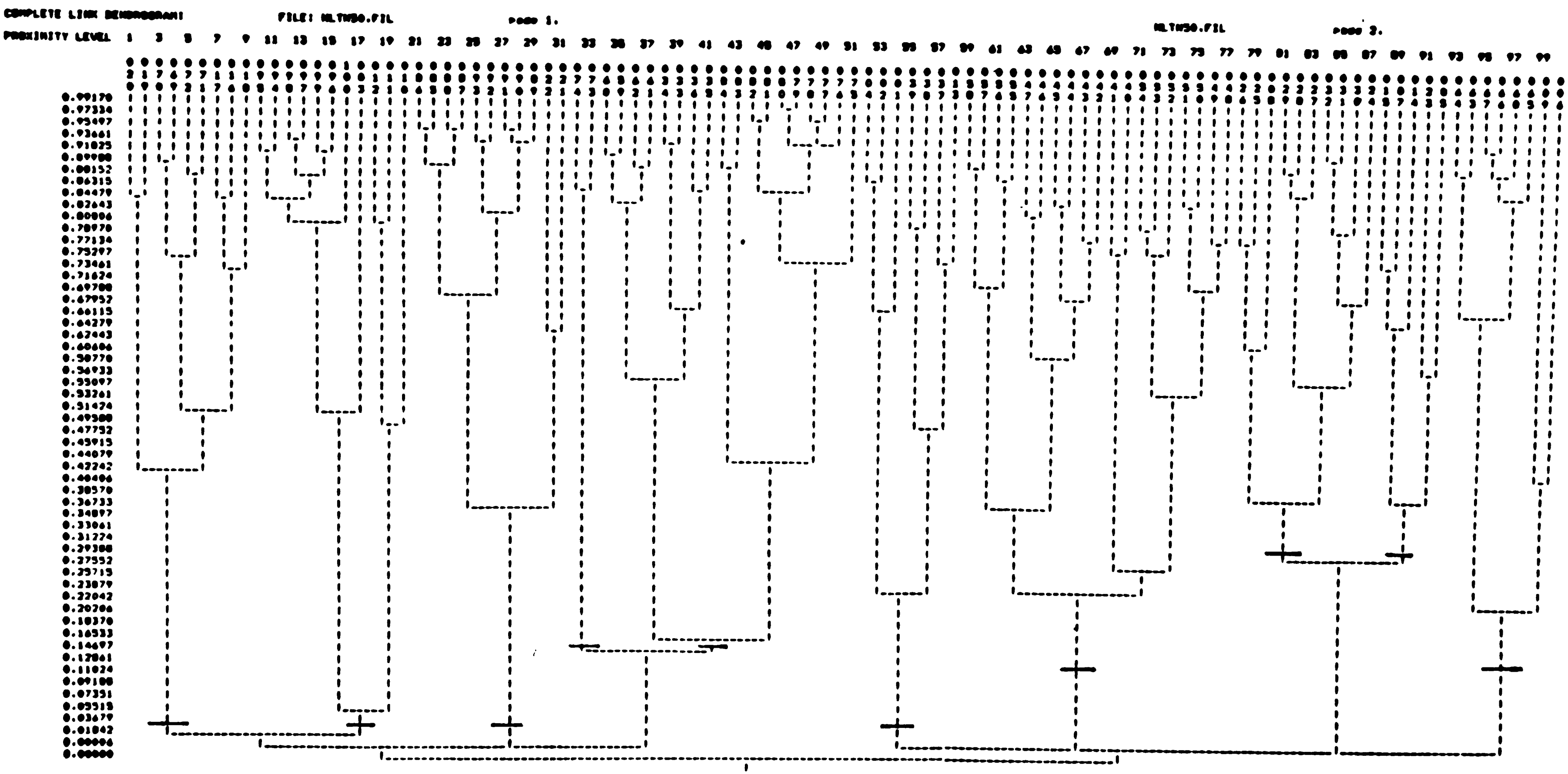
APPENDIX D (4)

DENDROGRAM OF 100 FEATURES BASED ON TELEPHONE LTAS.



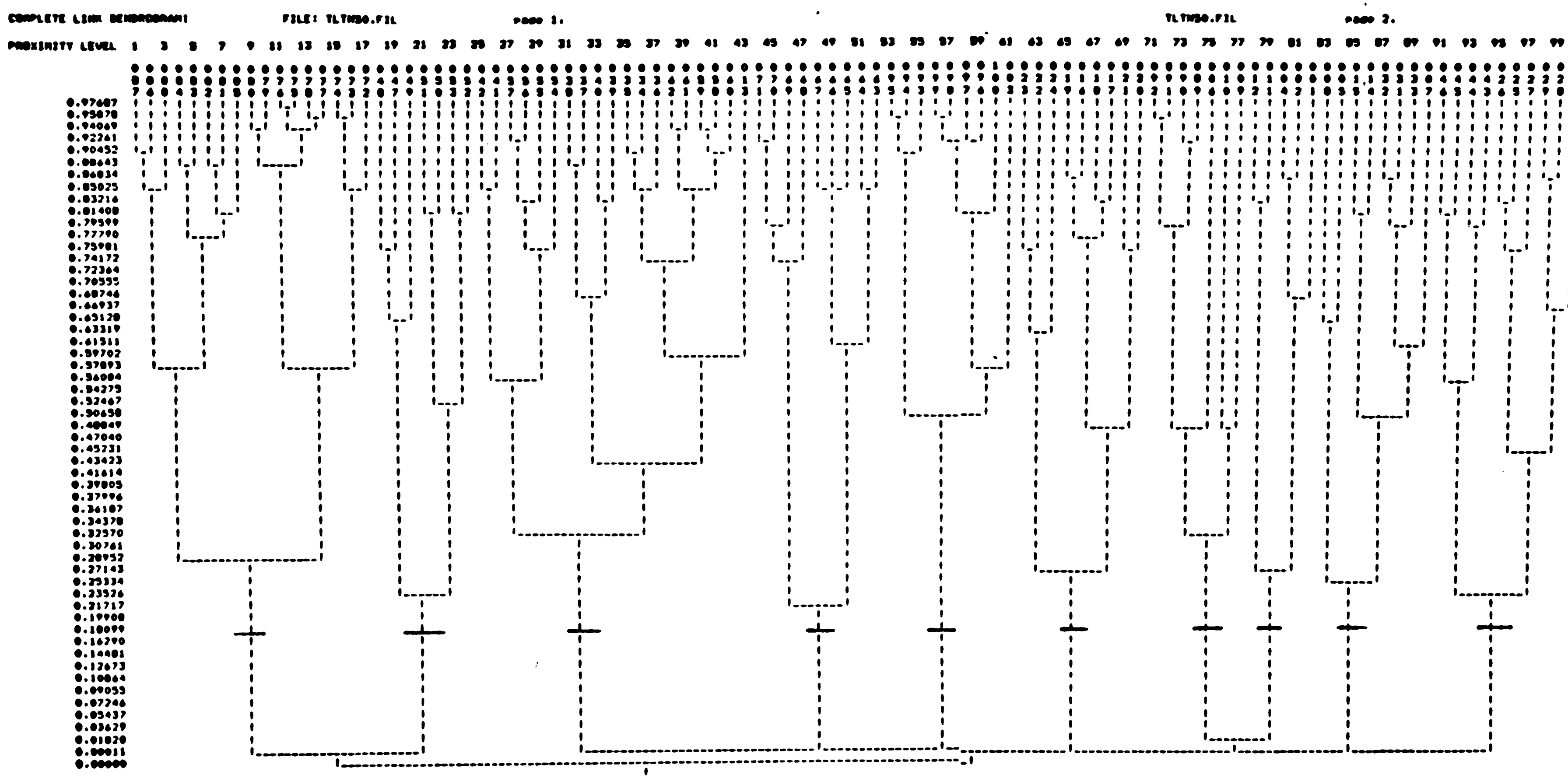
APPENDIX D (5)

DENDROGRAM OF 100 FEATURES BASED ON NORMAL LTAS.



APPENDIX D (6)

DENDROGRAM OF 100 FEATURES BASED ON LINEAR LTAS.



APPENDIX E

RAW DATA OF FFC FEATURES

APPENDIX E

RAW DATA OF FFC FEATURES: BY LINEAR TRANSMISSION

speaker number	sample number	$\bar{F}o$ (Hz)	σFo (Hz)	$\bar{\Delta}Fo$ (Hz)	$\bar{\Delta}Fo/\bar{F}o$ ratio	σAo ampl.	Fo(max) (Hz)	Fo(min) (Hz)	Fo(rng) (Hz)	$\bar{\Delta}Ao$ ampl.
	1	107.24	16.69	5.53	5.16	43.43	163.93	67.57	96.37	8761.86
	2	109.29	17.18	6.18	5.66	43.75	204.08	64.52	139.57	8643.10
1	3	106.77	13.86	4.18	3.92	43.16	172.41	53.76	118.65	8951.75
	4	110.73	16.39	5.05	4.56	45.59	161.29	57.47	103.82	9001.60
	5	109.21	17.60	5.62	5.14	42.06	178.57	82.64	95.93	8603.13
	1	89.66	13.47	3.69	4.12	49.17	125.00	70.92	54.08	9151.75
	2	90.08	11.50	3.74	4.15	44.28	138.89	68.49	70.40	9495.82
2	3	88.33	10.94	3.11	3.52	49.00	121.95	70.42	51.53	9320.30
	4	89.46	12.28	3.98	4.45	53.00	112.36	57.47	54.89	10542.20
	5	91.92	10.46	2.53	2.76	45.28	109.89	78.74	31.15	8993.58
	1	108.91	19.83	3.43	3.15	43.08	153.85	73.53	80.32	8752.73
	2	109.47	18.54	6.61	6.04	19.03	166.67	78.74	87.93	4783.08
3	3	112.35	19.84	4.81	4.28	50.04	166.67	69.44	97.22	9384.75
	4	110.15	22.19	5.04	4.57	46.57	156.25	79.37	76.88	9170.75
	5	109.00	22.40	4.87	4.47	47.06	158.73	49.50	109.23	9280.34
	1	127.56	24.14	6.09	4.77	40.71	217.39	90.91	126.48	8228.55
	2	130.94	28.20	6.13	4.68	45.72	181.82	66.67	115.15	9009.06
4	3	131.27	19.17	5.15	3.92	43.74	178.57	104.17	74.40	8677.66
	4	131.03	20.33	7.47	5.70	101.62	185.19	93.46	91.73	3744.43
	5	125.73	26.98	6.07	4.83	46.27	208.33	90.09	118.24	8950.38
	1	107.33	22.14	5.91	5.50	46.77	153.85	45.05	108.80	9072.82
	2	110.56	18.51	5.86	5.30	50.61	151.52	86.21	65.31	9663.06
5	3	111.06	22.03	5.69	5.13	48.33	151.52	53.76	97.75	9467.44
	4	111.69	23.96	7.56	6.77	46.65	370.37	74.63	295.74	8978.88
	5	109.81	21.11	6.09	5.54	43.14	175.44	44.25	131.19	8624.11
	1	165.30	33.61	10.77	6.51	20.24	277.78	83.33	194.44	5174.55
	2	165.14	31.63	8.45	5.12	41.26	227.27	86.21	141.07	8482.85
6	3	164.52	40.46	9.25	5.62	44.99	277.78	98.04	179.74	8697.50
	4	173.73	39.26	9.94	5.72	40.25	263.16	88.50	174.66	8245.51
	5	160.64	42.05	9.12	5.68	44.97	270.27	71.43	198.84	8908.37
	1	116.62	24.00	4.53	3.88	47.22	181.82	80.00	101.82	8860.46
	2	117.27	28.89	4.72	4.02	48.08	185.19	66.67	118.52	9619.97
7	3	113.69	19.98	4.18	3.68	43.95	153.85	86.96	66.89	8684.46
	4	115.08	20.08	4.13	3.59	45.07	151.52	80.65	70.87	8772.21
	5	115.15	22.72	4.26	3.70	45.31	166.67	88.50	78.17	8894.06
	1	118.43	33.15	6.38	5.38	47.12	232.56	80.65	151.91	8947.70
	2	117.41	27.95	6.57	5.60	46.17	181.82	85.47	96.35	8767.81
8	3	119.79	36.01	4.70	3.92	44.83	192.31	84.03	108.27	8580.54
	4	119.46	32.73	5.73	4.79	41.50	192.31	80.65	111.66	8450.04
	5	115.22	21.98	5.07	4.40	45.26	161.29	84.75	76.54	8803.62
	1	115.00	28.89	4.70	4.09	53.14	185.19	75.76	109.43	10158.54
	2	110.27	25.38	4.82	4.37	47.78	185.19	74.07	111.11	9009.67
9	3	108.88	18.64	3.90	3.58	50.73	151.52	79.37	72.15	9409.29
	4	112.15	24.50	4.27	3.80	42.30	166.67	89.29	77.38	8383.58
	5	107.63	20.12	3.60	3.35	48.00	156.25	78.12	78.12	9300.37
	1	146.24	43.08	7.24	4.95	44.10	256.41	99.01	157.40	8879.20
	2	148.26	46.44	8.46	5.71	43.23	294.12	68.97	225.15	8156.65
10	3	151.89	40.30	6.64	4.37	40.76	238.10	104.17	133.93	7961.03
	4	158.83	48.06	8.54	5.38	42.78	344.83	94.34	250.49	8422.73
	5	156.70	45.91	7.94	5.07	41.93	357.14	94.34	262.80	8485.80

APPENDIX E

RAW DATA OF FFC FEATURES: BY NORMAL TRANSMISSION

speaker number	sample number	$\bar{F}o$ (Hz)	σFo (Hz)	$\bar{\Delta}Fo$ (Hz)	$\bar{\Delta}Fo/\bar{F}o$ (ratio)	σAo ampl.	Fo(max) (Hz)	Fo(min) (Hz)	Fo(rng) (Hz)	$\bar{\Delta}Ao$ ampl.
	1	106.87	11.52	4.49	4.20	41.54	126.58	75.76	50.82	8612.94
	2	108.29	12.62	4.31	3.98	41.83	147.06	90.91	56.15	8371.87
1	3	109.11	11.35	3.90	3.58	42.37	135.14	90.09	45.05	8640.17
	4	107.85	12.72	5.04	4.67	47.56	138.89	71.94	66.95	9021.36
	5	107.88	12.91	5.13	4.76	38.67	149.25	84.03	65.22	8045.95
	1	84.67	10.00	2.87	3.40	50.49	98.04	67.57	30.47	9486.71
	2	90.79	17.94	3.93	4.33	53.35	116.28	55.25	61.03	10147.71
2	3	91.83	14.39	3.02	3.29	58.51	113.64	71.94	41.69	10640.63
	4	88.39	13.06	3.79	4.28	48.70	123.46	72.99	50.46	9866.24
	5	93.77	12.85	3.86	4.12	53.71	117.65	72.46	45.18	9906.37
	1	105.26	15.30	4.97	4.72	42.07	129.87	49.02	80.85	7994.99
	2	109.99	15.40	4.90	4.46	43.06	138.89	78.12	60.76	8594.69
3	3	115.17	14.48	3.93	3.41	47.74	140.85	97.09	43.76	9141.12
	4	111.88	13.82	3.84	3.44	46.56	147.06	90.91	56.15	8908.32
	5	104.27	26.21	4.84	4.64	42.67	136.99	45.05	91.94	9057.27
	1	132.93	34.67	4.94	3.72	43.01	188.68	100.00	88.68	8491.22
	2	121.32	18.62	4.21	3.47	44.99	151.52	89.29	62.23	9071.54
4	3	127.80	34.89	4.96	3.88	45.05	196.08	89.29	106.79	9001.69
	4	122.78	26.37	5.66	4.61	45.34	172.41	96.15	76.26	9241.12
	5	126.17	37.82	5.35	4.24	44.81	217.39	83.33	134.06	8633.23
	1	109.47	24.46	5.25	4.80	45.57	147.06	88.50	58.56	9136.10
	2	109.90	18.17	5.76	5.24	50.19	131.58	90.09	41.49	9117.34
5	3	106.87	11.27	4.80	4.49	57.30	128.21	78.12	50.08	10205.92
	4	107.34	13.71	5.73	5.34	44.06	133.33	86.21	47.13	8772.67
	5	106.76	24.63	5.41	5.07	56.02	131.58	45.87	85.71	10280.46
	1	164.97	33.15	8.41	5.10	42.06	232.56	108.70	123.86	8170.47
	2	167.78	25.10	7.95	4.74	42.25	212.77	131.58	81.19	8447.16
6	3	170.50	39.29	8.22	4.82	42.63	243.90	126.58	117.32	8434.58
	4	171.24	31.82	7.94	4.64	42.10	256.41	131.58	124.83	8257.78
	5	162.60	38.34	7.27	4.47	47.92	270.27	126.58	143.69	8852.94
	1	117.38	21.91	3.81	3.24	48.10	156.25	92.59	63.66	9380.95
	2	118.60	19.42	4.05	3.42	49.48	158.73	91.74	66.99	9309.45
7	3	118.57	21.61	4.56	3.85	48.24	151.52	95.24	56.28	9451.14
	4	118.72	21.21	3.97	3.34	46.74	149.25	94.34	54.91	9089.74
	5	112.41	15.34	4.47	3.98	47.95	135.14	92.59	42.54	8823.71
	1	113.28	23.39	5.44	4.80	55.73	153.85	90.09	63.76	10916.22
	2	111.72	19.59	5.14	4.60	44.92	147.06	92.59	54.47	8686.57
8	3	127.53	39.21	5.74	4.50	42.38	192.31	98.04	94.27	8182.26
	4	116.75	32.64	5.01	4.29	49.61	161.29	86.96	74.33	9662.65
	5	112.59	28.71	4.32	3.84	55.10	161.29	86.96	74.33	10539.89
	1	114.59	28.86	3.85	3.36	46.13	161.29	85.47	75.82	8442.39
	2	120.09	26.85	4.33	3.61	48.06	158.73	81.30	77.43	9319.74
9	3	106.59	21.04	4.36	4.09	51.52	147.06	84.75	62.31	10265.07
	4	114.73	25.68	3.54	3.08	54.27	161.29	86.21	75.08	9850.66
	5	107.75	21.50	4.19	3.88	44.59	149.25	74.07	75.18	8694.71
	1	160.76	52.81	6.26	3.89	42.55	256.41	116.28	140.13	8352.17
	2	146.77	50.57	7.43	5.07	42.94	294.12	112.36	181.76	8050.48
10	3	157.68	52.30	6.45	4.09	44.54	250.00	114.94	135.06	8914.43
	4	157.99	50.07	6.77	4.28	43.08	243.90	116.28	127.62	8617.10
	5	151.56	55.50	5.93	3.91	42.41	250.00	93.46	156.54	8536.33

APPENDIX E

RAW DATA OF FFC FEATURES: BY TELEPHONE TRANSMISSION

speaker number	sample number	$\bar{F}o$ (Hz)	σFo (Hz)	$\overline{\Delta F}o$ (Hz)	$\overline{\Delta F}o/\bar{F}o$ (ratio)	σAo ampl.	Fo(max) (Hz)	Fo(min) (Hz)	Fo(rng) (Hz)	$\overline{\Delta A}o$ ampl.
	1	107.41	21.46	5.29	4.93	45.14	136.99	49.26	87.73	8552.70
	2	109.07	12.96	4.51	4.13	49.42	138.89	74.07	64.81	9637.16
1	3	110.94	19.38	6.87	6.19	45.18	151.52	59.88	91.63	9015.24
	4	107.75	18.81	5.53	5.14	43.04	147.06	44.64	102.42	8968.29
	5	108.75	13.72	5.79	5.33	44.14	140.85	84.75	56.10	8357.97
	1	87.89	14.65	3.74	4.25	48.81	123.46	40.98	82.47	9445.13
	2	91.62	14.45	4.93	5.38	50.22	135.14	61.35	73.79	9881.80
2	3	90.07	13.22	4.12	4.58	50.46	114.94	68.97	45.98	9614.30
	4	87.95	13.60	4.01	4.56	53.07	121.95	44.84	77.11	10119.20
	5	90.48	17.98	4.47	4.94	55.89	142.86	64.10	78.75	10010.96
	1	106.51	25.01	4.77	4.48	51.77	147.06	45.45	101.60	10059.78
	2	110.28	18.47	4.32	3.91	46.83	149.25	78.74	70.51	8950.81
3	3	113.42	18.17	4.74	4.18	41.28	166.67	88.50	78.17	8340.65
	4	107.50	23.61	4.69	4.36	48.25	147.06	46.95	100.11	9309.29
	5	112.63	21.00	4.45	3.95	48.63	151.52	75.76	75.76	9403.09
	1	126.64	26.89	5.74	4.53	44.00	188.68	79.37	109.31	8607.10
	2	128.27	30.66	6.88	5.36	48.98	192.31	65.79	126.52	9158.48
4	3	129.49	31.27	6.18	4.77	48.59	188.68	85.47	103.21	9666.28
	4	128.04	26.58	5.64	4.40	39.82	188.68	78.74	109.94	8031.53
	5	127.49	31.59	5.19	4.07	48.52	200.00	92.59	107.41	9213.97
	1	109.01	23.13	5.24	4.80	42.24	172.41	77.52	94.89	8524.42
	2	111.30	19.18	4.81	4.32	46.51	172.41	81.30	91.11	8980.70
5	3	112.34	20.02	5.44	4.84	46.04	161.29	84.75	76.54	9611.59
	4	110.29	27.92	6.75	6.12	50.52	166.67	47.85	118.82	9532.09
	5	112.44	24.89	5.88	5.23	52.23	158.73	47.39	111.34	10119.64
	1	165.98	42.21	9.16	5.52	43.66	256.41	71.94	184.47	9073.06
	2	165.37	37.07	8.73	5.28	42.66	285.71	98.04	187.68	8232.82
6	3	169.01	41.91	9.86	5.84	45.49	277.78	90.09	187.69	8986.43
	4	168.65	39.11	10.43	6.18	40.62	263.16	90.09	173.07	8412.08
	5	161.90	38.33	9.36	5.78	43.60	270.27	64.94	205.34	8372.09
	1	115.97	24.94	4.52	3.90	45.13	192.31	88.50	103.81	8982.04
	2	118.82	26.35	4.84	4.08	47.23	185.19	89.29	95.90	8893.48
7	3	115.74	24.03	5.58	4.82	45.87	158.73	75.76	82.97	9061.41
	4	120.14	27.20	5.65	4.70	48.72	178.57	91.74	86.83	9427.97
	5	116.20	21.59	4.61	3.97	48.05	172.41	89.29	83.13	9538.86
	1	120.21	37.86	5.95	4.95	47.92	212.77	91.74	121.02	9418.07
	2	117.39	35.26	5.51	4.69	48.07	200.00	52.63	147.37	8961.88
8	3	120.33	35.35	5.78	4.81	46.55	188.68	73.53	115.15	8968.28
	4	120.33	38.26	6.34	5.27	47.73	250.00	75.76	174.24	9348.98
	5	114.78	34.96	6.26	5.45	47.14	188.68	44.44	144.23	9310.05
	1	112.60	27.63	4.66	4.14	50.84	169.49	84.03	85.46	8916.34
	2	112.49	25.69	5.15	4.58	49.98	169.49	78.74	90.75	9965.02
9	3	111.21	22.89	4.77	4.29	46.76	185.19	86.21	98.98	9126.73
	4	114.17	24.45	4.22	3.70	48.39	161.29	89.29	72.00	9202.25
	5	108.19	20.40	4.59	4.24	52.75	161.29	78.12	83.17	10054.63
	1	151.49	42.67	7.30	4.82	45.72	263.16	112.36	150.80	9392.55
	2	150.60	46.24	6.64	4.41	47.69	277.78	114.94	162.84	9158.49
10	3	156.42	45.91	8.83	5.65	51.99	344.83	107.53	237.30	10142.53
	4	156.95	40.71	7.16	4.56	45.83	263.16	119.05	144.11	9048.48
	5	161.80	37.00	7.27	4.49	46.02	232.56	117.65	114.91	9215.12

APPENDIX F

A SAMPLE OUTPUT: PARTIAL COMPUTER PRINT-OUT
OF VOICE IDENTIFICATION OPERATION

(Cross-transmission by the composite of FFC and LTAS)

APPENDIX F

A SAMPLE OUTPUT: PARTIAL COMPUTER PRINT-OUT OF VOICE IDENTIFICATION OPERATION

(Cross-transmission by the composite of FFC and LTAS)

```
RUN MAKRO3
DEBUG TY
PROGRAM MAKRO3:  FOR EXPERIMENT III.

UNKNOWN MASTER FILE IN FREQ.(1440)   /DKO:TELE50.FIL
KNOWN MASTER FILE IN FREQ.(1440)     DKO:NORM50.FIL
EXT NAME OF FREQ. FILE (FRI OR FRL) =FRL

UNKNOWN MASTER FILE FROM TIME(1440)  /DKO:TELE50.FIL
KNOWN MASTER FILE FROM TIME(1440)    /DKO:NORM50.FIL

PROVIDE INFO. FOR FREQUENCY FEATURES.
Feature selection:  Respond by typing 1 or 0 .
Feature # 1 ?1
Feature # 2 ?1
Feature # 3 ?1
Feature # 4 ?1
Feature # 5 ?1
Feature # 6 ?1
Feature # 7 ?1
Feature # 8 ?1
Feature # 9 ?1
Feature #10 ?1

PROVIDE INFO. FOR TIME FEATURES.
Feature selection:  Respond by typing 1 or 0 .
Feature # 1 ?1
Feature # 2 ?1
Feature # 3 ?1
Feature # 4 ?1
Feature # 5 ?1
Feature # 6 ?1
      :                     :
NORMALIZATION BY Z-TRANSFORM.
NORMALIZATION BY Z-TRANSFORM.
NORMALIZATION BY Z-TRANSFORM.
NORMALIZATION BY Z-TRANSFORM.

EUCLIDIAN DISTANCES BETWEEN PATTERN 1 OF UNKNOWN SPEAKER  1
AND PATTERN I OF KNOWN SPEAKER J.

 3.79441(1, 1)   4.91311(2, 1)   4.20709(3, 1)   5.00701(4, 1)   3.93085(5, 1)
 4.08275(1, 2)   6.34234(2, 2)   7.20040(3, 2)   6.39060(4, 2)   6.49731(5, 2)
 5.11102(1, 3)   5.40975(2, 3)   5.07131(3, 3)   5.54474(4, 3)   5.91521(5, 3)
 4.82182(1, 4)   5.00284(2, 4)   5.09771(3, 4)   9.00474(4, 4)   4.96042(5, 4)
 3.65242(1, 5)   3.78518(2, 5)   4.06050(3, 5)   7.18745(4, 5)   4.16289(5, 5)
 7.17218(1, 6)   5.00491(2, 6)   6.44472(3, 6)   6.60467(4, 6)   6.22737(5, 6)
 6.62104(1, 7)   6.67190(2, 7)   7.14393(3, 7)   5.67001(4, 7)   6.04476(5, 7)
 4.94539(1, 8)   4.71212(2, 8)   4.75529(3, 8)   4.75444(4, 8)   6.60449(5, 8)
 5.34724(1, 9)   4.79603(2, 9)   4.81142(3, 9)   4.02743(4, 9)   4.39234(5, 9)
 6.69952(1,10)   6.00855(2,10)   5.97150(3,10)   8.14595(4,10)   8.17440(5,10)

PATTERN 1 OF UNKNOWN SPEAKER  1 IS IDENTIFIED WITH
PATTERN 1 OF   KNOWN SPEAKER  5
DISTANCE =   3.65242

EUCLIDIAN DISTANCES BETWEEN PATTERN 2 OF UNKNOWN SPEAKER  1
AND PATTERN I OF KNOWN SPEAKER J.

 3.09505(1, 1)   3.90831(2, 1)   3.74300(3, 1)   4.30422(4, 1)   4.09332(5, 1)
 4.40738(1, 2)   5.07721(2, 2)   5.90785(3, 2)   6.10325(4, 2)   6.12070(5, 2)
 5.40659(1, 3)   5.38422(2, 3)   5.79901(3, 3)   6.81640(4, 3)   5.08064(5, 3)
 5.23583(1, 4)   6.40207(2, 4)   5.39993(3, 4)   9.18440(4, 4)   6.08484(5, 4)
 4.05038(1, 5)   4.70290(2, 5)   4.97050(3, 5)   7.66373(4, 5)   6.53745(5, 5)
 7.39954(1, 6)   5.20401(2, 6)   6.63019(3, 6)   7.21497(4, 6)   6.05267(5, 6)
      :                            :

      :                            :
 6.91821(1, 5)   6.84329(2, 5)   7.19021(3, 5)   6.02563(4, 5)   6.94033(5, 5)
 5.74309(1, 6)   4.84404(2, 6)   3.92109(3, 6)   4.99420(4, 6)   4.99353(5, 6)
 6.37884(1, 7)   6.86807(2, 7)   7.70874(3, 7)   7.00920(4, 7)   7.07673(5, 7)
 4.27109(1, 8)   4.84714(2, 8)   4.03932(3, 8)   4.39920(4, 8)   5.09700(5, 8)
 6.77076(1, 9)   5.70404(2, 9)   7.32020(3, 9)   6.21750(4, 9)   6.71193(5, 9)
 2.78854(1,10)   3.84010(2,10)   3.43297(3,10)   2.78801(4,10)   3.30630(5,10)

PATTERN 4 OF UNKNOWN SPEAKER 10 IS IDENTIFIED WITH
PATTERN 4 OF   KNOWN SPEAKER 10
DISTANCE =   2.78801

EUCLIDIAN DISTANCES BETWEEN PATTERN 5 OF UNKNOWN SPEAKER 10
AND PATTERN I OF KNOWN SPEAKER J.

 6.47683(1, 1)   6.61380(2, 1)   7.05059(3, 1)   7.07447(4, 1)   6.31944(5, 1)
 7.70304(1, 2)   8.69603(2, 2)   9.72373(3, 2)   9.10203(4, 2)   8.98541(5, 2)
 6.09797(1, 3)   7.25701(2, 3)   6.70694(3, 3)   7.14241(4, 3)   6.04320(5, 3)
 5.50329(1, 4)   6.77010(2, 4)   6.56151(3, 4)   9.00479(4, 4)   6.39035(5, 4)
 6.67975(1, 5)   7.10395(2, 5)   7.30064(3, 5)   6.43834(4, 5)   7.13043(5, 5)
 5.78226(1, 6)   4.00742(2, 6)   4.39452(3, 6)   5.17126(4, 6)   4.76437(5, 6)
 4.75124(1, 7)   6.63103(2, 7)   7.03900(3, 7)   6.03236(4, 7)   4.90035(5, 7)
 4.12004(1, 8)   5.42022(2, 8)   4.61390(3, 8)   4.49373(4, 8)   5.07797(5, 8)
 6.90760(1, 9)   4.61977(2, 9)   6.96182(3, 9)   6.15144(4, 9)   6.45915(5, 9)
 3.62464(1,10)   3.43430(2,10)   4.29714(3,10)   3.52392(4,10)   3.63004(5,10)

PATTERN 5 OF UNKNOWN SPEAKER 10 IS IDENTIFIED WITH
PATTERN 2 OF   KNOWN SPEAKER 10
DISTANCE =   3.43430

Summary by Nearest Neighbor Rule:

Correct = 46,   Incorrect =  4,   Rate(%) = 92.00

Summary by Minimum Set Distance Rule:
Unknown Speaker   Known Speaker        Set Distance
---------------------------------------------------
       1                1                 4.46974
       2                2                 4.40411
       3                3                 3.90844
       4                4                 3.35491
       5                5                 4.09082
       6                6                 3.21302
       7                7                 3.40836
       8                8                 3.74909
       9                9                 3.33932
      10               10                 3.53943
---------------------------------------------------

Correct =10,   Incorrect = 0,   Rate(%) =100.00
```

APPENDIX G

VOICE IDENTIFICATION RESULTS: OPERATIONS 1 THROUGH 24

APPENDIX G(1)

Unknown speakers by telephone IDS; known speakers by normal IDS

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Identification decision Result	by the minimum set Dist. with(j)	Result
1	1	1.7777 (3,9)	wrong		
	2	1.5292 (4,1)	correct		
	3	1.8175 (5,9)	wrong	2.3586 (1)	correct
	4	1.0640 (3,5)	wrong		
	5	1.6315 (3,5)	wrong		
2	1	1.8673 (1,2)	correct		
	2	2.1925 (3,2)	correct		
	3	1.9993 (3,2)	correct	3.4640 (2)	correct
	4	1.6759 (3,2)	correct		
	5	1.6037 (3,2)	correct		
3	1	1.4836 (1,5)	wrong		
	2	1.2609 (1,5)	wrong		
	3	1.1477 (1,5)	wrong	2.6344 (1)	wrong
	4	1.5970 (1,1)	wrong		
	5	2.0373 (5,3)	correct		
4	1	1.4600 (1,4)	correct		
	2	1.8592 (1,4)	correct		
	3	1.6361 (2,4)	correct	2.2995 (4)	correct
	4	2.6929 (5,8)	wrong		
	5	1.9339 (1,5)	wrong		
5	1	1.5335 (2,3)	wrong		
	2	1.9991 (4,8)	wrong		
	3	1.5074 (5,1)	wrong	2.0586 (5)	correct
	4	1.3402 (2,3)	wrong		
	5	1.2415 (5,5)	correct		
6	1	1.5413 (1,6)	correct		
	2	2.2373 (5,1)	wrong		
	3	2.2056 (1,8)	wrong	2.8626 (6)	correct
	4	2.1080 (1,6)	correct		
	5	1.5771 (1,10)	wrong		
7	1	1.7083 (3,5)	wrong		
	2	1.6675 (4,4)	wrong		
	3	1.9032 (4,7)	correct	2.6385 (7)	correct
	4	1.6261 (5,7)	correct		
	5	1.6496 (3,7)	correct		
8	1	1.8931 (5,8)	correct		
	2	0.9828 (3,8)	correct		
	3	1.8621 (3,5)	wrong	2.8592 (4)	wrong
	4	1.8286 (4,5)	wrong		
	5	1.7547 (3,8)	correct		
9	1	2.0119 (5,7)	wrong		
	2	1.7101 (2,9)	correct		
	3	2.4109 (3,7)	wrong	2.5586 (1)	wrong
	4	1.2693 (5,9)	correct		
	5	1.7384 (4,3)	wrong		
10	1	1.7599 (3,10)	correct		
	2	1.3788 (4,10)	correct		
	3	1.9221 (4,10)	correct	2.8297 (10)	correct
	4	1.7877 (5,10)	correct		
	5	1.9251 (5,10)	correct		

i = sample index — Number of correct I.D. = 26 — = 7

j = speaker index — Rate = 52 % — = 70 %

APPENDIX G(2)

Unknown speakers by telephone IDS; known speakers by linear IDS.

