

The WireLess TelNet voiXtreme voice system

Description

The WireLess TelNet voice system allows the PDA to say texts and to perform ASR inputs using a “standard” telnet application (VT / 5250).

TTS System

Text data to be said (by TTS). May be supplied by:

- TnVT Print commands (**ESC 5i / ESC 4i**)
- Tags (beginning and end) found in displayed text screen data.
- “Voice Triggers” containing constant and variable texts.

ASR System

Voice input from user. The ASR system recognizes the user's voice and converts it into text data using “grammars”. Data input is processed by the local application, and then sent to host following programming configuration.

Grammars are recognition algorithms designed to produce text inputs.

Available grammars are:

- “Digits”: accepts decimal digits from “0” to “9”, and “Cancel” or “Repeat” orders. *Cancel* and *Repeat* keywords are processed by voice library and host transmission is performed following input algorithm.
- “Commands”: accepts “OK”, “Cancel” and “Repeat” orders. *OK*, *Cancel* and *Repeat* keywords are processed by voice library, and host transmission is performed following input algorithm.

Keywords change according to the language selected for the ASR

Voice Triggers

Voice triggers are defined in three sections:

- Conditions, match needed to perform an event
- TTS, to give the user a voice instruction
- ASR, to get data from user

“Voice triggers” will start the voice input process. Their execution is controlled by the presence of specific conditions on the emulated screen, such as:

- Text found at a specific place of the screen
- Cursor position

One or more triggers could be defined to perform voice events. Voice events will be TTS only, ASR only, and TTS + ASR.

In the Voice Trigger, a TTS “voice announce” may be defined to be said before the ASR recognition process starts. These announce may be a mix of constant text (defined inside the configuration file) and variable text (read from the emulated screen). Variable text will be defined as begin position and length. Three numeric values separated by commas, row and column position in screen, and text length delimited by curl brackets ({r,c,l}) i.e. {10,7,14} for a text located at row 10, column 7 with 14 characters length.

For the ASR recognition it is possible to define the ASR grammar to be used and the ASR algorithm.

Standard voice grammars Digits and Commands are included with the voiXtreme libraries, and custom grammars will be added.

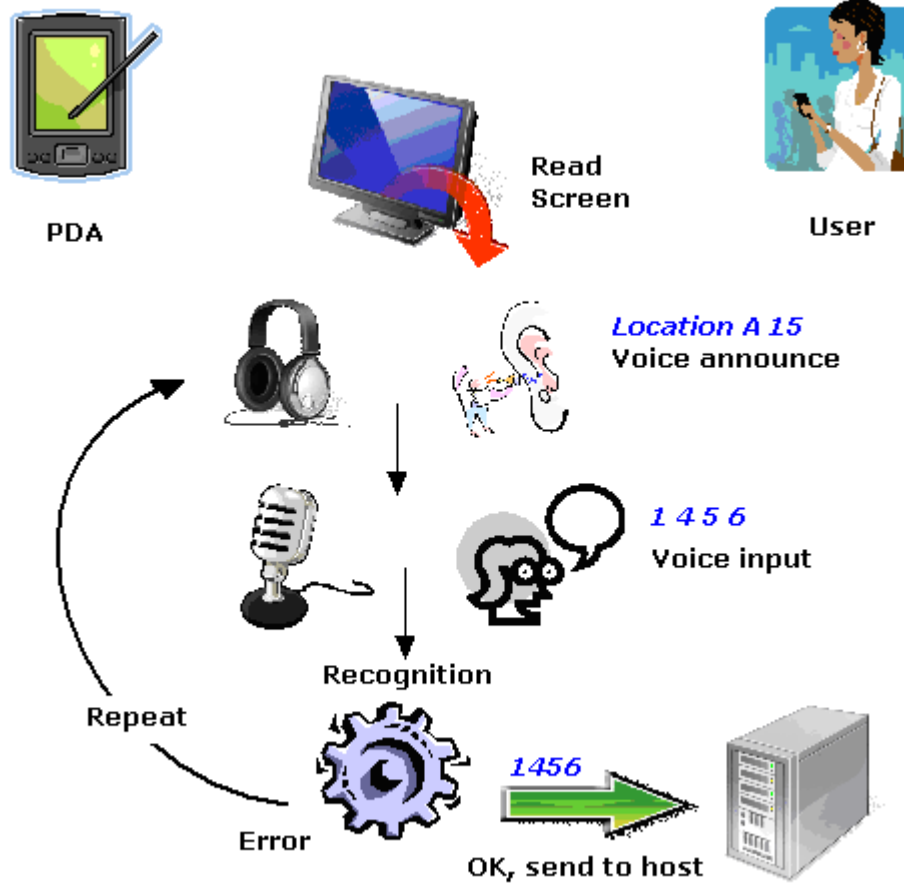
The ASR Algorithm could be:

