1. Introduction 1

2. Hardware Interface 2

2.1. PCMCIA 2.0 VoiceCard 2

2.1.1. General 2

2.1.2. Establishing the interface 2

2.1.3. API Calling convention 2

2.1.4. Handshaking 2

2.1.5. Popup Client synchronisation 2

2.1.6. Tablet 3

2.2. Internal (PC /SE) 4

2.2.1. Interface summary 4

2.2.2. Handshaking 4

2.2.3. Tablet 4

2.3. Stand Alone (SA) 5

2.3.1. Connector 5

2.3.2. Data protocol 5

2.3.3. Handshaking 5

3. Control codes 6

3.1. Command code summary 6

3.2. Return code summary 8

3.3. Text to Speech control code summary 10

4. Speech Commands 12

4.1. Speech Control 12

4.1.1. Speak now 12

4.1.2. Speak (small delay) 12

4.1.3. Fast word reading 12

4.2. Mode Resets 12

4.2.1. [t] text reading mode reset 12

4.2.2. [p] Phoneme Reading Mode reset 13

4.2.3. [x] Dictionary entering mode 13

4.3. Text Normalizer Resets 13

4.3.1. alpha literal 14

4.3.2. digit literal 14

4.3.3. prosodic punctuation literal 14

4.3.4. whitespace literal 14

4.3.5. arithmetic pronunciation 14

4.3.6. full number 14

4.3.7. forced lower case 15

4.3.8. control character pronunciation 15

4.3.9. time pronunciation 15

4.3.10. abbreviation expansion 15

4.4. Field Resets 15

4.5. Homograph marker 15

4.6. Special Punctuation Marks for Changing Prosody 16

4.7. Voice Characteristic Resets 16

4.7.1. Speaking Rate 16

4.7.2. Voiced excitation function 16

4.7.3. Unvoiced gain 17

4.7.4. Fundamental frequency 17

4.7.5. Pitch topline 17

4.8. Index Markers 17

4.9. Changing the form of the Reset Commands 17

4.9.1. Changing the lead-in character 18

4.9.2. Double the lead-in character 18

5. T-T-S (tm) Text Normalizer Performance 19

5.1. Overview 19

5.2. Pronouncing Numbers 20

5.3. Pronouncing Letters and Words 22

5.4. Homographic Spellings 23

5.5. Interpreting Punctuation 23

6. T-T-S (tm) American English Phonemes 25

6.1. Boundaries and Silence 25

6.2. Precisely Specifying Stress and Pitch 25

6.3. Transcription Conventions 26

6.4. Symbols for Phonemic Transcription of Words 26

6.4.1. Consonant phonemes: 26

6.4.2. Vowel phonemes (as they are pronounced in stressed

syllables): 27

Introduction

Keynote GOLD synthesizers are available in four formats:

In a type II PCMCIA computer card. (Voicecard Model)

For IBM PC and compatible computers using the industry standard

8 bit bus, (PC model)

For selected models of the Toshiba range of portable laptop and

notebook style computers, using the internal modem connection,

(SE model)

As a stand alone synthesizer with a serial RS-232 connection to

the host computer. (SA model)

This document describes the Keynote GOLD speech synthesizer

interface and communication protocol.

Hardware Interface

PCMCIA 2.0 VoiceCard

General

To use the Voicecard, the supplied software driver must be

running.

Establishing the interface

An INTERRUPT 2Fh function is provided to establish an interface

to the voice card driver.

To minimize conflicts with other INT 2F users, the Voicecard

driver may use any multiplex number in the range 20h to 3fh. To

locate the Voicecard multiplex number, execute the Get Voicecard

API interrupt call repeatedly with AH set to each multiplex

number in turn.

INTERRUPT 2Fh, Function 5643 Get VoiceCard API address

Registers at call: AH=multiplex ID (range 20h to 3fh)

AL=00

BX=4B47h 'KG' (Keynote Gold)

CX=5643h 'VC' (Voice Card)

Return: AL = FFh

BX = 4F4Bh 'OK'

ES:DI = Voice card API function address.

The correct multiplex number is indicated by BX='OK'. To send

commands to the VoiceCard, execute a FAR call to the address

returned in ES:DI.

API Calling convention

The driver API function uses the 'C' calling convention. It

takes a single byte parameter pushed onto the stack, and returns

a char type in AL.

The API function preserves the segment registers and SI, DI, BP,

but may destroy the contents of any other register.

The caller must remove the pushed parameter from the stack when

the VoiceCard driver returns.

Handshaking

Keynote asynchronous commands may be sent at any time.

Keynote synchronous commands, and output text may be sent at any

time. However, the driver will only accept text, and text

synchronous commands if the driver's internal buffer is not

full. Character input which is accepted returns 0. Character

input which is not accepted will return with a BUFFER\_FULL

return code (01h).

The asynchronous commands are listed on page 6, synchronous and

text commands on page 10.

Popup Client synchronisation

The VoiceCard API supplies a busy status flag (page 6). To avoid

possible deadlock, TSR VoiceCard clients must check this flag

before popping up.

The VoiceCard driver is interrupt driven and uses very little

processing time. The API is available at all times and the

driver is interruptable. System interrupts are left on during

VoiceCard processing.

However, under these conditions it is possible for a pop-up TSR

to pop-up during the VoiceCard processing. If this occurs, and

the pop-up screen reader then waits for speech to finish, the

screen reader will wait forever. Therefore before popping up,

TSRs should check the VoiceCard busy status, and pop up only if

the status return value is zero.

VoiceCard client programs may output text or any synthesizer

commands regardless of the Busy Status value as the API is

available for use at all times.

Tablet

An INTERRUPT 2Fh function is provided to locate the IO address

for the tablet interface.

INTERRUPT 2Fh, Function 5643 Get VoiceCard Tablet address

Registers at call: AH=multiplex ID (range 20h to 3fh)

AL=01 (get tablet IO address)

BX=4B47h 'KG' (Keynote Gold)

CX=5643h 'VC' (Voice Card)

Return: AL = FFh

BX = 4F4Bh 'OK'

DX = Voice card tablet address in IO space.