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	1.6555 (4,3)	wrong		
	2	2.1921 (2,3)	wrong		
	3	1.8780 (2,1)	correct	2.7161 (1)	correct
	4	1.7904 (5,9)	wrong		
	5	1.7447 (2,4)	wrong		
2	1	1.3595 (5,2)	correct		
	2	1.6832 (1,2)	correct		
	3	2.1007 (4,2)	correct	2.6780 (2)	correct
	4	2.3101 (4,2)	correct		
	5	1.9775 (4,2)	correct		
3	1	2.1425 (5,1)	wrong		
	2	1.8025 (4,5)	wrong		
	3	1.6916 (5,3)	correct	2.4885 (3)	correct
	4	1.9905 (2,1)	wrong		
	5	2.3532 (5,3)	correct		
4	1	1.4871 (1,4)	correct		
	2	1.7267 (2,4)	correct		
	3	1.9639 (5,7)	wrong	2.6164 (4)	correct
	4	2.2227 (3,5)	wrong		
	5	1.0971 (3,5)	wrong		
5	1	1.5373 (1,3)	wrong		
	2	2.1961 (5,8)	wrong		
	3	1.6682 (4,3)	wrong	2.3964 (3)	wrong
	4	1.6418 (4,9)	wrong		
	5	1.7190 (3,9)	wrong		
6	1	1.7451 (1,6)	correct		
	2	1.6678 (4,10)	wrong		
	3	1.3089 (1,6)	correct	2.7962 (8)	wrong
	4	1.9476 (4,8)	wrong		
	5	1.9751 (2,10)	wrong		
7	1	1.4434 (5,7)	correct		
	2	1.6036 (2,7)	correct		
	3	1.4319 (3,7)	correct	3.0197 (7)	correct
	4	1.3947 (2,7)	correct		
	5	1.8256 (4,7)	correct		
8	1	1.6241 (1,4)	wrong		
	2	1.7658 (3,6)	wrong		
	3	1.2124 (3,8)	correct	2.7587 (4)	wrong
	4	2.1096 (5,8)	correct		
	5	1.2452 (1,4)	wrong		
9	1	2.1398 (2,9)	correct		
	2	1.7909 (5,1)	wrong		
	3	2.3345 (5,9)	corrent	2.8691 (1)	wrong
	4	1.3858 (5,1)	wrong		
	5	1.9867 (5,9)	correct		
10	1	1.4806 (5,10)	correct		
	2	1.9687 (1,10)	correct		
	3	1.7139 (2,10)	correct	2.4750 (10)	correct
	4	1.5606 (5,10)	correct		
	5	1.8269 (5,10)	correct		

i = sample index Number of correct I.D. = 28 = 6

j = speaker index Rate = 56 % = 60 %

APPENDIX G (3)
Unknown speakers by telephone LTAS known speakers by normal LTAS

Unknown Speaker (j)	Sample (i)	Known: Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	2.3676 (3,6)	wrong		
	2	2.4635 (1,1)	correct		
	3	2.2858 (5,5)	wrong	3.0963 (6)	wrong
	4	2.2326 (1,1)	correct		
	5	2.6622 (1,5)	wrong		
2	1	3.5451 (2,7)	wrong		
	2	2.2639 (4,2)	correct		
	3	3.5966 (3,2)	correct	4.0967 (2)	correct
	4	3.1266 (4,2)	correct		
	5	3.1417 (4,2)	correct		
3	1	1.9319 (3,3)	correct		
	2	1.7198 (3,9)	wrong		
	3	2.3838 (3,3)	correct	2.4546 (3)	correct
	4	2.3213 (4,3)	correct		
	5	2.2562 (3,4)	wrong		
4	1	2.5638 (1,4)	correct		
	2	2.1204 (5,4)	correct		
	3	2.3327 (1.4)	correct	2.9174 (4)	correct
	4	2.4394 (1,4)	correct		
	5	2.8335 (4,5)	wrong		
5	1	1.3061 (1,5)	correct		
	2	2.1310 (2,5)	correct		
	3	2.1986 (1,5)	correct	2.5603 (5)	correct
	4	2.2406 (4,7)	wrong		
	5	2.2180 (5,5)	correct		
6	1	2.2113 (1,6)	correct		
	2	1.7672 (1,6)	correct		
	3	1.4844 (1,6)	correct	2.5684 (6)	correct
	4	2.2084 (1,6)	correct		
	5	2.3095 (1.6)	correct		
7	1	1.8280 (3,3)	wrong		
	2	2.0575 (4,7)	correct		
	3	2.1566 (3,9)	wrong	2.5092 (3)	wrong
	4	2.2262 (4.7)	correct		
	5	1.5700 (5,5)	wrong		
8	1	1.1033 (2,8)	correct		
	2	1.2544 (2,8)	correct		
	3	1.3869 (4,8)	correct	2.0834 (8)	correct
	4	1.2812 (4,8)	correct		
	5	1.3601 (1,8)	correct		
9	1	2.4932 (3,10)	wrong		
	2	1.9309 (5,9)	correct		
	3	1.9735 (5,9)	correct	2.8942 (9)	correct
	4	2.7607 (3,4)	wrong		
	5	1.9386 (4,9)	correct		
10	1	2.5930 (4,8)	wrong		
	2	1.4026 (1,10)	correct		
	3	2.1133 (1,10)	correct	2.6961 (8)	wrong
	4	2.2182 (4,10)	correct		
	5	2.0961 (1,8)	wrong		

i = sample index Number of correct I.D. = 35 = 7
j = speaker index Rate = 70 % = 70 %

APPENDIX G(4)

Unknown speakers by telephone LTAS known speakers by linear LTAS

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	2.3288 (1,5)	wrong		
	2	2.7100 (1,7)	wrong		
	3	2.4538 (5,5)	wrong	3.1008 (1)	correct
	4	2.7360 (3,1)	correct		
	5	2.5760 (1,5)	wrong		
2	1	3.2513 (1,7)	wrong		
	2	2.2536 (4,2)	correct		
	3	4.4801 (2,2)	correct	3.6387 (7)	wrong
	4	3.6024 (4,2)	correct		
	5	3.3782 (4,2)	correct		
3	1	1.5379 (3,3)	correct		
	2	2.0567 (2,3)	correct		
	3	2.3577 (3,3)	correct	2.0171 (3)	correct
	4	1.7694 (3,3)	correct		
	5	1.5903 (2,3)	correct		
4	1	2.3857 (4,7)	wrong		
	2	2.3144 (1,4)	correct		
	3	2.0629 (4,7)	wrong	3.1618 (4)	correct
	4	2.1618 (3,5)	wrong		
	5	2.2359 (4,5)	wrong		
5	1	1.3720 (1,5)	correct		
	2	1.7971 (4,5)	correct		
	3	1.9521 (5,5)	correct	2.5686 (5)	correct
	4	1.9209 (4,5)	correct		
	5	2.0852 (5,5)	correct		
6	1	2.1020 (2,6)	correct		
	2	2.0708 (2,6)	correct		
	3	1.5473 (1,6)	correct	2.4420 (6)	correct
	4	1.3797 (2,6)	correct		
	5	1.9305 (2,6)	correct		
7	1	2.1016 (1,3)	wrong		
	2	2.4252 (5,7)	correct		
	3	2.6531 (4,7)	correct	2.6905 (3)	wrong
	4	2.8138 (4,3)	wrong		
	5	2.8822 (5,7)	correct		
8	1	2.0032 (3,8)	correct		
	2	1.7706 (3,8)	correct		
	3	2.2102 (3,8)	correct	2.4422 (8)	correct
	4	2.1054 (3,8)	correct		
	5	1.7313 (5,8)	correct		
9	1	1.9003 (1,9)	correct		
	2	1.7539 (3,9)	correct		
	3	1.7371 (3,9)	correct	2.5414 (9)	correct
	4	2.2954 (2,6)	wrong		
	5	2.2359 (3,9)	correct		
10	1	2.2030 (4,8)	wrong		
	2	1.9102 (3,8)	wrong		
	3	2.3798 (4,8)	wrong	3.3879 (8)	wrong
	4	2.0649 (3,8)	wrong		
	5	1.9134 (4,8)	wrong		

i = sample index Number of correct I.D. = 33 = 7

j = speaker index Rate = 66 % = 70 %

APPENDIX G(5)

Both unknown and known speakers by telephone IDS.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	Identification decision by the minimum set Dist. with(j)	Result
1	1	1.6114 (5,8)	wrong		
	2	1.5670 (1,3)	wrong		
	3	1.6030 (4,9)	wrong	2.3475 (9)	wrong
	4	1.3874 (5,9)	wrong		
	5	2.6277 (2,7)	wrong		
2	1	3.1701 (2,7)	wrong		
	2	2.4383 (3,2)	correct		
	3	2.1099 (4,2)	correct	3.1701 (2)	correct
	4	1.9481 (5,2)	correct		
	5	1.9378 (4,2)	correct		
3	1	1.5671 (2,1)	wrong		
	2	1.5846 (3,3)	correct		
	3	1.4708 (2,3)	correct	2.4615 (3)	correct
	4	1.9546 (5,9)	wrong		
	5	1.7956 (3,3)	correct		
4	1	1.3381 (3,4)	correct		
	2	1.7394 (1,4)	correct		
	3	1.3125 (1,4)	correct	2.3490 (8)	wrong
	4	2.2637 (5,4)	correct		
	5	1.7677 (3,5)	wrong		
5	1	0.9872 (4,5)	correct		
	2	1.9600 (2,4)	wrong		
	3	1.4251 (1,5)	correct	2.4753 (1)	wrong
	4	0.9874 (1,5)	correct		
	5	1.6024 (4,5)	correct		
6	1	1.7376 (4,6)	correct		
	2	1.8982 (1,1)	wrong		
	3	1.8080 (4,6)	correct	2.5765 (10)	wrong
	4	1.8392 (1,6)	correct		
	5	2.3214 (1,10)	wrong		
7	1	1.9659 (4,1)	wrong		
	2	1.6277 (5,1)	wrong		
	3	1.8614 (4,7)	correct	2.4696 (7)	correct
	4	1.7286 (2,7)	correct		
	5	1.7756 (5,9)	wrong		
8	1	1.9961 (1,4)	wrong		
	2	1.7430 (5,8)	correct		
	3	1.8135 (1,4)	wrong	2.3234 (8)	correct
	4	1.3860 (1,4)	wrong		
	5	1.6114 (1,1)	wrong		
9	1	1.8976 (3,9)	correct		
	2	1.6164 (4,9)	correct		
	3	1.8860 (1,9)	correct	2.3502 (9)	correct
	4	1.5055 (4,5)	wrong		
	5	1.3874 (4,1)	wrong		
10	1	1.7663 (4,10)	correct		
	2	2.0853 (3,10)	correct		
	3	1.9005 (2,6)	wrong	2.3203 (10)	correct
	4	1.6498 (5,10)	correct		
	5	1.0461 (4,10)	correct		

i = sample index Number of correct I.D. = 29 = 6
j = speaker index Rate = 58 % = 60 %

APPENDIX G(6)

Both unknown and known speakers by normal IDS.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	1.5889 (5,1)	correct		
	2	1.7836 (1,5)	wrong		
	3	1.4125 (4,1)	correct		
	4	1.5359 (3,1)	correct		
	5	1.7123 (1,1)	correct	2,4946 (3)	wrong
2	1	2.3320 (5,2)	correct		
	2	2.6436 (3,2)	correct		
	3	2.0921 (5,2)	correct		
	4	2.5106 (3,2)	correct		
	5	2.2155 (3,2)	correct	3.1067 (2)	correct
3	1	1.3037 (2,5)	wrong		
	2	1.2637 (1,5)	wrong		
	3	1.7365 (1,3)	correct		
	4	1.7695 (1,1)	wrong		
	5	1.8087 (2,1)	wrong	2.4236 (4)	wrong
4	1	1.8115 (2,5)	wrong		
	2	1.6386 (3,4)	correct		
	3	1.7621 (2,4)	correct		
	4	2.0575 (3,4)	correct		
	5	2.1941 (3,5)	wrong	2.3642 (4)	correct
5	1	1.2637 (2,3)	wrong		
	2	1.3037 (1,3)	wrong		
	3	1.6134 (1,9)	wrong		
	4	1.7977 (2,5)	correct		
	5	1.7198 (1,5)	correct	1.0858 (5)	correct
6	1	2.4533 (2,6)	correct		
	2	2.3671 (2,8)	wrong		
	3	2.5581 (4,6)	correct		
	4	2.5285 (5,6)	correct		
	5	2.6519 (4,6)	correct	2.8170 (6)	correct
7	1	2.0259 (3,7)	correct		
	2	2.5177 (3,7)	correct		
	3	2.0802 (1,7)	correct		
	4	2.0575 (5,7)	correct		
	5	2.1809 (4,7)	correct	3.1426 (7)	correct
8	1	2.4965 (2,8)	correct		
	2	1.8168 (4,3)	wrong		
	3	1.8904 (3,9)	wrong		
	4	2.2565 (3,4)	wrong		
	5	2.1575 (3,3)	wrong	3.2555 (9)	wrong
9	1	1.6134 (3,5)	wrong		
	2	2.2370 (3,9)	correct		
	3	1.7958 (3,1)	wrong		
	4	2.0560 (5,9)	correct		
	5	2.0016 (4,9)	correct	2.3946 (9)	correct
10	1	1.8376 (5,10)	correct		
	2	1.9327 (4,10)	correct		
	3	2.5047 (2,6)	wrong		
	4	1.9870 (2,10)	correct		
	5	2.0426 (1,10)	correct	2.6479 (10)	correct

i = sample index Number of correct I.D. = 32 = 7

j = speaker index Rate = 64 % = 70 %

APPENDIX G (7)

Both unknown and known speakers by linear IDS.

Speaker (j)	Sample (i)	Identification decision: by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	2.0848 (1,4)	wrong		
	2	2.2532 (5,1)	correct		
	3	1.8320 (1,3)	wrong		
	4	2.2975 (3,1)	correct		
	5	2.0145 (2,9)	wrong	2.8420 (1)	correct
2	1	1.8041 (3,2)	correct		
	2	1.6173 (3,2)	correct		
	3	1.6716 (2,2)	correct		
	4	1.8506 (2,2)	correct		
	5	2.6898 (4,2)	correct	2.8871 (2)	correct
3	1	1.8320 (3,1)	wrong		
	2	2.1846 (1,3)	correct		
	3	1.9999 (1,3)	correct		
	4	1.9483 (1,5)	wrong		
	5	2.2215 (5,5)	wrong	2.5094 (1)	wrong
4	1	1.8429 (3,5)	wrong		
	2	1.6622 (2,5)	wrong		
	3	1.2410 (4,4)	correct		
	4	1.4460 (3,4)	correct		
	5	2.0644 (3,4)	correct	2.4883 (4)	correct
5	1	1.8700 (1,3)	wrong		
	2	1.6622 (2,4)	wrong		
	3	1.8429 (1.4)	wrong		
	4	1.7624 (3,9)	wrong		
	5	1.9707 (4,5)	correct	2.4894 (5)	correct
6	1	2.3903 (4,6)	correct		
	2	2.1771 (5,6)	correct		
	3	2.2580 (1,8)	wrong		
	4	2.0699 (5,6)	correct		
	5	2.0536 (2,6)	correct	2.7320 (6)	correct
7	1	2.1139 (3,7)	correct		
	2	1.6166 (5,7)	correct		
	3	2.2373 (1,7)	correct		
	4	2.4234 (2,7)	correct		
	5	1.8216 (2,7)	correct	2.7158 (5)	wrong
8	1	1.8702 (3,8)	correct		
	2	2.0453 (4,8)	correct		
	3	1.8159 (1,8)	correct		
	4	1.8968 (5,8)	correct		
	5	1.6918 (4,8)	correct	2.3813 (8)	correct
9	1	2.8487 (5,9)	correct		
	2	1.9816 (4,9)	correct		
	3	1.6758 (3,4)	wrong		
	4	1.8581 (2,9)	correct		
	5	1.7386 (2,4)	wrong	2.6370 (4)	wrong
10	1	2.0094 (5,10)	correct		
	2	1.6545 (4,10)	correct		
	3	1.7508 (5,10)	correct		
	4	1.7779 (2,10)	correct		
	5	1.6576 (2,10)	correct	2.0094 (10)	correct

i = sample index   Correct rate = 70 %   = 70 %

j = speaker index

APPENDIX G (8)
Both unknown and known speakers by telephone LTAS

Unknown Speaker (j)	Sample (i)	Known — Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	1.6922 (5,2)	Correct		
	2	1.2325 (3,1)	correct		
	3	1.2868 (2,1)	correct		
	4	2.4086 (2,1)	correct		
	5	1.5688 (1,1)	correct	2.8117 (1)	correct
2	1	2.6566 (4,2)	correct		
	2	2.0052 (5,2)	correct		
	3	2.5504 (5,2)	correct		
	4	1.6415 (5,2)	correct		
	5	1.8465 (4,2)	correct	2.8150 (2)	correct
3	1	1.2647 (4,3)	correct		
	2	2.0079 (5,3)	correct		
	3	1.8817 (4,3)	correct		
	4	1.1413 (1,3)	correct		
	5	1.5908 (1,3)	correct	2.2982 (3)	correct
4	1	0.6117 (3,4)	correct		
	2	1.9871 (4,4)	correct		
	3	0.6660 (1,4)	correct		
	4	1.4275 (3,4)	correct		
	5	1.9974 (4,4)	correct	1.9974 (4)	correct
5	1	1.9415 (2,5)	correct		
	2	1.7755 (4,5)	correct		
	3	1.4398 (5,5)	correct		
	4	1.7212 (2,5)	correct		
	5	1.3164 (3,5)	correct	2.4355 (5)	correct
6	1	1.1004 (3,6)	correct		
	2	1.7214 (3,6)	correct		
	3	1.2239 (1,6)	correct		
	4	1.3003 (3,6)	correct		
	5	1.5774 (4,6)	correct	1.8204 (6)	correct
7	1	1.1491 (4,7)	correct		
	2	0.8082 (3,7)	correct		
	3	0.9317 (2,7)	correct		
	4	0.9440 (1,7)	correct		
	5	1.7216 (3,7)	correct	1.7216 (7)	correct
8	1	1.6059 (4,8)	correct		
	2	1.3370 (4,8)	correct		
	3	1.2555 (4,8)	correct		
	4	1.4606 (3,8)	correct		
	5	2.0294 (1,8)	correct	2.0294 (9)	correct
9	1	1.6242 (2,9)	correct		
	2	1.4756 (3,9)	correct		
	3	1.3522 (2,9)	correct		
	4	1.9257 (5,9)	correct		
	5	1.4617 (3,9)	correct	2.2893 (9)	correct
10	1	1.9326 (3,10)	correct		
	2	1.5187 (3,10)	correct		
	3	1.5730 (2,10)	correct		
	4	1.6352 (5,10)	correct		
	5	1.5421 (2,10)	correct	1.9890 (10)	correct

i = sample index   Number of correct I.D. = 50   = 10
j = speaker index   Rate = 100 %   = 100 %

APPENDIX G(9)
Both unknown and known speakers by normal LTAS

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	Identification decision by the minimum set Dist. with(j)	Result
1	1	1.4748 (3,1)	correct		
	2	1.5367 (1,1)	correct		
	3	1.5983 (1,1)	correct		
	4	1.9904 (1,1)	correct		
	5	1.7460 (1,5)	wrong	2.1580 (1)	correct
2	1	2.3024 (4,2)	correct		
	2	2.4877 (5,2)	correct		
	3	1.9504 (4,2)	correct		
	4	1.7453 (3,2)	correct		
	5	2.4345 (1,2)	correct	2.9751 (2)	correct
3	1	0.6638 (5,3)	correct		
	2	1.1126 (5,3)	correct		
	3	0.6853 (1,3)	correct		
	4	1.7203 (1,3)	correct		
	5	1.7181 (1,3)	correct	1.7203 (3)	correct
4	1	0.7979 (4,4)	correct		
	2	1.1831 (5,4)	correct		
	3	0.8601 (1,4)	correct		
	4	0.9213 (1,4)	correct		
	5	1.2374 (2,4)	correct	1.9373 (4)	correct
5	1	1.1742 (2,5)	correct		
	2	1.2977 (1,5)	correct		
	3	1.8747 (2,5)	correct		
	4	1.2829 (2,5)	correct		
	5	2.1601 (3,5)	correct	2.1601 (5)	correct
6	1	1.1445 (3,6)	correct		
	2	1.1783 (3,6)	correct		
	3	1.0902 (1,6)	correct		
	4	1.6034 (1,6)	correct		
	5	1.4621 (1,6)	correct	1.6034 (6)	correct
7	1	1.4135 (2,7)	correct		
	2	1.5370 (1,7)	correct		
	3	1.4139 (2,7)	correct		
	4	1.8349 (5,7)	correct		
	5	1.6299 (4,7)	correct	2.1398 (7)	correct
8	1	1.4018 (3,8)	correct		
	2	1.3827 (3,8)	correct		
	3	1.5878 (2,8)	correct		
	4	1.4099 (3,8)	correct		
	5	1.4210 (3,8)	correct	1.4210(8)	correct
9	1	1.8797 (2,9)	correct		
	2	1.7563 (1,9)	correct		
	3	1.5394 (4,9)	correct		
	4	1.5537 (5,9)	correct		
	5	1.4302 (4,9)	correct	2,4886 (9)	correct
10	1	1.3632 (5,10)	correct		
	2	1.5876 (5,10)	correct		
	3	2.0014 (4,10)	correct		
	4	1.7963 (3,10)	correct		
	5	1,4171 (1,10)	correct	2.3288(10)	correct

i = sample index   Number of correct I.D. = 49   = 10
j = speaker index   Rate = 98 %   = 100 %

APPENDIX G(10)

Both unknown and known speakers by linear LTAS

Unknown Speaker (j)	Sample (i)	Known — Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	Identification decision by the minimum set: Dist. with(j)	Result
1	1	1.6216 (4,1)	correct		
	2	2.5226 (1,1)	correct		
	3	1.1635 (5,1)	correct		
	4	1.4345 (5,1)	correct		
	5	1.0019 (3,1)	correct	2.5994 (1)	correct
2	1	1.7803 (2,2)	correct		
	2	2.0002 (1,2)	correct		
	3	1.9460 (4,2)	correct		
	4	1.7437 (3,2)	correct		
	5	1.8164 (3,2)	correct	2.0713 (2)	correct
3	1.	1.0323 (3,3)	correct		
	2	1.3142 (5,3)	correct		
	3	0.9088 (1,3)	correct		
	4	1.1386 (1,3)	correct		
	5	1.1092 (2,3)	correct	1.5842(3)	correct
4	1	1.8273 (3,4)	correct		
	2	1.6194 (5,4)	correct		
	3	1.2902 (4,4)	correct		
	4	1.1721 (5,4)	correct		
	5	1.2264 (4,4)	correct	1.9545 (4)	correct
5	1	1.1776 (2,5)	correct		
	2	1.1068 (4,5)	correct		
	3	1.7464 (4,5)	correct		
	4	1.3118 (2;5)	correct		
	5	1.9072 (2,8)	wrong	2.2466 (5)	correct
6	1	1.1335 (3,6)	correct		
	2	1.5933 (1,6)	correct		
	3	1.0792 (1,6)	correct		
	4	1.2828 (1,6)	correct		
	5	1.6387 (3,6)	correct	1.7635 (6)	correct
7	1	1.1346 (3,7)	correct		
	2	1.5510 (5,7)	correct		
	3	1.0468 (5,7)	correct		
	4	1.6726 (5,7)	correct		
	5	1.1702 (3,7)	correct	1.6726 (7)	correct
8	1	1.6215 (2,8)	correct		
	2	1.5671 (1,8)	correct		
	3	1.5418 (4,8)	correct		
	4	1.6650 (5,8)	correct		
	5	1.4599 (4,8)	correct	1.9684(8)	correct
8	1	2.0199 (2;9)	correct		
	2	1.3856 (5,9)	correct		
	3	1.3715 (2,9)	correct		
	4	1.7271 (5,9)	correct		
	5	1.2622 (2,9)	correct	2.1987 (9)	correct
9	1	1.8211 (4,10)	correct		
	2	1.7614 (4,10)	correct		
	3	1.9976 (5,10)	correct		
	4	1.9664 (2,10)	correct		
	5	1.9681 (1,10)	correct	2.3926 (10)	correct
i = sample index		Number of correct I.D. = 49			= 10
j = speaker index			Rate = 98 %		= 100 %

APPENDIX G(11)

Unknown speakers by telephone FFC; known speakers by normal FFC.

Unknown Speaker (j)	Sample (i)	Known: Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	1.1428 (2,3)	wrong		
	2	1.3478 (3,3)	wrong		
	3	0.8861 (3,1)	correct	1.6879 (1)	correct
	4	1.2316 (2,3)	wrong		
	5	1.5837 (5,1)	correct		
2	1	1.2170 (1,2)	correct		
	2	1.1113 (4,2)	correct		
	3	1.2383 (5,2)	correct	2.5539 (2)	correct
	4	1.2113 (5,2)	correct		
	5	1.1452 (3,5)	wrong		
3	1	0.8706 (2,8)	wrong		
	2	0.5634 (4,3)	correct		
	3	1.3086 (2,1)	wrong	3.0001 (7)	wrong
	4	1.3544 (5,9)	wrong		
	5	0.9565 (2,1)	wrong		
4	1	1.2479 (1,4)	correct		
	2	1.4083 (4,4)	correct		
	3	0.9067 (4,8)	wrong	2.8700 (8)	wrong
	4	2.3305 (1,4)	correct		
	5	1.4067 (2,7)	wrong		
5	1	0.6568 (3,3)	wrong		
	2	0.7151 (4,3)	wrong		
	3	1.1855 (3,7)	wrong	2.9900 (7)	wrong
	4	1.8224 (5,5)	correct		
	5	1.2930 (5,5)	correct		
6	1	1.9926 (1,6)	correct		
	2	1.5376 (1,6)	correct		
	3	1.5371 (1,6)	correct	2.7445 (6)	correct
	4	1.4852 (1,6)	correct		
	5	2.0072 (1,6)	correct		
7	1	1.0438 (2,4)	wrong		
	2	1.9616 (4,7)	correct		
	3	0.7965 (2,8)	wrong	3.2124 (9)	wrong
	4	0.7700 (3,5)	wrong		
	5	0.7624 (1,7)	correct		
8	1	1.3854 (4,8)	correct		
	2	1.5792 (3,4)	wrong		
	3	1.2232 (3,4)	wrong	1.9745 (8)	correct
	4	1.3309 (5,4)	wrong		
	5	1.6731 (5,3)	wrong		
9	1	1,2714 (2,7)	wrong		
	2	0.5531 (3,9)	correct		
	3	0.9244 (2,4)	wrong	2.0333 (7)	wrong
	4	0.7184 (4,7)	wrong		
	5	1.1367 (3,9)	correct		
10	1	1.2418 (3,10)	correct		
	2	1.2231 (3,10)	correct		
	3	3.1401 (5,6)	wrong	2.9788 (10)	correct
	4	1.4187 (4,10)	correct		
	5	2.2191 (4,10)	correct		

i = sample index Number of correct I.D. = 26 = 5

j = speaker index Rate = 52 % = 50 %

APPENDIX G(12)
Unknown speakers by telephone FFC; known speakers by linear FFC.