- *Single* (Confirm=No). The engine recognizes the user input, then send it to host.
- *Single with repetition* (Repeat=Yes). The engine recognizes the user input, repeats the input to the user by TTS, then send it to host.
- *Validate* (Confirm=Yes). The engine recognizes the user input, repeats the input to user and recognizes a command said by user (“*accept*” / “*cancel*”).
 - If *cancel*, it restarts the trigger from start (voice announce).
 - If *accept*, it sends user input to host.
- *Validate with repetition* (Repeat=Yes). Same as Validate, the validation command is echoed to user.

Input Modes:

The following images describe the different input modes.

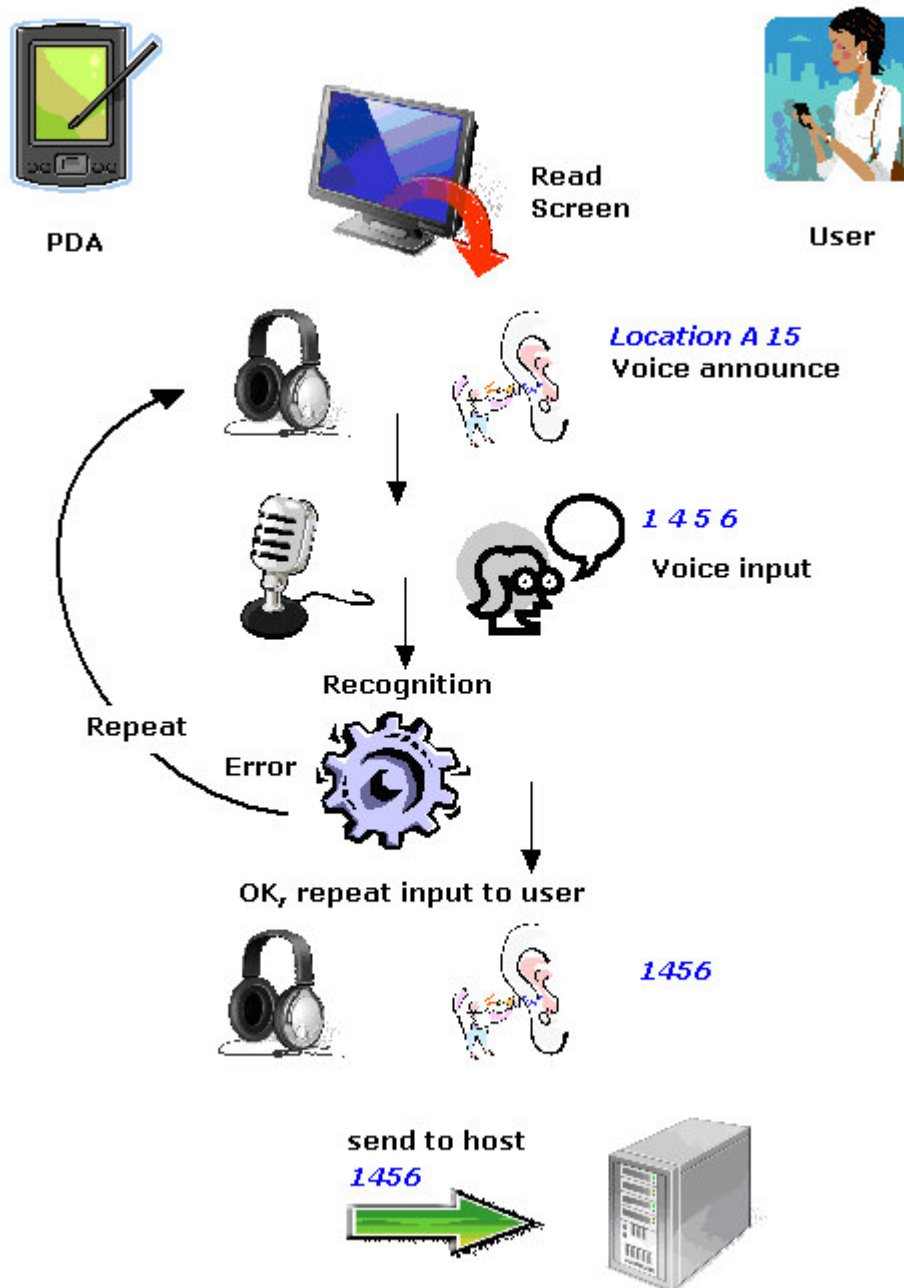
Single Voice input



Single Voice input algorithm:

- 1/ A defined trigger matches among screen data and cursor position.