The value in DX is the IO address for the tablet trigger port.

The tablet status port is at AX+4. This call can not be made

until the multiplex ID has been found using the Get Voicecard

API interrupt 2Fh call.

Port Address Data

Tablet Trigger Port (Write Only) AX

Tablet status port (Read Only) AX+4 Bit 0 - Tablet left

button Bit 1 - Tablet Right button Bit 2 - Tablet X Bit 3 -

Tablet Y Bits 4-7 Reserved

Writing arbitrary data to the tablet port triggers the tablet

hardware. Data read from this port is undefined. The tablet

status bits (X and Y) go low immediately, and return to a logic

"1" level after a time period proportional to the position at

which the tablet is being touched.

Touching the tablet at the top left hand corner produces a short

pulse, touching it at the lower right hand corner produces a

long pulse. The maximum delay before the status bits reset

(return to a high level) is usually less than 12ms. The tablet

may not be retriggered less than 4 micro seconds after both x

and y bits have reset.

Internal (PC /SE)

Interface summary

In the following table "input" and "output" are from the point

of view of the host computer.

Speech port Tablet port

PC Address Jumper in lower position 2a8 2ac

PC Address Jumper in upper position 2e8 2ec

SE Address 3f8 3a8

Input Bit 7 - Speech buffer handshake (1=full) Bits 0-6 Status

return value Bit 0 - Tablet Y Bit 1 - Tablet X Bit 2 - Tablet

right button Bit 3 - Tablet left button Bit 4 - Speech port

handshake (1=not ready, 0=ready) Bit 5-7 not used

Output 8 bit speech data Trigger tablet

Handshaking

Two levels of handshaking are used:

Data may only be transferred to the speech card when the port

handshaking bit is low. This handshaking operates on a

character-by-character basis.

In addition text, and text synchronous commands, may only be

sent when the text buffer is not full.

The asynchronous commands are listed on page 6, synchronous and

text commands on page 10.

Tablet

Writing arbitrary data to the tablet port triggers the tablet

hardware. The tablet status bits (X and Y) go low immediately,

and return to a logic "1" level after a time period proportional

to the position at which the tablet is being touched.

Touching the tablet at the top left hand corner produces a short

pulse, touching it at the lower right hand corner produces a

long pulse. The maximum delay before the status bits reset

(return to a high level) is usually less than 12ms. The tablet

may not be retriggered less than 4 micro seconds after both x

and y bits have reset.

Stand Alone (SA)

Connector

Pin 2 Transmit data

Pin 3 Receive data

Pin 5 Common

The SA requires a "Straight through" cable to connect to a

standard 9 pin IBM PC serial connector.

Data protocol

Rom version prior to version 7: 38400 Baud 8 data bits 1 stop

bit No parity Rom version 7 and later: 9600 Baud 8 data bits 1

stop bit No parity

Handshaking

Keynote asynchronous commands may be sent at any time.

Text, and text synchronous commands may only be sent when the

XON condition is active.

Control codes

Command code summary

The commands in the following table may be sent at any time,

independent of speech buffer handshaking status.

Command Send Reply Comment

(hex) VoiceCard PC/SE SA Ver<=6 SA Ver>=7

Reset 01,.... Total restart. Second byte selects language

English Spanish French German Italian Dutch Japanese Default

01,EF 01,EE 01,ED 01,EC 01,EB 01,EA 01,E9 01,E0 - - - - - - -

- F7 F7 F7 F7 F7 F7 F7 F7 17 17 17 17 17 17 17 17 A null

should be sent after this command, as PC and SE cards discard

the following character. For SA, assume XOFF until reset

acknowledge is received.

Return Status 02,.... Select status returned by

synthesizer.

Index 02,80 index index+80 index+1C Return current index

marker, range 0-63(hex)

Serial no. 1 02,81 00 00 80

Serial no. 2 02,82 00 00 80

Serial no. 3a 02,83 00 00 80

Self test report 02,84 00 00 80

operating error word 02,85 00 00 80

version no. 02,86 Version Version Version +80

exist code 1 02,87 7A 7A FA

exist code 2 02,88 58 58 D8

SA status 02,89 - Packet, see below

SA low battery acknowledge 02,8A - - - Stops SA "charge

battery" announcement, and low battery status transmission

language set 1 02,8B Packet, see below Returns languages

available

language set 2 02,8C 00 00 80

default language 02,8D Packet, see below Returns the

power on default language.

Busy Status 02,8E 00 or non zero - - VoiceCard only. TSR

clients must not pop up while the busy flag is non zero.

Instant (asynchronous) speech control commands These commands

affect speech as soon as possible.

Cutoff 03 - F5 15 Immediately silence speech, and clear

buffers. For SA, assume XOFF until cutoff acknowledge is

received.

Speaking rate 04,(80-FF) - - -

Volume (not PC) 05,(80-FF) - - -

Other commands

Tone () 06, Duration, Count high, Count low. - F1 12

Reply sent when tone ends.

Reserved 07

Tablet scan (SA only) 08 - Packet, see below

Set wakeup timer (SA only) 09,(81-FF) - F2 14 Set wakeup

timer period. SA sends wakeup code at regular intervals.

Stop wakeup timer 09,80 - - -

The tone command is 6 bytes long. Duration parameter is one

byte. Count parameters are two byte, most significant byte

first. Counter runs at 2.5Mhz, so Count H = Count L = 1250(dec)

gives a 1KHz tone.

Return code summary

Return VoiceCard PC/SE SA Ver<=6 SA Ver>=7 Comment

Status return range 00-7F 00-7F 80-FF

Index range 00-63 80-E3 1C-7F

Status return packets

Language Set 1 (single byte) Bit 0 - English Bit 1 - Spanish

Bit 2 - French Bit 3 - German Bit 4 - Italian Bit 5 - Dutch Bit

6 - Japanese As for PC Bit 7=0 As for PC Bit 7=1 Sent in

response to language set 1 request. Bits are set (1) if the

language is available in the synthesizer.

