FEXLTTS-SOFFR-01-03 (Ver. 1.3)

OKI middle ware for Speech Control Processor

French Text To Speech Ver 1.0 User's Manual

July 12, 2001

Modification History

1) 30-Mar-2000 : modify of speed rate range

NOTICE

1. The information contained herein can change without notice owing to product and/or technical improvements. Before using the product, please make sure that the information being referred to is up-to-date.

2. The outline of action and examples for application circuits described herein have been chosen as an explanation for the standard action and performance of the product. When planning to use the product, please ensure that the external conditions are reflected in the actual circuit, assembly, and program designs.

3. When designing your product, please use our product below the specified maximum ratings and within the specified operating ranges including, but not limited to, operating voltage, power dissipation, and operating temperature.

4. Oki assumes no responsibility or liability whatsoever for any failure or unusual or unexpected operation resulting from misuse, neglect, improper installation, repair, alteration or accident, improper handling, or unusual physical or electrical stress including, but not limited to, exposure to parameters beyond the specified maximum ratings or operation outside the specified operating range.

5. Neither indemnity against nor license of a third party's industrial and intellectual property right, etc. is granted by us in connection with the use of the product and/or the information and drawings contained herein. No responsibility is assumed by us for any infringement of a third party's right which may result from the use thereof.

6. The products listed in this document are intended for use in general electronics equipment for commercial applications (e.g., office automation, communication equipment, measurement equipment, consumer electronics, etc.). These products are not authorized for use in any system or application that requires special or enhanced quality and reliability characteristics nor in any system or application where the failure of such system or application may result in the loss or damage of property, or death or injury to humans.
   Such applications include, but are not limited to, traffic and automotive equipment, safety devices, aerospace equipment, nuclear power control, medical equipment, and life-support systems.

7. Certain products in this document may need government approval before they can be exported to particular countries. The purchaser assumes the responsibility of determining the legality of export of these products and will take appropriate and necessary steps at their own expense for these.

8. No part of the contents contained herein may be reprinted or reproduced without our prior permission.

9 MS-DOS, Microsoft, Windows, and WindowsNT are registered trademarks of Microsoft Corporation.

Contents

1 Introduction

The French Text To Speech system correctly synthesises the majority of French texts. It is sometimes necessary, however, to modify the text to make it compatible with the constraints given in the following paragraphs before submitting it to the Text To Speech process.

2 User interface description

Data transmission/receipt between MSM7630 and the host processor is called the user interface. Section of interface type is determined by the settings of the configuration register, explained below. Data means text data, dictionary data and control codes.

2.1 Reading the configuration register

When MSM7630 starts up, it reads external configuration register values and makes user interface and other environment settings

The user interface to be used is determined by the configuration register value (see table 2-1). Therefore the serial port and parallel port cannot be used in parallel.

Table 2-1

Register Value	*Interface*
000	2400bps serial port
001	4800bps serial port
010	9600bps serial port
011	19200bps serial port
100	Micro-controller interface

The configuration register is connected to pins D[26:24]. Pull-up 10K register gives register value "1",also Pull-down 10K register gives value "0". (when the bus capacitance is 100pF)

Determine the value of each register so that the bus will stabilize within 18micro second.

2.2 Individual Interface description

2.2.1 Serial port interface

When a serial port interface is selected by the configuration register (when register value is set to 000,001,010 or 011), the data transmit/receive specification is as follows:

Data Format	8bit, no parity, 1stop bit
Transfer Rate	Selectable from 2400, 4800, 9600 or 19200bps
Busy Control	RTS Control

The diagram below shows a serial port interface example.

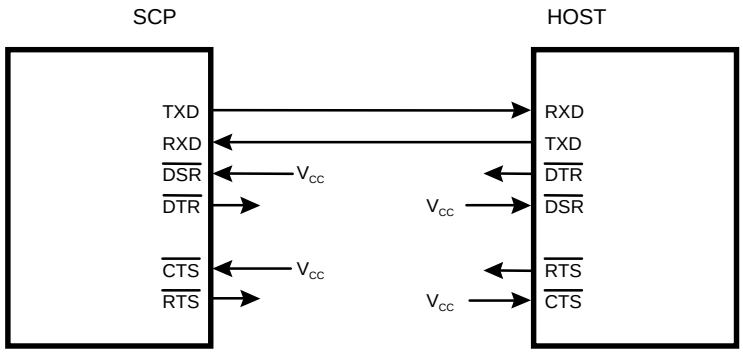


Figure 2-1

Be sure that the ports have sufficient drive capability.

The transmit/receive process from the host is as follows.

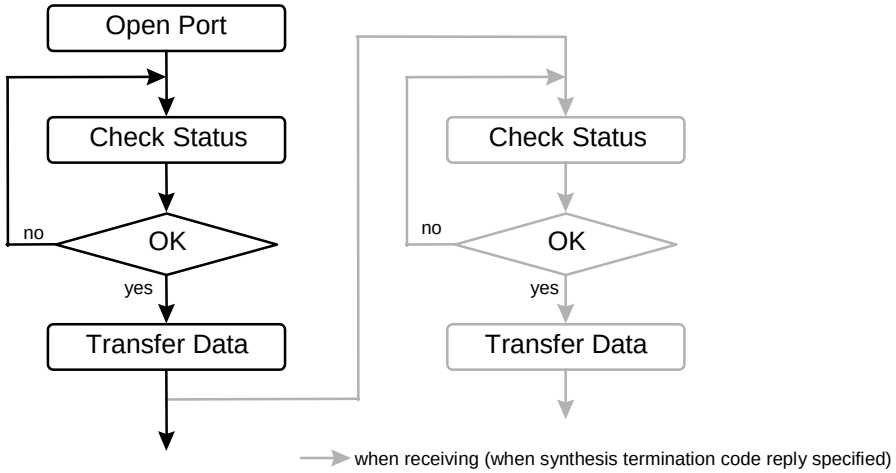


Figure. 2-2

The RTS pin will output "0" during reset and immediately after its release. When the serial port cannot accept data, or in other words when the serial port buffer (1Kbyte) has become full, the RTS pin output will change to "1". When the serial port can accept data, the RTS pin will output "0".

Because RTS is controlled by software, tens of clock may pass from output of the stop bit until RTS rises. However, RTS is set to become invalid when 128 bytes remain in the receive buffer, so there will be no worry about overrun.

There is no standard time interval from the rise of RTS to the fall of the start bit.

2.2.2 Micro-controller Interface

When a micro-controller interface is selected by the configuration register (when register value is set to 100), the data transmit/receive specification is as follows.

8-Bit data port	PD
Status	PIBF, POBF
Control	PCS, PA, PWR, PRD

Table 2-2

PCS,PA,PWR,PRD	Operation
1xxx	Not operating
0x11	Not operating
0010	PIBF,POBF = output, PD = high-impedance
0110	PIBF,POBF = high-impedance, PD = output
0001	Prohibited input
0101	Write to PD
0x00	Prohibited input

x: Don't care

For example, to access from a host CPU, connect as shown in the falling diagram.