Unknown Speaker (j)	Sample (i)	Known: Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	1.7007 (1,1)	correct		
	2	1.0547 (3,9)	wrong		
	3	1.8466 (2,1)	correct	1.9787 (3)	wrong
	4	1.5381 (1,1)	correct		
	5	2.1778 (5,1)	correct		
2	1	1.4248 (2,2)	correct		
	2	1.5386 (4,2)	correct		
	3	0.9873 (1,2)	correct	2.3223 (2)	correct
	4	1.0944 (4,2)	correct		
	5	2.1796 (4,2)	correct		
3	1	0.9494 (2,8)	wrong		
	2	0.8041 (4.7)	wrong		
	3	2.2421 (4,9)	wrong	2.4319 (7)	wrong
	4	0.8285 (5,3)	correct		
	5	1.8538 (5,1)	wrong		
4	1	1.4131 (1,4)	correct		
	2	1.3132 (1,8)	wrong		
	3	0.9596 (5,4)	correct	1.7576 (8)	wrong
	4	2.8226 (2,3)	wrong		
	5	0.9950 (3,8)	wrong		
5	1	0.4113 (3,9)	wrong		
	2	1.0196 (3,7)	wrong		
	3	0.8494 (4,3)	wrong	1.8548 (7)	wrong
	4	1.8840 (1,5)	correct		
	5	1.6649 (3,5)	correct		
6	1	1.2306 (5,6)	correct		
	2	2.2216 (4,6)	correct		
	3	1.2181 (4,6)	correct	3.0534 (6)	correct
	4	1.7259 (1,6)	correct		
	5	2.0479 (2,10)	wrong		
7	1	0.9909 (4,9)	wrong		
	2	0.8129 (4,9)	wrong		
	3	0.9098 (5,8)	wrong	1.7310 (9)	wrong
	4	0.8707 (5,8)	wrong		
	5	0.6740 (5,7)	correct		
8	1	0.9648 (1,8)	wrong		
	2	1.6674 (2,8)	correct		
	3	0.5557 (4,8)	correct	1.8843 (8)	correct
	4	1.1638 (1,8)	correct		
	5	1.6222 (5,5)	wrong		
9	1	1.2716 (1,7)	wrong		
	2	0.8444 (1,9)	correct		
	3	0.7246 (1,7)	wrong	1.8902 (7)	wrong
	4	0.8177 (4,7)	wrong		
	5	1.5685 (3,9)	correct		
10	1	0.3579 (1,10)	correct		
	2	1.4299 (1,10)	correct		
	3	2.8851 (4,10)	correct	3.0185 (10)	correct
	4	1.0524 (3,10)	correct		
	5	1.0439 (3,10)	correct		

i = sample index   Number of correct I.D. = 29   = 4
j = speaker index   Rate = 58 %   = 40 %

APPENDIX G(13)
Both unknown and known speakers by telephone FFC.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Identification decision Result	by the minimum set Dist. with(j)	Result
1	1	1.2070 (4,1)	correct		
	2	1.1192 (1,5)	wrong		
	3	1.8451 (1,3)	wrong		
	4	1.3057 (1,1)	correct		
	5	1.9226 (5,3)	wrong	2.3325 (9)	wrong
2	1	1.6466 (4,3)	wrong		
	2	1.7131 (3,2)	correct		
	3	1.3898 (2,1)	wrong		
	4	1.8547 (5,2)	correct		
	5	1.5961 (4,2)	correct	1.9433 (2)	correct
3	1	1.1875 (3,7)	wrong		
	2	1.0107 (2,5)	wrong		
	3	1.4440 (5,3)	correct		
	4	1.6466 (1,2)	wrong		
	5	1.2807 (1,4)	wrong	2.3315 (7)	wrong
4	1	1.3795 (5,3)	wrong		
	2	1.5756 (5,8)	wrong		
	3	1.0004 (4,7)	wrong		
	4	1.6816 (1,4)	correct		
	5	1.1115 (2,7)	wrong	2.0275 (8)	wrong
5	1	0.8933 (5,7)	wrong		
	2	0.6366 (3,9)	wrong		
	3	1.2392 (3,7)	wrong		
	4	1.9150 (5,8)	wrong		
	5	1.9986 (4,5)	correct	1.7363 (7)	wrong
6	1	1.3464 (3,6)	correct		
	2	1.9976 (5,6)	correct		
	3	1.6476 (1,6)	correct		
	4	1.9962 (5,6)	correct		
	5	1.6168 (1,6)	correct	1.9976 (6)	correct
7	1	0.8063 (2,7)	correct		
	2	0.7863 (3,9)	wrong		
	3	1.1875 (1,3)	wrong		
	4	1.0004 (3,4)	wrong		
	5	0.8933 (1,5)	wrong	1.6709 (7)	correct
8	1	1.2497 (3,4)	wrong		
	2	1.4292 (3,8)	correct		
	3	1.5328 (2,8)	correct		
	4	1.7872 (1,8)	correct		
	5	1.5636 (2,8)	correct	2.2270 (8)	correct
9	1	1.2398 (2,7)	wrong		
	2	1.5109 (5,9)	correct		
	3	0.6366 (2,5)	wrong		
	4	0.9424 (5,7)	wrong		
	5	1.2523 (2,9)	correct	1.4046 (7)	wrong
10	1	1.0207 (4,10)	correct		
	2	1.1375 (4,10)	correct		
	3	3.6933 (3,6)	wrong		
	4	0.9220 (1,10)	correct		
	5	1.0978 (4,10)	correct	3.9223 (7)	wrong
i = sample index		Number of correct I.D. = 24			= 4
j = speaker index			Rate = 48 %		= 40 %

APPENDIX G(14)

Both unknown and known speakers by normal FFC.

Unknown Speaker (j)	Sample (i)	Known / Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	0.8283 (2,3)	wrong		
	2	0.9473 (4,5)	wrong		
	3	1.4701 (2,5)	wrong	2.3165 (5)	wrong
	4	1.2608 (2,3)	wrong		
	5	1.3943 (2,3)	wrong		
2	1	1.8760 (5,2)	correct		
	2	1.1495 (1,5)	wrong		
	3	2.1275 (5,2)	correct	2.1275 (2)	correct
	4	1.2417 (5,2)	correct		
	5	1.1429 (4,2)	correct		
3	1	2.0839 (5,3)	correct		
	2	0.8283 (1,1)	wrong		
	3	0.7673 (4,3)	correct	1.2440 (7)	wrong
	4	0.6685 (3,3)	correct		
	5	1.7827 (1,3)	correct		
4	1	1.1918 (3,4)	correct		
	2	0.6282 (4,7)	wrong		
	3	1.2954 (1,4)	correct	1.4211 (7)	wrong
	4	1.4253 (2,8)	wrong		
	5	1.3122 (3,4)	correct		
5	1	1.1495 (2,2)	wrong		
	2	1.4701 (3,1)	wrong		
	3	1.4735 (5,2)	wrong	2.1218 (7)	wrong
	4	0.9473 (2,1)	wrong		
	5	2.2079 (2,2)	wrong		
6	1	1.2855 (3,6)	correct		
	2	1.7038 (4,6)	correct		
	3	0.8738 (4,6)	correct	1.8526 (6)	correct
	4	1.1382 (3,6)	correct		
	5	1.7863 (4,10)	wrong		
7	1	0.6297 (2,7)	correct		
	2	0.5310 (1,7)	correct		
	3	0.9912 (2,7)	correct	1.2240 (7)	correct
	4	0.6257 (1,7)	correct		
	5	1.1836 (3,3)	wrong		
8	1	1.8702 (3,5)	wrong		
	2	1.0018 (2,3)	wrong		
	3	1.6076 (1,4)	wrong	1.8001 (1)	wrong
	4	1.4763 (3,7)	wrong		
	5	1.2645 (3,9)	wrong		
9	1	1.3790 (4,7)	wrong		
	2	0.9993 (2,7)	wrong		
	3	1.2645 (5,8)	wrong	1.8749 (7)	wrong
	4	1.5850 (2,7)	wrong		
	5	1.3876 (2,3)	wrong		
10	1	0.9932 (3,10)	correct		
	2	2.5705 (4,10)	correct		
	3	0.7290 (4,10)	correct	25705 (10)	correct
	4	1.0303 (3,10)	correct		
	5	1.4040 (1,10)	correct		

i = sample index   Number of correct I.D. = 24   = 4

j = speaker index   Rate = 48 %   = 40 %

APPENDIX G(15)

Both unknown and known speakers by linear FFC.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	Identification decision by the minimum set Dist. with(j)	Result
1	1	1.0832 (4,1)	correct		
	2	1.2584 (1,1)	correct		
	3	1.0433 (4,1)	correct	1.9375 (5)	wrong
	4	0.9005 (5,3)	wrong		
	5	1.1143 (1,1)	correct		
2	1	0.7391 (2,2)	correct		
	2	0.8428 (1,2)	correct		
	3	0.8131 (1,2)	correct	1.8741 (2)	correct
	4	1.6192 (2,2)	correct		
	5	1.2699 (3,2)	correct		
3	1	0.8031 (5,9)	wrong		
	2	4.2353 (5,1)	wrong		
	3	0.8754 (2,9)	wrong	1.6365 (9)	wrong
	4	0.6028 (5,8)	wrong		
	5	0.9005 (4,1)	wrong		
4	1	0.9243 (5,4)	correct		
	2	1.3800 (2,7)	wrong		
	3	1.5194 (5,7)	wrong	1.7212 (8)	wrong
	4	7.2062 (2,8)	wrong		
	5	1.2256 (1,4)	correct		
5	1	0.8874 (5,5)	correct		
	2	1.2873 (4,3)	wrong		
	3	0.8830 (1,5)	correct	2.3478 (1)	wrong
	4	3.7324 (2,10)	wrong		
	5	0.7887 (1,5)	correct		
6	1	3.6106 (4,6)	correct		
	2	1.6442 (4,6)	correct		
	3	1.2236 (4,6)	correct	2.5136 (10)	wrong
	4	1.1199 (3,6)	correct		
	5	1.2410 (2,10)	wrong		
7	1	0.8159 (2,9)	wrong		
	2	0.9658 (1,9)	wrong		
	3	0.5861 (4,7)	correct	1.3057 (7)	correct
	4	0.4873 (3,7)	correct		
	5	0.4876 (3,7)	correct		
8	1	1.3840 (5,4)	wrong		
	2	1.2177 (5,4)	wrong		
	3	1.2593 (4,8)	correct	1.6033 (8)	correct
	4	1.1899 (5,4)	wrong		
	5	0.6028 (4,3)	wrong		
9	1	0.9658 (2,7)	wrong		
	2	0.8159 (1,7)	wrong		
	3	0.5846 (5,9)	correct	1.6800 (9)	correct
	4	0.5988 (5,7)	wront		
	5	0.4809 (3,9)	correct		
10	1	1.6610 (3,10)	correct		
	2	1.2410 (5,6)	wrong		
	3	1.3597 (1,10)	correct	2.8418 (10)	correct
	4	0.7244 (5,10)	correct		
	5	0.6256 (4,10)	correct		

i = sample index Number of correct I.D. = 28 = 5

j = speaker index Rate = 56 % = 50 %

APPENDIX G(16)
Unknown speakers by telephone(IDS+FFC); known speakers by normal(IDS+FFC).

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	3.4344 (1,5)	wrong		
	2	3.3339 (1,5)	wrong		
	3	3.0029 (3,1)	correct	3.8780 (5)	wrong
	4	2.7733 (1,1)	correct		
	5	2.9467 (5,3)	wrong		
2	1	4.0001 (1,8)	wrong		
	2	2.3910 (4,2)	correct		
	3	3.9761 (3,2)	correct	4.3086 (2)	correct
	4	3.7089 (4,2)	correct		
	5	3.5406 (3,2)	correct		
3	1	3.8548 (2,9)	wrong		
	2	3.5361 (3,3)	correct		
	3	4.6579 (4,3)	correct	3.6655 (5)	wrong
	4	3.8641 (4.5)	wrong		
	5	2.2600 (2,3)	correct		
4	1	2.8104 (1,7)	wrong		
	2	2.0866 (4,4)	correct		
	3	3.0011 (4,8)	wrong	4.0164 (4)	correct
	4	3.0998 (1,4)	correct		
	5	3.1235 (1,4)	correct		
5	1	2.0202 (1,9)	wrong		
	2	3.0375 (4,9)	wrong		
	3	3.6108 (2,8)	wrong	4.6013 (9)	wrong
	4	3.3980 (2,5)	correct		
	5	2.1731 (5,5)	correct		
6	1	2.9056 (1,10)	wrong		
	2	2.2243 (1.6)	correct		
	3	2.1478 (1,6)	correct	3.4627 (6)	correct
	4	3.6614 (1,6)	correct		
	5	3.0109 (1,10)	wrong		
7	1	2.1084 (3,3)	wrong		
	2	2.2581 (4,3)	wrong		
	3	3.0555 (5,7)	correct	3.4177 (7)	correct
	4	2.1761 (5,7)	correct		
	5	3.5554 (5,7)	correct		
8	1	2.9450 (4,8)	correct		
	2	3.0108 (4,8)	correct		
	3	2.0001 (4,8)	correct	4.0514 (8)	correct
	4	2.5805 (4,8)	correct		
	5	3.1222 (4,8)	correct		
9	1	3.2222 (2,9)	correct		
	2	1.8033 (3,1)	wrong		
	3	3.0005 (5,9)	correct	3.7224 (1)	wrong
	4	2.9989 (2,4)	wrong		
	5	2.5412 (3,9)	correct		
10	1	2.9899 (1,10)	correct		
	2	2.3706 (1,6)	wrong		
	3	3.9989 (2,6)	wrong	4.0197 (10)	correct
	4	3.9203 (4,10)	correct		
	5	3.1369 (4,10)	correct		

i = sample index Number of correct I.D. = 31 = 6
j = speaker index Rate = 62 % = 60 %

APPENDIX G(17)
Unknown speakers by telephone(IDS+FFC); known speakers by linear(IDS+FFC).

Unknown Speaker (j)	Sample (i)	Known — Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	3.2256 (1,5)	wrong		
	2	3.2207 (5,9)	wrong		
	3	3.3550 (5,5)	wrong	3.8575 (1)	correct
	4	3.2401 (4,1)	correct		
	5	4.3913 (2,5)	wrong		
2	1	4.4399 (1,7)	wrong		
	2	2.7287 (4,2)	correct		
	3	4.7167 (2,2)	correct	4.8425 (7)	wrong
	4	3.7650 (4,2)	correct		
	5	4.0204 (4,2)	correct		
3	1	2.4843 (4,1)	wrong		
	2	2.7575 (1,3)	correct		
	3	3.5950 (1,3)	correct	3.7410 (3)	correct
	4	2.1226 (3,3)	correct		
	5	3.1720 (1,4)	wrong		
4	1	2.8339 (1,4)	correct		
	2	2.6766 (2,3)	wrong		
	3	2.9562 (1,4)	correct	3.6610 (8)	wrong
	4	3.6870 (1,4)	correct		
	5	3.3761 (4,3)	wrong		
5	1	2.8435 (5,8)	wrong		
	2	3.2175 (2,7)	wrong		
	3	2.8178 (2,1)	wrong	4.0610 (8)	wrong
	4	3.5123 (1,5)	correct		
	5	3.3593 (1,5)	correct		
6	1	3.1294 (2,6)	correct		
	2	3.5379 (2,6)	correct		
	3	2.5508 (3,6)	correct	3.8011 (6)	correct
	4	2.4671 (1,10)	wrong		
	5	3,3762 (1,10)	wrong		
7	1	2.6171 (1,3)	wrong		
	2	2.6221 (5,7)	correct		
	3	3.0940 (4,3)	wrong	4.2421 (7)	correct
	4	2.9949 (4,3)	wrong		
	5	2.9599 (5,3)	wrong		
8	1	2.4948 (1,8)	correct		
	2	2.8349 (3,8)	correct		
	3	2.3684 (4,8)	correct	3.2303 (8)	correct
	4	3.2630 (4,8)	correct		
	5	3.1400 (4,8)	correct		
9	1	2.8160 (1,9)	correct		
	2	2.3419 (3,1)	wrong		
	3	2.3018 (3,9)	correct	2.9851 (5)	wrong
	4	2.8760 (4,1)	wrong		
	5	2.7312 (3,9)	correct		
10	1	3.5804 (1,6)	wrong		
	2	2.8889 (1,10)	correct		
	3	3.7659 (4,6)	wrong	4.3465 (10)	correct
	4	3.0866 (1,10)	correct		
	5	3.4282 (1,10)	correct		

i = sample index
j = speaker index

Number of correct I.D. = 28 = 6
Rate = 56 % = 60 %

APPENDIX G (18)
Unknown speakers by telephone(LTAS+FFC); known speakers by normal(LTAS+FFC)

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	Identification decision by the minimum set Dist. with(j)	Result
1	1	3.6524 (1,5)	wrong		
	2	3.0951 (1,1)	correct		
	3	3.3446 (1,1)	correct	4.4697 (1)	correct
	4	2.3071 (1,1)	correct		
	5	3.8588 (5,1)	correct		
2	1	4.2840 (1,2)	correct		
	2	2.6401 (4,2)	correct		
	3	4.5476 (3,2)	correct	4.4841 (2)	correct
	4	3.3295 (4,2)	correct		
	5	3.3492 (4,2)	correct		
3	1	3.4609 (5,3)	correct		
	2	3.0524 (3,3)	correct		
	3	2.8537 (1,3)	correct	3.9884 (3)	correct
	4	2.7594 (1,3)	correct		
	5	2.9426 (5,3)	correct		
4	1	2.9590 (1,4)	correct		
	2	3.0474 (5,4)	correct		
	3	2.8607 (1,4)	correct	3.5549 (4)	correct
	4	2.7287 (1,4)	correct		
	5	3.7803 (1,4)	correct		
5	1	2.8110 (1,5)	correct		
	2	2.6593 (1,5)	correct		
	3	3.2131 (2,5)	correct	4.0908 (5)	correct
	4	3.4112 (1,3)	wrong		
	5	3.5637 (1,5)	correct		
6	1	3.2555 (3,6)	correct		
	2	2.9076 (2,6)	correct		
	3	2.3935 (3,6)	correct	3.2130 (6)	correct
	4	3.2903 (3,6)	correct		
	5	2.8991 (3,6)	correct		
7	1	2,2171 (1,3)	wrong		
	2	2.2879 (4,7)	correct		
	3	2.4743 (3,9)	wrong	3.4856 (7)	correct
	4	2.2801 (4,7)	correct		
	5	2.1019 (5,7)	correct		
8	1	3.3397 (2,8)	correct		
	2	1.8287 (2,8)	correct		
	3	2.1621 (4,8)	correct	3.7699 (8)	correct
	4	1.9583 (4,8)	correct		
	5	3.3247 (3,8)	correct		
9	1	3.2522 (1,9)	correct		
	2	2.4636 (5,9)	correct		
	3	2.8187 (3,9)	correct	3.3593 (9)	correct
	4	3.5689 (5,9)	correct		
	5	2.6800 (4,9)	correct		
10	1	3.5119 (1,10)	correct		
	2	3.9226 (5,10)	correct		
	3	2.5923 (1,10)	correct	3.5294 (10)	correct
	4	2.7880 (4,10)	correct		
	5	3.4344 (2,10)	correct		

i = sample index — Number of correct I.D. = 46 — = 10
j = speaker index — Rate = 92 % — = 100 %

APPENDIX G(19)
Unknown speakers by telephone(LTAS+FFC); known speakers by normal(LTAS+FFC)

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	3.6944 (1,1)	correct		
	2	3.6321 (1,1)	correct		
	3	3.0198 (1,1)	correct	3.8780 (1)	correct
	4	2.5733 (3,1)	correct		
	5	3.5973 (5,1)	correct		
2	1	4.2680 (1,2)	correct		
	2	2.5219 (4,2)	correct		
	3	4.4614 (3,2)	correct	4.3086 (2)	correct
	4	3.6374 (3,2)	correct		
	5	3.8467 (3,2)	correct		
3	1	2.8549 (2,3)	correct		
	2	2.3732 (3,3)	correct		
	3	3.5636 (4,3)	correct	3.6655 (3)	correct
	4	3.0286 (4,3)	correct		
	5	2.8259 (2,3)	correct		
4	1	2.8514 (1,4)	correct		
	2	2.7566 (4,4)	correct		
	3	3.0358 (4,4)	correct	4.0164 (4)	correct
	4	3.3737 (4,4)	correct		
	5	3.3922 (1,4)	correct		
5	1	2.8025 (1,5)	correct		
	2	2.5935 (4,5)	correct		
	3	3.1862 (2,8)	wrong	4.6013 (5)	correct
	4	3.3000 (2,5)	correct		
	5	2.5674 (5,5)	correct		
6	1	2.9766 (1,6)	correct		
	2	2.3424 (1,6)	correct		
	3	2.1369 (1,6)	correct	3.4627 (6)	correct
	4	2.6614 (1,6)	correct		
	5	3.0599 (1,6)	correct		
7	1	2.4258 (3,3)	worng		
	2	2.2711 (4,7)	correct		
	3	2.5062 (5,7)	correct	3.4177 (7)	correct
	4	2.7195 (5,7)	correct		
	5	2.2268 (5,7)	correct		
8	1	2.3775 (4,8)	correct		
	2	2.6494 (4,8)	correct		
	3	2.0882 (4,8)	correct	4.0514 (8)	correct
	4	2.7563 (4,8)	correct		
	5	3.0276 (4,8)	correct		
9	1	3.2060 (2,9)	correct		
	2	2.7576 (3,9)	correct		
	3	2.4444 (5,9)	correct	3.7224 (9)	correct
	4	3.2542 (2,4)	wrong		
	5	2.3321 (3,9)	correct		
10	1	3.3501 (1,10)	correct		
	2	2.1988 (1,10)	correct		
	3	4,8714 (2,10)	correct	4.0197 (10)	correct
	4	2.6331 (4,10)	correct		
	5	3.5376 (4,10)	correct		

i = sample index	Number of correct I.D. = 47		= 10
j = speaker index		Rate = 96 %	= 100 %

APPENDIX G (20)
Unknown speakers by telephone SPT; known speakers by normal SPT.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	Identification decision by the minimum set Dist. with(j)	Result
1	1	0.7505 (5,9)	wrong		
	2	0.8211 (1,9)	wrong		
	3	0.8084 (1,9)	wrong	0.8710 (6)	wrong
	4	0.8783 (1,9)	wrong		
	5	0.8454 (1,9)	wrong		
2	1	0.9380 (1,9)	wrong		
	2	0.9160 (1,9)	wrong		
	3	1.0769 (1,9)	wrong	1.2068 (6)	wrong
	4	1.1695 (1,9)	wrong		
	5	1.0225 (1,9)	wrong		
3	1	0.8005 (1,9)	wrong		
	2	0.7916 (1,9)	wrong		
	3	0.8082 (1,9)	wrong	0.9076 (6)	wrong
	4	0.7805 (1,9)	wrong		
	5	0.8075 (1,9)	wrong		
4	1	0.7306 (1,9)	wrong		
	2	0.8192 (1,9)	wrong		
	3	0.7532 (1,9)	wrong	0.9006 (6)	wrong
	4	0.7757 (1,9)	wrong		
	5	0.7336 (1,9)	wrong		
5	1	0.8386 (1,9)	wrong		
	2	0.8176 (1,9)	wrong		
	3	0.9402 (1,9)	wrong	0.0402 (6)	wrong
	4	0.8506 (1,9)	wrong		
	5	0.9102 (1,9)	wrong		
6	1	0.8087 (4,6)	correct		
	2	0.7775 (4,10)	wrong		
	3	0.8298 (1,9)	wrong	0.9203 (6)	wrong
	4	0.8365 (4,10)	wrong		
	5	0.8410 (1,9)	wrong		
7	1	0.7924 (1,9)	wrong		
	2	0.7608 (1,9)	wrong		
	3	0.7669 (5,9)	wrong	0.8838 (9)	wrong
	4	0.7860 (5,9)	wrong		
	5	0.7667 (5,9)	wrong		
8	1	0.8214 (1,9)	wrong		
	2	0.8072 (1,9)	wrong		
	3	0.7628 (3,10)	wrong	0.8422 (6)	wrong
	4	0.8004 (1,9)	wrong		
	5	0.7753 (5,9)	wrong		
9	1	0.9158 (1,5)	wrong		
	2	0.8481 (1,9)	correct		
	3	0.8718 (1,9)	correct	1.0581 (9)	correct
	4	0.8567 (1,9)	correct		
	5	0.8462 (1,9)	correct		
10	1	0.8779 (1,9)	wrong		
	2	0.8536 (1,9)	wrong		
	3	0.8698 (1,9)	wrong	0.8500 (6)	wrong
	4	0.7421 (4,6)	wrong		
	5	0.9261 (1,9)	wrong		

i = sample index　　Number of correct I.D. = 5　　= 2
j = speaker index　　Rate = 10 %　　= 20 %

APPENDIX G (21)
Unknown speakers by telephone SPT; known speakers by linear SPT.

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	0.7425 (4,9)	wrong		
	2	0.7632 (3,9)	wrong		
	3	0.7670 (2,9)	wrong	0.7999 (6)	wrong
	4	0.7952 (1,9)	wrong		
	5	0.7854 (1,9)	wrong		
2	1	0.7750 (1,9)	wrong		
	2	0.7708 (1,9)	wrong		
	3	0.8332 (1,9)	wrong	0.9578 (9)	wrong
	4	0.9017 (1,9)	wrong		
	5	0.8112 (1,9)	wrong		
3	1	0.7599 (3,6)	wrong		
	2	0.7501 (3,9)	wrong		
	3	0.7722 (3,6)	wrong	0.7566 (6)	wrong
	4	0.7405 (5,6)	wrong		
	5	0.7564 (3,6)	wrong		
4	1	0.7285 (5,7)	wrong		
	2	0.7926 (2,9)	wrong		
	3	0.7335 (2,9)	wrong	0.8018 (6)	wrong
	4	0.7369 (4,4)	correct		
	5	0.7153 (5,9)	wrong		
5	1	0.7892 (2,9)	wrong		
	2	0.7974 (2,9)	wrong		
	3	0.8132 (1,9)	wrong	0.8600 (9)	wrong
	4	0.8164 (2,9)	wrong		
	5	0.8161 (1,9)	wrong		
6	1	0.7653 (3,6)	correct		
	2	0.7372 (3,6)	correct		
	3	0.7601 (3,6)	correct	0.8111 (6)	correct
	4	0.7764 (3,10)	wrong		
	5	0.8045 (5,6)	correct		
7	1	0.7697 (1,9)	wrong		
	2	0.7485 (5,9)	wrong		
	3	0.7462 (5,9)	wrong	0.8462 (6)	wrong
	4	0.7892 (5,9)	wrong		
	5	0.7385 (5,9)	wrong		
8	1	0.7873 (2,9)	wrong		
	2	0.7739 (2,9)	wrong		
	3	0.7661 (4,9)	wrong	0.8074 (6)	wrong
	4	0.7746 (3,9)	wrong		
	5	0.7776 (3,6)	wrong		
9	1	0.8104 (1,9)	correct		
	2	0.8105 (3,9)	correct		
	3	0.8181 (3,9)	correct	0.8837 (9)	correct
	4	0.8273 (2,9)	correct		
	5	0.8299 (2,9)	correct		
10	1	0.8264 (3,6)	wrong		
	2	0.8421 (2,9)	wrong		
	3	0.8534 (2,9)	wrong	0.8269 (6)	wrong
	4	0.7768 (5,7)	wrong		
	5	0.8895 (2,9)	wrong		

i = sample index
j = speaker index

Number of correct I.D. = 10 = 2
Rate = 20 % = 20 %

APPENDIX G(22)
Both unknown and known speakers by telephone SPT.

Unknown Speaker (j)	Sample (i)	Known by the nearest-neighbor Dist. with(i,j)	Result	by the minimum set Dist. with(j)	Result
1	1	0.6572 (2,7)	wrong		
	2	0.6497 (3,1)	correct		
	3	0.6497 (2,1)	correct	0.6976 (7)	wrong
	4	0.6726 (2,1)	correct		
	5	0.6759 (2,1)	correct		
2	1	0.6785 (2,2)	correct		
	2	0.6785 (1,2)	correct		
	3	0.7168 (5,2)	correct	0.7367 (2)	correct
	4	0.7150 (5,2)	correct		
	5	0.7050 (4,2)	correct		
3	1	0.6564 (3,3)	correct		
	2	0.6804 (3,3)	correct		
	3	0.6564 (1,3)	correct	0.6950 (3)	correct
	4	0.6914 (1,1)	wrong		
	5	0.6704 (3,3)	correct		
4	1	0.6555 (3,4)	correct		
	2	0.6816 (4,8)	wrong		
	3	0.6477 (4,4)	correct	0.6874 (4)	correct
	4	0.6477 (3,4)	correct		
	5	0.6601 (4,4)	correct		
5	1	0.6475 (2,5)	correct		
	2	0.6475 (1,5)	correct		
	3	0.6903 (5,5)	correct	0.6827 (6)	wrong
	4	0.6622 (1,5)	correct		
	5	0.6800 (4,5)	correct		
6	1	0.6304 (3,6)	correct		
	2	0.6455 (1,6)	correct		
	3	0.6304 (1,6)	correct	0.6578 (6)	correct
	4	0.6377 (3,6)	correct		
	5	0.6578 (3,6)	correct		
7	1	0.6233 (2,7)	correct		
	2	0.6181 (5,7)	correct		
	3	0.6287 (5,7)	correct	0.6288 (7)	correct
	4	0.6201 (5,7)	correct		
	5	0.6181 (2,7)	correct		
8	1	0.6400 (4,8)	correct		
	2	0.6545 (1,8)	correct		
	3	0.6624 (5,8)	correct	0.6819 (8)	correct
	4	0.6400 (1,8)	correct		
	5	0.6583 (1,8)	correct		
9	1	0.6555 (4,9)	correct		
	2	0.6358 (4,9)	correct		
	3	0.6535 (2,9)	correct	0.6618 (9)	correct
	4	0.6246 (5,9)	correct		
	5	0.6246 (4,9)	correct		
10	1	0.6507 (3,10)	correct		
	2	0.6112 (3,10)	correct		
	3	0.6111 (2,10)	correct	0.6712 (10)	correct
	4	0.6640 (5,7)	wrong		
	5	0.6394 (3,10)	correct		

i = sample index   Number of correct I.D.= 46   = 8
j = speaker index   Rate= 92 %   = 80 %

APPENDIX G (23)
Both unknown and known speakers by normal SPT.

Unknown Speaker (j)	Sample (i)	Known — Identification decision — by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	0.5457 (5,1)	correct		
	2	0.6013 (1,1)	correct		
	3	0.6065 (2,1)	correct	0.6157 (1)	correct
	4	0.6104 (1,1)	correct		
	5	0.5700 (1,1)	correct		
2	1	0.6561 (2,2)	correct		
	2	0.6255 (4,2)	correct		
	3	0.6682 (2,2)	correct	0.6188 (2)	correct
	4	0.6235 (2,2)	correct		
	5	0.6588 (2,2)	correct		
3	1	0.5510 (4,3)	correct		
	2	0.5795 (5,3)	correct		
	3	0.5562 (5,3)	correct	0.5887 (3)	correct
	4	0.5634 (1,3)	correct		
	5	0.5801 (3,3)	correct		
4	1	0.6160 (2,4)	correct		
	2	0.5936 (5,4)	correct		
	3	0.6232 (5,4)	correct	0.6713 (4)	correct
	4	0.6713 (1,4)	correct		
	5	0.6175 (2,4)	correct		
5	1	0.5919 (4,8)	wrong		
	2	0.6238 (4,5)	correct		
	3	0.6039 (1,5)	correct	0.6384 (5)	correct
	4	0.6116 (5,5)	correct		
	5	0.6028 (1,8)	wrong		
6	1	0.6198 (2,6)	correct		
	2	0.6318 (5,6)	correct		
	3	0.6410 (5,6)	correct	0.6410 (6)	correct
	4	0.6299 (5.6)	correct		
	5	0.6422 (4,6)	correct		
7	1	0.6497 (2,7)	correct		
	2	0.6263 (3,7)	correct		
	3	0.6028 (5,7)	correct	0.6611 (7)	correct
	4	0.6274 (5,7)	correct		
	5	0.6267 (3,7)	correct		
8	1	0.5844 (4,8)	correct		
	2	0.6016 (4,8)	correct		
	3	0.6369 (5,8)	correct	0.6369 (8)	correct
	4	0.5805 (5,3)	wrong		
	5	0.5937 (1,8)	correct		
9	1	0.6652 (5,9)	correct		
	2	0.5975 (3,9)	correct		
	3	0.6095 (2,9)	correct	0.6773 (5)	wrong
	4	0.6051 (2,9)	correct		
	5	0.6636 (3,9)	correct		
10	1	0.6236 (2,10)	correct		
	2	0.6360 (1,10)	correct		
	3	0.6324 (2,10)	correct	0.6488 (10)	correct
	4	0.6476 (1,10)	correct		
	5	0.6253 (1,10)	correct		

i = sample index　　Number of correct I.D. = 47　　= 9
j = speaker index　　Rate = 94 %　　= 90 %

APPENDIX G(24)
Both unknown and known speakers by linear SPT.