Default language (single byte) 01 - English 02 - Spanish 03 -

French 04 - German 05 - Italian 06 - Dutch 07 - Japanese As for

PC Bit 7=0 As for PC Bit 7=1 Sent in response to default

language request. Return code indicates default language used

by synthesizer at power on.

SA Status (single byte) - Bit 0 - Battery low Bit 1 - Tone

active Bit 2 - Buffer full Bit 3 - Charger on Bits 4-6 not used

Sent in response to SA status request. Bits are set (1) if the

condition is true. Buffer full corresponds to XOFF condition

SA touch tablet (3 byte) - First byte: Bit 0 - Y7 Bit 1 - Y8

Bit 2 - X7 Bit 3 - X8 Bit 4 - Switch 2 Bit 5 - Switch 1 Bit 6 -

not used Second byte: Bits 0-6 X coordinate Third byte: Bits

0-6 Y coordinate Three byte packet sent in response to an SA

tablet request. X and Y coordinates are given as 9 bit numbers.

(0,0) is in the lower right corner. The switch bits are set if

the switch is being pressed. X and Y are 1FF when the tablet is

not being touched.

Asynchronous Return Codes (SA only)

Low battery - F0 10 Battery voltage low. This code is

retransmitted every 30 seconds until acknowledged

XON - F3 11

Tone finished - F1 12 Sent when tone stops sounding

XOFF - F4 13

Wakeup - F2 14 Sent at requested intervals

Cutoff acknowledge - F5 15 Assume XON after this

Power on - F6 16 Assume XON after this

Reset acknowledge - F7 17 Assume XON after this

Goodbye - FF 18 Sent 3 minutes after the first low battery

indication. SA will switch off after sending this code.

Text to Speech control code summary

The following commands are inserted in the text to be spoken.

The commands are acted on as they are reached during speech.

They are therefore known as "synchronous" commands because they

are synchronized to the speech output.

In the following table:

The "space" character (hex 20) is shown as

Where brackets {} delineate numbers, they are not included in

the text sent to the synthesizer. For example[r100] is the

default speech rate command.

Control characters are delineated by <>. For example <CR> is the

carriage return character

Command Code Comment

Reading

Speak now ~| or 1F(hex) Commence speaking previous text.

Small delay } <CR> Commence speech, insert a small delay

Fast word reading [~2,1] Read immediately without waiting for

punctuation

Voice

Speaking rate [r{percentage}] Percentage of normal speed,

range -90 to 200.

Voiced excitation [e{num}] {num} from 2 to 7. Whisper = 2,

normal = 3

Unvoiced gain [u{num}] {num} in units of 0.75db, range -50 to

20.

Fundamental freq [f{freq}] {freq} in Hz, range 39 to 200

Pitch topline [h{num}] Alters pitch range of voice. {num} in

range -300 to 100

Index

Index marker [i{num}] {num} in range 0 - 99(dec)

or 1D(hex),0-63(hex) where the second byte is the index mark

number.

Text Normaliser "x" is 0 (off) or 1 (on)

Alpha literal [n1,x] Pronounce names of letters (spelling)

default: off

Digit literal [n2,x] Pronounce digits individually. default:

off

Prosidic punctuation literal [n3,x] Pronounce commas, periods

etc. default: off

Whitespace literal [n4,x] Pronounce space, carriage return,

tab. default: off

arithmetic pronunciation [n5,x] Read mathematical texts.

Default: off

full number [n6,x] Inhibit grouping of digits, read full

numbers. default: off

forced lower case [n7,x] Pronounce uppercase letter groups as

words. Default:off

control character pronunciation [n8,x] Pronounce control

characters. Default: on

time pronunciation [n9,x] Pronounce times of day. (e.g. 8:00 =

eight o'clock) Default: on

abbreviation expansion [n10,x] Default: on

Pronounciation

homograph mark ~ Place immediately ahead of word to select

alternate pronunciation, e.g. minute ~minute

Prosidy

comma-like pause ]

period-like pause }

stress ~+ Stress the following word

de-stress ~- Remove stress from the following word

emphasis ~! Emphasize the following word (de-stress the rest

of the sentence)

question ~? Produce rising intonation at the end of the

sentence

Mode switching

Text mode [t] Normal reading mode

Phoneme mode [p] Following text read according to phoneme

rules (See appendix)

Dictionary mode [x] Begin dictionary entering mode.

Speech Commands

The following commands are inserted as ASCII characters in the

text input to the synthesizer. They are acted on as the speech

passes that point in the text. In the commands below

indicates space character.

Speech Control

Keynote GOLD will normally begin speaking after punctuation.

However in cases where no punctuation is output to the card, or

where individual words must be spoken separately the following

commands can be used to initiate speech.

Speak now

~|

This command initiates speech immediately after the next word.

Words must be delimited by spaces or punctuation. The speak now

command disrupts prosody (intonation) contouring, and should

normally be used only for word-by-word pronunciation. A small

delay is incurred between words using this command. See "fast

word reading" below.

Speak (small delay)

} <CR>

This command may be used to initiate speech in the absence of

punctuation. Speech begins immediately, and a small comma-like

pause is inserted at the command.

Fast word reading

[~2,1] is ON. The default [~2,0] is OFF.

Keynote GOLD normally waits for punctuation before it begins to

speak a phrase, in order to use the punctuation marks to develop

prosody contouring. This command set Keynote GOLD to commence

speaking each word as soon as the immediately following word is

received.

This may be used as an alternative to the speak now command

where a number of words must be spoken quickly without

intonation.

Mode Resets

The Keynote GOLD speech synthesizers use the T-T-S(tm) text to

speech system from Berkeley Speech Technologies, Inc. T-T-S

converts ASCII text which is sent to it into spoken output. You

can modify the default speech which Keynote GOLD produces using

T-T-S speech resets.

Resets are typed within square brackets[] when incorporated in

ASCII text sent to Keynote GOLD. Reset variables are enclosed

within curly brackets {} in this documentation but the curly

brackets are not included in text sent to Keynote GOLD. The

format of the resets can be changed; see the end of this section.