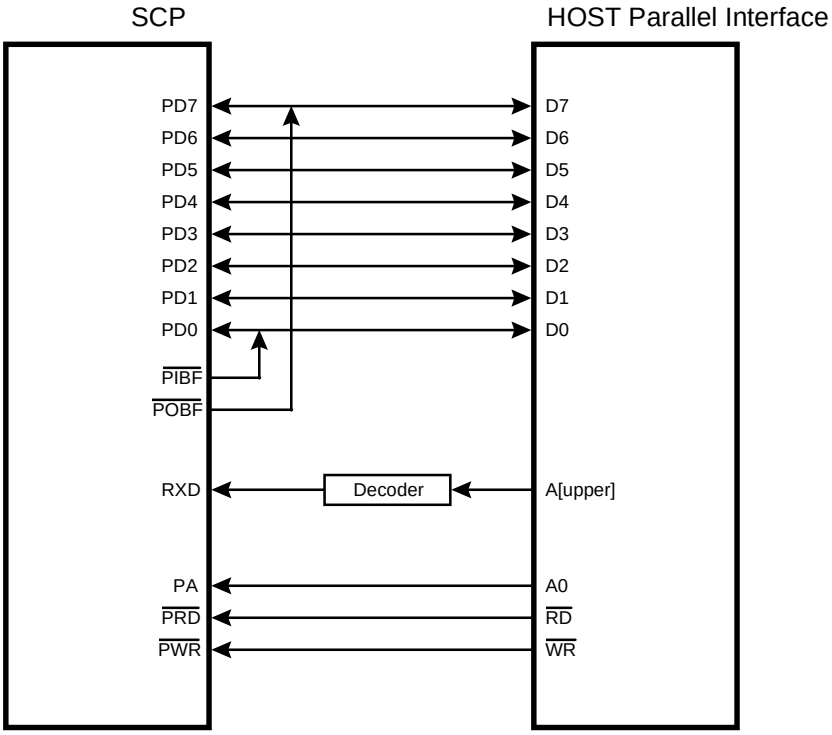


Figure 2-3

In the above case, PIBF (write buffer bit) and POBF (read buffer bit) are connected wire-OR to data port bits 7 and 0 respectively, so the relation between address, status, and data is as follows.

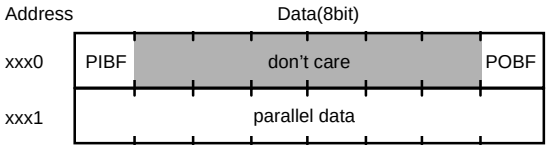


Figure 2-4

The data transfer process is as follows. The "xxx" indicates a MSM7630 parallel port address.

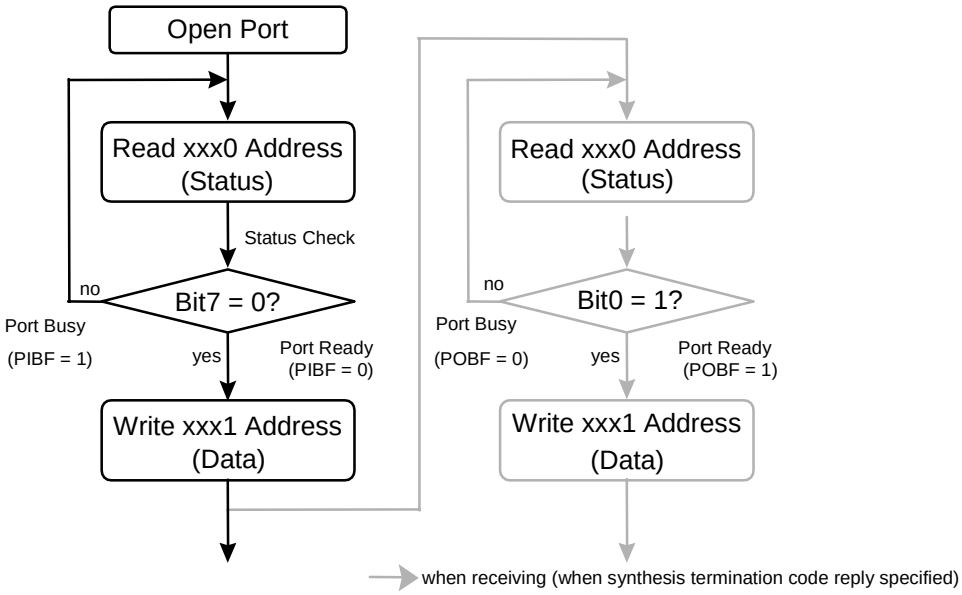


Figure 2-5

For a parallel port when a synthesis termination code replay is specified, the termination code might be missed unless the port is polled until a sentence has been transferred and the termination code accepted.

2.2.3 MSM7630 Start-up Sequence

MSM7630 operates under the following sequence when reset is applied. Make reference to the flow chart, when designing a text to speech synthesiser device that uses MSM7630.

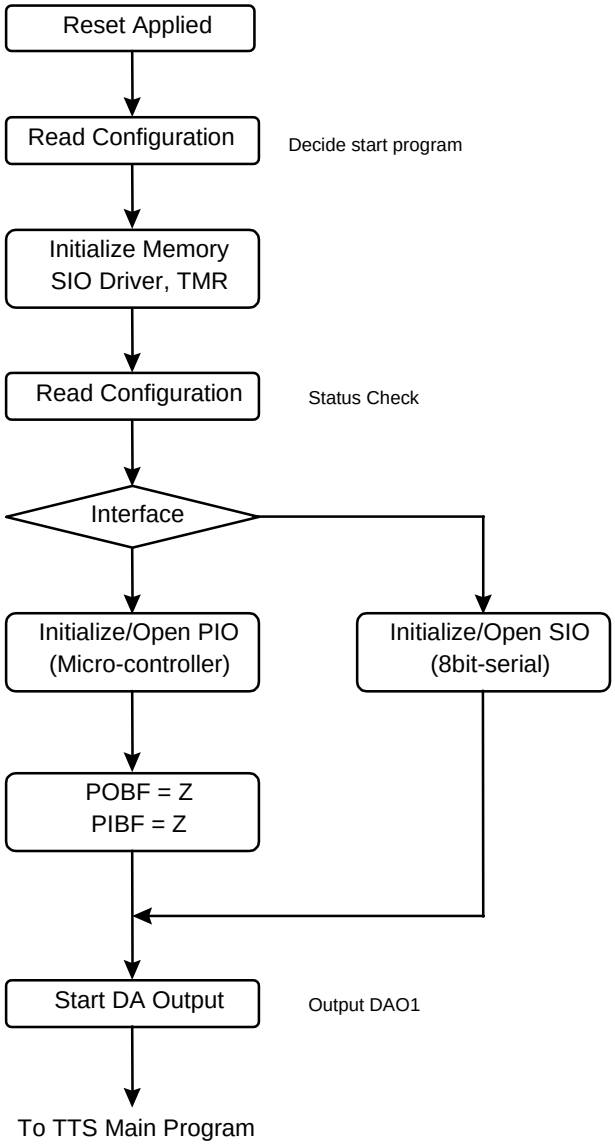


Figure 2-6

ROM accesses are granted immediately after reset. A[23:1] will fluctuate at this time Cache reads are performed, so in particular the three low-order bits will continuously change.

Active signals at this time will be as follows.

A[23:1](especially A[3:1]), ROM, RD

Next the configuration register value will be read, and the DRAM used will be set. This starts DRAM refresh, so the following signals will become active.

RAS, CAS0, CAS1

Next the SIO drive will be initialized. For male phoneme simplex data, the mode will be set, the configuration register value will be read again, and the interface used will be set.

Based on these settings, the following signals will become active.

8-Bit serial interface	RTS,{TXD}
Micro-controller interface	{POBF, PIBF, PD}

However, these signals might not be seen as active for data.

Finally initialization of DA register (internal) values will begin, and DAO1 pin output voltage will become active 1.5Volt. Control will then jump to the main routine. After this the individual interface will wait for input.

The above start-up sequence needs about 700mSec. MSM7630 does not perform self-diagnostic as part of its start-up process.

3 Text To Speech program specification

3.1 Operating Mode

MSM7630 has the operating modes shown in the table below. The operating mode is selected by an operating mode specification (refer to the control code/command listing in Appendix Table). The default mode is text to speech synthesis mode. When in this mode, input sentences can be output as synthesized speech.

Table 3-1

Mode	*Function*
0	Text To Speech synthesis mode
1	Unused
2	Unused
3	Exception dictionary read mode

Control codes and commands are provided to control MSM7630 operation. The validity of control codes and commands differs depending on the operating mode. The table below gives a summary of control codes and commands.

Table 3-2

Category		*Function*
Level1 control codes	Escape codes	Valid except in exception dictionary read mode. These codes primarily set the initial operating state of MSM7630.
Level2 control codes	Text-related	Valid in text-to-speech synthesis mode. These code primarily control how sentences are read.
Level3 control codes	Text-related	Valid except in exception dictionary read mode. These codes primarily control speech quality.
Commands	Control codes	Valid in text-to-speech synthesis mode. Commands control the speech synthesis sequence.