- 2/ The TTS section of trigger is processed to say the voice announce to the user.
- 3/ Waits end of TTS announce.
- 4/ The ASR section of trigger is processed to perform the ASR input, using the grammar indicated in the section (Digits)..
- 5/ End of recognition:
 - Repeat, goes to (2).
 - Error (no recognition, low reliability level, min / max) goes to (2)
 - Cancel, sends to Host the Cancel string, then exit.
 - Digits, sends to Host the input followed by Terminator string, then exit.

Single + Repeat Voice input



Single + repeat Voice input algorithm:

- 1/ A defined trigger matches among screen data and cursor position.
- 2/ The TTS section of trigger is processed to say the voice announce to the user.

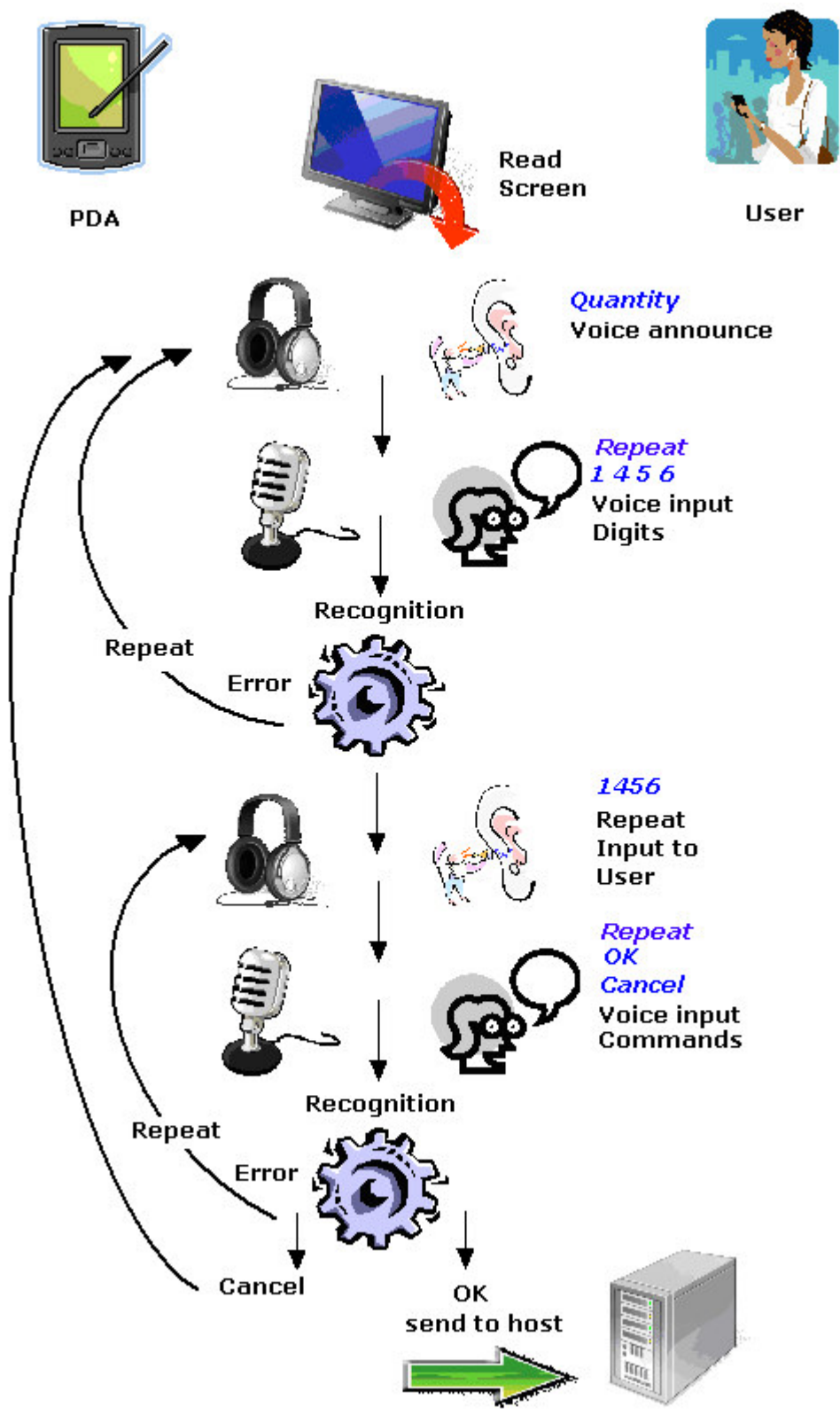
3/ Waits end of TTS announce.