[t] text reading mode reset

Keynote GOLD initializes to Text Reading Mode. The operation of

Keynote GOLD in Text Reading Mode is controlled by T-T-S Text

Normalizer Resets. Voice characteristic resets may be used to

alter many aspects of the output to create a wide range of

character voices and special effects.

Use the text reading mode reset to return to text reading mode

after invoking other modes.

[p] Phoneme Reading Mode reset

The most exact way of representing the pronunciation of a spoken

word is by using phonemes, the special symbols which have been

devised for that purpose, and which appear in standard

dictionaries of English. The phoneme reading mode is used for

this purpose.

To enter Phoneme Reading mode enter

[p]

anywhere in the text. Text following [p] will be read according

to the phoneme symbols on pages 25 - 27. Each phoneme symbol

must be preceded by a space.

Keynote GOLD speech systems have the standard phonemes of

American English. Phonemes unique to other languages are not

available, consequently words of other languages or dialects

will be pronounced in the same way as a typical American speaker

would, not the way a native speaker would. For this reason

Keynote GOLD may refuse to speak certain non-English, or non

Spanish, combinations of phonemes.

[x] Dictionary entering mode

This mode is used to connect any desired phonemic pronunciation

with a designated English letter sequence (normally a word) in

the RAM resident User Dictionary (UD). After a dictionary entry

has been created, the specified phonemic pronunciation is used

every time the word appears in the input text. Using the UD

entries can be pronounced completely differently from the way

they appear in the text, providing a method for automatic

expansion of special abbreviations.

Entries are added to the User Dictionary by associating an

English spelling with its phonemic transcription. To add an

entry, type both the word and the phonemic transcription of how

you want it pronounced in the following form:

[x] places Keynote GOLD in Dictionary entering mode.

ssss is the English spelling, typed without spaces.

p p p p is its pronunciation in phonemes, with each

symbol preceded by a space.

[t] returns to Text Reading Mode.

Each word that you enter must be preceded by a separate "[x]". A

return to text reading mode is only required after the last

entry has been entered, but "[t]" and a carriage return may be

placed at the end of each entry.

The user dictionary mode allows entry of a special phonemic

pronunciation for a specified ASCII sequence into a RAM based

dictionary, so that the special pronunciation is invoked when

the sequence appears in the Keynote GOLD input text.

As the user dictionary is in RAM memory, all definitions will be

lost when the synthesizer is switched off.

Text Normalizer Resets

The T-T-S text normalizer converts everything in the ASCII text

stream input to Keynote GOLD into letters, to be pronounced as

words if possible. It knows about the format and conventions of

written English, and makes decisions based on an analysis of the

way special sequences most often will be pronounced, taking into

account semantic information available in the text.

However, because different types of text use different formats

and conventions the text normalizer resets are provided to allow

you to control the pronunciation of a text. These resets operate

only in text reading mode.

To return T-T-S to its default condition the reset [n0,0] or

simply [n] is used.

alpha literal

[n1,1] is ON. [n1,0] the default, is OFF.

If the alpha literal reset is on, the letters of the alphabet

will be pronounced as their names. For example, "hello" is

spoken as "h-e-l-l-o".

digit literal

[n2,1] is ON. [n2,0] the default, is OFF.

Groups of numbers are spoken according to certain conventions.

For instance "2409 Telegraph" is pronounced "twenty four oh nine

telegraph". With the digit literal switch on it, would be "two

four zero nine" instead.

prosodic punctuation literal

[n3,1] is ON. [n3,0] the default, is OFF.

T-T-S normally does not read punctuation marks aloud, but uses

them to create the prosody or intonation pattern of the

sentence. The punctuation literal pronounces them, as in

"Charles comma Prince of Wales", but does not pronounce the

non-prosodic marks such as the apostrophe in "don't".

whitespace literal

[n4,1] is ON. Default is [n4,0], OFF.

This reset causes white space characters such as space, carriage

return and tab to be pronounced.

arithmetic pronunciation

[n5,1] is ON. Default is [n5,0] OFF.

This reset is used for reading mathematical texts. It gives a

variant pronunciation to a special group of characters:

\* becomes "times" instead of "asterisk"

/ becomes "over" instead of "slash"

~ becomes "approximately" instead of "tilde"

! is pronounced "factorial"

- becomes "minus"

It also produces a special pronunciation of formulas in cases

such as

1.34 E-6 "one point three four times ten to the minus six"

2^8 "two to the eight"

full number

[n6,1] is ON. Default is [n6,0] OFF.

This reset inhibits the normal grouping of numbers of less than

four digits, pronouncing them as full numbers instead. For

instance, "2409" becomes "two thousand four hundred nine"

instead of "twenty four oh nine".

forced lower case

[n7,1] is ON. Default is [n7,0] OFF.

Sequences of upper case letters are normally pronounced as the

names of the letters: "IBM" is pronounced "I-B-M". This reset is

used to pronounce them as words, for instance "AFTRA" would be

pronounced as the word "aftra". Some common letter groups such

as DEC are in the system dictionary and are normally pronounced

as words anyway.

control character pronunciation

[n8,1] is ON, the default. [n8,0] turns it OFF.

T-T-S normally pronounces control characters except for those

used for asynchronous commands.

time pronunciation

[n9,1] is ON, the default. [n9,0] is OFF.

Normally T-T-S will pronounce times of day, such as "8:00",

"eight o'clock". With this reset OFF "8:00" would be pronounced

literally "eight zero zero"

abbreviation expansion

[n10,1] is on, the default. [n10,0] is off.

T-T-S normally expands common abbreviations. The expansions are

chosen depending on context, and the condition of the field

resets (see below for field resets). This reset prevents

abbreviations being expanded. It does not affect abbreviations

expansions stored in the user dictionary.

Field Resets

Field resets are used by T-T-S and the Text Normalizer to select

the appropriate pronunciation for the context in the case of

ambiguous spellings and abbreviations. For instance "NE" is

pronounced "Northeast" in a street address, but "Nebraska" when

it is the name of a state.

Field resets are coded in the form [z{code},?] where the "?" can

be 1 for ON or 0 for OFF. These are the Field Marker Resets

which are recognized by T-T-S:

[z1,?] personal name

[z3,?] organization name

[z4,?] street address (room number, floor, apartment etc.)