3.1.1 Text To Speech synthesis mode

In this mode, sentences are input and then speech synthesised. MSM7630 detects a termination in the input text (by a termination character) and starts the speech synthesizing operation.

Returning synthesis termination code

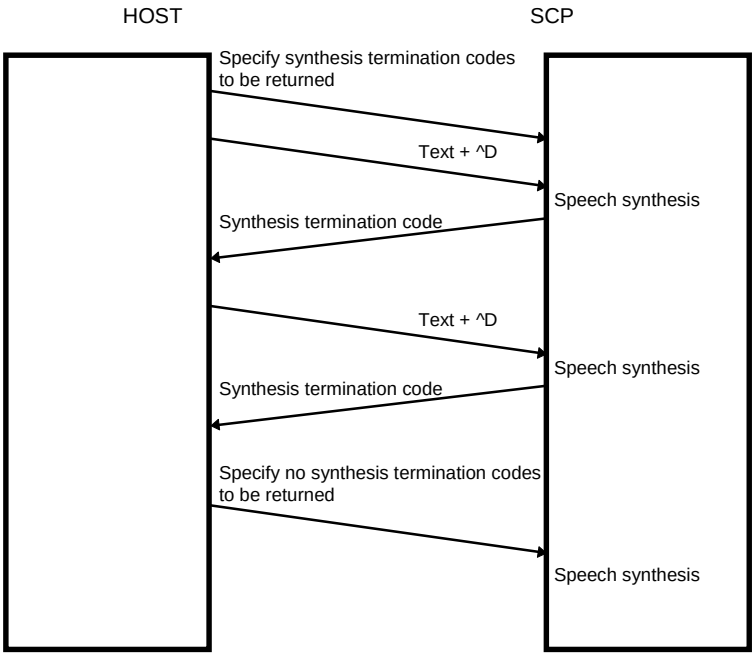
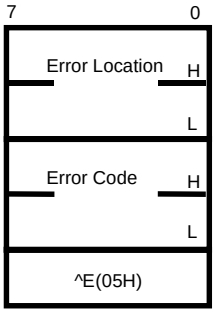


Figure 3-1

In the text to speech synthesis process, MSM7630 normally just synthesizes speech from accepted test, and does not return anything, so a host cannot inspect MSM7630 software status.

For these case MSM7630 can be made to return a synthesis termination code each time synthesis processing of s sentence completes (each time the synthesized sound is output) by specifying that a synthesis termination code is to be returned (refer to "Control Codes/Commands (1) Level 1").

When a synthesis termination code has been specified to be returned, only the response request code ^D (04H), not the termination characters, will be recognized as a terminator. The host appends the response request code ^D (04H) to each sentence of text and sends the sentence to MSM7630. The host then must not send further text or Level 1 control codes until MSM7630 returns the synthesis termination code. MSM7630 will return the synthesis termination code when output of synthesized sound ends. After the synthesis termination code has been returned, the host can immediately send the next text. Fig. 3-1 shows the sequence when return of synthesis termination codes has been specified, and Fig. 3-2 shows the format of the synthesis termination code.



Error location is 2-Byte binary data
=FFFFH:normal termination
not equal FFFFH:
indicates location where text analysis failed as number of bytes from start of text or from previous ^D.

Error code is data that indicates the cause of the error. It will be FFFFH for normal termination.

Figure 3-2

3.1.2 Exception Dictionary Read Mode

In this mode, an exception dictionary created by a utility that runs on the host is downloaded into the devices. An exception dictionary is not appended to the previously sent user dictionary, but entirely overwrites it. An exception dictionary that has been sent cannot then be read.

3.1.2.1 Dictionary transfer procedure for serial and microcontroller interfaces

After the host has specified exception dictionary read mode (refer to "Control Codes/Commands (1) Level 1"), it will receive an ACK (06H) code from MSM7630, and then will send the exception dictionary. After MSM7630 receives the exception dictionary, it performs a BCC check and, based on the result, sends a termination response of ACK (06H) for normal termination or NACK (15H) for abnormal termination. After it sends the termination response, MSM7630 will automatically transfer to its default operating mode ("text-to-speech synthesis mode).

3.1.2.2 Time-out

In exception dictionary read mode, MSM7630 will monitor the time interval between character transmissions. When the interval timer times out (about one second), MSM7630 will transfer to text-to-speech synthesis mode. It will not inform the host.

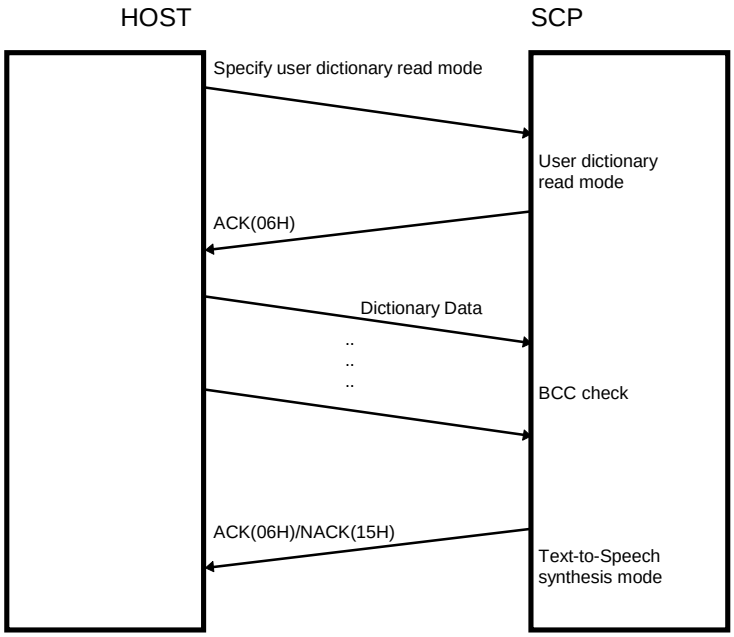
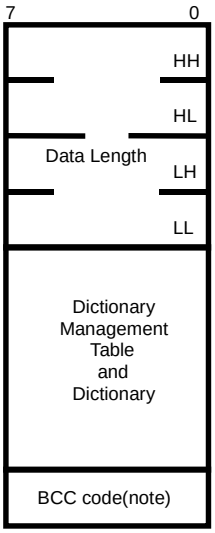


Figure 3-3



Note:The BCC code (1 byte) is for the exclusive OR of all data in the dictionary management table and the dictionary.

Figure 3-4

3.1.3 Hardware sound output (busy signal)

Busy signal should be given while sound output. Busy signal is active low level.

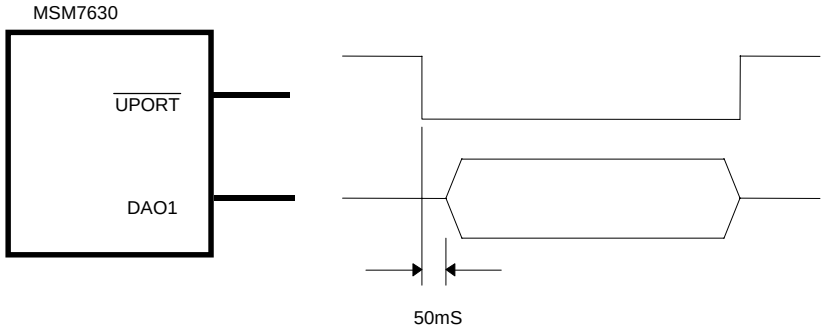


Figure 3-5

3.2 Control Codes Specifications

Control codes are sent by the host to control MSM7630's speech synthesis operations before starting. Some are sent alone, and some are sent inserted anywhere between sentences or words in the text.

3.2.1 Level1 Control Code

Level 1 control codes are output before the text file to set the operating state of MSM7630. Text characters are specified in half size capitals to follow the escape code (1BH). Lists the Level 1 control code

Table 3-3

Level 1 Control Code	*Description*
1	Code format
2	Operating mode
3	Synthesis termination code

3.2.1.1 Code format

Specifies the code format of input text. The word *dos* refers to IBM extended characters.