Unknown Speaker (j)	Sample (i)	Known Identification decision by the nearest-neighbor: Dist. with(i,j)	Result	by the minimum set: Dist. with(j)	Result
1	1	0.7017 (2,5)	wrong		
	2	0.6498 (3,1)	correct		
	3	0.7354 (5,1)	correct	0.7069 (3)	wrong
	4	0.6544 (5,1)	correct		
	5	0.6366 (3,1)	correct		
2	1	0.6701 (4,2)	correct		
	2	0.6811 (5,2)	correct		
	3	0.7462 (4,1)	wrong	0.7852 (2)	correct
	4	0.9683 (5,2)	correct		
	5	0.6670 (4,2)	correct		
3	1	0.6376 (3,3)	correct		
	2	0.7594 (5,3)	correct		
	3	0.7364 (1,3)	correct	0.6650 (3)	correct
	4	0.6382 (1,3)	correct		
	5	0.6557 (2,3)	correct		
4	1	0.7596 (5,4)	correct		
	2	0.9718 (4,4)	correct		
	3	0.6945 (4,4)	correct	0.6945 (4)	correct
	4	0.6705 (2,4)	correct		
	5	0.6608 (1,4)	correct		
5	1	0.6756 (2,5)	correct		
	2	0.6531 (5,8)	wrong		
	3	0.6562 (2,8)	wrong	0.6926 (5)	correct
	4	0.6772 (1,5)	correct		
	5	0.6595 (2,5)	correct		
6	1	0.6624 (5,6)	correct		
	2	0.6519 (5,6)	correct		
	3	0.7284 (5,6)	correct	0.6624 (6)	correct
	4	0.6432 (5,6)	correct		
	5	0.6297 (3,6)	correct		
7	1	0.6612 (2,7)	correct		
	2	0.9065 (5,7)	correct		
	3	0.7341 (4,7)	correct	0.7416 (8)	wrong
	4	0.6354 (3,7)	correct		
	5	0.6479 (2,7)	correct		
8	1	0.9161 (3,8)	correct		
	2	0.6562 (3,5)	wrong		
	3	0.6575 (1,8)	correct	0.6880 (8)	correct
	4	0.6597 (5,8)	correct		
	5	0.6531 (2,5)	wrong		
9	1	0.7194 (3,9)	correct		
	2	0.7390 (5,9)	correct		
	3	0.6458 (2,9)	correct	0.7194 (9)	correct
	4	0.6525 (5,9)	correct		
	5	0.6403 (2,9)	correct		
10	1	0.6600 (5,10)	correct		
	2	0.7810 (3,10)	correct		
	3	0.6773 (2,10)	correct	0.6971 (10)	correct
	4	0.6971 (5,10)	correct		
	5	0.6732 (1,10)	correct		

i = sample index
j = speaker index

Number of correct I.D. = 44   = 8
Rate = 88 %   = 80 %

APPENDIX H

LIST OF FORTRAN SOFTWARE

APPENDIX H

LIST OF FORTRAN SOFTWARE

Main Program	Usage
AUTOCM*	Digitization, automatic speech compression.
POZCMP*	Detection and concatenation of pauses and audio play-back of the concatenated pauses.
SHORT8*	Segmentation of digitized speech and generation of short-term spectra. Uses FFT algorithm with 256(2^8) sampled points.
IDSLTS*	Computes IDS (intensity deviation spectrum) and LTS (long-term averaged spectrum).
DENCOR*	Computes the Pearson correlation coefficients, then creates a similarity matrix of these coefficients as input data file to the clustering program, COMPLE.
COMPLE**	Major program to prepare the complete-link dendrogram.
FRATIO*	Computes F-ratios of the features within each cluster of IDS and LTS formed by complete-link dendrogram.
FFFRAT*	Computes F-ratios of FFC features.
UPICK*	Interactive peak detecting algorithm. Uses a light pen to detect Fo from digitized speech signal in the time domain, then creates a file of Fo contour.
FFPICK*	Computes FFC features from a file of Fo contour.
NAKRO1 NAKRO2 NAKRO3* NAKRO4	Executes the cross-transmission voice identification operations.
NAKROA NAKROB NAKROC	Executes the within-transmission voice identification operations.
NAKROP PLOTIL	Plotting programs.

Continued.

APPENDIX H

LIST OF FORTRAN SOFTWARE (continued)

TOSI1**	Creates choral speech and choral spectrum. Uses FFT algorithm with 4096 sampled points.
SAMMON**	Nonlinear projection program.

Subprograms

ZNORM*	Standardization routine. Uses the z-transformation.
SUBF*	Routine to compute F-ratios.
SETDIS*	Routine to find minimum set distance.
IDA**	Routine to perform A/D and D/A conversions (written in assembly language).
FILFIL*	Creates a master file of file names.
RDFILE*	Reads file names from a master file.
FFT10** FFT12**	Fast Fourier Transform(written partially in assembly language). FFT10 is called by Program SHORT8, and FFT12 is called by TOSI1.

System softwares

Graphic processing routines
Fortran IV compiler
and other essential system routines.

---

* Listings of the Fortran source codes are attached in the following pages. These softwares were designed and written by the author.

** Laboratory software used on PDP 11/40 at Speech and Hearing Sciences Research Laboratory (Director, Dr. Oscar Tosi), Audiology and Speech Sciences, Michigan State University.

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C       AUTOCM:  AUTOMATIC SPEECH COMPRESSOR - FIRST PART -
C
C       Written by     : Hirotaka Nakasone
C       Last modified : 11-May-83
C
C       Department of Audiology and Speech Sciences
C       Michigan State University
C       East Lansing
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C234567890 234567890 234567890 234567890 234567890 234567890 234567890
C
        DOUBLE PRECISION PROG
        INTEGER*4 IN4,JB4,LE4
        INTEGER IN2(2),IBUF(500)
        INTEGER BUFF1(5120),BUFF2(5120)
        INTEGER SFILE(4),OFILE(4),WER,FILCNT,TP
        LOGICAL*1 FILNAM(6),YORN,SI,ITAB,IB
        COMMON /PASS3/ ISWICH,NSEC,SFILE,TP,HAP,SUMMAX,SUMMIN,OFILE
        COMMON /BUFFR/ BUFF1,BUFF2
        COMMON /FIL/ FILNAM,FILCNT
        EQUIVALENCE (IN4,IN2)
        DATA ITAB/"013/,IB/"007/,SUMMAX/0.0/
        DATA SFILE(1)/3RDK0/,SFILE(4)/3RSND/,SUMMIN/0.0/
        DATA OFILE(1)/3RDK1/,OFILE(2)/3RCMP/,OFILE(3)/3RRSD/
        DATA OFILE(4)/3RSND/,IN2/256,0/
        DATA PROG/12RDK1HNAUTOSAV/


 1200   FORMAT('+',1A1)
 1220   FORMAT(1X,1A1)
        CALL INIT(IBUF,500)
        CALL SCROL(1,1000)
        TYPE 1220, ITAB
C       TYPE 1220, IB
        CALL APNT(0.,730.,0,-3,0,4)
        CALL VECT(900.,0.)
        CALL APNT(200.,860.,0,-8)
        CALL SUBP(1)
        CALL TEXT(' PROGRAM AUTOCM')
        CALL ESUB
        TYPE 1200, IB
C       CALL OFF(1)
        CALL SCROL(20,570,8)
 400    FORMAT(1A1)
        TYPE 215
 215    FORMAT(' Enter file name ( 6 letters ) : ',$)
        READ(5,405) (FILNAM(M),M=1,6)
 405    FORMAT(6A1)
 110    TYPE 220
 220    FORMAT(' How many seconds ( in I3 FMT ) ? : ',$)
        ACCEPT 410, NSEC
 410    FORMAT(I3)
        IF(NSEC.GE.1 .AND. NSEC.LE.118) GO TO 105
           TYPE 225
 225       FORMAT(' Duration out of range. Try again.')
        GO TO 110
```

```
105     NBLKS = NSEC*40+1
        ICHAN = IGETC()
        IF(ICHAN .LT.0) STOP ' Channel not available'
        NLET = IRAD50(6,FILNAM,SFILE(2))
        IF(NLET.NE.6) STOP 'Bad RAD50 conversion'
        MBLKS = IENTER(ICHAN,SFILE,NBLKS)
        IF(MBLKS.LT.0) GO TO 115
        TYPE 555
555     FORMAT(' ENTER AP(0.0 - 1.0) >',$)
        READ(5,559) HAP
559     FORMAT(F6.4)
        TP = 10
        WER = 0
        FILCNT = NSEC*2
C               --- Digitization ---
        CALL QICH(ICHAN,FILCNT,BUFF1,BUFF2,WER)

        IF(WER .EQ. 0) GO TO 120
        TYPE 230,WER
230     FORMAT(/' Register 0 = ',o10)
        STOP ' WRITE ERROR IN OUTPUT FILE'
120     CALL CLOSEC(ICHAN)
        CALL IFREEC(ICHAN)
        GO TO 125
115     CALL CLOSEC(ICHAN)
        IF(IFREEC(ICHAN).NE.0) STOP' Error IFREEC next to 115'
        IF(MBLKS .EQ. -1) STOP 'NO CHANNEL OPEN'
        TYPE 235,NBLKS,NSEC
235     FORMAT(' Too long for current free blocks available.'/
        &,1X,I4,' blocks needed for ',I3,' second(s).'/
        &' Want to try again by reducing # of sec (Y/N) ? ',$)
        ACCEPT 400,YORN
        IF(YORN.EQ.89) GO TO 110
        STOP 'PROGRAM EXITED'
125     WRITE(7,240) (FILNAM(M),M=1,6)
240     FORMAT(///,' A sound file just created : ',6A1,'.SND'/)
        TYPE 1200, IB
        ISWICH = 2

C                 --- Chaining to Program HNAUTO to complete
C                     automatic speech compression ---
        CALL CHAIN(PROG,ISWICH,34)

        CALL EXIT
C       Link information: *AUTOCM=AUTOCM,QICH,GTLIB,FUSS,SYSLIB/F
        END
*
```

```
CHHHHH HNAUTO: SECOND PART OF AUTOCM (AUTOMATIC SPEECH COMPRESSOR)
C
C
C       PROGRAM HNAUTO IS  DESIGNED TO PERFORM AUTOMATIC
C       SPEECH COMPRESSION BY APPLYING PAUSOMETRIC DEFINITION
C       OF PAUSES PROPOSED BY TOSI (1974)
C
C       Program HNAUTO is written on RT-11FB, V02c-02B, PDP11/40.
C
C       HNAUTO detects pauses in a sound input filemeasures,
C       eliminates pauses, compresses signals, and creats a
C       compressed speech file of pre-determined duration in
C       seconds.
C
C       Presently, the maximum input speech duration is 118 seconds,
C       and the maximum output (compressed speech) depends on the
C       free blocks available in one of two disks in operation.
C
C
C       To compile: /U/S/N:m switches
C       To link   : *HNAUTO=HNAUTO,IAUTO,CMPRES,HNLIB,SYSLIB/F
C
C       Author        : Hirotaka Nakasone
C       Date          : 03 APR, 1983
C
C       Department of Audiology and Speech Sciences
C       Michigan State University
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        DOUBLE PRECISION VCH,PCH
        COMMON /B../ COUNT,IOUT,NSIZE,IDONE,LAST
        COMMON /PASS3/ ISWICH,NSEC,SFILE,TP,AP,SUMMAX,SUMMIN,OFILE
        INTEGER*4 IN4,JB4,LE4
        INTEGER IHOLD(2560),IBUF(2560),SFILE(4),OFILE(4)
        INTEGER IOUT(2560),FLAG,TMPKS,TPKT,TP
        INTEGER TOTP,TS,IN2(2)
        EQUIVALENCE (IN4,IN2)
        LOGICAL*1 TM(8),YORN,NAME(6),IBEL,BLANK
        DATA SFILE(1)/3RDK0/,SFILE(4)/3RSND/
        DATA OFILE/3RDK1,3RCMP,3RRSD,3RSND/
        DATA PCH/12RDK1POZCMPSAV/
        DATA VCH/12RDK1POZV1 SAV/
        DATA IN2/256,0/,ISWICH/3/,IBEL/"007/,BLANK/' '/
3011    FORMAT('+',1A1)


        CALL RCHAIN(IF,ISWICH,34)
        NSIZE = 2560
        UP = 0.01
        DOWN = 0.01
        COUNT = 0.0
        NW = 10

Check from which progam this program is chained to.


        IF(IF.NE.-1) GO TO 4100
```

```
        IF(ISWICH.EQ.2) GO TO 4105
        IF(ISWICH.EQ.1) GO TO 4110
        STOP ' ***INVALID ISWITCH***'

4100    TYPE 4200
4200    FORMAT(' Enter input sound file name (6 letters) >',$)
        READ(5,4400) (NAME(M),M = 1,6)
4400    FORMAT(6A1)
        CALL IRAD50(6,NAME,SFILE(2))
1352    WRITE(7,1350)
        READ(5,1220) NSEC

4110    ICIN = IGETC()
        IF(ICIN .LT. 0) STOP ' NO INPUT CHANNEL.'
        ILK  = LOOKUP(ICIN,SFILE)
        IF(ILK .LT. 0) STOP ' BAD LOOKUP.'
          INBK = 0
          SUMMIN = 10000.
          SUMMAX = 0.
          IF(NSEC.LT.0) GO TO 1352
          IF(NSEC.GE.4) GO TO 7001
            MBLOCK = NSEC*40 - 10
            GO TO 5001
7001      MBLOCK = 150
          INBK = 40
5001      CALL TIME(TM)
          WRITE(7,1000) (TM(M),M=1,8)
5000      CONTINUE
            NWD = IREADW(NSIZE,IBUF,INBK,ICIN)
            IF (NWD .LT. 0) STOP ' READ ERROR.'
              JJ = 1
5010          SUM = 0.
              DO 5030 J=1,128
                  SUM = SUM + IABS(IBUF(JJ))
                  JJ = JJ + 1
5030          CONTINUE
              SUM = SUM/128.
              IF(SUM .LT. SUMMIN) SUMMIN=SUM
              IF(SUM .GT. SUMMAX) SUMMAX=SUM
              JJ = JJ - 64
              IF(JJ .LT. NSIZE) GO TO 5010
              INBK = INBK + NW
         IF ( INBK .LE. MBLOCK ) GO TO 5000
         CALL CLOSEC(ICIN)
         CALL IFREEC(ICIN)

         CALL TIME(TM)
         TYPE 3011,IBEL
         WRITE(7,2000) (TM(M),M=1,8)
         WRITE(7,5200) SUMMAX,SUMMIN


6030     WRITE(7,1100)
         READ(5,1200) HAP
         WRITE(7,1300)
         READ(5,1220) TP
C
C
4105     CALL TIME(TM)
         TYPE 2020,(TM(M),M=1,8)
         ICIN = IGETC()
         IF(ICIN .LT. 0) STOP ' NO INPUT CHANNEL.'
         ILK=LOOKUP(ICIN,SFILE)
         IF(ILK .LT. 0) STOP ' BAD LOOKUP.'
         TMSEC = 0.5 * FLOAT(ILK) * 25.6
         ICOUT=IGETC()
```

```
        IF(ICOUT .LT. 0) STOP ' NO OUTPUT CHANNEL'
        NB=IENTER(ICOUT,OFILE,-1)
        IF(NB.LT.O) STOP ' ENTER FAILED'
C                        ---INITIALIZE PARAMETERS---
        MBLOCK = NSEC*40 - 10
2999    COUNT = 0.0
        AP = (1.- HAP) * (SUMMAX - SUMMIN) + SUMMIN
        DO 5500 I = 1, NSIZE
           IBUF(I) = 0
           IOUT(I) = 0
           IHOLD(I)= 0
5500    CONTINUE
        INBK = 0
        IBKOUT = 0
        IDONE = 0
        LHOLD = 0
        LAST = 0
        TPKT = 0
        FLAG = 0
        IBEG = 1
        IEND = 0

1       CONTINUE
          NWD=IREADW(NSIZE,IBUF,INBK,ICIN)
        . IF(NWD .LT. 0) STOP ' READ ERROR.'
C                         -----STARTING PAUSOMETRY----
          KS = 0
          IKS = 0
          TMPKS = 0
          JK = 0
50        IPT = 0
60        SUM = 0.
        . DO 65 I =1,10
            JK = JK + 1
            SUM = SUM + IABS(IBUF(JK))
65        CONTINUE
          SUM = SUM / 10.

C                         -----TESTING AP PARAMETER-----
          IF(SUM .LE. AP) GO TO 70
            KS = KS + 1
          IF(FLAG .EQ. 0) GO TO 115
            ITOT = 10 * (KS-1+TMPKS)
          IF(ITOT .LE. 0) GO TO 8050
            IBEG = IEND - ITOT + 1
D           TYPE 8012,IEND, IBEG, ITOT
D8012       FORMAT(1X,'FROM 8010>>>IEND=',I7,' IBEG=',I7,' ITOT=',I7)
8010        CALL CMPRES(IBUF,IBEG,IEND,ITOT)
            IF(IDONE .LT. 1) GO TO 8050
            NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
            IBKOUT = IBKOUT + NW
            IF(IDONE .EQ. 2) GO TO 8010
8050      CONTINUE
          IKS = 0
           KS = 1
           FLAG = 0
           GO TO 120
115       IF(LHOLD .EQ. 0) GO TO 120
            LHOLD = 0
            IHBEG = 1
            IHEND = IHTOT
D         TYPE 8022,IHBEG,IHEND,IHTOT
D8022     FORMAT(1X,'FROM 8020>>> IHBEG=',I7,' IHEND=',I7,' IHTOT',I7)
8020      CALL CMPRES(IHOLD,IHBEG,IHEND,IHTOT)
          IF(IDONE .LT. 1) GO TO 120
          NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
```

```
        . IBKOUT = IBKOUT + NW
          IF(IDONE .EQ. 2) GO TO 8020
120      TPKT = 0
         TMPKS = IKS
         IF( JK .LT. NSIZE ) GO TO 50
         ITOT = (KS+TMPKS) * 10
         IBEG = JK - ITOT + 1
         IEND = JK
D        TYPE 8072,IBEG,IEND,ITOT
D8072    FORMAT(1X,' FROM 8070>>> IBEG=',I7,' IEND=',I7,' ITOT=', I7)
8070     CALL CMPRES(IBUF,IBEG,IEND,ITOT)
         IF(IDONE.LT.1) GO TO 4
         NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
         IBKOUT = IBKOUT + NW
         IF(IDONE.EQ.2) GO TO 8070
         GO TO 4
C
C                                ---END OF PROCESSING FOR IBUF < AP--
C
70       IPT = IPT + 1
C                                -----TESTING TP PARAMETER-----
           IF((IPT +TPKT) .LT. TP) GO TO 80
             IEND = JK - IPT*10
D            TYPE 72,IEND,COUNT/10.
D72          FORMAT(1X,' FROM D72, IEND = ',I7,' COUNT =',F10.2)
             FLAG = 1
             LHOLD= 0
             IKS = 0
             GO TO 85
80         IKS = IKS + 1
85         CONTINUE
           IF(JK .LT. NSIZE) GO TO 60
           TPKT = TPKT + IPT
           IF(TPKT .GT. 30000) GO TO 90
           IF(FLAG.EQ.1) GO TO 5
           IH=JK-IPT*10
           LHOLD = 1
           IHTOT = IPT * 10
           DO 9020 J=1,IHTOT
             IH = IH + 1
             IHOLD(J)=IBUF(IH)
9020       CONTINUE
5          ITOT = (KS+TMPKS)*10
           IF(ITOT .LE. 0) GO TO 4
           IEND = JK - IPT*10
           IBEG = IEND - ITOT + 1
8100       CALL CMPRES(IBUF,IBEG,IEND,ITOT)
           IF(IDONE.LT.1) GO TO 4
           CALL IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
           IBKOUT=IBKOUT+NW
           IF(IDONE .EQ. 2) GO TO 8100
4         INBK = INBK + NW
          IF(INBK .LE. MBLOCK) GO TO 1
          IF(IDONE.LT.1) CALL IWRITW(LAST,IOUT,IBKOUT,ICOUT)
          GO TO 9050
                                 -----END OF INPUT SIGNAL-----
C
90        TYPE 93,TPKT
93        FORMAT(' Warning! Excessively long pause of',I5,' msec so
          & far detected.'/' Check AP or TP parameter, or input sound
          & file.'/' Run again.',//)
          CALL CLOSEC(ICIN)
          CALL IFREEC(ICIN)
          STOP' Interrupted exit'

C                --- Check compression status by UP and DOWN method
C                    by calling INTEGER FUNCKTION IAUTO ---
```

```
9050    IF(IAUTO(HAP,UP,DOWN,COUNT,TMSEC).LT.0) GO TO 2999
C                 Come here if compression satisfactory.

        CALL CLOSEC(ICOUT)
        CALL IFREEC(ICOUT)
        CALL CLOSEC(ICIN)
        CALL IFREEC(ICIN)
        TYPE 3010,IBEL
        CALL TIME(TM)
        WRITE (7,3020) (TM(M),M=1,8)
3020    FORMAT(1X,8A1,'  Signal compression completed.'/)
C
C         --- Audio playback of the compressed speech ---

        ICHAN = IGETC()
        NB = LOOKUP(ICHAN,OFILE)
        CALL JICVT(NB,JB4)
        CALL JMUL(IN4,JB4,LE4)
        TYPE 6696, IBEL,IBEL,IBEL
6696    FORMAT('+',3A1)

        COUNT = COUNT/10.
        TYPE 6000,COUNT,NB
        IWAIT = 60
6666    IER = IDA(ICHAN,6,400,100,LE4,,IWAIT,0)
        IWAIT = 0
        TYPE 7100
7100    FORMAT(' Hit <Return> to play again.'/' Otherwise type any
        & key, then hit <return>     >',$)
        ACCEPT 3010,YORN
        IF(YORN.EQ.BLANK) GO TO 6666
        CALL CLOSEC(ICHAN)
        IF(IFREEC(ICHAN).NE.0) STOP ' CLOSEC FAILED'
6000    FORMAT(' Output file name: CMPRSD.SND',/,' Total compress
        &ed signal duration = ',F10.1,' msec.(',I4,' Blocks)'//)
        ISWICH = 3
1000    FORMAT(1X,8A1,'  Searching the average peak amplitude.
        &Wait.',$)
1100    FORMAT(' Enter AP parameter ( 0.0 =< AP =< 1.0000 )   >',$)
1200    FORMAT(F10.4)
1220    FORMAT(I5)
1300    FORMAT(' Enter TP parameter ( 1 =< TP =< 10000 msec) >',$)
1350    FORMAT(' Enter the number of seconds ( 1=<SEC=<60 )   >',$)
2000    FORMAT(1X,8A1,'  Search ended.',/)
2020    FORMAT(1X,8A1,'  Compression in procee.  Wait.'/)
3010    FORMAT(1A1)
3040    FORMAT(F8.1)
5200    FORMAT(1X,'Pre-detection outputs :',//,1X,'Computed maximum
        & amplitude =',F10.4,/,1X,'Computed minimum amplitude =',
        &F10.4,5(/))

7777    CALL EXIT
        END
*
```

```
C**********************************************************
C                                                         *
C       INTEGER FUNCTION IAUTO(HAP,UP,DOWN,COUNT,TMSEC)   *
C                                                         *
C       Called by Program HNAUTO.                         *
C                                                         *
C       TMSEC = 0.5*FLOAT(NUBLK)*25.6  ! Threshold.       *
C                                                         *
C       COUNT = The number of sample points detected as   *
C               pauses ( devide by 10 to obtain msec.)    *
C                                                         *
C       IAUTO = 99,  IF COMPRESSION COMPLETED.            *
C                                                         *
C       IAUTO = -99, IF COMPRESSION NOT COMPLETED.        *
C                                                         *
C       Written by :  Hirotaka Nakasone                   *
C             Date :  11 May, 1983                        *
C**********************************************************

        FUNCTION IAUTO(HAP,UP,DOWN,COUNT,TMSEC)

        TEST = COUNT/10. - TMSEC
        THRESH = TMSEC * 0.1
        IF(TEST.GT.THRESH) GO TO 10
        IF(TEST.LT.0.0) GO TO 20
C               HERE, MISSION COMPLETED.
        IAUTO = 99
        RETURN
C               HERE, COMPRESSED SIGNAL TOO LONG, SO REDUCE
C               AP VALUE BY CURRENT DOWN VALUE.

10      HAP = HAP - DOWN
        UP = DOWN / 2.

D       TYPE 100,COUNT/10.,TEST,DOWN,HAP
100     FORMAT(1X,' Compressed too long; ',F10.1,' msec.',/,
        &' Difference = ',F10.1,' msec.  Ap is reduced by ',F7.5,/,
        &' Repeating compression by Ap = ',F7.5,//)

        IAUTO = -99
        RETURN

C               HERE, COMPRESSED SIGNAL TOO SHORT.  SO INCREASE THE
C               AP VALUE BY CURRENT UP VALUE

20      HAP = HAP + UP
        DOWN = UP / 2.

D       TYPE 200,COUNT/10.,TEST,UP,HAP
200     FORMAT(' Compressed too short; ',F10.1,' msec.',/,
        &' Difference = ',F10.1,' msec.  Ap is increased by',F7.5,/,
        &' Repeating compression by Ap = ',F7.5,//)

        IAUTO = -99

        RETURN
        END
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C     Program POZCMP (on RT-11FB, v02c-02B, PDP11/40)
C
C     POZCMP detects, measures, and stores pauses from running
C     speech. The detected pauses are concatenated so that it
C     can be reproducible through DAC either to a loud speaker
C     or to a tape recorder.
C
C     An input file resides in Disk 0 with .SND extension,
C     and an output file, also in Disk 0.
C
C     Subprogram called:
C        CMPRES: Used to fill temporary buffer
C        until it gets full, 10 blocks(2560 words)
C        IDA: DAC & ADC routine found in FUSS.LIB.
C
C     Program chaining(from) : PAUSOM
C         "        "   ( to) : POZV1, SIGCMP, DTOA
C     POZCMP can be run from Monitor directly by typing .RUN POZCMP.
C
C     To compile:    /U/S/N:m  switches, (m=14)
C     To link   :    *POZCMP=POZCMP,CMPRES,FUSS,SYSLIB/F
C
C     Written by :   Hirotaka Nakasone
C           Date :   27 March, 1982
C
C     Department of Audiology and Speech Sciences
C     Michigan State University
C     East Lansing, Michigan
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        DOUBLE PRECISION VCH,SCH
        COMMON/PASS3/ ISWICH,NSEC,SFILE,TP,AP,SUMMAX,SUMMIN,OFILE
        COMMON/B1/COUNT,IOUT,NSIZE,IDONE,LAST
        INTEGER*4 IN4,JB4,LE4
        INTEGER IHOLD(2560),IBUF(2560),SFILE(4),OFILE(4)
        INTEGER IN2(2),IOUT(2560),FLAG,TPKT,TP
        LOGICAL*1 TM(8),YORN,NAME(6),IBEL,BLANK
        EQUIVALENCE (IN4,IN2)
        DATA SFILE(1)/3RDK0/,SFILE(4)/3RSND/
        DATA OFILE/3RDK0,3RCMP,3RRSD,3RSND/
        DATA VCH/12RDK POZV1 SAV/,SCH/12RDK SIGCMPSAV/
        DATA ISWICH/2/,IBEL/"007/,BLANK/' '/

3011    FORMAT('+',1A1)
        CALL RCHAIN(IF,ISWICH,34)
        COUNT = 0.0
        NSIZE=2560
        NW = 10
        IF(IF.EQ.-1) GO TO 4100
        TYPE 4200
4200    FORMAT(' Enter input file name (6 letters) >',$)
        READ(5,4205) (NAME(M),M=1,6)
4205    FORMAT(6A1)
        CALL IRAD50(6,NAME,SFILE(2))
1352    WRITE(7,1350)
        READ(5,1220) NSEC
4100    IF(ISWICH.EQ.3) GO TO 6050
```

```
3010    FORMAT(1A1)
        ICIN = IGETC()
        IF(ICIN .LT. 0) STOP ' NO INPUT CHANNEL.'
        ILK  = LOOKUP(ICIN,SFILE)
        IF(ILK .LT. 0) STOP ' BAD LOOKUP.'
          INBK = 0
          SUMMIN = 10000.
          SUMMAX = 0.
          IF(NSEC.LT.0) GO TO 1352
          IF(NSEC.GE.4) GO TO 7001
            MBLOCK = NSEC*40 - 10
            GO TO 5001
7001      MBLOCK = 150
          INBK = 40
5001      CALL TIME(TM)
          WRITE(7,1000) (TM(M),M=1,8)
5000      CONTINUE
            NWD = IREADW(NSIZE,IBUF,INBK,ICIN)
            IF (NWD .LT. 0) STOP ' READ ERROR.'
              JJ = 1
5010          SUM = 0.
              DO 5030 J=1,128
                  SUM = SUM + IABS(IBUF(JJ))
                  JJ = JJ + 1
5030          CONTINUE
              SUM = SUM/128.
              IF(SUM .LT. SUMMIN) SUMMIN=SUM
              IF(SUM .GT. SUMMAX) SUMMAX=SUM
              JJ = JJ - 64
              IF(JJ .LT. NSIZE) GO TO 5010
              INBK = INBK + NW
        IF ( INBK .LE. MBLOCK ) GO TO 5000
        CALL CLOSEC(ICIN)
        CALL IFREEC(ICIN)
        CALL TIME(TM)
        WRITE(7,2000) (TM(M),M=1,8)
        TYPE 3011, IBEL
        WRITE(7,5200) SUMMAX,SUMMIN
        DO 5500 I = 1, NSIZE
          IBUF(I) = 0
          IOUT(I) = 0
          IHOLD(I)= 0
5500    CONTINUE
        GO TO 6030
6023    FORMAT(F10.4)
6030    WRITE(7,1100)
        READ(5,1200) AP
        AP = (1.- AP) * (SUMMAX - SUMMIN) + SUMMIN
        WRITE(7,1300)
        READ(5,1220) TP
C
C
6050    CALL TIME(TM)
        TYPE 2020,(TM(M),M=1,8)
C                           ---DIRECT ACCESS PROCEDURES FOR
C                           ---INPUT AND OUTPUT FILES---
        ICIN = IGETC()
        IF(ICIN .LT. 0) STOP ' NO INPUT CHANNEL.'
        ILK=LOOKUP(ICIN,SFILE)
        IF(ILK .LT. 0) STOP ' BAD LOOKUP.'
        ICOUT=IGETC()
        IF(ICOUT .LT. 0) STOP ' NO OUTPUT CHANNEL'
        NB=IENTER(ICOUT,OFILE,-1)
D       TYPE 1400,NB
D1400   FORMAT(1X,' NUMBER OF BLOCKS ALLOCATED FOR OUTPUT FILE =',I4)
        IF(NB.LT.0) STOP ' ENTER FAILED'
```

```
C                        ---INITIALIZE PARAMETERS---
          MBLOCK = NSEC*40 - 10
          COUNT = 0.0
          INBK = 0
          IBKOUT = 0
          IDONE = 0
          LHOLD = 0
          IH = 0
          LAST = 0
          TPKT = 0
          FLAG = 0
          IBEG = 1
C                        ---PAUSE COMPRESSION BEGINS---
1         CONTINUE
C
            NWD=IREADW(NSIZE,IBUF,INBK,ICIN)
            IF(NWD .LT. 0) STOP ' READ ERROR.'
C                        -----STARTING PAUSOMETRY----
            IKS = 0
            JK = 0
50          IPT = 0
60          SUM = 0.
            DO 65 I =1,10
              JK = JK + 1
              SUM = SUM + IABS(IBUF(JK))
65          CONTINUE
            SUM = SUM / 10.
C                        -----TESTING AP PARAMETER-----
            IF(SUM .LT. AP) GO TO 70
              IF(FLAG .EQ. 0) GO TO 120
              FLAG = 0
              IEND = JK - 10
              ITOT = IEND - IBEG + 1
              IF(LHOLD .EQ. 0) GO TO 8010
              LHOLD = 0
8020          CALL CMPRES(IHOLD,IHBEG,IHEND,IHTOT)
              IF(IDONE .LT. 1) GO TO 8010
D             TYPE 8022,IBKOUT,INBK
D8022         FORMAT(1X,' IWRITW WITH IBKOUT=',I5,' INBK=',I5)
              NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
D             TYPE 8024,NB
D8024         FORMAT(1X,' NUMBER OF BLOCK SIZE = ',I4)
              IBKOUT=IBKOUT+NW
              IF(IDONE .EQ. 2) GO TO 8020
8010          CALL CMPRES(IBUF,IBEG,IEND,ITOT)
              IF(IDONE .LT. 1) GO TO 120
D             TYPE 8012,IBKOUT,INBK
D8012         FORMAT(1X,'IWRITW WITH IBKOUT=',I5,' INBK=',I5)
              NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
D             TYPE 8014,NB
D8014         FORMAT(1X,' NUMBER OF BLOCK SIZE = ',I4)
              IBKOUT=IBKOUT+NW
              IF(IDONE .EQ. 2) GO TO 8010
120         TPKT = 0
            IF( JK .LT. NSIZE) GO TO 50
            GO TO 4
70          IPT = IPT + 1
C                            -----TESTING TP PARAMETER-----
            IF((IPT +TPKT) .LT. TP) GO TO 80
              IBEG = JK + 1 - (IPT * 10)
              FLAG = 1
80          IF(JK .LT. NSIZE) GO TO 60
            TPKT = TPKT + IPT
            IF(TPKT .GT. 30000) GO TO 90
            IF(FLAG.EQ.0) GO TO 8040
85          IF(LHOLD.EQ.0) GO TO 8030
```