[z5,?] city name

[z6,?] state or province, zip code

[z7,?] nation

[z] restores all fields to default.

Homograph marker

A number of English spellings, "homographs", can be pronounced

in more than one way, for different meanings. For example: wind,

expose, minute, duplicate, buffet. T-T-S assigns ambiguous

spellings their more common pronunciation. However if you place

a tilde (~) in front of the spelling some homographs will be

assigned their alternative pronunciations from the System

Dictionary. For example:

In less than a minute, ~minute quantities began to appear.

They tied a bow on the ~bow of the boat.

Special Punctuation Marks for Changing Prosody

The normal English punctuation marks such as period and comma

create prosodic (intonation) contours in sentences which are

read by T-T-S in Text Reading Mode. However, they do not provide

enough information to T-T-S to permit the kind of special

emphasis a human speaker might use for certain kinds of

sentences.

The following additional marks are used by T-T-S to allow you to

add emphasis, change the pitch contour, and introduce pauses.

Using these marks you can have T-T-S pronounce your text with

the desired prosodic nuances, yet retain the regular English

spelling of the words.

"]" produces a shorter comma-like pause;

"}" produces a shorter period like pause;

" ~+ " stresses the immediately following word, if it is not

already stressed.

" ~- " removes stress from the following word if it is not

already unstressed.

"~!" Emphasizes a particular word in a sentence. This is

done by stressing the immediately following word, and

de-stressing the words after that.

"~?" produces a question with a rising intonation.

The question mark in English speech is produced both with

and without a rising intonation, depending on context. This

prosody mark produces the rising intonation needed by

questions anticipating a yes or no answer.

For precise pronunciation of a phrase it is also possible to

enter the text completely as phonemes, in Phoneme Reading Mode.

In Phoneme Reading Mode, a wide variety of pitch and stress

contours can be described. See pages 25 - 27.

Voice Characteristic Resets

Voice characteristic resets can be used singly, or in

combination, to create a variety of "different people", to add

emphasis, excitement, and other personality characteristics to

the T-T-S speech. Voice resets can be used with either the text

reading mode, or the phoneme reading mode.

Speaking Rate

[r{percentage}]

This Reset controls the rate of speech (i.e. makes the voice

speak faster or slower). {Percentage} is a positive or negative

integer (in the range of -90 to about 200) representing a

percentage change to be applied to the default rate, which is 0.

Positive values increase sound durations by the given ratio,

producing slower speech. The maximum speech rate is about

400-600 words per minute.

Voiced excitation function

[e{num}]

This Reset changes the excitation function for the voice. {Num}

is from 2 to 7. The default is 3, which produces a precise

pronunciation; value 6 sounds "mellower". The value 2 gives an

entirely voiceless output (whispering).

Unvoiced gain

[u{num}]

This Reset increases or decreases the amplitude of voiceless

segments relative to voiced ones. It determines how prominent

the pronunciation of the sound "s" will be. {Num} can be a

positive or negative integer, in units of 0.75 dB. The default

is 0 and the range is about -50 to 20.

Fundamental frequency

[f{freq}]

This Reset determines the overall pitch of the voice. It affects

the inherent pitch characteristics of the speaker, but not the

intonation. {Freq} can be 39-200Hz in integer increments. The

default is 80. A zero value will cause a return to the default

value.

Pitch topline

[h{num}]

This Reset changes the pitch range of the voice by increasing or

decreasing the Hz value of the pitch topline. Raising the

topline makes the speaker's intonation sound more excited or

emphatic. {Num} can range from -100 to about 300. The higher the

number, the higher the topline. The default value is 0. Whenever

the fundamental frequency is changed, {num} is reset to the

default value.

Index Markers

For interactive display systems it is often important to know at

which point in previously input text the synthesizer is

currently speaking. Index markers have been provided for this

purpose. The index marker has the form

[i{num}]

where {num} is an integer between 0 and 127 inclusive. An

alternative code sequence for the Keynote GOLD SA is:

1D,80-E3 (hex, two byte command)

where the second byte represents an index from 0 to 99 inclusive.

After the Keynote GOLD SE or PC has been instructed to return

index markers the code returned from the lower 7 bits of the

synthesizer port will reflect the last index marker passed

during speech.

The Keynote Gold SA will asynchronously send the appropriate

index value to the host computer as an index marker is passed

during speech, and will repeat the transmission each time it is

sent a "return index" command.

For example, while speaking the text "The [i1]quick brown

[i2]fox is [i3] asleep[i0]." the index mark returned from

Keynote GOLD will change to 1 at the start of speaking the word

"quick", will change to 2 after speaking the word "brown" and

will change to 0 after speaking the word "asleep".

Index marks can be inserted in any numeric order. The default

index returned is 0. Index markers may only be inserted between

words.

Changing the form of the Reset Commands

Sometimes an ASCII file may contain sequences which could be

incorrectly interpreted as intentional reset commands. There

are two methods which you can use to avoid having such sequences

interpreted as Resets.

Changing the lead-in character

[c{dec}]

This command changes the lead-in character to the ascii code

{dec}. For example, to change the lead-in character to the

ampersand "&", which has an ASCII value of 38, send Keynote GOLD

"[c38]". All following resets will need to have the form &...].

To change back to "[" (ASCII 91) send Keynote GOLD "&91]".

Double the lead-in character

The Reset lead-in character will be interpreted as a normal

character in a text if it is doubled. For the standard lead-in

character '[' while [x] enters Dictionary Entering mode, the

sequence [[x] remains in Text Reading Mode, and is spoken.

T-T-S (tm) Text Normalizer Performance

Overview

This section provides full documentation for the performance of

the standard (default) Text Normalizer Module supplied with the

T-T-S program in the BeSTspeech TM Packages. Many decisions

made by the default Normalizer can be altered using the Resets

listed on page13 10.