Table 3-4

	Code format	*Description*
1	[ESC]C0	IBM dos (default)
2	[ESC]C1	ISO 8859-1
3	[ESC]CD	Return to default

3.2.1.2 Operating mode specification

Specifies the MSM7630's operating mode.

Table 3-5

	Code format	*Description*
1	[ESC]M0	Text-to-Speech synthesis mode (default)
2	[ESC]M1	reserved
3	[ESC]M2	reserved
4	[ESC]M3	Exception dictionary read mode

3.2.1.3 Synthesis termination codes returned/not returned

This feature specifies whether or not a synthesis termination code is to be returned after synthesis ends for each sentence. Since MSM7630 normally speech synthesizes the text it receives without returning anything, the host cannot inspect its status. Therefore, while the host shows text one character at a time on its display and sends the text to the MSM7630 for speech synthesis processing, the display and synthesized sounds may not be synchronized (since there is a process delay from text input to synthesis start). Synthesis termination codes are used to synchronize the host and MSM7630 processes.

Table 3-6

	Code format	*Description*
1	[ESC]E0	Do not return synthesis termination codes (default). (note 1) The terminating character will be recognized as the end of text. If text analysis is not possible, then the portion of text that cannot be analyzed will be skipped, but the speech synthesis process will be performed.
2	[ESC]E1	Return synthesis termination code. Instead of a terminating character, only the response request code ^D (04H) will be recognized as the end of text (note 2).
3	[ESC]ED	Return to default setting

Note 1: Fig. 3-2 shows the format of synthesis termination codes.

Note 2: The response request code is appended after the text's terminating character.

3.2.2 Level 2 Control Code

Level 2 control codes do not support this version.

3.2.3 Level 3 Control Code

Level 3 control codes can be inserted anywhere between words in the text, not just between sentence. They primarily voice quality, enabling fine control of voice quality for each word.

Table 3-7

Level 3 Control Code	*Description*
1	Pitch modification
2	Speed rate modification
3	Voice volume
4	Pause control
5	Modulated sound output

3.2.3.1 Pitch modification

This control permits to change the pitch in the text.

Table 3-8

	Code format	*Description*
1	{H+n}	n: From –100 (low) to +100 (high) . The default value is 0.
2	{HD}	Return to default setting

This control permits to modify the pitch of the synthetic voice.

The Control Name value is **pitch**, the Control Information value is surrounded with brackets and ranging from -100 (low) to +100 (high).

Example:

"Bonjour {H+10} monsieur {H-10} Dupont..."

3.2.3.2 Speed rate modification

This control permits to modify the speech rate of the text.

Table 3-9

	Code format	*Description*
1	{T+n}	n: From +100 (slow) to –40 (fast) . The default value is 0.
2	{TD}	Return to default setting

This control permits to modify the speech rate of the text.

The Control Name value is **duree**, the Control Information value is surrounded with brackets and ranging from -40 (fast) to +100 (slow).

Example:

"Bonjour {T+30} monsieur {T-30} Dupont..."

3.2.3.3 Voice volume

Specifies the loudness of voice.

Table 3-10

	Code format	*Description*
1	{P-n}	n: From –100 (min) to 0 (max) . The default value is 0.
2	{PD}	Return to default setting

3.2.3.4 Pause Control

This control allows a pause in the text.

Table 3-11

	Code format	*Description*
1	{p1000ms}	1000 millisecond pause in the text
2	{p1s}	1 second pause in the text
3	{p2mn}	2 minute pause in the text

This control inserts a break in the text.

The Control Name value is **pause**, the Control Information value is a duration surrounded with brackets and expressed either in millisecond or in second or in minute.

Example:

"Bonjour {p1000ms} monsieur..."

"Bonjour {p1s} monsieur..."

"Bonjour {p2mn} monsieur..."

3.2.3.5 Modulated sound output

Output modulated sounds.

Table 3-12

	Code format	*Description*
1	{B0}	500Hz, 160ms sine wave
2	{B1}	1kHz, 160ms sine wave
3	{B2}	2kHz, 160ms sine wave
4	{B3}	Chime 1 (short-long)
5	{B4}	Chime 2 (rising tone: short-short-short-long)
6	{B5}	Chime 3 (falling tone: short-short-short-long)

3.2.4 Command Specification

Commands are interrupting processes that are completely asynchronous with MSM7630's internal processes. Synthesis Stop, pause and restart are provided by commands. Commands are invalid in text-to speech synthesis, used primarily to control the sequence of speech synthesis. Commands are allocated to control codes below 0x20.

3.2.4.1 Stop

Stops the current text-to-speech synthesis process.

Table 3-13

	Code format	*Description*
1	^C(03H)	Stop the current Text-to-Speech synthesis process

The stop command causes MSM7630 to discard all text captured so far during synthesis, including speech synthesis parameters. MSM7630 will then return to an input wait state.

3.2.4.2 Initialize

Stops processing of the current operating mode. Returns all Level 1 to 3 Code settings (including mode specification) to their defaults.

Table 3-14

	Code format	*Description*
1	^R(12H)	Stop processing of the current operating mode

4 Rules to be applied

4.1 Sentence

4.1.1 Number of characters

A sentence contains less than 255 characters (including markers[1] and control codes[2]). If a sentence contains more than 255 characters without punctuation mark, it is truncated between two words to produce two or several sentences which will be less than 255 characters long[3].

4.1.2 Number of words

A sentence contains less than 25 words (excluding markers[4] and control codes[5]). If a sentence contains more than 25 words without punctuation mark, it is truncated between two words to produce two or several sentences which will be less than 25 words long.

4.1.3 Syntax

A sentence should always start with a capital letter and finish with a full stop. Note that a word at the beginning of a sentence will not have the same prosody that a word within the sentence.

---

[1] See chapter Markers

[2] See chapter Switches

[3] Overflow may be caused by the translation of numbers and acronyms. For example the number 033544628, which has 9 characters, will have 46 characters after translation.

[4] See chapter Markers

[5] See chapter Switches

If a sentence does not start with a capital letter, the system could :

- apply a prosody which does not correspond to the prosody of an initial word,

- consider that the previous word and the "." constitute an abbreviation.

Example:

Il avait doublé ce cap. la chaleur était étouffante.

will give

Il avait doublé ce capital la chaleur était étouffante.

instead of

Il avait doublé ce cap. La chaleur était étouffante.

4.2 Word

A word contains less than 26 characters. Longer words are truncated after the 26th character.

4.3 Character

A character must be coded IBM extended ASCII or ISO 8859-1[6].

Refer to appendix B for the translation of ASCII codes recognized by the system.

[6] *Depending on the platform*

4.4 Proper Names

The system uses the initial capital letter to distinguish between a proper and a common name.

Example:

le lot

and

4.5 Dash

le Lot

The presence of a dash has an influence on the selection of the grammatical class and may enable liaison between two words. The correct use of the dash is important.

4.5.1 Between words

In French, the main function of the dash is to constitute a unit. The units are either lexical or grammatical.

Lexical units:

- The presence of the dash enables the selection of the grammatical class of hyphenated words and may modify their pronunciation.

  Examples:

  après-midi, sans-gêne, sous-main, pousse-café, tire-botte, laissez-passer, rouge-gorge...

- It enables to identify a sequence of words which form a semantic unit.

  Examples: distinction between

  après-demain and *après demain*

  un pot-au-feu and *un pot au feu*

Paris-est and *Paris est...*

- It avoids difficult or impossible analysis.

  Examples:

  pêle-mêle, porc-épic, nu-tête...

- It is always present in proper names beginning with the word *saint*.

  Examples:

  Saint-Honoré, saint-Honoré, ST-Honoré, St-Honoré, st-Honoré

  will give

  Sain Tonoré

  but

  Saint Honoré

  will give

  sain honoré

Grammatical units:

- It is required to insure the liaison between the verb and the following pronoun.

  Examples:

  Est-ce vrai...

- It is also required between the pronoun and the adjective *même*.