4/ The ASR section of trigger is processed to perform the ASR input, using the grammar indicated in the section (Digits)..

5/ End of recognition:

- Repeat, goes to (2).
- Error (no recognition, low reliability level, min / max) goes to (2)
- Cancel, say Cancel to user, sends to Host the Cancel string, then exit.
- Digits, say the input to user, sends to Host the input followed by Terminator string, then exit.

Validate Voice Input (input digits)



Validate Voice input algorithm (digits):

- 1/** A defined trigger matches among screen data and cursor position.
- 2/** The TTS section of trigger is processed to say the voice announce to the user.
- 3/** Waits end of TTS announce.
- 4/** The ASR section of trigger is processed to perform the ASR input, using the grammar indicated in the section (Digits).
- 5/** End of recognition:
 - Repeat, goes to (2).
 - Error (no recognition, low reliability level, min / max) goes to (2)
 - **Cancel, goes to (9)**
 - Digits, goes to (6).
- 6/** Say the input to the user by TTS.
- 7/** Perform the ASR recognition, using the “Commands” grammar.
- 8/** End of recognition:
 - Repeat, goes to (6).
 - Error (no recognition, low reliability level, min / max) goes to (6)
 - Cancel, goes to (6).
 - OK, says OK to user, sends to Host the input followed by Terminator string, then exit.
- 9/** Say the CancelConfirmation phrase to the user by TTS.
- 10/** Perform the ASR recognition, using the “Commands” grammar..
- 11/** End of recognition:
 - Repeat, goes to (9)
 - Error (no recognition, low reliability level, min / max) goes to (9).
 - Cancel, says Cancel to user, goes to (2).
 - OK, says OK to user, sends to Host the Cancel string, then exits.

Validate Voice Input (input cancel)