To be read correctly, a text must be interpreted according to

the conventions of written English. This is the work of the

Text Normalizer. One of its primary functions is to assign an

unambiguous meaning to characters and constructions that could

be read in different ways in different circumstances. Here are

some examples:

1. A semicolon usually signals a prosodic pause; it is not

pronounced:

This semicolon is an example; the one in the sentence above

is also.

However, when a character is cited (enclosed in quotation

marks), it does not signal a prosodic pause but should be

pronounced. The sentence

The C language requires a ';' at the end of each statement.

should be pronounced:

"the cee language requires a semicolon at the end of each

statement."

2. The digit "2" contributes to a different pronunciation in

each of the following constructions:

200 "two hundred"

12 "twelve"

2nd "second"

20 "twenty"

3. In addition to ending sentences, periods have a number of

other functions. For example, they can:

Mark an abbreviation: "etc"

Be part of a file name: "command.com"

Mark an ellipsis: "Well..."

Be a silent decimal point: "$45.98"

Be a pronounced decimal point: "3.1416"

4. A string of uppercase letters is often pronounced as the

names of the letters, while the equivalent lowercase letters

would be pronounced as regular words:

POW "prisoner of war"

ID "Identification"

LA "Los Angeles"

RIP "rest in peace"

SAT "scholastic aptitude test"

PA "public address system"

(or if the State Field Reset is on, "Pennsylvania")

The Text Normalizer determines how ambiguous constructions like

those illustrated in (1) through (4) above should be pronounced.

How a particular construction should be pronounced often depends

on the format of the text and the conventions it uses. Because

different types of text use different sorts of conventions, BST

can make application-specific Normalizers on a custom basis.

For some applications, a Normalizer might not be needed at all.

The default BST Normalizer which is included in the T-T-S module

supplied with the BeSTspeech Developer's Package has a number of

features that would be useful for an application that needed to

read generalized sorts of text. Other features could be built

into more applications-specific systems.

BST's Text Normalizer also gives the user the opportunity to

change the way a text is pronounced through the use of various

pronunciation Resets. These Resets are mentioned throughout

this document and are discussed specifically on pages 12 - 18.

Pronouncing Numbers

T-T-S pronounces numbers -- i.e., sequences of digits in three

different ways:

1. Literally, as the names of the digits:

1234 "one two three four"

567 "five six seven"

9001 "nine zero zero one

2. In groups of two:

1234 "twelve thirty-four"

567 "five sixty-seven"

9001 "ninety oh one"

3. As full numbers:

1234 "one thousand two hundred thirty four"

567 "five hundred sixty-seven"

9001 "nine thousand one"

Each of these pronunciations is appropriate in different

circumstances. For example, pronouncing digits in groups of two

is appropriate for dates and addresses:

In 1985 in nineteen eighty-five"

357 Elmwood St. three fifty-seven elmwood street"

Pronunciation as a full number is appropriate for dollar amounts:

$1985 one thousand nine hundred eighty-five dollars"

$357.00 "three hundred fifty-seven dollars and no cents"

A literal pronunciation is appropriate for decimal amounts and

bank account numbers:

2.1985 "two point one nine eight five"

005237-1 "zero zero five two three seven, one"

The Normalizer pronounces numbers correctly in each type of

context. To do so, it uses the following conventions:

1. A string of digits will be pronounced literally if:

a. The string is five or more digits long:

1234567 "one two three four five six seven"

70083 "seven zero zero eight three"

b. The string follows a decimal point:

12.87 "twelve point eight seven"

3.1416 "three point one four one six"

c. The digit-literal pronunciation Reset is on:

123 "one two three"

1006 "one zero zero six"

(See page 14 for more information on the digit-literal Reset.)

2. A string of up to four digits will be pronounced as a full

number if:

a. It ends in "00" or "000":

800 "eight hundred"

1200 "twelve hundred"

3000.5 "three thousand point five"

b. It is a dollar amount:

$ 279 "two hundred seventy-nine dollars"

$1006 "one thousand six dollars"

c. The full number pronunciation Reset is on (while the

digit-literal Reset is off):

279 "two hundred seventy-nine"

1006 "one thousand six"

(See 14 for more information on the full number pronunciation

Reset.)

3. Otherwise, strings of up to four digits are pronounced in

groups of two:

279 "two seventy-nine"

1006 "ten oh six"

1881 "eighteen eighty-one"

990 "nine ninety"

4. If a number includes commas marking off thousands, millions,

billions, etc., it will be pronounced as a full number. T-T-S

can pronounce full numbers up to 9,999,999,999,999,999.

1,006 "one thousand six"

20,000,000 "twenty million"

8,622,401,699,127 "eight billion, six hundred

twenty-two million, four hundred one

thousand, six hundred ninety-nine point one

two seven"

5. Two decimal digits following a dollar amount will be

interpreted as cents if at all possible:

$35.01 "thirty-five dollars and one cent"

$.01 "one cent"

$8.98 "eight dollars and ninety-eight cents"

$8.98 million "eight point nine eight million dollars"

6. A digit followed by the appropriate suffix will be pronounced

as an ordinal:

1st "first"

11th "eleventh"

20th "twentieth"

2,000th "two thousandth"

53rd "fifty-third"

22nds "twenty-seconds"

The Normalizer recognizes two other special uses of numbers -

phone numbers and times of day - and pronounces them

appropriately.

7. Phone numbers

Phone numbers, social security numbers. bank account numbers

and other hyphenated numbers are pronounced literally, with a

prosodic pause at the hyphen.

841-5083 "eight four one, five zero eight three"

6-59802-1 "six, five nine eight zero two, one"

However, if a group of three or four digits ends in a string of

zeros, the zeros will be pronounced as a "hundred" or a

"thousand":

597-8000 "five nine seven, eight thousand"

333-4400 "three three three, forty-four hundred"

These same rules apply to area codes that are enclosed in

parentheses:

(800) 764-9009 "eight hundred, seven six four, nine zero

zero nine

(415) 841-5083 "four one five, eight four one, five zero

eight three"

Some hyphenated numbers are not pronounced literally in this

way. Dates and other short sequences of numbers that are

separated by hyphens are pronounced in groups of 2. For these

numbers, the hyphen is pronounced as "dash":

1985-86 "nineteen eighty-five dash eighty six"

figure 22-3 "figure twenty-two dash three"

8. Times of day

The Normalizer can read times of day of a 12-hour clock. It

will appropriately read hours, minutes, and seconds:

6:00 "six o'clock"

6:03:03 "six oh three and three seconds"

12:59:94.2 "twelve fifty-nine and ninety four point two seconds"

A special pronunciation Reset allows you to turn the

pronunciation of these numbers as times of day off and on. See

page 15.