  Examples:

  lui-même, moi-même mais ceux mêmes qui...

- It indicates the coordination without break.

  Examples:

  Une sourd-muet, Les années 1941-1942, vingt-deux, soixante-dix-huit...

4.5.2 Between digits

Without specific information about application, the dash between two digits is always pronounced *trait d'union.* According to the application, it is necessary to pre-process the text. The dash may denote either the mathematical minus sign, the classical dash or the conjunction *à* (example: *34-35 rue de Sèvres*).

4.6 Punctuation

Punctuation plays an crucial rôle in the texts analysed by the system. It is necessary to put a space character just after the punctuation marks.

4.6.1 List of pronunciations recognised by the system and their effects

Table 4-1

PUNCTUATION	*INTONATION*	*PAUSE*
.	falling	long
;	falling	long
:	rising	short
,	rising	short to medium (depending on the number of words)
...	very rising	long
!	falling	medium if one word, long otherwise
!!!!	rising	no pause
"([{)]}	rising	medium
?	falling (if the sentence is short), rising otherwise	medium

4.6.2 Automatic breaks

If a sentence contains too many words or two many characters without punctuation mark, the system automatically inserts a break in front of some words called break words. For example, a break in front of the phrase *après que* will take priority to a break between the word *après* and the word *que*.

4.6.3 The comma

• The comma is required to rise ambiguities in some sentences.

Example:

does

Il chante juste quand il veut.

mean

Il chante, juste quand il veut.

or

Il chante juste, quand il veut.

The system will choose the second option because the word *quand* is classified as a break word.

• The comma may be used to correct breaks automatically inserted.

Breaks can be automatically inserted according to the syntax analysis of the sentence. If a break is not correct, it can be avoided by the insertion of a comma in the sentence.

Example:

In the following sentence, the system automatically inserts a break :

La contribution sociale constitue donc un nouvel impôt (break) *sur les revenus qui sera payé par tout le monde.*

The break is not correct in this sentence but will be correct in the following sentence :

Il voulait constituer un nouvel impôt (break) *sur les revenus qui lui avaient été attribués.*

In the first sentence, a comma inserted between the word *revenus* and the conjunction *qui* avoids the insertion of a break.

4.6.4 Full stop

• A full stop is always considered as a punctuation mark if the following letter is capital and if the previous letter is not the last letter of an acronym or an abbreviation.

Examples:

S.A. Super.

will give the sentence

S A.

followed by the sentence

Super.

Ecrivez un N. C'est bien.

will give the sentence

Ecrivez un N.

followed by the sentence

C'est bien.

Ecrivez un M. C'est bien.

will give the sentence

Ecrivez un monsieur[7] *c'est bien.*

• A full stop followed by a small letter is not considered by the system as the end of a sentence. It is assumed to be part of the previous word.

Example:

édit. du livre.

will give

édition du livre.

and

C.I. directeur.

will give

C I directeur.

4.6.5 The succession of full stops

The succession of full stops is often used to denote an enumeration. This is not the case for the system which translates it by a very rising prosody followed by a pause.

Example:

Mais..... que voulez vous ?

To convey an enumeration it is better to use the abbreviation "*etc.*".

[7] *The capital letter after the mark has no effect because the abbreviation "*M.*" is quite always followed by a proper name.*

4.7 Acronyms and abbreviations

4.7.1 The acronyms

• Written in the A.B.C. form:

If the acronym is present in this form in the "Acronym" table, it is pronounced as indicated in the "Translation" column. It is spelled out otherwise[8].

Examples:

C.A.P.E.S.

will be pronounced

capès

because it is listed as an acronym processed by the system.

B.B.C.

will be pronounced

bébécé

because it is not listed.

• Written in the A B C form:

In this form, the acronym will be always spelled out.

• Written in the ABC form:

If the acronym is present in this form in the "Acronym" table, it will be pronounced as indicated in the "Translation" column. If it is not listed, it will be pronounced if it is pronounceable and it will be spelled out otherwise.

Examples:

FT

will be pronounced

[8] *To avoid ending a sentence prematurely, it is recommended not to put a full stop after the last letter of the acronym.*

France Telecom

because it is listed.

SPO

will be pronounced

spo

IFTGH

will be pronounced

iftg

WTGHJ

will be pronounced

doublevétégéachji

- **Prosody according to the acronym writing:**

FTGH or F T G H

will be pronounced

éf(lengthening) té(lengthening) gé(lengthening) ach(lengthening)

but

F.T.G.H.

will be pronounced

éftégéach

- **Conclusion:**

In order to make an acronym spelled out, it is always better to write it in the A.B.C or ABC form.

The system does not deal with acronyms and abbreviations. It will try to pronounce the acronym or the abbreviation as a normal word.

4.7.2 The abbreviations

- **Written in the abc form:**

If the abbreviation is present in this form in the "Abbreviation" table, it will be pronounced as indicated in the "Translation" column. If it is not listed, it will be pronounced if it is pronounceable and it will be spelled out otherwise.

Example:

3 av Général Leclerc

will be pronounced

3 avé Général Leclerc

because *av* is not an abbreviation of the system.

- **Written in the abc. form:**

If the abbreviation is present in this form in the "Abbreviation" table, it will be pronounced as indicated in the "Translation" column. In this case, some abbreviations accept that the following word begins with a capital letter without any beginning of sentence generated by the system. If it is not listed, it will be pronounced if it is pronounceable and it will be spelled out otherwise.

Examples:

pour cet art. du catalogue

will be pronounced

pour cet article du catalogue

because the abbreviation *art.* is an abbreviation of the system and it is followed by a small letter.

pour cet art. Les répétitions..

will be translated by the sentence

pour cet art

followed by the sentence

Les répétitions

because the abbreviation *art.* is followed by a capital letter.

- **Abbreviation of a measurement unit:**

If the measurement unit is present in this form in the "Abbreviation" table, it will be pronounced as indicated in the "Translation" column. However, contrary to other abbreviations, it is translated only if it is preceded by a number and a space character. It is spelled out otherwise. This processing is done to avoid conflicts between measurement units and code numbers.

Example:

5 km2

will be pronounced

cinq kilomètres carrés

- **Special case of the time:**

Examples :

8h04

will be pronounced

huit heures zéro quatre

8h04 mn

will be pronounced

huit heures zéro quatre minutes

8h04mn 15 sec

will be pronounced

huit heures zéro quatre minutes quinze secondes

8h04 mn 15 s

will be pronounced

huit heures zéro quatre èm èn quinze ès

- **Conflicts:**

An abbreviation of the system can be used to translate a word which has the same spelling but not the same pronunciation because it has a different meaning.

Example:

Melle

will be pronounced

Mademoiselle

even if Melle *refers to the city.*

4.7.3 List of acronyms and abbreviations of the system

See Appendix A.

4.7.4 List of acronyms and abbreviations of the user

List of abbreviations

See the *ACRONYME.RGS* ASCII file.

Adding or modifying an abbreviation

See the chapter "Abbreviation lexicon".

Note:

At the end of the sentence, the last point must be separated from the abbreviation by a space character.

4.8 Numeration

4.8.1 Numbers

A number is composed with less than 15 digits.

• Special case of the pronunciation of numbers in Belgian and in Switzerland

The numbers 70, 80 and 90 (pronounced respectively *soixante dix, quatre vingt* and *quatre vingt dix* in the usual case) can be respectively pronounced *septante*, *huitante* and *nonante* or *septante, quatrevingt* and *nonante*[9].

• Numbers from 1 to 15 digits

Examples:

2345678987

will be pronounced

deux milliards trois cent quarante cinq millions six cent soixante dix huit mille neuf cent

quatre vingt sept

311.994 or *311.994*

will be pronounced

trois cent onze mille neuf cent quatre vingt quatorze

[9] *Depending on the platform*

- **Decimal Numbers**

Example:

311,994

will be pronounced

trois cent onze virgule neuf cent quatre vingt quatorze

- **Numbers with leading zeros**

Examples:

0000006

will be pronounced

zéro zéro (break) zéro zéro (break) zéro zéro six

00123455

will be pronounced

zéro zéro (break) cent vingt trois mille quatre cent cinquante cinq

A break is automatically inserted after each pair of zeros.