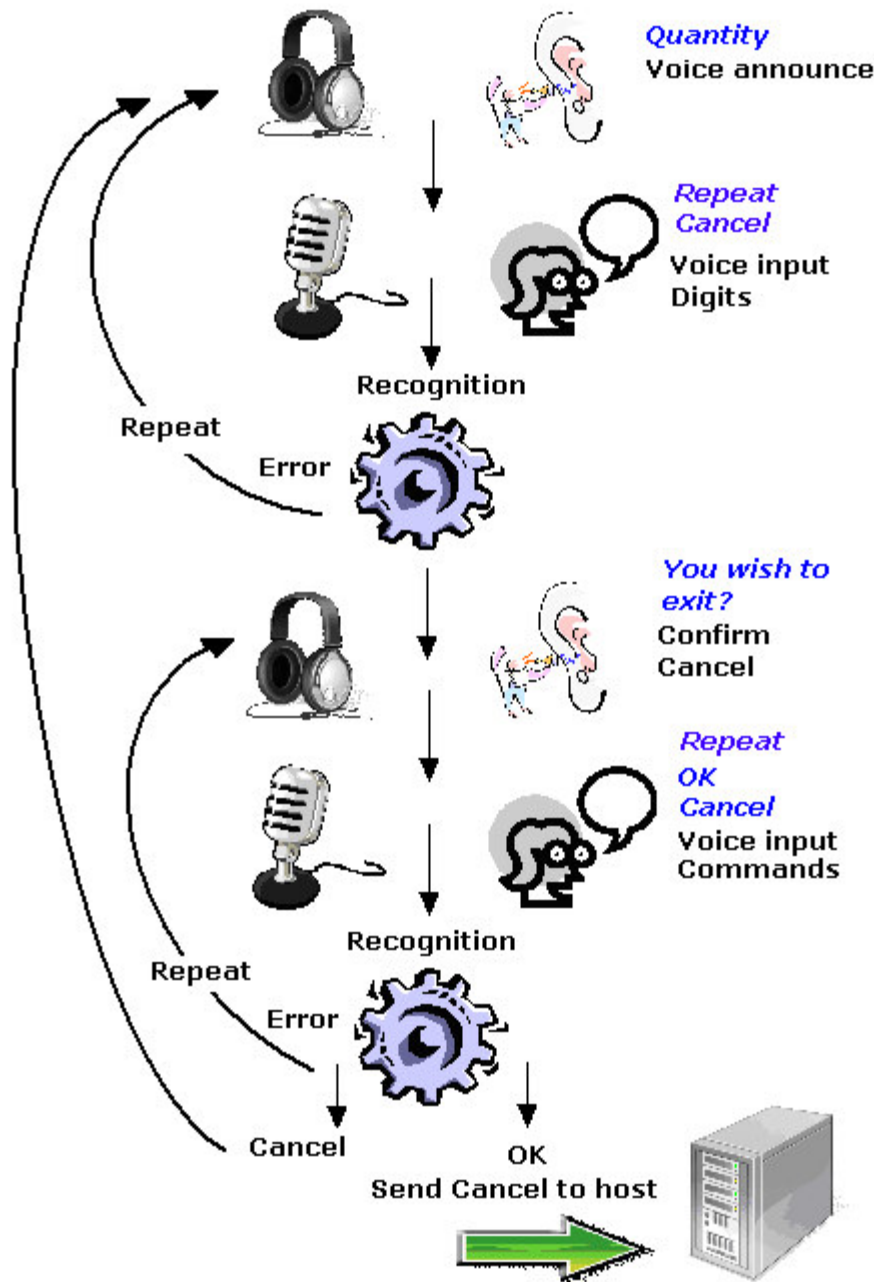
PDA



Read Screen



User



Validate Voice input algorithm (cancel):

- 1/** A defined trigger matches among screen data and cursor position.
- 2/** The TTS section of trigger is processed to say the voice announce to the user.
- 3/** Waits end of TTS announce.
- 4/** The ASR section of trigger is processed to perform the ASR input, using the grammar indicated in the section (Digits).
- 5/** End of recognition:
 - Repeat, goes to (2).
 - Error (no recognition, low reliability level, min / max) goes to (2)
 - Cancel, goes to (9)
 - Digits, goes to (6).
- 6/** Say the input to the user by TTS.
- 7/** Perform the ASR recognition, using the “Commands” grammar.
- 8/** End of recognition:
 - Repeat, goes to (6).
 - Error (no recognition, low reliability level, min / max) goes to (6)
 - Cancel, goes to (6).
 - OK, says OK to user, sends to Host the input followed by Terminator string, then exit.
- 9/** Say the CancelConfirmation phrase to the user by TTS.
- 10/** Perform the ASR recognition, using the “Commands” grammar..
- 11/** End of recognition:
 - Repeat, goes to (9)
 - Error (no recognition, low reliability level, min / max) goes to (9).
 - Cancel, says Cancel to user, goes to (2).
 - OK, says OK to user, sends to Host the Cancel string, then exits.

Read screen

Read Screen sample

In this sample a part of the screen (delimited by SAY: / :SAY) is read to user by TTS.

```
** Order Picking OPL-LOC  
  
Order Number   : 1259-AA0  
Type: Mandatory  
Bin Location    : A-L-008  
Bin-Loc ID:      
SAY: Go to ALPHA LIMA 8 :SAY
```

“go to alpha lima eight” will be heard.

The text delimited by the tags “SAY: - :SAY” is said to the user.

Print Escape commands

This sample shows a text sent to terminal inside a DEC-VT print command.

The escape character (0xc1B) is noted by <ESC>.

```
<ESC>[5iGO TO ALPHA LIMA 8<ESC>[4i
```

“go to alpha lima eight” will be heard.