```
             LHOLD = 0
             IHBEG=1
             IHEND=IHTOT
8000         CALL CMPRES(IHOLD,IHBEG,IHEND,IHTOT)
           IF(IDONE .LT. 1) GO TO 8030
D            TYPE 8002,IBKOUT,INBK
D8002        FORMAT(1X,'IWRITW WITH IBKOUT=',I5,' INBK=',I5)
             NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
D            TYPE 8004,NB
D8004        FORMAT(1X,' NUMBER OF BLOCK SIZE = ',I4)
             IBKOUT=IBKOUT + NW
             IF(IDONE .EQ. 2) GO TO 8000
8030       IF(FLAG.EQ.0) GO TO 4
           FLAG = 0
           IEND=JK
           ITOT=IPT*10
9000       CALL CMPRES(IBUF,IBEG,IEND,ITOT)
           IF(IDONE .LT. 1) GO TO 4
D          TYPE 9002,IBKOUT,INBK
D9002      FORMAT(1X,'IWRITW WITH IBKOUT=',I5,' INBK=',I5)
           NB = IWRITW(NSIZE,IOUT,IBKOUT,ICOUT)
D          TYPE 9004,NB
D9004      FORMAT(1X,' NUMBER OF BLOCK SIZE = ',I4)
           IBKOUT=IBKOUT+NW
           IF(IDONE .EQ. 2) GO TO 9000
           GO TO 4
8040       IH=NSIZE-IPT*10
           LHOLD = 1
           IHTOT = IPT * 10
           DO 9020 J=1,IHTOT
             IH = IH + 1
             IHOLD(J)=IBUF(IH)
9020       CONTINUE
           IF(INBK.LT.MBLOCK) GO TO 85
4          INBK = INBK + NW
         IF(INBK .LE. MBLOCK) GO TO 1
         IF(IDONE.LT.1) CALL IWRITW(LAST,IOUT,IBKOUT,ICOUT)
         GO TO 95
C                                  -----END OF INPUT SIGNAL-----
90       TYPE 93,TPKT
93       FORMAT(' Warning! Excessively long pause of ',I5,' msec so
         &far detected.'/' Check AP or TP parameter, or input sound
         & file'/' Run again.',//)

         CALL CLOSEC(ICIN)
         CALL IFREEC(ICIN)
         STOP ' Interrupted exit'

95       CALL CLOSEC(ICOUT)
         CALL IFREEC(ICOUT)
         CALL CLOSEC(ICIN)
         CALL IFREEC(ICIN)

         CALL TIME(TM)
         WRITE (7,3011) (IBEL,M=1,3)
         WRITE(7,7100) (TM(M),M=1,8)
1000     FORMAT(1X,8A1,'  Searching the average peak amplitude.'/)
1100     FORMAT(' Enter AP parameter( 0.0 =< AP =< 1.0000 )  >',$)
1200     FORMAT(F10.4)
1220     FORMAT(I5)
1300     FORMAT(' Enter TP parameter( 1 =< TP =< 10000 msec) >',$)
1350     FORMAT(' Enter the number of seconds ( 1=<SEC=<60 ) >',$)
2000     FORMAT(1X,8A1,'  Search ended.',/)
2020     FORMAT(T3,8A1,'  Pausé compression begins.  Wait. ',$)
3040     FORMAT(F8.1)
5200     FORMAT(1X,'Pre-detection outputs:',//,1X,'Computed maximum
```

```
      & amplitude =',F10.4,/,1X,'Computed minimum amplitude =',
      &F10.4,5(/))
7100  FORMAT(/,T3,8A1,' PAUSE compression completed.',//)
      COUNT=COUNT/10.
      ICHAN = IGETC()
      IN2(1)=256
      IN2(2)=0
      NB = LOOKUP(ICHAN,OFILE)
      TYPE 6000,COUNT,NB
6000  FORMAT(' Output file name: CMPRSD.SND',/,' Total compressed
      & pauses = ',F10.1,' msec(',I4,' blocks)'/)
      CALL JICVT(NB,JB4)
      CALL JMUL(IN4,JB4,LE4)
6720  IER = IDA(ICHAN,6,400,100,LE4,,120,60)
      TYPE 6710
6710  FORMAT(' Hit <Return> to play again.'/' Otherwise type any
      & key, then hit <Return>    >',$)
      ACCEPT 3010,YORN
      IF(YORN .EQ. BLANK) GO TO 6720
      CALL CLOSEC(ICHAN)
      IF(IFREEC(ICHAN).NE.0) STOP ' IFREEC failed'
      TYPE 6666
6666  FORMAT(' Try with other TP and AP (Y/N) ?',$)
      ACCEPT 3010,YORN
      IF(YORN.EQ.89) GO TO 6030
      ISWICH = 2
6695  TYPE 6670
6670  FORMAT(' Options (Type a letter of your choice below.)',///
      &,T5,' S --- to creat a compressed SIGNAL.',/
      &,T5,' H --- to get PRINT OUTS  of pauses and signals.',/
      &,T5,' Q --- to QUIT this program.',//,T5,'>',$)
      ACCEPT 3010,YORN
      IF(YORN.EQ.81) GO TO 6680
      ISWICH = 2
      IF(YORN.EQ.72) CALL CHAIN(VCH,ISWICH,34)
      IF(YORN.EQ.83) CALL CHAIN(SCH,ISWICH,34)
      TYPE 6690
6690  FORMAT(' Invalid choice. Try again !'/)
      GO TO 6695
6680  CONTINUE
      CALL EXIT
      END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C       Program CMPRES is called from AUTOCM, POZCMP, SIGCMP, and
C       POZV1.
C       CMPRES fills output buffer by the (pauses/signals) one
C       NSIZE (10 blocks = 2560 words) at a time.
C
C       Written by :  Hirotaka Nakasone
C       Last modified: 27-March, 1982
C       Department of Audiology and Speech Sciences
C       Michigan State University
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        SUBROUTINE CMPRES(INB,IB,IE,IT)
        COMMON/B1/COUNT,IOUT,NSIZE,IDONE,LAST
        INTEGER INB(2560),IOUT(2560)

D       TYPE 100,LAST,IDONE,IB,IE
D100    FORMAT(//,1X,' AT CMPRES ENTERANCE, LAST=',I7,' IDONE=',I2,/,
D       &T24,' IBEG=',I7,' IEND=',I7)
        IF(IT .LE. 0) GO TO 4
        IF((LAST+IT) .LE. NSIZE) GO TO 2
          ITEMP = NSIZE - LAST - 1
          T = 0.
          DO 1 I=IB,IB+ITEMP
            LAST = LAST + 1
            IOUT(LAST)=INB(I)
            IT = IT - 1
            T = T + 1.
1         CONTINUE
          COUNT=COUNT+T
          LAST = 0
          IDONE = 2
          IB = IB + ITEMP + 1
          GO TO 4
C
2       CONTINUE
        T = 0.
        DO 3 I = IB,IE
          LAST = LAST + 1
          IOUT(LAST)=INB(I)
          T = T + 1.
3       CONTINUE
        COUNT=COUNT+T
        IF(LAST.LT.NSIZE) GO TO 5
          LAST=0
          IDONE=1
          GO TO 4
5       CONTINUE
        IDONE=0
4       CONTINUE
D       TYPE 200,LAST,IDONE,IB,IE
D200    FORMAT(1X,' LEAVING CMPRSD WITH LAST =',I7,' IDONE=',I2,/,
D       &T23,' IBEG=',I7,' IEND=',I7,///)
        RETURN
        END.
*
```

```
CHHHH PROGRAM SHORT8:                                        HHHHHHHHHHHH
C
C     SHORT8 is designed to creat a set of parent files of
C     short-term spectra. A parent file can have as many as
C     400 short-term spectra. 2 spectra are stored in a 1-block
C     INTEGER*2.
C
C     Input sound file resides in DK1 with .SND extension.
C
C     Output short-term spectrum will be assigned .STS
C     extension name.
C
C     This output file is then stored in DK0.
C
C     To compile:
C        *SHORT8=SHORT8/U
C     To link:
C        *SHORT8=SHORT8,FFT10H,SHUFFL,TABE10,SYSLIB/F
C
C     Written by:  Hirotaka Nakasone
C           Date:  May 4, 1983
C        Updated:  May 6, 1983
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

      INTEGER FILEIN(5,6),FILOUT(5,5,6),NSEC(5),WCNT
      INTEGER IBUF(256),NSPT(5),DATAIN(4),DATOUT(4),LBUF(256)
      REAL TX(256)
      COMPLEX F(256)
      COMMON /FFTCOM/F
      LOGICAL*1 TM(8),BUG, YN
      DATA DATAIN(1)/3RDK0/,DATAIN(4)/3RSND/
      DATA DATOUT(1)/3RDK0/,DATOUT(4)/3RSTS/
      FUNIT = 5000./128.
      BUG = .FALSE.

D     TYPE 10
10    FORMAT(' DEBUG (Y/N) ?',$)
D     ACCEPT 11, YN
11    FORMAT(1A1)
D     IF(YN.EQ.89) BUG = .TRUE.
      TYPE 20
20    FORMAT(' Which disk has input file (1/0) ?',$)
      ACCEPT 25, NDISK
25    FORMAT(I1)
      IF(NDISK.EQ.1) CALL IRAD50(3,'DK1',DATAIN(1))
      TYPE 30
30    FORMAT(' Which disk will store output file (1/0) ?',$)
      ACCEPT 25, NDISK
      IF(NDISK.EQ.1) CALL IRAD50(3,'DK1',DATOUT(1))

      TYPE 100
100   FORMAT(' Number of input sound files(max=5)    >',$)
      ACCEPT 200,NFILES
200   FORMAT(I3)
      DO 300 I=1,NFILES
        TYPE 400,I
```

```
400       FORMAT(1X,'Type input file #',I2,' (6A1)  > ',$)
          READ(5,500) (FILEIN(I,J),J=1,6)
500       FORMAT(6A1)
          TYPE 510
510       FORMAT(' Number of seconds of this input file   > ',$)
          ACCEPT 200,NSEC(I)
          TYPE 600
600       FORMAT(' Number of parent spectra from this file
         &(max=5)   > ',$)
          ACCEPT 200,NSPT(I)
          TYPE 700
700       FORMAT(' Type all parent spectra names (6A1) below.'/)
          DO 800 K=1,NSPT(I)
            TYPE 900,K
900         FORMAT(T20,' For output parent file #',I2,'    > ',$)
            READ(5,500)(FILOUT(I,K,J),J=1,6)
800       CONTINUE
300     CONTINUE

C                       END OF INPUT PROCEDURES. NOW BEGIN LOOP FOR
C                       ALL INPUT FILES.

        NSAMP = 256
        DO 1100 II=1,NFILES
          CALL TIME(TM)
          TYPE 1110,(TM(M),M=1,8),(FILEIN(II,KK),KK=1,6)
1110      FORMAT(1X,8A1,' BEGIN INPUT FILE  DK0: ',6A1,'.SND'/)
          CALL CVRAD1(DATAIN,FILEIN,II)
          ICH = IGETC()
          IF(LOOKUP(ICH,DATAIN).LT.0) STOP 'BAD LOOKUP'
          KSIZE = (NSEC(II)/NSPT(II))*40
          IF(KSIZE.LE.0) STOP ' WRONG KSIZE'
          MBLOCK = KSIZE-4
          NBLK = 0
          DO 1200 JJ= 1,NSPT(II)
            IH = 0
            NST = 0
            CALL CVRAD2(DATOUT,FILOUT,II,JJ)
            ICOUT = IGETC()
            NB = IENTER(ICOUT,DATOUT,-1)
C
5111        NWD = IREADW(256,IBUF,NBLK,ICH)
D           KBLK=NBLK
            IF(NWD.LT.0) STOP ' READ ERROR'
C
            DO 110 J=1,256
              TR = FLOAT(IBUF(J))/100.
              F(J)=CMPLX(TR,0.0)
110         CONTINUE


D           TYPE 3, KBLK
D3          FORMAT(1X,' NBLK =',I4)
C                             --- NOW CALL THE FFT ALGORITHM ---
C
            CALL FFT10H(8)
C
C                             --- CONVERT RETURNED REAL AND IMAGINARY
C                                 COMPONENTS TO ABSOLUTE VALUES ---

            DO 7010 I= 1,128
              TX(I)=CABS(F(I))
              LBUF(IH+I)=IFIX(TX(I))
              FREQ = FUNIT * (I-1)
D             IF(BUG) TYPE 7014,I,TX(I),FREQ,IH+I,LBUF(IH+I)
D7014         FORMAT(6X,I4,7X,E15.4,3X,6X,F8.1,I4,6X,I5)
```

```
7010        CONTINUE
C
          IH = IH + 128
          IF(IH.LE.128) GO TO 7030
             IH = 0
C                                 WRITE 1 BLOCK OF STS
             CALL IWRITW(256,LBUF,NST,ICOUT)
C                                 UPDATE NUMBER OF BLOCKS
             NST = NST + 1
7030       NBLK = NBLK + 4
           IF(NBLK.LE.MBLOCK) GO TO 5111
           MBLOCK = MBLOCK + KSIZE

          CALL CLOSEC(ICOUT)
          IF(IFREEC(ICOUT).NE.0) STOP 'ERROR IN IFREEC'
          CALL TIME(TM)
          TYPE 7050,(TM(M),M=1,8),(FILOUT(II,JJ,KK),KK=1,6)
7050      FORMAT(T20,8A1,T45,'DK1:',6A1,'.STS'/)
1200    CONTINUE

        CALL CLOSEC(ICH)
        IF(IFREEC(ICH).NE.0) STOP ' ERROR IN IFREEC(ICH)'
        TYPE 1250,II,(FILEIN(II,KK),KK=1,6)
1250    FORMAT(1X,'<< END OF INPUT FILE #',I2,'    ',6A1,'.SND'/)
1100    CONTINUE
        TYPE 1300
1300    FORMAT(20(/),1X,'SUMMARY OF OUTPUT FILES IN',/
        &,1X,'CATEGORY #     SHORT-TERM SPECTRA FILE NAME'/)


        DO 1310 I=1,NFILES
          DO 1320 J = 1,NSPT(I)
            TYPE 1330,I,(FILOUT(I,J,K),K=1,6)
1330        FORMAT(4X,I3,T20,6A1,'.STS')
1320      CONTINUE
1310    CONTINUE
C
C
        CALL EXIT
        END
C

        SUBROUTINE CVRAD1(OUTFIL,INFILE,I2)
        INTEGER OUTFIL(4),INFILE(5,6)
        LOGICAL*1 DUMMY(6)
        DO 1 I=1,6
          DUMMY(I) = INFILE(I2,I)
1       CONTINUE
        IER=IRAD50(6,DUMMY,OUTFIL(2))
D       TYPE 10,IER,(DUMMY(M),M=1,6)
D10     FORMAT(1X,'NUMBER OF CHAR. CONVERTED(IN)=',I3,' : ',6A1,/)
        RETURN
        END
C
        SUBROUTINE CVRAD2(OUTFIL,INFILE,I2,J2)
        INTEGER OUTFIL(4),INFILE(5,5,6)
        LOGICAL*1 DUMMY(6)
        DO 1 I=1,6
          DUMMY(I) = INFILE(I2,J2,I)
1       CONTINUE
        IER = IRAD50(6,DUMMY,OUTFIL(2))
D       TYPE 10,IER,(DUMMY(M),M=1,6)
D10     FORMAT(1X,'NUMBER OF CHAR. CONVERTED(OUT)=',I3,' : ',6A1,/)
        RETURN
```

```
C  FILE   IDSLTS.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C  PROGRAM IDSLTS:
C
C  IDSLTS is designed to generate several different types of spectra,
C  from a parent short-term spectra.
C
C  (1)  Mean Deviation spectrum, defined by
C
C                D(k) = [ SUM [ S(i,j,k) - m(i,j,k) ]**2 ] / I
C
C                  for i = 1, 2, 3, ..., I
C                  I = number of short-term spectra
C                  for j = 1,2,3, ..., J
C                  J = number of frequency components, and
C                  k = 1,2, ..., K,  k is index of kth pattern.
C                  m(i,j,k) = ( SUM [s(i,j,k)] ) / I
C
C  (2)  Intensity deviation spectrum (IDS), defined by
C
C                C(k) = [ SUM [ABC( S(i,j,k) - m(i,j,k) )] ] / m(i,j,k)
C
C                  where m and i defined above.
C
C  (3)  Long-term averaged spectrum <LTS), defined by
C
C                L(k) = [ SUM [S(i,j,k)] ] / I
C
C                  where i defined above.
C
C  All the spectra defined above are derived from the same set of
C  parent short-term spectra.
C
C  Input        Short-term spectra file with ext. STS
C  Output       (1) Intensity deviation spectra (IDS) with ext. IDS
C               (2) Long-term averaged spectra (LTS)  with ext. LTS
C
C  Subprogram called:  FUNCTION NCHECK for file input verification.
C
C  To compile        :  IDSLTS = IDSLTS/U
C  To link           :  IDSLTS = IDSLTS,NCHECK,SYSLIB/F
C
C  Programmed by     :  Hirotaka Nakasone
C  Date              :  10 Oct., 1982
C  Last modified     :  14 Oct., 17, May,1983
C
C  Department of Audiology and Speech Sciences
C  Michigan State University
C  East Lansing, Michigan
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

      COMMON /NAME/ FNAMES
      DOUBLE PRECISION DEXT
      REAL SUM1(128),AIDS(128)
      INTEGER IBUF(256),EXT(4)
      LOGICAL*1 FNAMES(15,100),LDATE(9),LTIME(8),LDEV(3),BUG
      LOGICAL*1 LEXT(3),MFILE(15),FMT(60),L12(12),LN(9)
      LOGICAL*1 ED(3),EL(3),ES(3),NEED,LEX(3,5),LFILE(15),YN
      LOGICAL*1 NORMD(3),NORML(3)
```

```
        DATA MFILE(1)/'D'/,MFILE(2)/'K'/,MFILE(4)/':'/,MFILE(11)/'.'/
        DATA DEXT/3RSTS/,EXT(4)/3RSTS/
        DATA ED(1)/'I'/,ED(2)/'D'/,ED(3)/'S'/
        DATA EL(1)/'L'/,EL(2)/'T'/,EL(3)/'S'/
        DATA ES(1)/'S'/,ES(2)/'T'/,ES(3)/'S'/

1       FORMAT(1A1)
2       FORMAT(6A1)
3       FORMAT(1A1,3A1)
4       FORMAT(15A1)
5       FORMAT(I3)
6       FORMAT(12A1)
7       FORMAT(14A1)
8       FORMAT(9A1,3A1)
9       FORMAT(1X,/)
10      FORMAT(1X,8A1,3X,9A1,3X,'PROGRAM:   IDSLTS.'///)

        NORMD(1) = ED(1)
        NORMD(2) = ED(2)
        NORMD(3) = 78
        NORML(1) = EL(1)
        NORML(2) = EL(2)
        NORML(3) = 78


        CALL DATE(LDATE)
        CALL TIME(LTIME)
        WRITE(7,10) LTIME,LDATE

        BUG = .FALSE.
        TYPE 15
15      FORMAT(' DEBUG(Y/N) ?',$)
        ACCEPT 1,YN
        IF(YN.EQ.89) BUG = .TRUE.

        TYPE 20
20      FORMAT(' MASTER FILE OF FILE NAMES AVAILABLE(Y/N) ?',$)
        ACCEPT 1, YN
        IF(YN.NE.89) GO TO 200

C TO HERE TO READ THE MASTER FILE NAME OF FILES.
        NEED = .FALSE.
        TYPE 30
30      FORMAT(' ENTER MASTER FILE NAME [DEV:FILNAM.EXT]   >',$)
        READ(5,7) (MFILE(M),M=1,14)
        MFILE(15) = 0


C OPEN CHNN. FOR FILE INPUT.
        CALL ASSIGN(12,MFILE,14,'OLD')
        READ(12,50) (LFILE(L),L=1,14),NUMF,FMT
50      FORMAT(14A1,I3,60A1)
        DO 60 I = 1,NUMF
           READ(12,7) (FNAMES(M,I),M=1,14)
D          WRITE(7,65) (FNAMES(M,I),M=1,14),I
D65        FORMAT(1X,14A1,5X,I3)
60      CONTINUE
        CALL CLOSE(12)
        GO TO 250


C TO HERE IF THE MASTER FILE NAME IS NOT AVAILABLE.
200     CONTINUE
        NEED = .TRUE.
        TYPE 100
100     FORMAT(1X,'TOTAL NUMBER OF PARENT STS FILES ?',$)
```

```
        READ(5,5) NUMF
        TYPE 110
110     FORMAT(1X,'ENTER ALL PARENT STS FILE NAMES BELOW.'//)

C  LOOP FOR FILE NAME INPUT.
        DO 120 I = 1,NUMF
           WRITE(7,130) I
130        FORMAT(' STS FILE NAME #',I3,'  ',$)
           IF(LCHECK(DEXT,ICHAN,I).LT.0) STOP 'ERROR LCHECK'
           CALL CLOSEC(ICHAN)
           IF(IFREEC(ICHAN).NE.0) STOP '*ERROR IFREEC*'
120     CONTINUE

C BEGIN MAIN LOOP TO COMPUTE IDS AND LTS.
250     CONTINUE
        KOUNT = 0
        DO 300 I = 1,NUMF

C                                             INITIALIZE
                DO 305 J = 1, 127
                   SUM1(J) = 0.0
                   AIDS(J) = 0.0
305             CONTINUE

C                                             OPEN FOR A FILE
          ICHAN = IGETC()
          CALL IRAD50(3,FNAMES(1,I),EXT(1))
          CALL IRAD50(6,FNAMES(5,I),EXT(2))
          CALL IRAD50(3,FNAMES(12,I),EXT(4))
          CALL R50ASC(9,DEXT,LN(1))
          MAXBLK = LOOKUP(ICHAN,EXT)
          IF(MAXBLK .LT. 0) STOP '*BAD LOOKUP*'
          NBLK = 0
            JH = 0

310       NWORD = IREADW(256,IBUF,NBLK,ICHAN)
          IF(NWORD.LT.0) STOP'ERROR IN IREADW'

CDEBUGDEBUGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGDEBUG
          IF(.NOT. BUG) GO TO 315
          IF(NBLK.GE.2) GO TO 315
          TYPE 1100
1100      FORMAT(1X,//,'  STS#1      STS#2'/)
               DO 312 IB = 2,128
                  WRITE(7,1000) IBUF(IB),IBUF(IB+128)
1000              FORMAT(1X,I6,3X,I6)
312            CONTINUE
          TYPE 9
CDEBUGDEBUGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGDEBUG

315       CONTINUE
          DO 320 J = 1,127
               SUM1(J) = SUM1(J) + FLOAT(IBUF(J+1+JH))
320       CONTINUE
          JH = JH + 128
          IF(JH.LE.128) GO TO 315
          JH = 0
          NBLK = NBLK + 1
          IF(NBLK.LT.MAXBLK) GO TO 310
          DSTS = FLOAT(MAXBLK*2)

C                        COMPUTE THE AVERAGE INTENSITY OF EACH
C                        FREQUENCY COMPONENT.
```

```
          DO 330 J = 1, 127
             SUM1(J) = SUM1(J)/DSTS
330       CONTINUE
C
C         SUM1 NOW HOLDS THE AVERAGE INTENSITY.
C         RESET NBLK = 0
          NBLK = 0
            JH = 0
350       NWORD = IREADW(256,IBUF,NBLK,ICHAN)
355       CONTINUE
          DO 360 J = 1,127
             AIDS(J) = AIDS(J) + ABS(FLOAT(IBUF(J+1+JH))-SUM1(J))
360       CONTINUE
          JH = JH + 128
          IF(JH.LE.128) GO TO 355
          JH = 0
          NBLK = NBLK + 1
          IF(NBLK .LT. MAXBLK) GO TO 350

          DO 365 J = 1,127
               AIDS(J) = AIDS(J) / SUM1(J)
365       CONTINUE

C                                                 END OF ONE STS.  NOW,
C                                                 WRITE THE RESULTS.
          TYPE 9
          CALL WTSPEC(127,AIDS,I,ED)
          CALL WTSPEC(127,SUM1,I,EL)
          CALL NORPER(127,AIDS)
          CALL WTSPEC(127,AIDS,I,NORMD)
          CALL NORPER(127,SUM1)
          CALL WTSPEC(127,SUM1,I,NORML)

C                                                 CLOSE CHANNEL.
          CALL CLOSEC(ICHAN)
          IF(IFREEC(ICHAN).NE.0) STOP 'ICHAN NOT FREED.'
          KOUNT = KOUNT + 1
300     CONTINUE

C                                                 END OF ALL STS'S
        TYPE 9
        TYPE 9
        CALL TIME(LTIME)
        WRITE(7,400) LTIME,KOUNT
400     FORMAT(1X,8A1,//,' TOTAL NUMBER OF STS FILES PROCESSED =',I3,
        &///)

        DO 405 I = 1,3
               LEX(I,1) = ED(I)
               LEX(I,2) = NORMD(I)
               LEX(I,3) = EL(I)
               LEX(I,4) = NORML(I)
               LEX(I,5) = ES(I)
405     CONTINUE
        N = 4
        IF(NEED) N = 5
        YN = 0
        DO 420 I = 1,N
           WRITE(7,430) (LEX(J,I),J=1,3)
430        FORMAT(1X,'MASTER FILE NAME OF ',3A1,' FILES
        & [DEV:FILNAM.EXT]  >',$)
           READ(5,7) (MFILE(M),M=1,14)
           IF(MFILE(1).EQ.48) GO TO 420
           MFILE(15) = 0
           CALL ASSIGN(61,MFILE,14,'NEW')
           WRITE(61,440) (MFILE(M),M=1,14), KOUNT
```

```
440         FORMAT(14A1,I3,'  (5E15.7)')
            DO 450 II = 1,KOUNT
               WRITE(61,7)(FNAMES(M,II),M=1,11),(LEX(M,I),M=1,3)
450         CONTINUE
            CALL CLOSE(61)
420      CONTINUE
         CALL EXIT
         END


CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
         SUBROUTINE WTSPEC(NFREQ,SDATA,IX,E3)
C        Written by: Hirotaka Nakasone                                    H
C        Date      : 17 May 1983                                          H
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
         COMMON /NAME/ FNAMES
         LOGICAL*1 FNAMES(15,100),E3(3)
         REAL SDATA(NFREQ)

         DO 1 I = 1,3
            FNAMES(I+11,IX)= E3(I)
1        CONTINUE

         CALL ASSIGN(61,FNAMES(1,IX),14,'NEW')
         WRITE(61,3) (SDATA(J),J=1,NFREQ)
3        FORMAT(5E15.7)
         CALL CLOSE(61)
         WRITE(7,4) (FNAMES(L,IX),L=1,14)
4        FORMAT(1X,14A1)
         RETURN
         END

CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C FILE:  NORPER.FOR
C SUBROUTINE NORPER(NFREQ,SDATA)
C
C NORPER IS DESIGNED TO NORMALIZE IDS AND LTS FILES.
C NORMALIZED DATA WILL BE REPRESENTED BY NUMBERS
C BETWEEN 0 AND 100.
C SDATA IS A REAL ARRAY DIMENSIONED BY NFREQ.
C EACH DATUM WILL BE DEVIDED BY THE SUM OF NFREQ VALUES.
C
C IT HAS BEEN IMPLEMENTED AS A PART OF PRE-PROCESSING
C OF SPEECH DATA FOR A PH.D. DISSERTATION RESEARCH.
C ON THE PDP 11/40, DEPARTMENT OF AUDIOLOGY AND SPEECH
C SCIENCES, MICHIGAN STATE UNIVERSITY.
C
C WRITTEN BY HIROTAKA NAKASONE, 19 MAY 1983.
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
         SUBROUTINE NORPER(NFREQ,SDATA)
         REAL SDATA(NFREQ)
         TOTAL = 0.
         DO 1 I = 1, NFREQ
                 TOTAL = TOTAL + SDATA(I)
1        CONTINUE
         DO 2 I = 1, NFREQ
                 SDATA(I) = SDATA(I)/TOTAL
2        CONTINUE
         RETURN
         END
*
```

```
C     FILE = DENCOR.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C PROGRAM DENCOR.FOR
C
C DENCOR COMPUTES THE SPEARMAN PRODUCT MOMENT CORRELATION COEFFICIENTS
C (ABSOLUTE) OF ALL PAIRRING OF FREQUENCY COMPONENTS OF SPECTRA, I AND J,
C OVER NUMF FILES.
C
C A ((NDATA)*(NDATA-1)/2) COR. MATRIX WILL BE USED AS INPUT TO
C A CLUSTERING PROGRAM (COMPLETE LINK) TO SEARCH GROUPINGS BY
C FREQUENCY COMPONENTS.  A NDATA IS THE TOTAL NUMBER OF DISCRETE
C FREQUENCIES, EACH REPRESENTING ABOUT 39 HZ, FROM 117 TO 3978 HZ.
C THE OUTPUT OF DENCOR IS A SIMILARITY MATRIX, I.E., THE GREATER
C THE VALUE IN THE MATRIX IS, THE MORE SIMILAR TO ONE ANOTHER.
C
C SUBPROGRAMS:
C   FUNCTION LCHECK (H.NAKASONE):  FILE VERIFICATION ROUTINE.
C   FUNCTION INDEX:  MATRIX FORMATTING ROUTINE.
C
C PROGRAMMED AS A PART OF FEATURE EXTRACTION PROCEDURE FOR A
C PH.D. DISSERTATION (COMPUTER VOICE IDENTIFICATION BY USING
C INTENSITY DEVIATION SPECTRA AND FUNDAMENTAL FREQUENCY CONTOUR)
C BY THE AUTHOR.
C
C WRITTEN BY :HIROTAKA NAKASONE, 22-MAY-1983
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES, MSU.
C
C TO COMPILE:    DENCOR=DENCOR/UC/DJ
C TO LINK   :    DENCOR=DENCOR,SYSLIB/F
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        COMMON /NAME/FNAMES
        COMMON /NUMF/N
        DOUBLE PRECISION EXT,PROG
        REAL DMAT(4950),DAT(30,100)
        LOGICAL*1 FNAMES(15,100),YN,LNAME(15),FMT(60)
        LOGICAL*1 NEXT(3)
        DATA LNAME(1)/'D'/,LNAME(2)/'K'/,LNAME(4)/':'/
        DATA LNAME(10)/'.'/
        DATA PROG/12RDK1ORDPR SAV/

        LNAME(15) = 0
1       FORMAT(1A1)
2       FORMAT(I3)
3       FORMAT(6A1)
4       FORMAT(14A1)
6       FORMAT(15A1)
5       FORMAT(12A1)
8       FORMAT(1X,' WAIT.')
9       FORMAT(1X,/)

        N = 100
        DO 10 I = 1, 4950
           DMAT(I) = 0.
10      CONTINUE

C  READ DATA FILES FROM A MASTER FILE.
```

```
300     TYPE 210
210     FORMAT(' ENTER MASTER FILE NAME (DEV:FILNAM.EXT) >',$)
        READ(5,4) (LNAME(L),L=1,14)
        CALL ASSIGN(13,LNAME,14,'OLD')

        READ(13,410) (LNAME(L),L=1,14),NUMF,(FMT(M),M=1,60)
410     FORMAT(14A1,I3,60A1)
        DO 420 I = 1, NUMF
           READ(13,4) (FNAMES(M,I),M=1,14)
           WRITE(7,425) (FNAMES(M,I),M=1,14)
425        FORMAT(1X,14A1)
           CALL ASSIGN(12,FNAMES(1,I),14,'OLD')
           READ(12,FMT) DUM, DUM
           READ(12,FMT) (DAT(I,J),J=1,100)
           CALL CLOSE(12)
420     CONTINUE
        CALL CLOSE(13)

C   COMPUTE SUM(N) AND SSQ(N) FOR N FREQUENCIES OVER NUMF SPECTRA.