4.8.2 Time

Examples:

8:04

will be pronounced

huit heures zéro quatre

8:04'15"

will be pronounced

huit deux points zéro quatre cent quinze

4.8.3 Date

Examples:

31/05/94

will be pronounced

trente et un (break) zéro cinq (break) quatre vingt quatorze

31 05 94

will be pronounced

trente et un (break) zéro cinq (break) quatre vingt quatorze

31.05.94

will be pronounced

trente et un point zéro cinq point quatre vingt quatorze

4.8.4 Telephone number

Telephone numbers are pronounced in the same way as the dates.

4.8.5 Roman number

Roman numbers are translated by the system. They are composed with the capital

4.9 Using the "/"

letters "I, V, X, L, C, D, M".

- **As a number separator**

It inserts a break.

Example:

31/05/94

will be pronounced

trente et un (break) zéro cinq (break) quatre vingt quatorze

- **As a separator of acronyms in capital letters**

It inserts a break and forces the acronym to be spelled out.

Example:

PML/MVT

will be pronounced

péèmél (break) èmvété

- **In front of an acronym in capital letters**

It forces the acronym to be spelled out.

Example:

/LEV

will be pronounced

èleuvé

- **As a separator of words in small letters**

It is replaced by a space character.

- **As a separator between a number and a word**

It avoids the grammatical liaison.

Examples:

2 allée de la Rochelle

will be pronounced

deux zallée de la Rochelle

2/ allée de la Rochelle

will be pronounced

deux (break) allée de la Rochelle

4.10 List of ASCII Codes recognized by the system

See Appendix B for the list of ASCII codes recognized by the system.

5 User lexicons

5.1 Exceptions lexicon

The characters in the user lexicon files must be coded in IBM extended ASCII.

5.1.1 Using the lexicon

The exception lexicon permits the user to modify the pronunciation of a word or a group of consecutive words. Some French or foreign words, which are badly pronounced in accordance with the basic rules for French pronunciation, can be added in the exception lexicon.

The pronunciation writing uses a pseudo-orthographic method : it consists of writing the pronunciation with French alphabetical codes. For example, the pronunciation of the English word "Windows" can be written "ouinedauze".

5.1.2 Adding an entry to the lexicon file

With a text editor the user can add an entry to the exception lexicon. The exception lexicon is a file called USERCNVX.EXC in the installation directory. The maximum length of the lexicon depends on the available RAM resources[10].

An exception and its pronunciation must be written on one line (less than 256 characters long). An exception can contain five words and it is necessary to write the same number of words in the exception field than in the pronunciation field.

Writing punctuation marks in the exception field is forbidden. It is therefore impossible to add an abbreviation or an acronym in the exception lexicon. It is not necessary to respect the alphabetic order. Finally the look-up words are case-sensitive, unless the option "/i" is specified.

[10] *Depending on the platform*

Key characters list:

The characters "//" indicate a comment which stop at the end of the line.

The character ":" separates the exception field from the field of its pronunciation.

The field between the characters "<" and ">" is the pronunciation field.

The characters "##" indicate a word boundary in the exception pronunciation field.

The characters "/i" are optional and permit to ignore case.

Example:

```
// Beginning of the file
Windows : <ouinedauze> /i // foreign proper name
Bush : <bouche> /i // foreign proper name
John Kenedy : <djone##kénédi> /i // foreign proper name
// End of the file
```

Note:

5.2 Abbreviations lexicon

After modifications the exception lexicon must be reloaded in memory.

5.2.1 Using the lexicon

If the abbreviation is written in the left column of the file, it will be translated as indicated in the right column. The translation writing of abbreviations uses a pseudo-orthographic method. For example, the translation of the abbreviation "I.H.M." can be written "interface-homme-machine".

5.2.2 Adding an entry to the lexicon file

With a text editor the user can add a entry to the abbreviation lexicon. The abbreviation lexicon is a file called ACRONYME.RGS in the installation directory. The maximum length of the lexicon depends on the available RAM resources[11].

An abbreviation and its translation must be written on one line (less than 256 characters long).

It is not necessary to respect the alphabetic order. Finally the look-up words are case-sensitive.

Key characters list:

The characters "//" indicate a comment which stop at the end of the line.

The space character or the tabulation separates the abbreviation field from the field of its translation.

The character "-" indicates a word boundary in the abbreviation translation field.

Example:

```
// Beginning of the file
I.H.M.      interface-homme-machine
Dr.         doctor
// End of the file
```

Note:

After modifications the abbreviation lexicon must be reloaded in memory.

[11] *Depending on the platform*

6 Appendix A: List of acronyms and abbreviations of the system

6.1 List of current abbreviations

Conventions:

- The italic bold characters indicate no processing difference between lower and upper case.

- The characters between square brackets are optional.

Abbreviation	*Translation*	*Variant/Comment*
$ ***kg***[.]	dollar le kilogramme	
arr., Arr.	arrondissement	*a.r.r*
art., Art.	article	*a.r.t*
av., Av.	avenue	*a.v*
bat., Bat.	batiment	*b.a.t*
bd[.], ***bld***[.], boul., Boul.	boulevard	*b.d, b.l.d, b.o.u.l*
c.-***à***-***d***[.]	c'est à dire	
cal.	calorie	*c.a.l*
cap.	capital	*c.a.p*
cdt[.]	commandant	*c.d.t*
cf[.]	voir	*c.f*
ch[.]-***l***[.]	chef-lieu	
chap., Chap.	chapitre	*c.h.a.p*
CIE/	compagnie	*c.i.e*

cont.	contenu	*c.o.n.t*
cos.	cosinus	*c.o.s*
dép., Dép., ***dpt***[.]	département	*d.p.t*
diam.	diamètre	*d.i.a.m*
Dr[.]	docteur	
E.***U***.***A***[.]	Etats Unis d'Amérique	
e.***v***[.]	en ville	
E.-et-L[.]	Eure et Loire	
ed[.], ***éd***[.], édit.	édition	
env.	environ	*e.n.v*
ep.	épaisseur	*e.p*
etc[.]	etcétéra	*e.t.c*
ex., Ex.	exemple	*e.x*
FKg[.]	francs kilogramme	
F[.]***T***[.]***T***[.]***C***[.]	francs toutes taxes comprises	
faub[.], ***fbg***[.]	faubourg	*f.a.u.b*
fig.	figure	*f.i.g*
fr[.], ***frs***[.]	francs	*f.r, f.r.s*
francs kg[.]	francs le kilogramme	
G.-***B***[.]	Grande-Bretagne	
H.-***Rh***[.]	Haut-Rhin	
H[.]***T***[.]	hors taxes	
ibid.	ibidem	*i.b.i.d*
id.	idem	*i.d*
jr[.], ***jrs***[.]	jours	*j.r, j.r.s*
lat.	latitude	*l.a.t*
lg[.]	longueur	*l.g*
lieut[.]	lieutenant	
log.	logarithme	*l.o.g*
Melle	mademoiselle	
Mgr	monseigneur	*m.g.r*
min.	minimum	*m.i.n*

mn	minute	following context
Mr.	monsieur	
Mrs.	messieurs	
n/	Notre	if followed by space
N.	Nord or n	following context
n[.]***b***[.]	nota bene	
N.-***E***[.], ***N***.***E***[.]	nord-est	
N.-***O***[.]	nord-ouest	
N[.]***S***[.]***J***[.]***C***[.]	notre seigneur Jesus Christ	
No[.], ***Nr***[.], n°	numéro	
ns	nous or nanosecondes	following context
O. ***K***[.], ***O***.***K***[.]	oké	
p.***a***[.]	parents	
P.***U***[.]	prix unitaire	
P.***U***[.]***H***[.]***T***[.]	prix unitaire hors taxes	
prof.	professeur	
R.-***et***-***D***[.]	recherche et développement	
ref., REF/	référence	
S.-E[.]	sud-est	
S.-et-M.	Seine et Marne	
S.-et-O[.]	Seine et Oise	
S.-O[.]	sud-ouest	
S/CHEF	sous chef	
S/FIL	sans fil	
S/LIEUTENANT	sous lieutenant	
S/SOL	sous sol	
sec.	seconde	*s.e.c*
sem.	semaine	*s.e.m*
sept.[.]	septembre	
slts	salutations	*s.l.t.s*
St[.]	saint	*s.t*
T[.]***T***[.]***C***[.]	toutes taxes comprises	
tang.	tangente	*t.a.n.g*

TEL/, ***tel***, ***tél***	téléphone	*t.e.l*
th de la ville	théatre de la ville	
th de poche	théatre de poche	
th français	théatre français	
TLX/, ***tlx***[.], TWX/, ***twx***[.]	télex	*t.l.x, t.w.x*
tt cft[.]	tout confort	
U.***S***[.]	américain	
v/	votre	if followed by space
V/DEMANDE	votre demande	
var.	variante	*v.a.r*
viz.	à savoir	*v.i.z*
vol.	volume	*v.o.l*
volts h[.]***t***[.]	volts haute tension	

6.2 List of acronyms

Two written forms of acronyms are processed by the system[12].