Configuration File

Here below there's the description of the configuration file that we are designing.

The configuration file follows a “dot-ini” format with sections (delimited by brackets), keys (keywords inside the section, at left of "=" sign) and values (text at right of "=" sign). Sections and keys are in special font: **SECTION**. Values are in green.

Global definitions about the voice engine

TTS (Text-to-speech) definitions section

[TTS_DEF]

This section groups the global Text-To-Speech (TTS) settings, and enables the screen or printer to process voice actions.

VoiceSpeed=70

Default speed of TTS engine. User will modify engine speed.

TtsPrinterOn=Yes

Enables the TnVT printer commands (Esc 5 I) to be captured by TTS engine instead of sending them to printer. Text sent to printer will be said out loud to user by TTS.

TtsScreenOn=Yes

Allows the PDA to say by TTS (Text To Speech) some texts displayed on the screen. Enables the display screen data inspection for TTS beginning and end tags. Looks on the display for the given patterns (beginning and end) and, by TTS, reads aloud the delimited text.

Text delimiters in display definitions

TtsScrSayTagBeg=SAY:

Beginning pattern of the text to look for on emulated terminal display to be read out loud by TTS.

TtsScrSayTagEnd=:SAY

End pattern of the text to look for on emulated terminal display to be read out loud by TTS.

TtsScrSplTagBeg=SPL:

Beginning pattern of the text to look for on emulated terminal display to be spelled by TTS.

TtsScrSpITagEnd=:SPL

End pattern of the text to look for on emulated terminal display to be spelled by TTS.

User messages

Some messages that will be said to user at specific conditions.

MsgWelcome=Bonjour, je suis telle nette.

Message to say at startup.

MsgConnecting=Connexion?

Message to say when connecting.

MsgConnected=Connecté!

Message to say when TelNet connects to host.

MsgDisconnect=Fin connexion.

Message to say at end of TelNet connection.

Keys to control voice engine by user

KeyRepeat=2070

Key to repeat last text speech.

KeyVolUp=2071

KeyVolDn=2072

Keys to change the volume.

KeySpeedUp=2073

KeySpeedDn=2074

Keys to change the TTS speed.

Special character translator options

TranslatorUse=No

TranslatorName=application\WTnVt\Fnt_PC850.cfg

Character translator used for special character codifications. Allows customizing character-by-character translations.

ASR (Automatic Speech Recognition) definitions section

[ASR_DEF]

This section groups the ASR engine settings.

ReliabilityLvl=4000

Minimum reliability level necessary to accept word recognition. When a word is

recognized, ASR engine returns a “reliability level” which is the confidence level that the engine assigns to the recognition. A low reliability level denotes a confusing recognition that will be rejected.

CustomGrammar=custom

Name of the grammar referenced by the input mode parameter WLASR_GRAMMAR_CUSTOM.

ResCancel=Annuler

This is the keyword of the "commands" grammar the user should say to cancel an action.

ResAccept=Accepter

This is the keyword of the "commands" grammar the user should say to accept an action.

ResRepeat=Répéter

This is the keyword of the "commands" grammar the user should say to ask for a repetition (voice announce before ASR) during an ASR recognition.

SyncTts=Yes

This option forces the ASR engine (speech by user) to wait until the TTS engine ends (speech by PDA) before starting user speech recognition.

Voice Triggers

Voice triggers are several sections that allow to perform voice actions following conditions found on the emulated screen. 3 sections represent a trigger:

- The conditions section (conditions to satisfy for running the trigger actions) [VOICETRIGGER_XX].
- The TTS section (the TTS actions to do) [VOICETRIGGER_TTS_XX]
- The ASR section (the ASR actions to do) [VOICETRIGGER_ASR_XX]

One (at least) or more triggers will be defined.

Trigger 1 ...

[VOICETRIGGER_01]

[VOICETRIGGER_TTS_01]

[VOICETRIGGER_ASR_01]

Trigger 2 ...

[VOICETRIGGER_02]

[VOICETRIGGER_TTS_02]

[VOICETRIGGER_ASR_02]

Trigger 3 ...

[VOICETRIGGER_03]

[VOICETRIGGER_TTS_03]

[VOICETRIGGER_ASR_03]

...

Voice trigger conditions section

Defines the conditions to look for in the emulated screen to run the trigger.