        I1 = 3
        I2 = 102
        RN = FLOAT(NUMF)
        ND = 0
        DO 620 I=1,N-1
           DO 610 J=I+1,N
              SUMX = 0.
              SUMY = 0.
              SUMXX = 0.
              SUMYY = 0.
              SUMXY = 0.
              DO 600 K=1,NUMF
                 SUMX = SUMX + DAT(K,I)
                 SUMY = SUMY + DAT(K,J)
                 SUMXX= SUMXX+ DAT(K,I)*DAT(K,I)
                 SUMYY= SUMYY+ DAT(K,J)*DAT(K,J)
                 SUMXY= SUMXY+ DAT(K,I)*DAT(K,J)
600           CONTINUE

              XBAR = SUMX / RN
              YBAR = SUMY / RN
              SDX = (SUMXX/RN - XBAR*XBAR)**0.5
              SDY = (SUMYY/RN - YBAR*YBAR)**0.5
              SDXSDY = SDX*SDY
              TEST = SUMXY/RN - XBAR*YBAR
              IF(SDXSDY .GT. 1.E-10) GO TO 605
                R = 1.0
                GO TO 606
605           R = ABS(TEST / SDXSDY)
606           DMAT(INDEX(I,J)) = R
              ND = ND+ 1
CD            WRITE(7,666) R,I,J
CD666         FORMAT(1X,'R=',E15.7,' FOR ',I3,' VS.',I3,' DMAT')
610        CONTINUE
620     CONTINUE

        CALL ASSIGN(13,'PROXTP.DAT',10,'NEW')
        WRITE(13,700) N
700     FORMAT(I3,'-1'/ ' (10F8.5)')
        DO 710 I = 1, N-1
           WRITE(13,720) (DMAT(INDEX(I,J)),J=I+1,N)
710     CONTINUE

720     FORMAT(10F8.5)
        CALL CLOSE(13)
```

```
      TYPE 800
800   FORMAT(' EXITING DENCOR.  HAVE YOU SET TTY WIDTH=132 ?'
     &/' IF NOT, INTERRUPT DENCOR AND DO FOLLOWS.'
     &//'  .SET TTY WIDTH=132 <RET>
     &'/'  .RUN ORDPR <RET>'//' IF YOU HAVE, JUST WAIT.'/////)

C          --- PROGRAM CHAINING TO ORDPR (INITIAL PART OF
C              HIERARCHICAL CLUSTERING BY COMPLETE-LINK METHOD ---

      CALL CHAIN(PROG,,0)
      CALL EXIT
      END


      INTEGER FUNCTION INDEX(I,J)
C
      COMMON /NUMF/ N
      INTEGER ROW,COL
C
      ROW = MINO(I,J)
      COL = MAXO(I,J)
      INDEX = (ROW-1)*N-ROW*(ROW+1)/2+COL
      RETURN
      END

*
```

```
C FILE:         FRATIO.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C FRATIO IS DESIGNED TO COMPUTE A SET OF F-RATIOS OF IDS AND LTS
C FEATURES WHICH ARE SENT AS SUBSETS BY THE PRECEDING CLUSTERING
C PROGRAMS.
C
C INPUT:
C
C (1) A MASTER FILE WHICH CONTAINS NAMES OF IDS FILES.
C (2) HNLIST.TMP, TEMPORARY FILE CREATED BY THE PROGRAM
C     DENCOR, AND DENDRO.  HNLIST.TMP HOLDS INFORMATION
C     ABOUT CLUSTERING GROUPS, AND NODES (OR, DISCRETE FRE-
C     QUENCY COMPONENTS) IN EACH CLUSTERING.
C
C OUTPUT:
C
C (1) F-RATIO'S SUMMARIZED IN THE TABLE WITH OTHER INFO'S.
C
C WRITTEN BY:  HIROTAKA NAKASONE
C        DATE:  30-MAY-83
C
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES,
C MICHIGAN STATE UNIVERSITY, EAST LANSING, MICHIGAN.
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        COMMON /SENSE/ X, F
        COMMON /SENSEP/ NODLST
        REAL A(5,10,100), X(5,10,30), F(30)
        INTEGER NODLST(100),IEND(20),NG,NP
        LOGICAL*1 FNAMES(14,50),LNAME(14),FMT(60),YN
        LOGICAL*1 MYNAME(10)
        DATA NG/10/,NP/5/,NS/10/,MAXNOD/30/
        DATA LNAME(1)/'D'/,LNAME(2)/'K'/,LNAME(3)/'0'/,LNAME(4)/':'/

CHECK IF CHAINED TO, OR RUN FROM THE KEYBOARD.

        CALL RCHAIN(IF,MYNAME,5)

        IF(IF.EQ.-1) WRITE(7,1010) (MYNAME(L),L=1,10)
        IF(IF.NE.-1) TYPE 1020
1010    FORMAT(1X,/////,' PROGRAM FRATIO! FILE SENT =    ',10A1,//)
1020    FORMAT(1X,/////,' PROGRAM FRATIO!'//)

5       FORMAT(1A1)
999     FORMAT(1X,/)
        CALL ASSIGN(12,'HNLIST.TMP',10,'OLD')
        READ(12,10) NODES
10      FORMAT(I3,/)
        READ(12,20) (NODLST(I),I=1,NODES)
20      FORMAT(20I4)
        CALL CLOSE(12)
D       WRITE(7,22) (NODLST(I),I=1,NODES)
D22     FORMAT(1X,20I4)
        IF(IF.EQ.-1) GO TO 3000
        TYPE 30
30      FORMAT(' Master file name (DEV:6A1.EXT) ?',$)
```

```
        READ(5,40) (LNAME(L),L=1,14)
40      FORMAT(14A1)
        GO TO 3500

3000    CONTINUE
        DO 3510 IM = 1,10
                LNAME(IM+4) = MYNAME(IM)
3510    CONTINUE

3500    TYPE 50
50      FORMAT(' Number of clusters ?',$)
        ACCEPT 60, NCLUST
60      FORMAT(I2)

        TYPE 70
70      FORMAT(' Enter ending node location for each clustter.'//)

        IBEG = 1
        DO 80 I = 1, NCLUST
95         TYPE 90, I
90         FORMAT(' For cluster  ',I2,'  >',$)
           ACCEPT 61, IEND(I)
61         FORMAT(I3)
           NC = IEND(I) - IBEG + 1
           IF(NC.LE.MAXNOD) GO TO 99
                TYPE 98
98              FORMAT(' NUMBER OF NODES OUT OF RANGE.')
                GO TO 95
99         TYPE 100, NC
100        FORMAT(' Total number of nodes = ',I2,'  OK ?',$)
           ACCEPT 5, YN
           IF(YN.EQ.78) GO TO 95
           IBEG = IEND(I) + 1
80      CONTINUE
D       TYPE 999
D888    FORMAT(10X,'X          A            I     J     K'/)
        CALL ASSIGN(13,LNAME,14,'OLD')
        READ(13,410) (LNAME(L),L=1,14),NUMF,(FMT(M),M=1,60)
410     FORMAT(14A1,I3,60A1)


        I = 1
        J = 1
        DO 420 IX = 1, NUMF
           READ(13,40) (FNAMES(M,IX),M=1,14)
           CALL ASSIGN(12,FNAMES(1,IX),14,'OLD')
           READ(12,FMT) DUMMY,DUMMY,(A(I,J,K),K=1,100)
D          TYPE 999
           CALL CLOSE(12)
           I = I + 1
           IF(I.LE.NP) GO TO 420
              I = 1
              J = J + 1
420     CONTINUE
        CALL CLOSE(13)

        IBEG = 1
        KCLUST = 0
D       TYPE 999
D       TYPE 888
800     KCLUST = KCLUST + 1
           ND  = IEND(KCLUST) - IBEG + 1
           KNT = 0
           DO 500 K = IBEG, IEND(KCLUST)
              KNT = KNT + 1
              DO 600 J = 1, NG
```

```
                DO 700 I = 1, NP
                   X(I,J,KNT) = A(I,J,NODLST(K))
D        WRITE(7,777) X(I,J,KNT),A(I,J,NODLST(K)),I,J,K,NODLST(K),KNT
D777     FORMAT(1X,F14.5,F14.5,5X,3I5,'    NODLST(K)=',I3,' KNT=',I3)
700             CONTINUE
600          CONTINUE
500       CONTINUE
D         TYPE 999


          CALL SUBFR(NP,NG,ND)
          CALL PRTFR(F,KCLUST,ND,IBEG,IEND(KCLUST))

          IBEG = IEND(KCLUST) + 1

       IF(KCLUST.LT.NCLUST) GO TO 800
       STOP' DONE.  NORMAL EXIT.'
       END


CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
       SUBROUTINE PRTFR(F,KCL,ND,IB,IE)
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

       COMMON /SENSEP/ NODLST
       INTEGER NODLST(100)
       REAL F(30)

D      WRITE(7,50) IB,IE
D50    FORMAT(1X,' IB =',I4,'  IE =',I4,$)
       FINTV = 5000./128.
       IFREQ = INT(FINTV*2.)
D      WRITE(7,60) FINTV,IFREQ
D60    FORMAT('+','FINTV =',F8.2,' IFREQ =',I5,/)
       TYPE 1, KCL
1      FORMAT(1X,///,T5,'F-RATIO TABLE FOR CLUSTER NO. ',I2,/)
       TYPE 2
2      FORMAT(1X,T5,'FREQUENCY(Hz)',T20,'SEQUENCE # ',T40,
       &'F-RATIO'/,T5,'-------------',T20,'-----------',T40,
       &'-------'/)

       K = 0
       DO 5 I = IB, IE
          K = K + 1
D         TYPE 55,NODLST(I)
D55       FORMAT(' NODLST(I)=',I5)
          KF = IFREQ + INT( FINTV * FLOAT(NODLST(I)) + 0.5)
          WRITE(7,10) KF, NODLST(I),F(K)
10        FORMAT(6X,I4,T23,I3,T33,F14.3)
5      CONTINUE
       TYPE 20
20     FORMAT(1X,//,T5,'________________________________________'/)
       RETURN
       END


CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
       SUBROUTINE SUBFR(NP,NG,ND)
C      COMPUTATION PART FOR FRATIOS
C      BY H. NAKASONE

CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

       COMMON /SENSE/ X, F
       REAL X(5,10,30), SUM(10),SSQ(10),F(30)
```

```
C       ND:  HOLDS NUMBER OF DIMENSIONS IN EACH NP.
C       NP:  HOLDS NUMBER OF PATTERNS IN EACH NG.
C       NG:  HOLDS NUMBER OF SPEAKERS (OR GROUPS).

        DFB = FLOAT(NG - 1)
        DFW = FLOAT(NG*NP - NG)
        GN = FLOAT(NP*NG)
        SN = FLOAT(NP)

        DO 1 K = 1, ND
           GSUM = 0.
           GSSQ = 0.
           T = 0.

           DO 2 J = 1, NG
              SUM(J) = 0.
              SSQ(J) = 0.

              DO 3 I = 1, NP
                 SUM(J) = SUM(J) + X(I,J,K)
                 SSQ(J) = SSQ(J) + X(I,J,K)*X(I,J,K)
3             CONTINUE

              GSUM = GSUM + (SUM(J)*SUM(J))/SN
              GSSQ = GSSQ + SSQ(J)
              T = T + SUM(J)

2          CONTINUE

           SSTOT = GSSQ - (T*T)/GN
           SSBET = GSUM - (T*T)/GN
           SSWIT = SSTOT - SSBET
           BGMS = SSBET / DFB
           WGMS = SSWIT / DFW


D          TSWG = 0.
D          DO 4 J = 1, NG
D             TSWG = TSWG + ( SSQ(J) - (SUM(J)*SUM(J))/ SN )
D4         CONTINUE
D          WRITE(7,10) SSTOT,SSBET,SSWIT
D10        FORMAT(' SSTOT=',F14.3,' SSBET=',F14.3,' SSWIT=',F14.3)
D          WRITE(7,12) BGMS,WGMS,TSWG
D12        FORMAT(' BGMS=',F14.3,' WGMS=',F14.3,' TSWG=',F14.3)
           F(K) = 9999.9091
           IF(WGMS.GT.0.) F(K) = BGMS / WGMS
1       CONTINUE
        RETURN
        END

*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C FILE:  UPICK.FOR
C
C UPICK IS DESIGNED TO DETECT FUNDAMENTAL FREQUENCIES IN A
C RUNNING SPEECH BY INTERACTIVE METHOD. THIS PROGRAM MUST BE
C CHAINED TO PROGRAM HPICK, THEN TO FFPICK TO COMPLETE
C MEASUREMENT OF 9 FEATURES OF FFC.
C
C WRITTEN BY:  H. NAKASONE
C       DATE:  26-JUN-83
C
C AS A PART OF SERIES OF SOUND PROCESSING SOFTWARES
C USED IN THE PH.D. DISSERTATION BY THE AUTHOR.
C
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C MICHIGAN STATE UNIVERSITY
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        COMMON /SENSE/ MAXPIC,NBLOCK,NAME12,NSR
        COMMON /ICSPEC/ ISPEC(39)
        COMMON /BUFF/ NBUF,XMAX,XMIN,YMAX,YMIN,YRANGE
        DOUBLE PRECISION HPROG, EXT
        INTEGER*4 JLEN
        INTEGER NDEV, NEXT, NBUF(3600)
        LOGICAL*1 NAME6(6), NAME12(12), YN, BEL, ONE
        LOGICAL*1 ZERO, CHA, REPT, VTAB, DOLBY

        DATA EXT/3RSND/, NSR/10000/, IFILTR/3/, BEL/"007/
        DATA ONE/.TRUE./, ZERO/.FALSE./, REPT/.FALSE./, VTAB/"013/
        DATA HPROG/12RDK1HPICK SAV/

        KNIT = 3600
        TYPE 1
1       FORMAT(' PROGRAM UPICK: '///,1X,' Want DOLBY?',$)
        ACCEPT 77,YN
77      FORMAT(A1)
        DOLBY = .FALSE.
        IF(YN.EQ.89) DOLBY = .TRUE.


500     CALL INIT(NBUF,KNIT)
        CALL SCROL(1,1000)
        TYPE 2000, VTAB
2000    FORMAT(1X,A1)
        CALL SCROL(2,100)
        CALL APNT(0.,200.,1,-5,0,4)
        CALL LVECT(1023.,0.)

        CALL APNT(0.,700.,1,-7,0,1)
        CALL SUBP(1)
        CALL TEXT(' READY ANALOG INPUT.',2,' WHEN READY,
     & HIT RETURN KEY TO START.')
        CALL ESUB
        CALL OFF(1)

        CALL APNT(0.,800.,0,-7,1,1)
```

```
      CALL SUBP(2)
      CALL TEXT(' CONVERSION IN PROCESS. ')
      CALL ESUB
      CALL OFF(2)

      ICHAN = IGETC()
      IF(ICHAN.LT.0) STOP' ICHAN NOT AVAILABLE'
      TYPE 2
2     FORMAT(' NUMBER OF SECONDS (I3) ?',$)
      ACCEPT 3, NSEC
      LEN = NSR/256
      RSEC = FLOAT(NSEC)
      SR = FLOAT(NSR)
      LENB = LEN * NSEC + 1

      TYPE 10
10    FORMAT(1X,'ENTER OUTPUT FILE NAME  >',$)
20    IF(ICSI(ISPEC,EXT,,,0).NE.0) GO TO 20
      IF(ISPEC(16).NE.0) GO TO 30
         TYPE 22
22       FORMAT(' ?HARDWARE ERROR?'//)
      CALL EXIT
30    IF(IFETCH(ISPEC(16)).NE.0) STOP'FETCHING ERROR'

      NBLOCK = IENTER(ICHAN,ISPEC(16),-1)
      IF(NBLOCK.GE.LENB) GO TO 100
         TYPE 40, LENB, NBLOCK
40       FORMAT(1X,'NOT ENOUGH BLOCK SIZE.'//'
      & BLOCK SIZE REQUESTED = ',I4,' AND ALLOCATED =',I4)
      CALL EXIT
100   TYPE 7, BEL
      CALL R50ASC(12,ISPEC(16),NAME12(1))
      CALL JAFIX(SR*RSEC+.5, JLEN)
      CALL ON(1)
      CALL BIT12(ONE)
210   CHA = ITTINR()
      IF(CHA.NE.13) GO TO 210
      CALL BIT12(ZERO)
      CALL OFF(1)
      CALL ON(2)
      TYPE 7, BEL
      I=IDA(ICHAN,7,INT(0.04*SR),INT(0.01*SR),JLEN,,60,0)
      IF(I.NE.0) TYPE 225, I
225   FORMAT(' ERROR IDA    I = ',I4)
      CALL CLOSEC(ICHAN)
      CALL IFREEC(ICHAN)
      CALL OFF(2)
      TYPE 250
250   FORMAT(1X,//,' SOUND INPUT PROCEDURE COMPLETED.  WAIT...'//)
      CALL INIT(NBUF,KNIT)
      NWORD = 9
C                    --- CREATE SUBPICTURES OF SPEECH WAVES  ---

      CALL SUBPIC(DOLBY)
C                    --- CHAINING TO PROGRAM HPICK ---
      CALL CHAIN(HPROG,MAXPIC,NWORD)

3     FORMAT(I3)
7     FORMAT('+',A1)

C TO COMPILE, UPICK=UPICK/U
C TO LINK,    UPICK=UPICK,FUSS,BIT12,SUBPIC,GTLIB,SYSLIB/F

      END
*
```

```
C FILE   SUBROUTINE:  SUBPIC.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C Routine SUBPIC creats a set of display subpictures of
C speech wave and command characters required by the succeeding
C program HPICK (main part of the interactive peak picking
C technique).  SUBPIC is called from program UPICK.
C
C H. NAKASONE
C JUN 27, 1983
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        SUBROUTINE SUBPIC(DOLBY)
        COMMON /SENSE/ MAXPIC, NBLOCK, NAME12, NSR
        COMMON /BUFF/ NBUF, XMAX, XMIN, YMAX, YMIN, YRANGE
        DOUBLE PRECISION DUMMY
        INTEGER IBUF(1024), NBUF(3600)
        LOGICAL*1 BEL, YN, DONE, NAME12(12), NAMOUT(10)
        LOGICAL*1 VTAB,NAME15(15),NAMER(6),DOLBY
        DATA VTAB/"013/, NSR/10000/
        DATA NAME12(10)/'S'/,NAME12(11)/'N'/,NAME12(12)/'D'/
        DATA NAME12(1)/'D'/,NAME12(2)/'K'/,NAME12(3)/'0'/
        DATA NAME15(1)/'D'/,NAME15(2)/'K'/,NAME15(3)/'0'/
        DATA NAME15(4)/':'/,NAME15(5)/'T'/,NAME15(6)/'M'/
        DATA NAME15(7)/'P'/,NAME15(11)/'.'/,NAME15(12)/'D'/
        DATA NAME15(13)/'P'/,NAME15(14)/'Y'/
        XMIN   = 0.
        YMIN   = -573.
        XMAX   = 1023.
        YMAX   = 450.
        YRANGE = 400.
        XEND   = 1022.
        NOISE  = 400
        KSIG   = 6000
        LION   = 8000
        IPRO   = 7001
        IRPT   = 7002
        ICOR   = 7003
        IJUMP  = 7004
        ISTART = 7005
        LWAIT  = 9000
        IDOT   = 9500
        IDOTC  = 9600
        ISIG   = 8500
        KNIT   = 3600
        NPTS   = 1024
        CALL SCROL(1,1000)
        TYPE 840,VTAB
840     FORMAT(1X,A1)
        CALL SCROL(1,12)
        CALL SCAL(XMIN,YMIN,XMAX,YMAX)
        CALL APNT(XMIN,YMAX,1,-8,0,1)
        CALL SUBP(LION)
        CALL LVECT(XMAX,0.,1)
        CALL LVECT(0.,-2.*YMAX,1)
        CALL LVECT(-XMAX,0.,1)
```

```
      CALL LVECT(0.,2.*YMAX,1)
      CALL APNT(0.,0.,0,-1,0,1)
      CALL LVECT(1023.,0.)
      CALL ESUB
C     CALL ERAS(LION)
C     CALL CMPRS
      CALL SAVE('DKO:FRAMER.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      YBOT = -(YMAX+65.)

      CALL APNT(150.,YBOT,1,-8,0,1)
      CALL SUBP(ICOR)
      CALL TEXT('?Correction?')
      CALL ESUB
C     CALL ERAS(ICOR)
C     CALL CMPRS
      CALL SAVE('DKO:CORECT.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(800.,YBOT,1,-8,0,1)
      CALL SUBP(IRPT)
      CALL TEXT('?Repeat?')
      CALL ESUB
C     CALL ERAS(IRPT)
C     CALL CMPRS
      CALL SAVE('DKO:REPEAT.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(550.,YBOT,1,-8,0,1)
      CALL SUBP(IPRO)
      CALL TEXT('?Proceed?')
      CALL ESUB
C     CALL ERAS(IPRO)
C     CALL CMPRS
      CALL SAVE('DKO:PROCED.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(0.,YBOT-26.,-1,-8,0,1)
      CALL SUBP(ISTART)
      CALL TEXT(' Hit frame')
      CALL ESUB
C     CALL ERAS(ISTART)
C     CALL CMPRS
      CALL SAVE('DKO:STARTR.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(350.,YBOT,1,-8,0,1)
      CALL SUBP(IJUMP)
      CALL TEXT('?Jump?')
      CALL ESUB
C     CALL ERAS(IJUMP)
C     CALL CMPRS
      CALL SAVE('DKO:JUMPER.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(0.,YBOT,-1,-8,0,1)
      CALL SUBP(ISIG)
      CALL TEXT(' Hit wave')
      CALL ESUB
C     CALL ERAS(ISIG)
C     CALL CMPRS
      CALL SAVE('DKO:SIGNAL.DPY')
      CALL INIT(NBUF,KNIT)
      CALL SCAL(XMIN,YMIN,XMAX,YMAX)
      CALL APNT(300.,0.,-1,-8)
```

```
         CALL SUBP(LWAIT)
         CALL TEXT(' W        A         I         T')
         CALL ESUB
C        CALL ERAS(LWAIT)
C        CALL CMPRS
         CALL SAVE('DK0:WAITER.DPY')
         CALL INIT(NBUF,KNIT)
         CALL SCAL(XMIN,YMIN,XMAX,YMAX)

50       SR = FLOAT(NSR)
         ICHAN = IGETC()
         CALL IRAD50(12,NAME12(1),DUMMY)
         NBLOCK = LOOKUP(ICHAN,DUMMY)
         IF(NBLOCK.LT.-1) STOP ' BAD LOOKUP'
         IF(NBLOCK.EQ.-1) STOP ' FILE NOT FOUND'

         RSEC = 256.* FLOAT(NBLOCK)/SR
D        TYPE 60, NBLOCK, RSEC
D60      FORMAT(1X,I4,' BLOCKS(= ',F8.4,' SEC) FOR THIS FILE.'///)
         MAXBLK = NBLOCK - (NPTS/256)

         NBLK = 0
         KSIG = 6000
         NAMSIG = 0
1000     NW = IREADW(NPTS,IBUF,NBLK,ICHAN)
         NAMSIG = NAMSIG + 1
         MAX = 0
CHECK IF NOISE REDUCTION BY DOLBY DESIRED.
         IF(DOLBY) GO TO 53
COME HERE IF NO DOLBY REQUESTED.
         DO 1550 I = 1, 1024
            IF(IABS(IBUF(I)).GT.MAX) MAX = IABS(IBUF(I))
1550     CONTINUE
         GO TO 55
COME HERE TO DO DOLBY.
53       DO 1553 I = 1, 1024
            ITEMP = 0
            IF(IBUF(I).GT.NOISE) ITEMP = IBUF(I) - NOISE
            IF(IBUF(I).LT.-NOISE)ITEMP = IBUF(I) + NOISE
            IBUF(I) = ITEMP
            IF(IABS(ITEMP).GT.MAX) MAX = IABS(ITEMP)
1553     CONTINUE

55       FAC = 1.0
         IF(MAX.GT.400) FAC = 400. / MAX

         DO 560 I = 1, NPTS
                 IBUF(I) = IBUF(I) * FAC
560      CONTINUE

         CALL APNT(0.,0.,9,-8,0,1)
         CALL SUBP(KSIG)
         CALL LVECT(0.,FLOAT(IBUF(1)))
         DO 565 I = 2,NPTS
            DY = FLOAT(IBUF(I) - IBUF(I-1))
            CALL LVECT(1.,DY)
565      CONTINUE
         CALL ESUB

         ND = NAMSIG / 100
         NAME15(8) = ND + "60
         ND = MOD(NAMSIG,100)
         NAME15(9) = ND / 10 + "60
         ND = MOD(ND,10)
         NAME15(10) = ND + "60
         NAME15(15) = 0
```

```
C       CALL ERAS(KSIG)
C       CALL CMPRS
        CALL SAVE(NAME15)

D       WRITE(7,700) NAME15, KSIG, NAMSIG
D700    FORMAT(1X,'DONE FOR ',15A1,3X,' KSIG=',I5,
D       &'  AND NAMSIG=',I5,/)

        CALL INIT(NBUF,KNIT)
        CALL SCAL(XMIN,YMIN,XMAX,YMAX)
300     NBLK = NBLK + 4
        IF(NBLK.LE.MAXBLK) GO TO 1000
        MAXPIC = NAMSIG
        CALL CLOSEC(ICHAN)
        CALL IFREEC(ICHAN)
        RETURN
        END

*
```

```
TT:=DKO:BIT12.FOR


C FILE: BIT12.FOR
C HHHHHHHHHHHHHHHHHHHH
C SUBROUTINE TO ACTIVATE/DEACTIVATE A BIT 12 FOR
C SPECIAL KEY MODE.
C WRITEN BY H. NAKASONE
C 08-JUN-1983
CHHHHHHHHHHHHHHHHHHHHHH

        SUBROUTINE BIT12(KON)
        LOGICAL*1 KON
C DEACTIVATE SPECIAL KEY MODE FOR KON = .FALSE..
        IF(.NOT.KON) CALL IPOKE("44,IPEEK("44).AND."167777)
C ACTIVATE SPECIAL KEY MODE FOR KON = .TRUE..
        IF(KON) CALL IPOKE("44,IPEEK("44).OR."10000)
        RETURN
        END
 *
```

```
C FILE   HPICK.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C HPICK is designed to measure Fo's (fundamental frequency)
C in speech signal directly from the time domain by the use
C of 'Interactive peak detecting technique'.
C
C Input:  File of display subpictures created by UPICK.
C Output: Data file which contains amplitudes(in absolute value)
C         and periods(in number of sampled points).
C
C Requirement:  (1) Light pen attached to the CRT.
C               (2) This program must be chained to FFPICK to complete
C                   a FFC feature file.
C H. NAKASONE
C JUN 27 1983
C Dept. of Audiology and Speech Sciences, MSU, East Lansing, MI.
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        COMMON /FFC/ KT, NAMOUT
        COMMON /SENSE/ MAXPIC, NBLOCK, NAME12, NSR
        COMMON /BUFF/ NBUF, XMAX, XMIN, YMAX, YMIN, YRANGE
        COMMON /ICSPEC/ ISPEC(39)
        DOUBLE PRECISION DUMMY, FPROG
        INTEGER IBUF(1024), NBUF(4000), NAMP(1500)
        INTEGER NTIME(100), NTAO(1500)
        LOGICAL*1 BEL, YN, DONE, NAME12(12), YNEND, NAMOUT(10)
        LOGICAL*1 DEBUG,VTAB, NAME(15), KNTP(4),LNAME(5)
        EQUIVALENCE (X0,X),(Y0,Y),(M0,M),(N0,N)
        DATA DUMMY/12RDK1TESTERSND/,VTAB/"013/
        DATA BEL/"7/,NAME12(10)/'S'/,NAME12(11)/'N'/,NAME12(12)/'D'/
        DATA NAME12(1)/'D'/,NAME12(2)/'K'/,NAME12(3)/'0'/
        DATA NAMOUT(10)/'C'/,NAMOUT(9)/'F'/,NAMOUT(8)/'F'/
        DATA NAMOUT(7)/'.'/
        DATA NSR/10000/
        DATA NAME(1)/'D'/,NAME(2)/'K'/,NAME(3)/'0'/,NAME(4)/':'/
        DATA NAME(5)/'T'/,NAME(6)/'M'/,NAME(7)/'P'/,NAME(11)/'.'/
        DATA NAME(12)/'D'/,NAME(13)/'P'/,NAME(14)/'Y'/
        DATA FPROG/12RDK1FFPICKSAV/

        CALL RCHAIN(IF,MAXPIC,9)

C INITIALIZE.
        DO 333 I = 1, 1500
           NAMP(I) = 0
333        NTAO(I) = 0
        DO 334 I = 1, 100
334        NTIME(I) = 0

        IF(IF.EQ.-1) GO TO 2222
        TYPE 2224
2224    FORMAT(' NUMBER OF DISPLAY PICTURES ?',$)
        ACCEPT 2226,MAXPIC
2226    FORMAT(I3)
        TYPE 2228
2228    FORMAT(' FILE NAME (6A1)  >',$)
        READ(5,2230) (NAME12(M),M=4,9)
2230    FORMAT(6A1)

2222    XMIN   = 0.
        YMIN   = -573.
        XMAX   = 1023.
        YMAX   = 450.
        YRANGE = 400.
        XEND   = 1022.
```

```
        KSIG   = 6000
        LION   = 8000
        IPRO   = 7001
        IRPT   = 7002
        ICOR   = 7003
        IJUMP  = 7004
        ISTART = 7005
        IDOT   = 9500
        IDOTC  = 9600
        ISIG   = 8500
        KNIT   = 4000
        NPTS   = 1024

        ENCODE(3,10,KNTP) MAXPIC
10      FORMAT(I3)
        KNTP(4) = 0

        DEBUG = .FALSE.
D       TYPE 3010
D3010   FORMAT(' DEBUG ?',$)
D       ACCEPT 805,YN
D       IF(YN.EQ.89) DEBUG = .TRUE.


50      SR = FLOAT(NSR)
        DO 35 I = 4, 9
           NAMOUT(I-3) = NAME12(I)
35      CONTINUE

        TIMOLD = 0.
        TAOOLD = 0.
        AMPOLD = 0.
        KT = 0
        NW = 0
        NBLK = 0
        KOUNT = 0

1111    CONTINUE
        CALL INIT(NBUF,KNIT)
        CALL SCROL(1,1000)
        TYPE 840,VTAB
840     FORMAT(1X,A1)
        CALL SCROL(1,12)
        CALL SCAL(XMIN,YMIN,XMAX,YMAX)

        CALL SUBP(LION)
        CALL RSTR('DKO:FRAMER.DPY')
        CALL ESUB
        CALL OFF(LION)

        YBOT = -(YMAX+65.)