The conventions used by the Text Normalizer give the user a

great deal of control over how numbers are to be pronounced:

The default pronunciation for long sequences (five or more

digits) is digit literal. To pronounce a long sequence as a

full number, use commas to delimit thousands, millions,

billions, etc.

The default pronunciation for short sequences (up to four

digits) is the one appropriate for dates, addresses, and a

variety of other uses -- in groups of two. To pronounce these

sequences as full numbers, use the full number pronunciation

Reset. To pronounce them literally, use the digit-literal

pronunciation Reset. (See page 14)

Pronouncing Letters and Words

The Text Normalizer decides which sequences of letters are

words, which are abbreviations, and which should be pronounced

literally as the names of the letters.

1. Abbreviations

The Text Normalizer expands abbreviations where it is

appropriate to do so:

Prof. Smith "professor smith"

63 ft. 11in. "sixty-three feet eleven inches"

a, b, c, d, etc. "ey, bee, cee, dee, etcetera"

It can match the same abbreviation spelling to more than one

full word:

Dr. Jones Dr. "doctor jones drive"

Sr. Castro, Sr. "senor castro, senior"

St. Agnes St. "saint agnes street"

Pt. Lookout "point lookout"

5 pt. "five pints"

2. Pronouncing letters as their names

The Text Normalizer recognizes when a group of letters should

be pronounced literally, as the names of the letters. A

sequence of letters will be pronounced literally if:

a. The sequence lacks any of the six vowel letters (a, e, i, o,

u,

lp record "el pee record"

fm radio "ef em radio"

pH "pee aitch"

55 mph "fifty-five em pee aitch"

b. The sequence includes only uppercase letters.

USA "yu ess ey"

OK "oh kay"

IRS "aye ar ess"

KFTU "kay ef tee yu"

There are some exceptions to this rule that the Normalizer

knows about, for example:

NATO "nato"

UNESCO "unesco"

MS-DOS "em ess dos"

The forced lowercase Resets can be used to ensure that

uppercase letters are pronounced as a regular word, rather than

as the names of the letters. See page 15.

c. The sequence consists of a single letter:

o's "ohs"

A) "ey"

y-coordinate "wye coordinate"

program.c "program dot cee"

d. The sequence consists of just two letters that do not stand

alone as an independent word:

76in8 "seventy-six aye en eight"

file.ri "file dot ar aye"

Homographic Spellings

Some words have more than one pronunciation, for example:

read record moderate

entrance close wound

project invalid resume

T-T-S will give these words their more frequent pronunciation.

To give them their other pronunciation, simply precede the

spelling with a tilde (~):

He went in the front entrance.

His paintings ~entrance me.

It opens an old wound.

The clock needs to be ~wound.

Interpreting Punctuation

The Normalizer interprets the significance of various

punctuation marks, pronouncing them only when they are used in

special ways. For example, a period will be pronounced in the

following sorts of constructions, although it is pronounced

differently in each one:

command.com "command dot com"

9.51 "nine point five one"

=%.$ "equals percent period dollar sign"

In none of these cases will the period be taken to mark a

prosodic break (as a sentence-final period does).

Unless they are used in a special way, like the periods

illustrated above, punctuation marks are normally not

pronounced. However, you can have T-T-S pronounce them by

turning on the punctuation-literal pronunciation Reset. See

page 14 for more information.

T-T-S interprets punctuation marks according to the standard and

accepted conventions of written English. For the most part, the

user does not need to be concerned with the decisions the Text

Normalizer is making. However, there are three conventions the

Normalizer uses that must be kept in mind:

1. The 2-space convention

In deciding whether a period signifies the end of a sentence,

the Normalizer may, on occasion, make use of the typing

convention that sentences are separated by at least two spaces.

2. End-of-line hyphens

T-T-S assumes that all end-of-line hyphens mark true word

boundaries. Texts prepared for T-T-S should not divide words at

the end of a line.

3. Periods in abbreviations

Some abbreviations are spelled the same as words that are not

abbreviations. For example:

in ("inches")

fig ("figure")

tab ("table")

apt ("apartment")

no ("number")

Jan ("January")

chap ("chapter")

For these spellings to be considered abbreviations, they must

be followed immediately by a period. The Text Normalizer uses

the period to decide on the correct pronunciation. T-T-S will

pronounce most abbreviations correctly even when the period is

missing, but apho

period is always needed after an abbreviation

that is spelled like a word:

It moved 6 in one day. "it moved six in one day."

It moved 6 in. one day. "it moved six inches one day."

apt 2B "apt two bee"

apt. 2B "apartment two bee"

No Carolina tobacco "no carolina tobacco"

No. Carolina tobacco "north carolina tobacco"

T-T-S (tm) American English Phonemes

Additional Phonemic Symbols for Transcribing Long Passages

Some users might want to transcribe long passages phonemically

for reading in Phoneme Mode. This technique provides very

precise control of pronunciation of words and of intonation

contours. Full transcriptions give excellent results for

messages that are often repeated, and which can use special

prosodic contours to convey additional meaning in particular

applications. For example, warning messages in an alarm system

might use specially transcribed messages to signal extreme

urgency through prosodic emphasis.

Boundaries and Silence

In phoneme-reading mode the boundaries between words and larger

prosodic units must be marked. The following boundary symbols

are used:

$W word boundary

$C a major prosodic boundary

$P a minor prosodic boundary

A prosodic pause can be inserted into the speech stream by using

the symbol for silence:

sl silence

Each "sl" symbol represents about 80 milliseconds of silence.

Longer periods of silence can be obtained by concatenating more

than one "sl". Utterances that are transcribed without "sl"

will be pronounced with the words run together as if in a single

phrase.