- Form (1) : C.N.E.T[.]
- Form (2) : CNET

There is no processing difference between lower and upper case.

The form (1) will be spelled in the usual case. The form (2) will be pronounced in accordance with the basic rules for French pronunciation.

If an acronym belongs to the following list, it will be pronounced as indicated in the second column.

Conventions:

/ãl → [en]suite

/y/ → [u]ne

/g/ → [g]are

/j/ → [j]e

Acronym	*Translation*
A.E.F.	*aeèf*
A.S.A.	*aza*
A.S.S.E.D.I.C.	*asédik*
C.A.P.E.S.	*kapès*
C.A.P.E.T.	*kapèt*
C.E.D.E.X.	*sedèks*
C.I.D.	*sid*
C.I.I.	*sédezi*
C.N.E.S.	*knès*
C.N.E.T.	*knèt*
C.N.E.T.L.A.N.N.	*knètlanion*

[12] *See chapter Using the "/"*

Acronym	*Translation*
C.N.I.T.	*knit*
C.O.D.E.R.	*kodèr*
D.E.U.G.	*dœg*
D.E.U.S.T.	*dœst*
D.I.N.	*din*
D.I.P.A.S.	*dipas*
D.O.M.	*dom*
D.V.O.R.A.K.	*dvorak*
E.A.O.	*eao*
E.C.U.	*éky*
E.D.F.	*edéèf*
E.G.F.	*ejéèf*
E.N.A.	*éna*
E.N.E.	*eène*
E.N.S.	*eènès*
E.O.R.	*éoèr*
E.T.A.	*etéa*
E.T.D.A.M.	*ètdam*
F.E.N.	*fèn*
F.I.D.E.S.	*fidès*
F.N.A.C.	*fnak*
F.T.	*frãstélékom*
G.I.F.A.S.	*jifas*
G.L.A.M.	*glam*
I.D.I.	*idi*
I.F.O.P.	*ifop*
I.F.R.E.S.	*ifrès*
I.G.A.M.E.	*igam*
I.G.A.S.	*igas*
I.G.E.N.	*ijèn*
I.N.A.	*ina*
I.N.E.D.	*inèd*

Acronym	*Translation*
I.N.F.R.E.P.	*infrèp*
I.N.R.A.	*inra*
I.N.S.E.E.	*insé*
I.N.S.E.R.M.	*insèrm*
I.P.E.S.	*ipès*
I.P.S.O.S.	*ipsos*
I.R.A.	*ira*
I.R.E.T.	*irèt*
J.O.C.	*jok*
M.R.A.P.	*mrap*
O.N.U.	*ony*
O.P.E.P.	*opèp*
O.R.S.E.C.	*orsèk*
O.T.A.N.	*otã*
O.V.N.I.	*ovni*
P.A.L.	*pal*
P.F.A.T.	*péfat*
S.A.M.U.	*samy*
S.E.C.A.M.	*sékam*
S.I.D.	*sid*
S.I.D.A.	*sida*
S.M.A.G.	*smag*
S.M.I.C.	*smik*
S.M.I.G.	*smig*
S.M.I.P.	*smip*
S.N.E.S.	*snès*
S.N.E.S.U.P.	*snèsyp*
S.O.F.R.E.S.	*sofrès*
T.G.V.	*téjévé*
T.O.M.	*tom*
Z.A.D.	*zad*

6.3 List of measurement units

Conventions:

- α means a number

- The italic bold characters indicate no processing difference between lower and upper case.

Abbreviation	*Translation*
α cm	α centimètres
α cm2, α cm²	α centimètres carrés
α cm3	α centimètres cubes
α dag	α décagrammes
α dal	α décalitre
α dam	α décamètres
α DM	α Deutsh Mark
α dm	α décimètres
α dm2, α dm²	α décimètres carrés
α dm3	α décimètres cubes
α F	α francs
α FF	α Francs Français
α g	α grammes
α gr	α grammes
α h, αh	α heures
α ha	α hectares
α hm	α hectomètres
α hm2, α hm²	α hectomètres carrés
α j	α jours
α jan	α janvier
α ***k***m	α kilomètres

Abbreviation	*Translation*
α ***k***m2, α ***k***m²	α kilomètres carrés
α m	α mètres
α m2, α m²	α mètres carrés
α mm	α millimètres
α mm2, α mm²	α millimètres carrés
α mm3	α millimètres cubes
α s	α secondes
α sec	α secondes
α sept	α septembre
dag., ***d.a.g***	décagramme
dal., ***d.a.l***	décalitre
dam., ***d.a.m***	décamètre
dB, d.b	décibel
degré[s] C., ° C.	degrés celcius
degré[s] F., ° F.	degrés Farenheit
degré[s] K., ° K.	degrés Kelvin
GHz, g.h.z	giga Hertz
grs, g.r.s	grammes
ha., ***h.a***	hectare
hW, h.w	hecto Watts
Hz, h.z	hertz
kbit, k.b.i.t	kilobits
kg, k.g	kilogramme
khz, k.h.z	kilo Hertz
kJ, k.j	kilo Joules
Km/h, Kmh, k.m.h	kilomètre heure
kms, k.m.s	kilomètre seconde
kV, k.v	kilo Volts
kW, k.w	kilo Watts
kWh, k.w.h	kilo Watts heure
mA., m.a	milliampère
mhz, m.h.z	méga Hertz

Abbreviation	*Translation*
mo, m.o	mégaoctet
mV, m.v	millivolts
ns, n.s	nous ou nanoseconde

7 APPENDIX B: List of ASCII codes translated

7.1 7 bits ASCII characters

Decimal ASCII code	*Character*	*Recognised as / translated by*
0	^@ (NUL)	ignored
1	^A (SOH)	marker
...	...	ignored
9	^I (HT)	separator of word
10	^J (LF)	separator of word
...	...	ignored
13	^M (CR)	separator of word
...	...	ignored
32		separator of word
33	!	exclamation point (mark) / pause
34	"	ignored
35	#	ignored
36	$	sign / *dólar*
37	%	sign / *por ciento*
38	&	sign / *y*
39	'	ignored
40	(	punctuation / pause or separator of phone number
41	)	punctuation / pause or separator of phone number

Decimal ASCII code	*Character*	*Recognised as / translated by*
42	*	sign / *estrella*
43	+	sign / *más*
44	,	punctuation / pause or decimal comma / *coma*
45	-	punctuation / pause or hyphen or sign / *menos*
46	.	punctuation / pause or date separation
47	/	date or phone number separation
48	0	digit zero
49	1	digit one
50	2	digit two
51	3	digit three
52	4	digit four
53	5	digit five
54	6	digit six
55	7	digit seven
56	8	digit eight
57	9	digit nine
58	:	colon / pause or time separation
59	;	semicolon / pause
60	<	ignored
61	=	sign / *es igual*
62	>	ignored
63	?	question mark / pause
64	@	ignored
65	A	A capital letter
66	B	B capital letter
67	C	C capital letter
68	D	D capital letter
69	E	E capital letter
70	F	F capital letter