        CALL APNT(150.,YBOT,1,-8,0,1)
        CALL SUBP(ICOR)
        CALL RSTR('DKO:CORECT.DPY')
        CALL ESUB
        CALL OFF(ICOR)

        CALL APNT(800.,YBOT,1,-8,0,1)
        CALL SUBP(IRPT)
        CALL RSTR('DKO:REPEAT.DPY')
        CALL ESUB
        CALL OFF(IRPT)

        CALL APNT(550.,YBOT,1,-8,0,1)
        CALL SUBP(IPRO)
```

```
        CALL RSTR('DKO:PROCED.DPY')
        CALL ESUB
        CALL OFF(IPRO)

        CALL APNT(0.,YBOT-26.,-1,-8,0,1)
        CALL SUBP(ISTART)
        CALL RSTR('DKO:STARTR.DPY')
        CALL ESUB
        CALL OFF(ISTART)

        CALL APNT(350.,YBOT,1,-8,0,1)
        CALL SUBP(IJUMP)
        CALL RSTR('DKO:JUMPER.DPY')
        CALL ESUB
        CALL OFF(IJUMP)

        CALL APNT(0.,YBOT,-1,-8,0,1)
        CALL SUBP(ISIG)
        CALL RSTR('DKO:SIGNAL.DPY')
        CALL ESUB
        CALL OFF(ISIG)

1000    NW = NW + 1
        LASTK = 0
        RMAX = 0.0

        ND = NW / 100
        NAME(8) = ND + "60
        LNAME(1)= NAME(8)
        ND = MOD(NW,100)
        NAME(9) = ND / 10 + "60
        LNAME(2)= NAME(9)
        ND = MOD(ND,10)
        NAME(10) = ND + "60
        LNAME(3) = NAME(10)
        NAME(15) = 0
        LNAME(4) = 45
        LNAME(5) = 0

        CALL ERAS(5000)
        CALL APNT(860.,-440.,-5,-8,0,1)
        CALL SUBP(5000)
        CALL TEXT(LNAME)
        CALL APNT(914.,-440.,-5,-8)
        CALL TEXT(KNTP)
        CALL ESUB

        CALL SUBP(KSIG)
        CALL RSTR(NAME)
        CALL ESUB

        CALL ON(LION)

450     TX = -1.0
        X = 0.0
        LASTK = 0

550     K = 0
        KLION = 1000
        CALL ON(ISTART)

COMMENCE THE FIRST LIGHT PEN HIT TESTING ON THE FRAME.

        CALL ERAS(5050)
90      CALL LPEN(MO,NO,XO,YO)
        IF(M.NE.1 .OR. N.NE.LION .OR. X.LT.TX) GO TO 90
```

```
      IF(ABS(XEND-X).LE.1.0) GO TO 300
      TYPE 7, BEL
      CALL ON(ISIG)
      CALL OFF(ISTART)

      TX = X
      TY = Y
100   CALL LPEN(MO,NO,XO,YO)
      IF(M.NE.1 .OR. N.NE.KSIG .OR. X-TX.LT.0.0) GO TO 100
      CALL ON(ISTART)
101   K = K + 1
      EX = TX
      EY = TY
      KX = INT(X) + 1
      IX = KX
      IBX = IABS(IBUF(KX))

105   KX = KX + 1
      IF(IABS(IBUF(KX)).LT.IBX) GO TO 106
      IBX = IABS(IBUF(KX))
      GO TO 105
106   KX = KX - 1
      CALL APNT(TX,TY,-2,-3,0,1)
      CALL SUBP(K)
      CALL LVECT(X-TX,Y-TY)
      CALL ESUB
      NTIME(K) = KX
      IF(K.LT.2) GO TO 1008

      FREQ = SR / FLOAT(NTIME(K)-NTIME(K-1)+1)
      CALL ERAS(5050)
      CALL APNT(860.,410.,-1,-8,0,1)
      CALL NMBR(5050,FREQ,'F6.1')

1008  CALL ON(ICOR)
      TX = FLOAT(KX)
      TY = Y

110   CALL LPEN(MO,NO,XO,YO)
      IF(M.NE.1) GO TO 110
      IF(N.EQ.KSIG .AND. X-TX.GT.20.) GO TO 101
      IF(N.EQ.LION .AND. X-TX.GE.0.0) GO TO 160
      IF(N.NE.ICOR) GO TO 110
COME HERE IF N = ICOR TO CORRECT PREVIOUSLY DROWN VECTOR.
145   CALL ERAS(K)
      CALL CMPRS
      K = K -1
      TX = EX
      TY = EY
      CALL OFF(ICOR)
      CALL ON(ISIG)
      GO TO 100

COME HERE IF N = LION TO CHECK OPTIONS MADE.
160   DX = X - TX
      DY = Y - TY
      KLION = KLION + 1
      CALL APNT(TX,TY,0,-3,0,1)
      CALL SUBP(KLION)
      CALL LVECT(DX,DY)
      CALL ESUB
      TYPE 7, BEL
      CALL ON(IJUMP)
      CALL ON(IPRO)
      CALL OFF(ICOR)
      CALL ON(IRPT)
```

```
      CALL OFF(ISIG)
      CALL OFF(ISTART)
      TX = X
      TY = Y

180   CALL LPEN(MO,NO,XO,YO)
      IF(M.NE.1) GO TO 180
      IF(N.NE.IJUMP .AND. N.NE.IPRO .AND. N.NE.IRPT) GO TO 180
      CALL OFF(IJUMP)
      CALL OFF(IPRO)
      CALL OFF(IRPT)
      TYPE 7, BEL
      IF(N.EQ.IRPT) GO TO 250
      IF(K.GE.2) GO TO 200

      CALL ERAS(K)
      CALL ERAS(KLION)
      CALL CMPRS
      K = K - 1
      KLION = KLION - 1
      IF(N.EQ.IPRO) GO TO 300
      CALL OFF(IJUMP)
      CALL OFF(IPRO)
      CALL OFF(ISIG)
      CALL ON(ISTART)
      EX = TX
      EY = TY
      GO TO 90

200   KT = KT + 1
      IF(KT.GT.1500) GO TO 600
      NAMP(KT) = IABS(IBUF(NTIME(K+1)))
      NTAO(KT) = 0
      TIMNEW = FLOAT(NTIME(1))
      TAONEW = TIMOLD + TIMNEW
CD    IF(DEBUG) TYPE 2206,TIMOLD,TAOOLD
CD2206 FORMAT(' TIMOLD =',F8.0,'  TAOOLD =',F8.0)
CD    IF(DEBUG) TYPE 2205,TIMNEW,TAONEW
CD2205 FORMAT(' TIMNEW =',F8.0,'  TAONEW =',F8.0)
      IF(TAONEW.LE.0.) GO TO 205
      IF( (ABS(TAONEW-TAOOLD)/TAONEW).GT. 0.25) GO TO 205
          NTAO(KT) = INT(TAONEW)

205   CONTINUE
D     IF(DEBUG) TYPE 217, NAMP(KT),NTAO(KT),KT
      DO 210 J = 2, K
         KT = KT + 1
         NAMP(KT) = IABS(IBUF(NTIME(J)))
         NTAO(KT) = NTIME(J) - NTIME(J-1)
D        IF(DEBUG) TYPE 217, NAMP(KT),NTAO(KT),KT
210   CONTINUE
D217  FORMAT(' NAMP=',I6,'  NTAO=',I6,'  KT=',I4)
      LASTK = LASTK + K

250   IF(K.LE.0) GO TO 305
      DO 310 I = 1, K
         CALL ERAS(I)
310   CONTINUE
305   IF(KLION.LE.1000) GO TO 330
      DO 320 I = 1001, KLION
         CALL ERAS(I)
320   CONTINUE
330   CALL CMPRS

      IF(N.EQ.IRPT) GO TO 400
      IF(N.EQ.IPRO) GO TO 300
```

```
        IF(TX.GE. XEND-10.) GO TO 300

COME HERE IF N = IJUMP.
        CALL OFF(IJUMP)
        CALL OFF(IPRO)
        CALL OFF(ISIG)
        CALL ON(ISTART)
        GO TO 550

400     CALL OFF(IJUMP)
        CALL OFF(IRPT)
        CALL OFF(IPRO)
        CALL OFF(ICOR)
        CALL OFF(ISIG)
        KT = KT - LASTK
        GO TO 450

300     IF(NW .GE. MAXPIC) GO TO 600
          CALL OFF(IJUMP)
          CALL OFF(IRPT)
          CALL OFF(IPRO)
          CALL OFF(ICOR)
          CALL OFF(ISIG)
          CALL OFF(ISTART)
          CALL OFF(LION)
          CALL ERAS(KSIG)
          CALL CMPRS

          TIMOLD = 0.
          TAOOLD = 0.

          IF(N.NE.IPRO) GO TO 1234
            TIMOLD = XEND - FLOAT(NTIME(K))
            TAOOLD = FLOAT(NTAO(KT))
1234      CONTINUE
          KOUNT = KOUNT + 1
          TYPE 7, BEL
          IF(KOUNT.LE.5) GO TO 1000
          KOUNT = 0
          CALL INIT(NBUF,KNIT)
          GO TO 1111

600     CALL INIT(NBUF,KNIT)
        CALL FREE
        TYPE 605,NAMOUT
605     FORMAT(' WRITING FFC RECORD FOR: ',10A1,/)
        CALL ASSIGN(13,NAMOUT,10,'NEW')
        WRITE(13,610) (NAMOUT(M),M=1,10),KT,NSR
610     FORMAT(10A1,I5,I6,' (3I7)'/)
        DO 620 I = 1, KT
           WRITE(13,630) NAMP(I),NTAO(I),I
630        FORMAT(3I7)
620     CONTINUE
        CALL CLOSE(13)
        TYPE 820
        CALL CHAIN(FPROG,KT,10)
7       FORMAT('+',A1)
800     FORMAT('+',A1)
805     FORMAT(A1)
810     FORMAT(' OK (Y/N) ?',$)
820     FORMAT(' CHAINING TO FFPICK.'/)
606     FORMAT(10A1)
        END

*
```

```
C  FILE FFPICK.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C  FFPICK COMPUTES SEVERAL MEASURES BASED UPON FUNDAMENTAL
C  FREQUENCY CONTOURS.
C  INPUT:  FFC DATA FILE CONTAINING PERIODS (STORED IN THE NUMBER
C          OF SAMPLED POINTS) AND ABSOLUTE AMPLITUDES.
C  OUTPUT: HAN FILE CONTAINING ITS FILE NAME, FORMAT INFORMATION
C          OF 9 FEATURES (MEASUREMENTS).
C
C  WRITTEN BY  H. NAKASONE
C  DATE        23-JUN-83
C
C  DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C  MICHIGAN STATE UNIVERSITY
C  EAST LANSING, MICHIGAN
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        COMMON /FFC/ KT, NAMOUT
        INTEGER NTAO(1500), NAMP(1500)
        REAL    RTAO(1500), RAMP(1500)
        LOGICAL*1 NAMOUT(10),BEL,FILNAM(10),FMT(60)
        EQUIVALENCE (NAMOUT,FILNAM)
        DATA BEL/"007/
CHECK IF CHAINED FROM HPICK.  DATA SENT ARE STORED IN KT AND NAMOUT.
        CALL RCHAIN(IF,KT,10)

        IF(IF.EQ.-1) GO TO 30
        TYPE 10
10      FORMAT(' INPUT FFC FILE NAME (10A1) ?',$)
        READ(5,20) (FILNAM(M),M=1,10)
20      FORMAT(10A1)
30      CALL ASSIGN(13,FILNAM,10,'OLD')
40      READ(13,50) FILNAM,NT,NSR,FMT
50      FORMAT(10A1,I5,I6,60A1)
        TYPE 51
51      FORMAT(1X,'Computation on FFC in progress.   Wait.'/)

        DO 60 I=1,NT
           READ(13,FMT) NAMP(I),NTAO(I),J
           RAMP(I) = FLOAT(NAMP(I))
           RTAO(I) = FLOAT(NTAO(I))
60      CONTINUE

        CALL CLOSE(13)
        SR = FLOAT(NSR)
        IF(IF.EQ.-1) TYPE 65,KT,NT
65      FORMAT(' TOTAL PERIODS SENT =',I5,'  AND READ =',I5)

        K = 0
        TAOMAX = 0.
        TAOMIN = 500.
        DO 100 I = 1, NT
```

```
            IF(RTAO(I).LE.20. .OR. RTAO(I).GE.250.) GO TO 100
              K = K + 1
              RTAO(K) = RTAO(I)
              RAMP(K) = RAMP(I)
              IF(RTAO(K).LT.TAOMIN) TAOMIN = RTAO(K)
              IF(RTAO(K).GT.TAOMAX) TAOMAX = RTAO(K)
100      CONTINUE
         RK = FLOAT(K)
D        TYPE 110,NT,K
D110     FORMAT(' ORIGINAL TAO''S =',I4,'  NOZERO TAO''S =',I4)

COMPUTE AVERAGE TAO, AVERAGE AMP, ETC.

         AVETAO = 0.
         AVEAMP = 0.
         SSQFF = 0.
         SSQAMP = 0.

         DO 200 I = 1, K
           AVETAO = AVETAO + RTAO(I)
           AVEAMP = AVEAMP + RAMP(I)
           SSQFF = SSQFF + (SR/RTAO(I))*(SR/RTAO(I))
           SSQAMP = SSQAMP + RAMP(I)*RAMP(I)
200      CONTINUE

         AVETAO = AVETAO / RK
         AVEAMP = AVEAMP / RK
         SDFF   = SQRT(SSQFF/RK - (SR/AVETAO)*(SR/AVETAO))
         SDAMP   = SQRT(SSQAMP/RK - AVEAMP*AVEAMP)

         DBSDAM = 0.
         IF(SDAMP.GT.0.0) DBSDAM = 20.*ALOG10(SDAMP)
         DBAVAM = 0.
         IF(AVEAMP.GT.0.0) DBAVAM = 20.*ALOG10(AVEAMP)

         DLFF = 0.
         DAMP = 0.
         DO 300 I = 2, K
            F1 = SR / RTAO(I)
            F2 = SR / RTAO(I-1)
            DLFF = DLFF + ABS(F1 -F2)
            DAMP = DAMP + ABS(RAMP(I) - RAMP(I-1))
300      CONTINUE
         DLFF = DLFF / (RK-1.)
         DAMP = DAMP/(RK-1.)
         IF(DAMP.GT.0.) DAMPDB = 20.*ALOG10(DAMP)

         AVFF = 0.
         IF(AVETAO.GT.0.) AVFF = SR / AVETAO
         FFMX = SR / TAOMIN
         FFMN = SR / TAOMAX
         FFRG = FFMX - FFMN

COMPUTE %

         PERJIT = (DLFF/AVFF)*100.
         PERSIM = (DAMP/AVEAMP)*100.

         TYPE 7222
7222     FORMAT(1X,/)
         WRITE(7,7200) (FILNAM(M),M=1,10),K
7200     FORMAT(//,1X,'-----------------------------------------
     &----------------------------------'/
     &' File: ',10A1,/,' Number of periods detected
     & = ',I4,/,'----------------------------------------
     &-----------------------------------'//)
```

```
      WRITE(7,7300) AVFF,SDFF,DLFF,PERJIT
7300  FORMAT(' Summary of fundamental frequency (F.F.):',//,
     &' Mean F.F.                          =',F8.2,' (Hz)'/,
     &' Standard deviation of F.F.         =',F8.2,' (Hz)'/,
     &' Average variation of  F.F.(DF)     =',F8.2,' (Hz)'/,
     &' Ratio of DF to mean F.F            =',F8.2,' (%)'///)
      WRITE(7,7400) AVEAMP,DBAVAM,SDAMP,DBSDAM,DAMP,DAMPDB,PERSIM
7400  FORMAT(' Summary of amplitude of F.F. :',//,
     &' Mean amplitude                     =',F9.2,' (',F5.1,' dB)'/,
     &' Standard deviation of ampl.        =',F9.2,' (',F5.1,' dB)'/,
     &' Average variation of ampl.(DA)     =',F9.2,' (',F5.1,' dB)'/,
     &' Ratio of DA to mean ampl.          =',F9.2,' (%)'
     &/,'------------------------------------------------------
     &---------------',/)
      WRITE(7,7450) FFMX,FFMN,FFRG
7450  FORMAT(1X,'Maximum F.F.=',F8.2,'   Minimum F.F.=',F8.2,
     &'   Range F.F.=',F7.2)

CONVERT EXTENSION .FFC TO .HAN.
CHARACTER H = 78! A = 65; N = 72  BY ASCII.

      FILNAM(10) = 78
      FILNAM(9) = 65
      FILNAM(8) = 72

      CALL ASSIGN(13,FILNAM,10,'NEW')
      WRITE(13,400) (FILNAM(M),M=1,10)
400   FORMAT(10A1,' (8F6.2,F10.2)'/)
      WRITE(13,410)AVFF,SDFF,DLFF,PERJIT,PERSIM,FFMX,FFMN,FFRG,DAMP
410   FORMAT(8F6.2,F10.2)
      CALL CLOSE(13)
      TYPE 425,FILNAM
425   FORMAT(//////,1X,'Computed HAN file name:  ',10A1//)
      CALL EXIT
      END

*
```

```
C FILE:  FFFRAT.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C FFFRAT IS DESIGNED TO COMPUTE F-RATIO'S OF 9 FFC FEATURES
C
C INPUT:  DATA FILE WITH AN EXT. NAME 'HAN'.
C OUTPUT: DATA FILE WITH AN EXT. NAME 'TIM'.
C
C 11-JULY-83
C H.NAKASONE
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
      COMMON /SENSE/ X, F
      COMMON /PASS/ FMT
      DOUBLE PRECISION VN(9)
      REAL X(5,10,20), F(20)
      LOGICAL*1 FNAMES(14,50),MNAME(14),LNAME(14),FMT(60)
      LOGICAL*1 OUTEXT(3),OUTDEV(3),EXT(3),DEV(3)
      DATA OUTEXT(1)/'T'/,OUTEXT(2)/'I'/,OUTEXT(3)/'M'/
      DATA OUTDEV(1)/'D'/,OUTDEV(2)/'K'/,OUTDEV(3)/'1'/
      DATA VN(1)/'    AVFF'/,VN(2)/'    SDFF'/,VN(3)/'    DLFF'/
      DATA VN(4)/'  PERJIT'/,VN(5)/'  PERSIM'/,VN(6)/'   FFMAX'/
      DATA VN(7)/'   FFMIN'/,VN(8)/'   FFRNG'/,VN(9)/'    DAMP'/

      NP = 5
      NG = 10
      ND = 9

      TYPE 10
10    FORMAT(' ENTER MASTER FILE NAME (14A1)  >',$)
      READ(5,20) (MNAME(M),M= 1,14)
20    FORMAT(14A1)

      TYPE 30
30    FORMAT(' ENTER EXT NAME OF INPUT DATA FILE >',$)
      READ(5,35) (EXT(M),M=1,3)
35    FORMAT(3A1)
      TYPE 40
40    FORMAT(' ENTER DEV NAME OF INPUT DATA FILE >',$)
      READ(5,35) (DEV(M),M=1,3)

      CALL RDFILE(MNAME,FNAMES,EXT,DEV)

C TRANSFER DATA TO ARRAY X.
D     TYPE 90
D90   FORMAT(1X,'AVFF',T8,'SDFF',T16,'DLFF',T24,'PERJIT',
D    &T32,'PERSIM',T40,'FFMAX',T48,'FFMIN',T56,'FFRNG',
D    &T64,'DAMP'//)
      K = 0
      DO 100 I = 1, NG

        DO 200 J = 1, NP
          K = K + 1
          CALL ASSIGN(12,FNAMES(1,K),14,'OLD')
          READ(12,210) (LNAME(M),M=1,10),FMT
```

```
210          FORMAT(10A1,60A1,/)
             READ(12,FMT) (X(J,I,KD),KD=1,ND)
D            WRITE(7,220) (X(J,I,KD),KD=1,ND)
D220         FORMAT(1X,8F7.2,F10.2)
             CALL CLOSE(12)
200        CONTINUE
           TYPE 9
9          FORMAT(1X,/)
100      CONTINUE

D        TYPE 90

         CALL SUBF(NF,NG,ND)

         PAUSE ' ADJUST TO THE TOP OF NEW PAGE.'

         WRITE(7,300) (MNAME(M),M=1,14)
300      FORMAT(10X,'SUMMARY OF F-RATIO''S:             ',14A1,//,
         &10X,'FEATURE',15X,'F-RATIO',/,10X,'-------',15X,
         &'-------',/)

         DO 310 I = 1, ND
            WRITE(7,320) I,VN(I),F(I)
320         FORMAT(11X,'X(',I1,')',3X,A8,F12.3,/)
310      CONTINUE

         TYPE 330
330      FORMAT(10X,'------------------------------')
         TYPE 9
         TYPE 9
C   NORMALIZE
         CALL ZNORM(X,NF,NG,ND)

C   CHANGE EXTENSION NAME.
         DO 350 I = 1, 50
           DO 360 J = 1, 3
              FNAMES(J,I) = OUTDEV(J)
              FNAMES(J+11,I) = OUTEXT(J)
360        CONTINUE
350      CONTINUE

C   WRITE THE NORMALIZED DATA FILE WITH NEW EXT NAME.
         J = 0
         K = 1
         DO 400 I = 1, 50
            CALL ASSIGN(12,FNAMES(1,I),14,'NEW')
            WRITE(12,20) (FNAMES(M,I),M=1,14)
            J = J + 1
            IF(J.LE.NF) GO TO 410
            J = 1
            K = K + 1
410         CONTINUE
            DO 420 KD = 1, ND
               WRITE(12,425) X(J,K,KD)
425            FORMAT(F14.6)
420         CONTINUE
            CALL CLOSE(12)
400      CONTINUE
         CALL EXIT
         END
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C PROGRAM NAKRO3.FOR
C
C NAKRO3 IS DESIGNED TO EXECUTE VOICE IDENTIFICATION OPERATION FOR
C DESIGN 3 (CROSS-TRANSMISSION BY COMPOSITE PARAMETER OF FFC AND IDS
C (OR LTS).
C
C NAKRO3 REQUIRES 4 MASTER FILES, EACH CONTAINING 50 FILE NAMES OF
C PATTERNS.
C
C SUBPROGRAMS CALLED:  SBWEIT,RDFILE,ZNORM,AND SUBSET.
C
C 13-JULY-83
C H. NAKASONE
C
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C MICHIGAN STATE UNIVERSITY
C EAST LANSING, MICHIGAN
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        REAL TXU(5,10,20),TXK(5,10,20)
        REAL XU(5,10,20), XK(5,10,20),FWEIT(10),TWEIT(9)
        LOGICAL*1 MFILUF(14),MFILUT(14),MFILKF(14),MFILKT(14)
        LOGICAL*1 UFILEF(14,50),UFILET(14,50),KFILEF(14,50)
        LOGICAL*1 KFILET(14,50),FINEXT(3),TINEXT(3),INDEV(3)
        LOGICAL*1 LFILE(14),DEBUG,YN,BEL
        DATA TINEXT(1)/'T'/,TINEXT(2)/'I'/,TINEXT(3)/'M'/
        DATA INDEV(1)/'D'/,INDEV(2)/'K'/,INDEV(3)/'1'/
        DATA BEL/"007/,DEBUG/.FALSE./

7       FORMAT('+',A1)
8       FORMAT(A1)
9       FORMAT(1X,/)
14      FORMAT(3A1)

        NDF = 10
        NDT = 9
        ND = NDF + NDT
        NP = 5
        NG = 10

        DEBUG = .FALSE.
        TYPE 4500
4500    FORMAT(' DEBUG ?',$)
        ACCEPT 4510,YN
        IF(YN.EQ.89) DEBUG = .TRUE.
4510    FORMAT(A1).

        TYPE 1
1       FORMAT(' PROGRAM NAKRO3 - Experiment III (Cross-Transmission)'/
        &' A pattern contains features from Frequency and Time Domain.'/
/)
        TYPE 10
10      FORMAT(' UNKNOWN MASTER FILE FROM FREQ.(14A1)  >',$)
        READ(5,20) MFILUF
20      FORMAT(14A1)
        TYPE 30
30      FORMAT(' KNOWN MASTER FILE FROM FREQ.(14A1)    >',$)
        READ(5,20) MFILKF
```

```
      TYPE 40
40    FORMAT(' EXT NAME OF FREQ. FILE (FRI OR FRL) >',$)
      READ(5,50) FINEXT
50    FORMAT(3A1)

      CALL RDFILE(MFILUF,UFILEF,FINEXT,INDEV)
      CALL RDFILE(MFILKF,KFILEF,FINEXT,INDEV)
      TYPE 9
      TYPE 60
60    FORMAT(' UNKNOWN MASTER FILE FROM TIME(14A1) >',$)
      READ(5,20) MFILUT
      TYPE 70
70    FORMAT(' KNOWN MASTER FILE FROM TIME(14A1)   >',$)
      READ(5,20) MFILKT

      CALL RDFILE(MFILUT,UFILET,TINEXT,INDEV)
      CALL RDFILE(MFILKT,KFILET,TINEXT,INDEV)

      TYPE 9
      TYPE 80
80    FORMAT(' PROVIDE INFO. FOR FREQUENCY FEATURES.'/)
      CALL SBWEIT(FWEIT,NDF)

      TYPE 9
      TYPE 90
90    FORMAT(' PROVIDE INFO. FOR TIME FEATURES.'/)
      CALL SBWEIT(TWEIT,NDT)

C TRANSFER DATA FROM UFILE AND KFJILE TO XU AND XK, RESPECTIVELY.

      KT = 0
      DO 100 J = 1, NG

        DO 110 I = 1, NP
           KT = KT + 1
           CALL ASSIGN(12,UFILEF(1,KT),14,'OLD')
           READ(12,20) LFILE
           DO 120 K = 1, NDF
              READ(12,150) TXU(I,J,K)
150           FORMAT(F14.6)
              XU(I,J,K) = TXU(I,J,K) * FWEIT(K)
120        CONTINUE
           CALL CLOSE(12)

           CALL ASSIGN(13,KFILEF(1,KT),14,'OLD')
           READ(13,20) LFILE
           DO 130 K = 1, NDF
              READ(13,150) TXK(I,J,K)
              XK(I,J,K) = TXK(I,J,K) * FWEIT(K)
130        CONTINUE
           CALL CLOSE(13)
110      CONTINUE
100   CONTINUE

      CALL ZNORM(XU,NP,NG,NDF)
      CALL ZNORM(XK,NP,NG,NDF)
      CALL ZNORM(TXU,NP,NG,NDF)
      CALL ZNORM(TXK,NP,NG,NDF)

      KT = 0
      DO 5100 J = 1, NG

         DO 5110 I = 1, NP

            KT = KT + 1
            CALL ASSIGN(12,UFILET(1,KT),14,'OLD')
```

```
              READ(12,20) LFILE
              KNTW = 0
              DO 5120 K = 1+NDF, NDT+NDF
                 READ(12,150) TXU(I,J,K)
                 KNTW = KNTW + 1
                 XU(I,J,K) = TXU(I,J,K) * TWEIT(KNTW)
5120          CONTINUE
              CALL CLOSE(12)
              CALL ASSIGN(13,KFILET(1,KT),14,'OLD')
              READ(13,20) LFILE
              KNTW = 0
              DO 5130 K = 1+NDF, NDT+NDF
                 READ(13,150) TXK(I,J,K)
                 KNTW = KNTW + 1
                 XK(I,J,K) = TXK(I,J,K) * TWEIT(KNTW)
5130          CONTINUE
              CALL CLOSE(13)
5110       CONTINUE
5100     CONTINUE

202      KNTU = 1
         KNTUP= 1
         KCORR= 0

2000     IF(DEBUG) TYPE 205, KNTUP,KNTU
205      FORMAT(1X,//,' EUCLIDIAN DISTANCES BETWEEN PATTERN ',I1,'
         & OF UNKNOWN SPEAKER ',I2,/,' AND PATTERN I OF
         & KNOWN SPEAKER J.'//)

         XMIN = 1.E30
         ITEMP = 1
         JTEMP = 1

COMPUTE EUCLIDIAN DISTANCE

         DO 210 J = 1, NG

            DO 220 I = 1, NP

               EUCL = 0.0
               DO 230 K = 1, NDF+NDT
                  EUCL = EUCL + (XU(KNTUP,KNTU,K) - XK(I,J,K))**2.
230            CONTINUE
               EUCL = EUCL**0.5
               IF(EUCL.GT.XMIN) GO TO 250
               ITEMP = I
               JTEMP = J
               XMIN = EUCL

250            IF(DEBUG) WRITE(7,260) EUCL,I,J
260            FORMAT('+',F10.5,'[',I1,',',I2,']',$)
220         CONTINUE
            IF(DEBUG) TYPE 9
210      CONTINUE

         IF(KNTU.EQ.JTEMP) KCORR = KCORR + 1

         WRITE(7,280) KNTUP, KNTU,ITEMP,JTEMP,XMIN
280      FORMAT(' PATTERN ',I1,' OF UNKNOWN SPEAKER ',I2,
         &' IS IDENTIFIED WITH'/' PATTERN ',I1,' OF   KNOWN
         & SPEAKER ',I2,/,' DISTANCE =',F10.5,//)

         KNTUP = KNTUP + 1
         IF(KNTUP.LE.NP) GO TO 2000
         KNTUP = 1
         KNTU = KNTU + 1
```

```
        IF(KNTU.LE.NG) GO TO 2000

        KINCOR = 50 - KCORR
        PCOR = (FLOAT(KCORR)/50.) * 100.


        TYPE 3000, KCORR,KINCOR,PCOR
3000    FORMAT(1X,//,' Summary by Nearest Neighbor Rule:'//'
        & Correct =',I3,',  Incorrect =',I3,',   Rate(%) =',F6.2,////)

        TYPE 9
        TYPE 9

COMPUTE THE MINIMUM SET DISTANCE

        CALL SUBSET(XU,XK,NP,NG,ND)

        TYPE 9
        TYPE 9
        IF(.NOT.DEBUG) GO TO 9999
D       TYPE 6000
D6000   FORMAT(' WANT DIFFERENT FEATURES COMBINATION ?',$)
D       ACCEPT 6008,YN
6008    FORMAT(A1)
D       IF(YN.NE.89) GO TO 9999
D       TYPE 6050
D6050   FORMAT(' FOR FREQUENCY DOMAIN:'/)
D       CALL SBWEIT(FWEIT,NDF)
D       TYPE 6060
D6060   FORMAT(' FOR TIME DOMAIN:'/)
D       CALL SBWEIT(TWEIT,NDT)


        CALL EXIT

        DO 6100 J = 1, NG
           DO 6200 I = 1, NP
              DO 6300 K = 1, NDF
                 XU(I,J,K) = TXU(I,J,K) * FWEIT(K)
                 XK(I,J,K) = TXK(I,J,K) * FWEIT(K)
6300          CONTINUE

              KNTW = 0
              DO 6400 K = 1+NDF, NDT+NDF
                KNTW = KNTW + 1
                XU(I,J,K) = TXU(I,J,K) * TWEIT(KNTW)
                XK(I,J,K) = TXK(I,J,K) * TWEIT(KNTW)
6400          CONTINUE
6200       CONTINUE
6100    CONTINUE

        GO TO 202

9999    CALL EXIT
        END

*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C PROGRAM NAKROA.FOR
C NAKROA IS DESIGNED TO EXECUTE WITHIN-TRANSMISSION VOICE
C IDENTIFICATION OPERATION.
C 12-JULY-83
C H. NAKASONE
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        REAL XU(5,10,10), DMAX(5,10), WEIGHT(10)
        LOGICAL*1 MFILU(14)
        LOGICAL*1 UFILE(14,50), INEXT(3),INDEV(3)
        LOGICAL*1 OUTEXT(3),OUTDEV(3), YN, DEBUG
        DATA INDEV(1)/'D'/,INDEV(2)/'K'/,INDEV(3)/'1'/
        DATA DEBUG/.FALSE./

8       FORMAT(A1)
9       FORMAT(1X,/)

        TYPE 10
10      FORMAT(' PROGRAM NAKROA:  VOICE I.D. WITHIN TRANSMISSION.'//)

        TYPE 15
15      FORMAT(' ENTER MASTER FILE NAME (14A1)        >',$)
        READ(5,20) MFILU
20      FORMAT(14A1)

        TYPE 25
25      FORMAT(' ENTER EXT NAME (FRI, FRL, OR TIM)    >',$)

C               FRI = IDS
C               FRL = LTS
C               TIM = FFC

        READ(5,22) INEXT
22      FORMAT(3A1)

        TYPE 26
26      FORMAT(' ENTER NUMBER OF DIMENSIONS           >',$)
        READ(5,27) ND
27      FORMAT(I3)

        CALL RDFILE(MFILU,UFILE,INEXT,INDEV)
        TYPE 29
29      FORMAT(' WANT PRINT OUT OF DISTANCE MATRIX    >',$)
        READ(5,8) YN
        IF(YN.EQ.89) DEBUG = .TRUE.