Precisely Specifying Stress and Pitch

It is possible to fine-tune the intonation contours of a

phonemically transcribed passage by the use of stress and pitch

markers. The best way to learn to use these is by experimenting

and listening carefully to the result.

Stress is indicated by a dollar sign ($). Pitch is indicated by

a pound sign (œ). These signs are followed by a digit that

indicates the level of stress or pitch. Higher numbers indicate

higher levels.

As noted earlier, primary stress in words can be be marked by

placing the symbol ' after the vowel of the most-stressed

syllable in a word. Using the ' symbol has the same effect as

using the stress level indicator $6.

The stress and pitch levels for secondary word stress " are: $5

œ4. The default stress level is $2 (for unstressed syllables).

When transcribing a full text, stress and pitch markers may be

used to specify utterance-level intonation, not just word-level

stress. A full range of stress markers (from $8 to $1) is

available in phoneme-reading mode, giving you the abilitv to

transcribe a wide variety of stress patterns:

$8 highest stress level

$6 equivalent to primary word stress

$5 equivalent to secondary word stress

$2 default (unstressed) level

$1 lowest stress level

Stress markers mainly affect the duration and amplitude of

syllables. The marker must immediately follow the vowel of the

syllable it marks. Unmarked vowel phonemes are assigned default

stress.

Pitch markers are also used to specify the intonation contour of

an utterance. So that many different kinds of contours can be

specified, a wide range of pitch targets is made available, from

œ10 to œ3:

œ10 highest

œ6 pitch target for primary stressed syllables

œ-3lowes

Pitch targets are associated with:

(1) Syllables, and

(2) "$C" and "$P" prosodic boundary markers.

On syllables, pitch markers immediately follow the stress

marker. If a syllable has no stress marker, the pitch target

immediately follows the vowel phoneme for the syllable. Pitch

targets on boundaries immediately follow the boundary marker.

Unmarked prosodic boundaries receive a default pitch target.

However, unlike stress, there is no default pitch marking for

syllables. The actual pitch levels for unmarked syllables are

interpolated from surrounding pitch targets.

Boundaries can be marked with as many as two pitch targets. Two

pitch targets should appear on boundaries that have words both

to the right and to the left. The first target will be the

final pitch level for the words that precede the boundary and

the second target will be the pitch onset for the words that

follow the boundary. The initial and final "$C" of a text

should each have only one pitch target.

Syllables can also have up to two pitch targets. However,

usually a single target is sufficient. Two pitch targets are

permitted on stressed syllables only to allow for very rapid

rises and falls in pitch.

Transcription Conventions

If you transcribe a text completely into phonemes to be read in

phoneme-reading mode it should start with a [p] reset and end

with a [t]. The first symbol in the transcription should be

"$C" or "$P". The transcription should end with the sequence

"sl $C;" (a silence, a prosodic boundary, and a semi-colon.) The

minor prosodic boundary "$P" can also be used at the end. A

default pitch target will be placed automatically if none is

specified following any prosodic boundary, but a different pitch

target may be specified instead if desired.

Symbols for Phonemic Transcription of Words

One symbol is used for each distinctive sound (phoneme) of

standard American English. In the lists below, each phoneme is

illustrated by a list of example words in which it appears and

arranged to demonstrate the contrasts between similar sounds.

Consonant phonemes:

w --- watt wet woo quit Duane wham

y --- yacht yet you use argue yam

h --- hot heard who hi ahoy ham

m --- sum ramp my limb ample moose

n --- sun rant nigh Lynn handle noose

ng --- sung rank drunk long ankle pinging

l --- lots stole feel sold lily fled

r --- rots store fear soared rare Fred

f --- fat half rough lift phase off

v --- vat have shove lived cover vivid

th --- booth author ether anthem thesis therapy

dh --- smooth other either rhythm these there

s --- sue bus lace recent city oxen

z --- zoo buzz lays resent zitty exact

ch --- batch chin hitch nature virtual church

jh --- badge gin Jeff soldier gradual judge

sh --- bash shin chef nation racial mission

zh --- beige measure vision fusion casual seizure

b --- bats robe baby beak obey amble

p --- pats rope puppy speak opaque ample

d --- door mad dime did buzzed road

t --- tore mat time strut bussed wrote

g --- got rag ogre Greg agog figs

k --- cot rack ocher quake pique fix

Vowel phonemes (as they are pronounced in stressed syllables):

i --- beet leak ease we ski eel

I --- bit lick is spirit hear\* ill

e --- bait lake came way steak ale

E --- bet Lech desk merry head el

ae --- bat lack ask graph had AI

u --- boot Luke dune move stew cooed

U --- put look bush lure tour could

o --- boat choke flow woe oboe code

O --- bought chalk flaw store\* long cawed

a --- pot lock spa mark starry cod

^ --- but luck done just hull cud

R --- Bert lurk earn mirth journey curd

ay --- bite like hire why eyes aisle

Oy --- boy join hoist coy oink oil

aw --- bout pound house cow ouch owl

Any of the vowel phonemes listed above for stressed syllables

can appear in unstressed syllables as well. For example, the

final syllable of "lucky" has the same vowel phoneme ("i") as

"keep". Those forms starred (\*) contain vowels which are

conventionally considered to be lax due to the following "r".

There is an additional vowel phoneme ("=") that only appears in

unstressed syllables.

= --- canal (1st syllable) support (1st syllable)

action (2nd syllable) tickle (2nd syllable)

Stress symbols in words are placed after the vowel of the

stressed syllable. Primary stress is a single quote ';

secondary stress is double quotes ". Boundaries between words

in multiword transcriptions are marked by the symbol $W.

Stress marks and phoneme symbols must always be preceded by a

space. If a phoneme symbol is made up of two characters ("sh")

they must be kept together. For example:

quiche k i ' sh

pizza p i ' t s =

fettuchine f E " t = ch i ' n i

three bedroom [x]3BR th r i " $W b E ' d r u m [t]

(UED abbreviation)

Keynote GOLD Speech Synthesizers