Decimal ASCII code	*Character*	*Recognised as / translated by*
71	G	G capital letter
72	H	H capital letter
73	I	I capital letter
74	J	J capital letter
75	K	K capital letter
76	L	L capital letter
77	M	M capital letter
78	N	N capital letter
79	O	O capital letter
80	P	P capital letter
81	Q	Q capital letter
82	R	R capital letter
83	S	S capital letter
84	T	T capital letter
85	U	U capital letter
86	V	V capital letter
87	W	W capital letter
88	X	X capital letter
89	Y	Y capital letter
90	Z	Z capital letter
91	[	punctuation / pause
92	\	ignored
93	]	punctuation / pause
94	^	ignored
95	_	ignored
96	`	ignored
97	a	a small letter
98	b	b small letter
99	c	c small letter
100	d	d small letter
101	e	e small letter
102	f	f small letter

Decimal ASCII code	*Character*	*Recognised as / translated by*
103	g	g small letter
104	h	h small letter or time
105	i	i small letter
106	j	j small letter
107	k	k small letter
108	l	l small letter
109	m	m small letter
110	n	n small letter
111	o	o small letter
112	p	p small letter
113	q	q small letter
114	r	r small letter
115	s	s small letter
116	t	t small letter
117	u	u small letter
118	v	v small letter
119	w	w small letter
120	x	x small letter
121	y	y small letter
122	z	z small letter
123	{	punctuation / pause
124	|	ignored
125	}	punctuation / pause
126	~	ignored
127		ignored

7.2 8 bits ASCII characters

Decimal ASCII code	*IBM extended Character*	*Recognised as / translated by*	*ISO* 8859-1 *Character*	*Recognised as / translated by*
128	Ç	Ç capital letter	▌	ignored
129	ü	ü small letter	▌	ignored
130	é	é small letter	‚	ignored
131	â	â small letter	*ƒ*	ignored
132	ä	a small letter	„	ignored
133	à	à small letter	…	ignored
134	å	a small letter	†	ignored
135	ç	ç small letter	‡	ignored
136	ê	ê small letter	ˆ	ignored
137	ë	ë small letter	‰	ignored
138	è	è small letter	Š	ignored
139	ï	ï small letter	‹	ignored
140	î	î small letter	Œ	ignored
141	ì	i small letter	▌	ignored
142	Ä	A capital letter	▌	ignored
143	Å	A capital letter	▌	ignored
144	É	É capital letter	▌	ignored
145	æ	é small letter	‘	ignored
146	Æ	É capital letter	’	ignored
147	ô	ô small letter	“	ignored
148	ö	o small letter	”	ignored
149	ò	o small letter	•	ignored
150	û	û small letter	–	ignored
151	ù	ù small letter	—	ignored
152	ÿ	y small letter	˜	ignored

Decimal ASCII code	*IBM extended Character*	*Recognised as / translated by*	*ISO* 8859-1 *Character*	*Recognised as / translated by*
153	Ö	O capital letter	™	ignored
154	Ü	U capital letter	š	ignored
155	¢	a small letter	›	ignored
156	£	sign / *livre sterling*	œ	ignored
157	¥	ignored	¦	ignored
158	₧	ignored	¦	ignored
159	*ƒ*	ignored	Ÿ	ignored
160	á	á small letter	NBSP	ignored
161	í	í small letter	¡	exclamation mark
162	ó	ó small letter	¢	ignored
163	ú	ú small letter	£	sign / *libra*
164	ñ	ñ small letter	¤	ignored
165	Ñ	Ñ capital letter	¥	ignored
166	ª	a small letter	¦	ignored
167	º	a small letter	§	ignored
168	¿	ignored	¨	ignored
169	⌐	ignored	©	ignored
170	¬	a small letter	ª	a small letter
171	½	sign / *un demi*	«	ignored
172	¼	sign / *un quart*	¬	a small letter
173	¡	a small letter	-	ignored
174	«	ignored	®	ignored
175	»	e small letter	¯	ignored
176	░	ignored	°	sign / *degré*
177	▒	ignored	±	a small letter
178	▓	ignored	²	sign / *puissance deux*
179	│	ignored	³	ignored
180	┤	ignored	´	ignored
181	╡	ignored	µ	sign / *micró*
182	╢	ignored	¶	a small letter
183	╖	ignored	·	a small letter

Decimal ASCII code	*IBM extended Character*	*Recognised as / translated by*	*ISO* 8859-1 *Character*	*Recognised as / translated by*
184	╕	ignored	¸	ignored
185	╣	ignored	¹	ignored
186	║	ignored	º	a small letter
187	╗	ignored	»	e small letter
188	╝	ignored	¼	sign / *un quart*
189	╜	ignored	½	sign / *un demi*
190	╛	ignored	¾	ignored
191	┐	ignored	¿	ignored
192	└	ignored	À	à small letter
193	┴	ignored	Á	á small letter
194	┬	ignored	Â	â small letter
195	├	ignored	Ã	A capital letter
196	─	ignored	Ä	A capital letter
197	┼	ignored	Å	A capital letter
198	╞	ignored	Æ	É capital letter
199	╟	ignored	Ç	Ç capital letter
200	╚	ignored	È	è small letter
201	╔	ignored	É	É capital letter
202	╩	ignored	Ê	ê small letter
203	╦	ignored	Ë	ë small letter
204	╠	ignored	Ì	i small letter
205	═	ignored	Í	í small letter
206	╬	ignored	Î	î small letter
207	╧	ignored	Ï	ï small letter
208	╨	ignored	Ð	D capital letter
209	╤	ignored	Ñ	Ñ capital letter
210	╥	ignored	Ò	o small letter
211	╙	ignored	Ó	ó small letter
212	╘	ignored	Ô	ô small letter
213	╒	ignored	Õ	O capital letter
214	╓	ignored	Ö	O capital letter

Decimal ASCII code	*IBM extended Character*	*Recognised as / translated by*	*ISO* 8859-1 *Character*	*Recognised as / translated by*
215	╫	ignored	×	x small letter
216	╪	ignored	Ø	O capital letter
217	┘	ignored	Ù	ù small letter
218	┌	ignored	Ú	ú small letter
219	█	ignored	Û	û small letter
220	▄	ignored	Ü	U capital letter
221	▌	ignored	Ý	Y capital letter
222	▐	ignored	Þ	ignored
223	▀	ignored	ß	ignored
224	α	a small letter	à	ignored
225	ß	a small letter	á	á small letter
226	Γ	a small letter	â	â small letter
227	π	ignored	ã	a small letter
228	Σ	ignored	ä	a small letter
229	σ	ignored	å	a small letter
230	µ	*sign / micró*	æ	é small letter
231	τ	ignored	ç	ç small letter
232	Φ	ignored	è	è small letter
233	Θ	ignored	é	é small letter
234	Ω	sign / *ohm*	ê	ê small letter
235	δ	ignored	ë	ë small letter
236	∞	ignored	ì	i small letter
237	ø	ignored	í	í small letter
238	ε	ignored	î	î small letter
239	∩	ignored	ï	ï small letter
240	≡	ignored	ð	a small letter
241	±	ignored	ñ	ñ small letter
242	≥	ignored	ò	o small letter
243	≤	ignored	ó	o small letter
244	⌠	ignored	ô	ô small letter
245	⌡	ignored	õ	o small letter

Decimal ASCII code	*IBM extended Character*	*Recognised as / translated by*	*ISO* 8859-1 *Character*	*Recognised as / translated by*
246	÷	ignored	ö	o small letter
247	≈	ignored	÷	a small letter
248	°	sign / *degré*	ø	a small letter
249	∙	ignored	ù	ù small letter
250	·	a small letter	ú	ú small letter
251	√	ignored	û	û small letter
252	ⁿ	ignored	ü	ü small letter
253	²	sign / *puissance deux*	ý	y small letter
254	■	ignored	þ	a small letter
255		ignored	ÿ	y small letter