        CALL SBWEIT(WEIGHT,ND)
        NP = 5
        NG = 10

C PLACE DATA FROM UFILE TO AN ARRAY XU.

        KT = 0
        DO 100 J = 1, NG

          DO 110 I = 1, NP
             KT = KT + 1
             CALL ASSIGN(12,UFILE(1,KT),14,'OLD')
             READ(12,20) LFILE
             DO 120 K = 1, ND
                READ(12,150) XU(I,J,K)
```

```
150             FORMAT(F14.6)
                XU(I,J,K) = XU(I,J,K) * WEIGHT(K)
120          CONTINUE
             CALL CLOSE(12)

110        CONTINUE
100     CONTINUE

C                     84 = CHARACTER T, FIRST OF TIM EXTENSION.
        IF(INEXT(1).NE.84) CALL ZNORM(XU,NP,NG,ND)

        KCOR1 = 0
        KCOR2 = 0

        KNTU = 0
1000    KNTU = KNTU + 1

        DO 1400 J = 1, NG
           DO 1410 I = 1, NP
                DMAX(I,J) = 0.
1410       CONTINUE
1400    CONTINUE

        KNTUP = 0
2000    KNTUP = KNTUP + 1

        IF(DEBUG) TYPE 205, KNTUP,KNTU
205     FORMAT(1X,//,' EUCLIDIAN DISTANCES BETWEEN PATTERN ',I1,'
        & OF UNKNOWN SPEAKER ',I2,/,' AND PATTERN I OF KNOWN
        & SPEAKER J.'//)

        XMIN = 1.E30
        ITEMP = 1
        JTEMP = 1

        DO 210 J = 1, NG

           RMAX = -1.E30
           DO 220 I = 1, NP

              EUCL = 0.0
              DO 230 K = 1, ND
                 EUCL = EUCL + (XU(KNTUP,KNTU,K)-XU(I,J,K))**2
230           CONTINUE
              EUCL = EUCL**0.5
              IF(EUCL.GT.RMAX) RMAX = EUCL
              IF(EUCL.GT.XMIN) GO TO 250
              ITEMP = I
              JTEMP = J
              XMIN = EUCL
250           IF(DEBUG) WRITE(7,260) EUCL,I,J
                GO TO 220

255           IF(DEBUG) TYPE 265
265           FORMAT('+',16X,$)

260           FORMAT('+',F10.5,'[',I1,',',I2,']',$)
220        CONTINUE
           IF(DEBUG) TYPE 9
           DMAX(KNTUP,J) = RMAX

210     CONTINUE

        IF(JTEMP.EQ.KNTU) KCOR1 = KCOR1 + 1
        TYPE 9
```

```
      WRITE(7,280) KNTUP, KNTU,ITEMP,JTEMP,XMIN
280   FORMAT(' THE PATTERN #',I1,' OF UNKNOWN SPEAKER ',I2,
      &' IS IDENTIFIED WITH'/' THE PATTERN #',I1,' OF    KNOWN
      & SPEAKER ',I2,/,' DISTANCE =',F10.5,//)

      IF(KNTUP.LT.NP) GO TO 2000

         CALL SETDIS(DMAX,KNTU,JTMP,SETMIN,NG)
         IF(JTMP.EQ.KNTU) KCOR2 = KCOR2 + 1

      IF(KNTU.LT.NG) GO TO 1000

      INCOR1 = NG * NP - KCOR1
      CRATE1 = KCOR1 * 2.
      INCOR2 = NG - KCOR2
      CRATE2 = KCOR2 * 10.

      TYPE 9
      TYPE 9
      TYPE 9

      WRITE(7,500) (MFILU(M),M=5,14)
500   FORMAT(' SUMMARY OF RESULTS [ MASTER FILE = ',10A1,' ]'//)
      WRITE(7,510) KCOR1,INCOR1,CRATE1
510   FORMAT(' BY THE NEAREST NEIGHBOR RULE:'/' CORRECT = ',I3,
      &',    INCORRECT = ',I3,'    RATE(%) = ',F6.2,/)

      WRITE(7,520) KCOR2,INCOR2,CRATE2
520   FORMAT(' BY THE MINIMUM SET DISTANCE RULE:'/' CORRECT = ',I3,
      &',    INCORRECT = ',I3,'    RATE(%) = ',F6.2,/)
      TYPE 9

      CALL EXIT
      END
*
```

*

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C PROGRAM RDCFET
C
C RDCFET CREATES A NEW DATA FILE OF OPTIMUM FEATURES OF
C IDS OR LTS PARAMETER.
C
C PART OF PRE-PROCESSING STAGE OF EXPERIMENTAL PROCEDURE
C (CHAPTER II).
C
C H. NAKASONE
C 12-JULY-83
C
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C MICHIGAN STATE UNIVERSITY
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        COMMON /PASS/ FMT
        REAL X(50,100), Y(10)
        INTEGER KFET(10)
        LOGICAL*1 FNAMES(14,50), MFILE(14),INEXT(3),INDEV(3)
        LOGICAL*1 OUTEXT(3),OUTDEV(3),BEL,FMT(60)
        DATA OUTDEV(1)/'D'/,OUTDEV(2)/'K'/,OUTDEV(3)/'1'/
        DATA BEL/"007/

8       FORMAT(A1)
9       FORMAT('+',4X,'*',$)
        NUMF = 50
1000    TYPE 10
10      FORMAT(' MASTER FILE NAME (14A1)  >',$)
        READ(5,20) MFILE
20      FORMAT(14A1)

        TYPE 30
30      FORMAT(' EXT NAME OF INPUT DATA FILE (3A1)  >',$)
        READ(5,40) INEXT
40      FORMAT(3A1)

        TYPE 45
45      FORMAT(' EXT NAME OF OUTPUT DATA FILE (3A1)  >',$)
        READ(5,40) OUTEXT

        TYPE 50
50      FORMAT(' DEV NAME OF INPUT DATA FILE (3A1)  >',$)
        READ(5,40) INDEV

        TYPE 6
6       FORMAT(1X,/)

        CALL RDFILE(MFILE,FNAMES,INEXT,INDEV,NUMF)
        TYPE 9

        DO 100 I = 1, NUMF
           CALL ASSIGN(12,FNAMES(1,I),14,'OLD')
           READ(12,FMT) DUM, DUM, (X(I,J),J=1,100)
           CALL CLOSE(12)
100     CONTINUE
```

```
        TYPE 9

        DO 110 I = 1, NUMF
           DO 120 J = 1, 3
              FNAMES(J,I) = OUTDEV(J)
              FNAMES(J+11,I) = OUTEXT(J)
120        CONTINUE
110     CONTINUE

C  X(50,100) HAS BEEN FILLED.

        TYPE 8, BEL,BEL,BEL

        TYPE 200
200     FORMAT(' ENTER NUMBER OF FEATURES  >',$)
        READ(5,210) NFET
210     FORMAT(I3)

        TYPE 220
220     FORMAT(' ENTER FEATURE NODES BELOW.'//)

        DO 230 I = 1, NFET
           TYPE 240, I
240        FORMAT(' NODE # OF THE FEATURE',I3,'  >',$)
           READ(5,210) KFET(I)
230     CONTINUE

C  NOW, KFET(I), I.E., INFORMATION ON OPTIMUM FEATURES,
C  HAS BEEN FURNISHED.

        DO 300 I = 1, NUMF
           DO 310 J = 1, NFET
                Y(J) = X(I,KFET(J))
310        CONTINUE
           CALL ASSIGN(12,FNAMES(1,I),14,'NEW')
           WRITE(12,20) (FNAMES(M,I),M=1,14)
           DO 320 J = 1, NFET
              WRITE(12,330) Y(J)
330           FORMAT(F14.6)
320        CONTINUE
           CALL CLOSE(12)
300     CONTINUE

        TYPE 9
        TYPE 8,BEL

        WRITE(7,400) NUMF,OUTEXT,OUTDEV(3)
400     FORMAT(1X,'DONE.  ',I3,1X,3A1,'  FILES STORED IN DISK ',A1,//)
        GO TO 1000
        CALL EXIT
        END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C       PROGRAM NAKROP
C       NAKROP IS TO PREPARE A PROXIMITY MATRIX AND A
C       CATEGORY NAME FILE AS INPUT DATA REQUIRED BY A
C       PROJECTION PROGRAM 'SAMMON'.
C
C       SUBPROGRAM REQUIRED:  NONE.
C       PROGRAMMS TO BE CHAINED: CHNROA,CHNRO2, & CHNRO3
C
C       WRITTEN BY:  HIROTAKA NAKASONE
C             DATE:  28 JULY, 1983
C
C       DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C       MICHIGAN STATE UNIVERSITY
C       EAST LANSING, MICHIGAN
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        REAL*8 PROGA,PROG2,PROG3
        INTEGER NCATEG(10),CATNAM(5,10)
        LOGICAL*1 NAMOUT(14),YN,IDATE(20),TITLE(80)
        DATA PROGA/12RDK1CHNROASAV/,PROG2/12RDK1CHNRO2SAV/
        DATA PROG3/12RDK1CHNRO3SAV/

        TYPE 5
5       FORMAT(' NAKROP : PRELIMINARY TO SAMMON''S PROJECTION.'//)
        TYPE 20
20      FORMAT(' ENTER DESCRIPTION OF THIS RUN.'/' >',$)
        READ(5,16) (TITLE(L),L=1,72)
16      FORMAT(72A1)

        TYPE 25
25      FORMAT(' NUMBER OF CATEGORIES ?',$)
        READ(5,200) NCAT
200     FORMAT(I2)
        NFILES = 0

        DO 30 I = 1, NCAT
           TYPE 31, I
31         FORMAT(1X,'CATEGORY # ',I2,'  >',$)
           READ(5,32) (CATNAM(J,I),J=1,5)
32         FORMAT(5A2)
           NCATEG(I) = 5
           NFILES = NFILES + 5
30      CONTINUE

        TYPE 350
350     FORMAT(' OUTPUT CATEGORY FILE NAME(14A1) >',$)
        READ(5,370) (NAMOUT(N),N=1,14)
370     FORMAT(14A1)

C       PREPARE CAT FILES IN TWO DIFFERENT NAMES.

        YN = .FALSE.
        CALL ASSIGN(11,'CNAMES.DAT',10,'NEW')
2200    WRITE(11,4000) NCAT
4000    FORMAT(I2)
        DO 4010 I = 1, NCAT
           WRITE(11,4020) NCATEG(I), (CATNAM(M,I),M=1,5)
4020       FORMAT(I2,5A2)
```

```
4010    CONTINUE
        CALL CLOSE(11)

        IF(YN) GO TO 2300
        YN = .TRUE.
        CALL ASSIGN(11,NAMOUT,14,'NEW')
        GO TO 2200

2300    TYPE 3100, NAMOUT
3100    FORMAT(1X,/,' CNAMES.DAT = ',14A1,///)

3310    TYPE 3300
3300    FORMAT(' OPTION:  CHOOSE 1, 2, OR 3.'///'
        & 1   FOR   WITHIN-TRANSMISSION PROJECTION.'/'
        & 2   FOR   CROSS-TRANSMISSION BY 1 PARAMETER.'/'
        & 3   FOR   CROSS-TRANSMISSION BY 2 PARAMETERS.'///'
        & ?',$)

        READ(5,3400) YN
3400    FORMAT(A1)
        IF(YN.EQ.49) CALL CHAIN(PROGA,,0)
        IF(YN.EQ.50) CALL CHAIN(PROG2,,0)
        IF(YN.EQ.51) CALL CHAIN(PROG3,,0)
        IF(YN.NE.49  .AND. YN.NE.50  .AND. YN.NE.51) GO TO 3310

        CALL EXIT
        END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C PROGRAM CHNRO2:
C
C CHNRO2 IS CHAINED FROM NAKROP. IT CAN BE ALSO RUN FROM THE KEY
C BOARD DIRECTLY.
C
C INPUT IS A MASTER DATA FILE NAME(S).
C
C OUTPUT IS A PROXIMITY MATRIX(UPPER HALF DIAGNAL).
C
C PURPOSE OF THIS PROGRAM IS TO PREPARE A SPECIFIC DATA FORMAT REQUIRED
C BY A LINEAR PROJECTION PROGRAM, 'SAMMON'.
C
C SUBPROGRAMS: SBWEIT,RDFILE,ZNORM,EUCMAT,INDEX.
C
C H. NAKASONE
C 28-JULY-83
C
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES
C MICHIGAN STATE UNIVERSITY
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        REAL*8 PROGS
        REAL XU(5,10,10), XK(5,10,10),WEIGHT(10)
        LOGICAL*1 MFILU(14),MFILK(14),LFILE(14)
        LOGICAL*1 UFILE(14,50), KFILE(14,50), INEXT(3),INDEV(3)
        LOGICAL*1 OUTEXT(3),OUTDEV(3),DEBUG,YN,NAMPRX(14)
        DATA INDEV(1)/'D'/,INDEV(2)/'K'/,INDEV(3)/'1'/
        DATA PROGS/12RDK1SAMMONSAV/

9       FORMAT(1X,/)
14      FORMAT(3A1)
        TYPE 10
10      FORMAT(' PROGRAM CHNRO2: CROSS-TRANSM. BY 1 DOMAIN.'//)
        TYPE 13
13      FORMAT(' EXT NAME OF INPUT FILE(FRI,FRL, OR TIM)  >',$)
        READ(5,14) INEXT
        TYPE 15
15      FORMAT(' MASTER FILE NAME FOR UNKNOWN (14A1)  >',$)
        READ(5,20) MFILU
20      FORMAT(14A1)

        TYPE 25
25      FORMAT(' MASTER FILE NAME FOR   KNOWN (14A1)  >',$)
        READ(5,20) MFILK

        TYPE 26
26      FORMAT(' ENTER NUMBER OF DIMENSIONS          >',$)
        READ(5,27) ND
27      FORMAT(I3)

        TYPE 28
28      FORMAT(' OUTPUT NAME FOR PROXIMITY MATRIX(14A1) >',$)
        READ(5,20) (NAMPRX(M),M=1,14)

        DEBUG = .FALSE.
        TYPE 4500
4500    FORMAT(' DEBUG ?',$)
        ACCEPT 4510,YN
        IF(YN.EQ.89) DEBUG = .TRUE.
4510    FORMAT(A1)

        CALL SBWEIT(WEIGHT,ND)
```

```
      CALL RDFILE(MFILU,UFILE,INEXT,INDEV)
      CALL RDFILE(MFILK,KFILE,INEXT,INDEV)

C  UFILE AND KFILE HAVE BEEN FILLED.

      NP = 5
      NG = 10

C TRANSFER DATA FROM UFILE AND KFJILE TO XU AND XK, RESPECTIVELY.

      KT = 0
      DO 100 J = 1, NG

        DO 110 I = 1, NP
          KT = KT + 1
          CALL ASSIGN(12,UFILE(1,KT),14,'OLD')
          READ(12,20) LFILE
          DO 120 K = 1, ND
            READ(12,150) XU(I,J,K)
150         FORMAT(F14.6)
            XU(I,J,K) = XU(I,J,K) * WEIGHT(K)
120       CONTINUE
          CALL CLOSE(12)

          CALL ASSIGN(13,KFILE(1,KT),14,'OLD')
          READ(13,20) LFILE
          DO 130 K = 1, ND
            READ(13,150) XK(I,J,K)
            XK(I,J,K) = XK(I,J,K) * WEIGHT(K)
130       CONTINUE
          CALL CLOSE(13)
110     CONTINUE
100   CONTINUE

      IF(INEXT(1).EQ.84) GO TO 200
C
C              --- STANDARDIZE DATA ---

      CALL ZNORM(XU,NP,NG,ND)
      CALL ZNORM(XK,NP,NG,ND)

      IF(DEBUG) TYPE 4210
4210  FORMAT(' NORMALIZATION DONE.'/)

200   CALL EUCMAT(XU,XK,NAMPRX,ND,NP,NG)
      IF(DEBUG) TYPE 4215
4215  FORMAT(' EUCMAT RETURNED.  NOW CHAINING TO DRVSAV.'//)

      CALL CHAIN(PROGS,,0)
      CALL EXIT
      END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C FILE:  FILFIL.FOR
C FILFIL IS TO CREATE A MASTER FILE OF FILE NAMES.
C MAXIMUM NUMBER OF FILES CAN BE STORED IN A MASTER
C FILE IS CURRENTLY 100.
C
C HIROTAKA NAKASONE, 22-MAY-1983
C DEPARTMENT OF AUDIOLOGY AND SPEECH SCIENCES, MSU.
C
C TO COMPILE:   FILFIL=FILFIL/U
C TO LINK    :   FILFIL=FILFIL,LCHECK,SYSLIB/F
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        COMMON /NAME/FNAMES
        COMMON /NUMF/NUMF
        DOUBLE PRECISION EXT
        LOGICAL*1 FNAMES(15,100),LETR(9),YN,LNAME(15),FMT(80)
        LOGICAL*1 NEXT(3)
        DATA LNAME(1)/'D'/,LNAME(2)/'K'/,LNAME(4)/':'/
        DATA LNAME(10)/'.'/

        LNAME(15) = 0
1       FORMAT(1A1)
2       FORMAT(I3)
3       FORMAT(6A1)
4       FORMAT(14A1)
5       FORMAT(12A1)
6       FORMAT(15A1)
7       FORMAT(3A1)
8       FORMAT(1X,' WAIT.')
9       FORMAT(1X,/)

C  TO HERE TO ENTER ALL FILES.

        TYPE 30
30      FORMAT(' ENTER NUMBER OF FILES (I3)   >',$)
        READ(5,40) NUMF
40      FORMAT(I3)
        TYPE 50
50      FORMAT(' SPECIFY EXT NAME (3A1)   >',$)
        READ(5,7) (NEXT(N),N=1,3)
        CALL IRAD50(3,NEXT(1),EXT)
        TYPE 60
60      FORMAT(' ENTER ALL FILES BELOW (6A1).',//)

        DO 70 I = 1,NUMF
75         TYPE 80,I
80         FORMAT(' FILE #',I3,'   >',$)
           IF(LCHECK(EXT,ICHAN,I).GT.0) GO TO 79
              TYPE 90
90            FORMAT(1X,'*FATAL ERROR*')
              STOP
79         CALL CLOSEC(ICHAN)
           CALL IFREEC(ICHAN)
70      CONTINUE

        IE = 5
        WRITE(7,100)
100     FORMAT(//,1X,' TABLE OF FILE NAMES '//)
```

```
          DO 120 J = 1, NUMF
               WRITE(7,130) (FNAMES(M,J),M=1,15),J
130            FORMAT(1X,15A1,'[',I3,']')
120       CONTINUE
          TYPE 9

145     TYPE 140
140     FORMAT(' OK ?',$)
        ACCEPT 1,YN
        IF(YN.EQ.89) GO TO 200

C         THEN DO SOME CORRECTION BUSINESS HERE.

        TYPE 150
150     FORMAT(' SPECIFY THE FILE NUMBER TO BE CORRECTED >',$)
        ACCEPT 2,I
155     TYPE 160,I
160     FORMAT(' ENTER THE CORRECT FILE NAME FOR #',I3,'  >',$)
        IF(LCHECK(EXT,ICHAN,I).LT.0) STOP'FATAL ERROR'
        CALL CLOSEC(ICHAN)
        CALL IFREEC(ICHAN)
        GO TO 145

200     TYPE 210
210     FORMAT(' ENTER MASTER FILE NAME(DEV:FILNAM.EXT) >',$)
        READ(5,4) (LNAME(L),L=1,14)
        CALL ASSIGN(12,LNAME,14,'NEW')
        WRITE(12,220) (LNAME(L),L=1,14),NUMF
220     FORMAT(14A1,I3,'  (5E15.7)')

        DO 230 I = 1,NUMF
           WRITE(12,4) (FNAMES(M,I),M=1,14)
230     CONTINUE
        CALL CLOSE(12)
        CALL EXIT
        END

*
```

```
C FILE:  ZNORM.FOR
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C SUBROUTINE ZNORM: TO STANDARDIZE FEATURE VALUES OF SPEECH
C PARAMETERS,IDS, FFC, AND LTS.
C Written by: Hirotaka Nakasone
C       Date: 11 JULY, 1983
C Dept. of Audiology and Speech Sciences, MSU, East Lansing, MI.
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        SUBROUTINE ZNORM(X,NP,NG,ND)
        REAL X(NP,NG,ND)
C       NP = NUMBER OF PATTERNS
C       NG = NUMBER OF GROUPS, OR SPEAKERS
C       ND = NUMBER OF DIMENSIONS, OR SAMPLES / SPEAKER

        NPAT = NP * NG
D100    TYPE 2
2       FORMAT(1X,'NORMALIZATION BY Z-TRANSFORM.'/)
        DO 10 J=1,ND
           RM = 0.0
           SD = 0.0

           DO 20 I = 1, NG
              DO 25 K = 1, NP
                RM = RM + X(K,I,J)
                SD = SD + X(K,I,J)*X(K,I,J)
25            CONTINUE
20         CONTINUE

           RM = RM/FLOAT(NPAT)
           SD = (SD/FLOAT(NPAT) - RM*RM)**0.5

           IF(SD.EQ.0.) GO TO 40
                DO 30 I = 1, NG
                 DO 35 K = 1, NP
                   X(K,I,J) = (X(K,I,J)-RM)/SD
35                CONTINUE
30              CONTINUE
                GO TO 10

40         CONTINUE
           DO 50 I = 1, NG
              DO 45 K = 1, NP
                X(K,I,J) = 0.0
45            CONTINUE
50         CONTINUE

10      CONTINUE
        RETURN
        END
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C SUBROUTINE RDFILE.FOR.
C RDFILE IS TO READ ALL THE FILES IN A MASTER FILE.
C 11 JULY 1983
C H. NAKASONE
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        SUBROUTINE RDFILE(MFILE,NAME50,EXT,DEV)
        COMMON /PASS/ FMT
        LOGICAL*1 MFILE(14),LNAME(14),FMT(60),NAME50(14,50)
        LOGICAL*1 EXT(3), DEV(3)

        CALL ASSIGN(13,MFILE,14,'OLD')
        READ(13,10) LNAME,NUMF,(FMT(M),M=1,60)
10      FORMAT(14A1,I3,60A1)

        DO 20 I = 1, NUMF
           READ(13,30) (NAME50(N,I),N=1,14)
30         FORMAT(14A1)

           DO 40 J = 1, 3
              NAME50(J+11,I) = EXT(J)
              NAME50(J,I) = DEV(J)

40         CONTINUE

20      CONTINUE

        CALL CLOSE(13)
        RETURN
        END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
         SUBROUTINE SUBF(NP,NG,ND)
C        COMPUTATION PART FOR FRATIO'S
C        BY H. NAKASONE
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
         COMMON /SENSE/ X, F
         REAL X(5,10,20), SUM(10),SSQ(10),F(20)

C        ND:  HOLDS NUMBER OF DIMENSIONS IN EACH NP.
C        NP:  HOLDS NUMBER OF PATTERNS IN EACH NG.
C        NG:  HOLDS NUMBER OF SPEAKERS (OR GROUPS).

         DFB = FLOAT(NG - 1)
         DFW = FLOAT(NG*NP - NG)
         GN = FLOAT(NP*NG)
         SN = FLOAT(NP)

         DO 1 K = 1, ND
            GSUM = 0.
            GSSQ = 0.
            T = 0.

            DO 2 J = 1, NG
               SUM(J) = 0.
               SSQ(J) = 0.

               DO 3 I = 1, NP
                  SUM(J) = SUM(J) + X(I,J,K)
                  SSQ(J) = SSQ(J) + X(I,J,K)*X(I,J,K)
3              CONTINUE

               GSUM = GSUM + (SUM(J)*SUM(J))/SN
               GSSQ = GSSQ + SSQ(J)
               T = T + SUM(J)

2           CONTINUE

            SSTOT = GSSQ - (T*T)/GN
            SSBET = GSUM - (T*T)/GN
            SSWIT = SSTOT - SSBET
            BGMS = SSBET / DFB
            WGMS = SSWIT / DFW


D           TSWG = 0.
D           DO 4 J = 1, NG
D              TSWG = TSWG + ( SSQ(J) - (SUM(J)*SUM(J))/ SN )
D4          CONTINUE
D           WRITE(7,10) SSTOT,SSBET,SSWIT
D10         FORMAT(' SSTOT=',F14.3,' SSBET=',F14.3,' SSWIT=',F14.3)
D           WRITE(7,12) BGMS,WGMS,TSWG
D12         FORMAT(' BGMS=',F14.3,' WGMS=',F14.3,' TSWG=',F14.3)
            F(K) = 9999.9091
            IF(WGMS.GT.0.) F(K) = BGMS / WGMS
1        CONTINUE
         RETURN
         END
```

```
C HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
        SUBROUTINE SUBSET(AU,BK,NP,NG,ND)
C
C TO COMPUTE MINIMUM SET DISTANCE OF SPEAKERS RECORDED
C THROUGH TWO DIFFERENT TRANSMISSION AND/OR RECORDING
C MEDIA.
C CALLED FROM NAKRO2 AND NAKRO3( both programs used in
C the final stage of voice I.D. in a Ph.D. dissertation.)
C
C 21-JUL-83
C H. NAKASONE
C HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        REAL AU(5,10,ND), BK(5,10,ND),SETMIN(10)

C                INITIALIZE

        KNTC = 0
        DO 1 I = 1, NG
           SETMIN(I) = 0.
1       CONTINUE

        TYPE 2
2       FORMAT(' Summary by Minimum Set distance Rule:'//
        &' Unknown Speaker ',T20,'Known Speaker',T40,
        &'Set Distance'/)

        TYPE 9
9       FORMAT(' ------------------------------------------------
        &-------'/)

        KUG = 0
100     KUG = KUG + 1

        KNG = 0
30      KNG = KNG + 1

        SMIN = 1.E20

        DO 5 JU = 1, NP

           SMAX = -1.E20
           DO 10 JK = 1, NP

               EUCL = 0.
               DO 20 K = 1, ND
                  EUCL = EUCL + (AU(JU,KUG,K)-BK(JK,KNG,K))**2.0
20             CONTINUE
               EUCL = EUCL**0.5

               IF(EUCL.LE.SMAX) GO TO 10
                 SMAX = EUCL
                 JKTEMP = JK
10         CONTINUE
           IF(SMAX.GE.SMIN) GO TO 5
           SMIN = SMAX
           JUTEMP = JU
5       CONTINUE

        SETMIN(KNG) = SMIN
        KGTEMP = JUTEMP
```

```
      IF(KNG.LT.NG) GO TO 30
      RMIN = SETMIN(1)
      KTEMP = 1

      DO 60 K = 2, NG
         IF(SETMIN(K).GT.RMIN) GO TO 60
         RMIN = SETMIN(K)
         KTEMP = K
60    CONTINUE

      IF(KUG.EQ.KTEMP) KNTC = KNTC + 1
      WRITE(7,70) KUG,KTEMP,RMIN
70    FORMAT(8X,I2,T25,I2,T39,F12.5,/)

      IF(KUG.LT.NG) GO TO 100
      TYPE 9
      P = KNTC * 10.
      TYPE 99,KNTC,NG-KNTC,P
99    FORMAT(1X,/,' Correct =',I2,',    Incorrect =',I2,',    Rate(%)
      & =',F6.2)

      RETURN
      END
*
```

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C        SUBROUTINE EUCMAT:  TO GENERATE A DISSIMILARITY
C        MATRIX BY USING EUCLIDIAN DISTANCE MEASURE.
C
C        CALLED FROM CHNRO2 AND CHNRO3.
C
C        WRITTEN BY:  HIROTAKA NAKASONE
C              DATE:  28 JULY, 1983
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        SUBROUTINE EUCMAT(XU,XK,NAMOUT,ND,NP,NG)

        COMMON /NFILES/ NFILES
        REAL XU(NP,NG,ND),XK(NP,NG,ND),X(50,20),DMAT(1300)
        LOGICAL*1 NAMOUT(14),YN

C       ARRANGE X ARRAY SO THAT ELEMENTS 1 TO 25 WILL CONTAIN
C       DATA OF  UNKNOWN SPEAKERS, AND 26 TO 50, OF KNOWN SPEAKERS.

        NFILES = NG*NP
        JU = 0
        DO 202 J = 2, 10, 2
           DO 205  I = 1, 5
                JU = JU + 1
                JK = JU + 25
                DO 210  K = 1, ND
                   X(JU,K) = XU(I,J,K)
                   X(JK,K) = XK(I,J,K)
210             CONTINUE
205       CONTINUE
202     CONTINUE

COMPUTE EUCLIDIAN DISTANCE.

        DO 300 I = 1, 49
           DO 310 J = I+1, 50
                SUM = 0.
                DO 320 K = 1, ND
                   SUM = SUM + (X(I,K)-X(J,K))**2.
320             CONTINUE
                DMAT(INDEX(I,J)) = SUM**0.5
310        CONTINUE
300     CONTINUE

CONTINUE TO WRITE THE RESULTS.

        YN = .FALSE.
        CALL ASSIGN(12,'PROXTP.DAT',10,'NEW')
5400    WRITE(12,1250) NFILES
1250    FORMAT(I3,' 1'/ '(10F8.3)' )

        DO 1400 I = 1, NFILES-1
           WRITE(12,1600) (DMAT(INDEX(I,J)),J=I+1,NFILES)
1400    CONTINUE
1600    FORMAT(10F8.3)
        CALL CLOSE(12)

        IF(YN) GO TO 2100
        YN = .TRUE.
        CALL ASSIGN(12,NAMOUT,14,'NEW')
        GO TO 5400
```

```
2100    WRITE(7,2200) NAMOUT
2200    FORMAT(1X,'TWO RECORDS,  PROXTP.DAT  AND ',14A1,'  DONE.'/)
        RETURN
        END

        INTEGER FUNCTION INDEX(I,J)
        COMMON /NFILES/ NFILES
        INTEGER ROW,COL

        ROW = MINO(I,J)
        COL = MAXO(I,J)
        INDEX = (ROW-1)*NFILES-ROW*(ROW+1)/2+COL
        RETURN
        END
```

*

```
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
C
C SUBROUTINE SETDIS:
C
C SETDIS IS TO FIND THE MINIMUM SET DISTANCE BETWEEN
C TWO SPEAKERS EACH REPRESENTED BY A VECTOR OF 9 OR
C 10 DIMENSIONS (FEATURES).
C SETDIS IS CALLED FROM VOICE IDENTIFICATION PROGRAMS,
C NAKRO4 AND NAKROA.
C
C 21-JUL-83
C H. NAKASONE
C
CHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH

        SUBROUTINE SETDIS(DMAX,KUG,JTMP,TMIN,N) .

        DIMENSION DMAX(5,10)

        JTMP = 1
        TMIN = 1.E30
        DO 20 J = 1, N
           SETMIN = DMAX(1,J)
           DO 10 I = 2, 5
                IF(DMAX(I,J).GT.SETMIN) GO TO 10
                SETMIN = DMAX(I,J)
10         CONTINUE
           IF(SETMIN.GT.TMIN) GO TO 20
                TMIN = SETMIN
                JTMP = J
20      CONTINUE

        TYPE 30,KUG,JTMP,TMIN
30      FORMAT(' BY SET DISTANCE RULE:'//
        & UNKNOWN ',I2,' IS IDENTIFIED WITH KNOWN ',I2,/,'
        & WITH THE MINIMUM SET DISTANCE =',F12.6,/)

        RETURN
        END
*
```