There are two kinds of conditions, the cursor position, on one hand, and three text conditions, on the other. Text conditions are excluding conditions (Logical AND) i.e. all the text conditions must be true to run the trigger. Conditions are:

- Text found on the emulated screen (on three places)
- Cursor position on screen

[VOICETRIGGER_01]

On=Yes

This trigger is enabled or not.

First text match on the emulated screen

Row_1=4

Row position to look for a text. 0 means all columns.

Col_1=4

Column position to look for a text. 0 means anywhere in the row.

Row=0 and Col=0 means anywhere in the screen.

Match_1=Show System Summary

Text to look for at Row_1 / Col_1 position. To consider ending spaces, text will be delimited by quotes "".

If no text to look for is supplied, the condition is not evaluated (TRUE).

Second text match on the emulated screen

Row_2=5

Col_2=4

Match_2=Main Menu

Third text match on the emulated screen

Row_3=0

Col_2=0

Match_2=

Cursor position condition

CursorAtRow=8

Cursor row position to evaluate. 0 Means in any row.

CursorAtCol=20

Cursor column position to evaluate. 0 Means in any column of the row.
Row=0 and Column=0, means anywhere in the screen.

Voice trigger TTS to do section**[VOICETRIGGER_TTS_01]**

These are the Text-To-Speech actions to do when the voice trigger runs. Constant text (supplied in the trigger) and variable text (extracted from emulated screen) will be said to user.

DoTts=Yes

Enables / disables TTS actions in this trigger.

TtsSay01=any constant text{r,c,l} another constant text {r,c,l} final text.

Combination of constant text (normal text in key) and variable text (delimited by curly brackets {}).

Variable text is denoted by three numeric values denoting the starting point of variable text by row and column, and the text length (row, column, length).

All the text is read as words. A special tag indicates parts of text to spell.

Special means of text tags <SPL></SPL> delimits text to spell, i.e. the text “*location* <SPL>AB8</SPL> *to go*” till be said “*location ey be eight*”.

TtsSay02= any constant text{r,c,l} another constant text {r,c,l} final text.

Several TTS actions are allowed.

Voice trigger ASR to do section.**[VOICETRIGGER_ASR_01]**

This is the ASR action to perform in this trigger:

DoAsr=Yes

Enables / disables ASR action for this trigger.

Grammar=digits

Grammar to use in this trigger action. The grammar is a set of rules and words that ASR engine use for speech recognition. Standard ASR engine includes two grammars:

commands (Cancel / Accept) and *digits* (0, 1, 2, 3, 4, 5, 6, 7, 8, 9).

Custom special grammars will be built.

LenMin=0

This is the minimum expected voice input length. Inputs lesser then this will be rejected, and a new input will be requested.

LenMax=32

This is the maximum expected voice input length. Inputs bigger than this will be rejected, and a new input will be requested.

Confirm=Yes

Uses the *Confirm* or the *Simple* voice input algorithm to perform ASR input action. See Input modes and figures below for ASR algorithm explanation.

Repeat=Yes

Uses or not the Repeat algorithm to perform ASR input action. See Input modes and figures below for ASR algorithm explanation.

CancelDo=Yes

Allow user to cancel a data input, and send cancel sequence to host.

CancelConfirmation=Voulez-vous quitter?

The question said to the user if a Cancel is said.

Cancel=\1B\41

A sequence sent to host computer when the user cancels the input without any data.

Terminator=\0D

This is a terminator string that will be sent to the host after the ASR data (added as trailing) to validate input (Enter in this case).

Sample configuration file

```
//-----  
// Tts definitions  
[TTS_DEF]  
VoiceSpeed=70  
TtsPrinterOn=Yes  
TtsScreenOn=No  
TtsScrSayTagBeg=SAY:  
TtsScrSayTagEnd=:SAY  
TtsScrSplTagBeg=SPL:  
TtsScrSplTagEnd=:SPL  
MsgWelcome=Bonjour, je suis telle nette.  
MsgDisconnect=Fin connexion.  
MsgConnected=Connecté!  
MsgConnecting=Connexion?.  
KeyRepeat=2070  
KeyVolUp=2071  
KeyVolDn=2072  
KeySpeedUp=2073  
KeySpeedDn=2074  
TranslatorUse=No  
TranslatorName=application\WtNvt\Fnt_PC850.cfg  
  
//-----  
// Asr definitions  
[ASR_DEF]  
ReliabilityLvl=4000  
ResCancel=Annuler  
ResAccept=Accepter  
ResRepeat=Répéter  
SyncTts=Yes  
  
//-----  
// Triggers  
//-----  
  
//-----  
// Voice trigger conditions  
[VOICETRIGGER_01]  
On=Yes  
Row_1=4  
Col_1=4  
Match_1=Show System Summary  
Row_2=0  
Col_2=0  
Match_2=  
CursorAtRow=0  
CursorAtCol=0  
  
// Voice trigger TTS to do  
[VOICETRIGGER_TTS_00]  
DoTts=Yes  
TtsSay01=Goto
```

```
TtsSay02=Location {1,4,7}  
TtsSay03=Please
```

```
// Voice trigger ASR todo  
[VOICETRIGGER_ASR_00]  
DoAsr=Yes  
Grammar=digits  
LenMin=4  
LenMax=4  
Confirm=Yes  
Repeat=Yes  
CancelDo=Yes  
CancelConfirmation=Voulez-vous quitter?  
Cancel=\1B\41  
Terminator=\0D
```

Screens and Triggers samples

This sample explains the triggers configuration to do a voice input for a picking operation Bin-Location and quantity input in two screens.

Bin location screen

```
** Order Picking OPL-LOC
User : 0234 John JUAREZ
Order Number : 1259-AA0
Line : 05 / 23
Type: Mandatory
Bin Location : AB- 15
Bin-Loc ID:  _
```

In this case a simple algorithm is used as the two digits location ID is known by system, then the input is send without confirmation, the host sends a new screen (if match) or sends the same (if not match).

PDA *“go to location ey be fifteen please”*

User *“five seven”*

PDA *“fifty seven”*

PDA to host *“57<ENTER>”*

```
[VOICETRIGGER_01]
```

```
On=Yes
```

```
Row_1=1
```

```
Col_1=1
```

```
Match_1=** Order Picking OPL-LOC
```

```
Row_2=0
```

```
Col_2=0
```

```
Match_2=
```

```
CursorAtRow=7
```

```
CursorAtCol=0
```

```
[VOICETRIGGER_TTS_01]
```

```
DoTts=Yes
```

```
TtsSay01=Go to
```

```
TtsSay02=Location <SPL>{6,19,2}</SPL>{6,22,3}
```

```
TtsSay03=Please
```

```
[VOICETRIGGER_ASR_01]
```

```
DoAsr=Yes
```

```
Grammar=digits
```

```
LenMin=2
```

```
LenMax=2
```

```
Confirm=No
```

```
Repeat=Yes
```

```
CancelDo=Yes
```

```
CancelConfirmation=Do you wish to exit?
```

```
Cancel=\1B
```

```
Terminator=\0D
```

Quantity at location screen

```
** Order Picking OPL-QTY
User : 0234 John JUAREZ
Order Number : 1259-AA0
Line : 05 / 23
Type: Mandatory
Bin Location : AB- 15
Product : 1238767653490
Quantity: 
```

In this case a validate algorithm is used to confirm input quantity. Only the last three digits of product are said.

PDA “product four hundred ninety, give quantity”

User “one four four”

PDA “one hundred and forty for”

User “accept”

PDA to host “144<ENTER>”

```
[VOICETRIGGER_02]
```

```
On=Yes
```

```
Row_1=1
```

```
Col_1=1
```

```
Match_1=** Order Picking OPL-QTY
```

```
Row_2=0
```

```
Col_2=0
```

```
Match_2=
```

```
CursorAtRow=8
```

```
CursorAtCol=0
```

```
[VOICETRIGGER_TTS_02]
```

```
DoTts=Yes
```

```
TtsSay01=Product {7,23,3}
```

```
TtsSay02=Give quantity
```

```
[VOICETRIGGER_ASR_02]
```

```
DoAsr=Yes
```

```
Grammar=digits
```

```
LenMin=1
```

```
LenMax=6
```

```
Confirm=Yes
```

```
Repeat=No
```

```
CancelDo=Yes
```

```
CancelConfirmation=Do you wish to exit?
```

```
Cancel=\1B
```

```
Terminator=\0D
```