VECSYS
Des Hommes de Parole°

Réf. : MS_RIS_5.01_001A

MEDIASPEECH®

REST INTERFACE SPECIFICATIONS

VERSION 5.01

Summary

Glossary

MS	MediaSpeech, Vecsys speech and language processing system.
Cookie	Parcels of text sent by a server to a Web client (usually a browser) and then sent back unchanged by the client each time it accesses that server.
<empty>	Used to describe an empty field.
Process	Task executed on MS. It means not only transcription.
BN	Broadcast News (TV or radio speech).
CTS	Conversational Telephone Speech
JobId	The job identifier is a string returned by MS to identify a process.
Gray fields	In the description of the API all grayed fields are optional fields.

1 Introduction.

MediaSpeech® (MS) is a product line dedicated for speech and language processing. It gives access to the Vecsys speech processing through a Web Service and REST interface.

These interfaces are available for the Lite and Factory specifications of MediaSpeech.

Users can upload audio file onto MS server, can monitor the transcription process and at last download the result.

This document explains the means for people to process audio file using MSF as a WEB service based on REST.

VECSYS provides to each user a login and password to access MSF. The user has access only to his uploaded audio files and their processes.

In the protocol description, the parameters in italic font with grey background are optional. In fat font is the default value.

2 Introduction to the REST API.

This chapter is dedicated to rules common to all functions listed in this document.

2.1   Speech and language processing.

To process a speech file, the user must first login into MSF, then upload his file (the currently supported format are **'wav','mp3','ogg','flv','bwf','mpg'**, **'wma'**, **'m4v'**, **'mp4'**, **'m4a'**, **'amr'**, **'mp2'**, **'gpg'**, **'aif'**, **'caf'**, **'alaw'**, **'al'**). If your file format is not supported, please call the technical support, its format may be added to MSF.

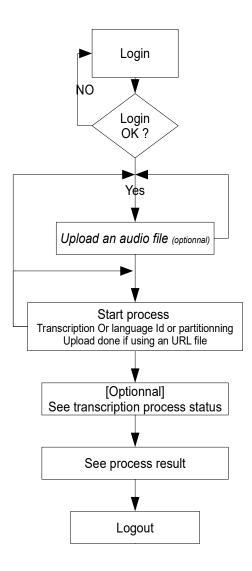Once uploaded or downloaded by MSF from an URL, the user can start the process (which can be transcription, alignment ...).

The user can upload all its files at once and then start the transcriptions, or upload then start the transcription sequenced.

It's possible to follow the progression of the process, and at last see the result of the process.

The user needs to logout, when he stops accessing MSF.

```
            ┌──────────────────┐
      ┌────▶│      Login       │
      │      └──────────────────┘
      │               │
 NO   │               ▼
      │         ◇ Login ◇
      └────────   OK ?
                    ◇
                    │
                   Yes
                    │
                    ▼
         ┌───────────────────────────────┐
         │ Upload an audio file (optionnal)│
         └───────────────────────────────┘
                    │
                    ▼
         ┌───────────────────────────────┐
         │         Start process          │
         │ Transcription Or language Id or partitionning │
         │ Upload done if using an URL file │
         └───────────────────────────────┘
                    │
                    ▼
         ┌───────────────────────────────┐
         │          [Optionnal]           │
         │ See transcription process status │
         └───────────────────────────────┘
                    │
                    ▼
         ┌───────────────────────────────┐
         │      See process result        │
         └───────────────────────────────┘
                    │
                    ▼
         ┌───────────────────────────────┐
         │           Logout               │
         └───────────────────────────────┘
```

2.2 Results

All functions return an array of strings composed by:

- An error code which gives the result of the function's call (0 means no error).
- An error message which gives the human readable error.
- Some additional data. Those data are specific to some functions (example: a Job identifier is returned when starting a new process).

```
status=xxx
message=bla bla


          -- additional optional data --
```

2.3 File name

Doc.: MSF_SpecificationsInterfaceRest.doc

Date : 26/03/2017

Some functions need a file name reference. This file can be an URL or a file located on MSF and previously download by MSF.

The URL file name must look like:

- http://srv/path/file.ext .
- ftp://srv/path/file.ext

A login and a password can be specified within the URL.

Example:

https://login:password@srv/path/file.ext

Be aware that using a non secure method, like FTP, may expose your login and password on the Internet.

You can setup your firewall to limit the access of your server to mediaspeech.com IP's (To get the list of IP's, just type "nslookup mediaspeech.com".

2.4 Callbacks

MSF use callback to give status information to its users. The callback is called when the status of a process is changing. When a job is completed the user has to download its result.

MSF does send the process results only when it's processing a multimedia stream.

The callback is an URL, its format must be:

- http://ip/module .
- http://ip/module?your_params=xxx .

MSF can adds one or more of the following information to the URL:

- JobId=xxx (xxx is a string representing the job identifier).
- Status=X (X: can be D for downloading, R for running, E for Ended, F for Finished (Ended but result has been downloaded)).
- Command=Cmd (Cmd: can be wget, trans, part, lid).
- ErrorCode=Err (Err is an integer).
- StreamId=xxx (xxx is a string representing the stream identifier).

On a status change MSF sends: JobId and Status.

On error MSF sends: JobId and Command and ErrorCode.

2.5 Encrypted files

Doc.: MSF_SpecificationsInterfaceRest.doc

Date : 26/03/2017

You can send to MSF an GPG encrypted audio file (functionnality only available for the transcription process).

MSF will store the encrypted file and will only decrypt the file to process it. The result will be encrypted with the same GPG key.

Your GPG key will not be stored into MSF and MSF keeps no uncrypted file after the transcription.

The file name MUST BE like:

audio_file_name.audio_extention.gpg

The gpgkey is the password used to encrypt your file with the following command:

gpg -c audio_file_name.audio_extention

2.6 Sessions

To use MSF you must provide with all functions calls the PHP session identifier provided when you login into MSF.

Be aware that most REST clients do not handle cookies by default. You must add the cookie support and keep the PHP session identifier within your client.

2.7 Common results.

The common part of all results is composed of 3 lines.

```
status=xxx
message=bla bla

result
```

The status : It's a digit (0 for no error).

The message : The message text may be empty, and may contain the error message if the status is not zero.

A blank line.

The result : May be empty. It's the result of the command.

3 Login

function	?function=login
Description	Provide a login and password to MSF in order to authenticate the user. The user MUST login into MSF and accept cookies to gain access to any functionality.

3.1 Arguments :

Name	Description	Submit method
username	Provide the username of the user's account.	POST
password	Provide the password of the user's account.	POST

3.2 Results.

The login returns a simple text.

The result of the login process is displayed in the message field.

```
status=xxx
message=bla bla
```

4 Upload a file.

The right way to upload a file is to use a FTP client. You can use it to access your home directory on MediaSpeech "ftp://mediaspeech.com".

The audio file must be stored in the '/media' directory.

```
jobid=The Job Id
```

The Job id is a string with digits, dots and characters.

VECSYS
Des Hommes de Parole®

6 Process status.

function	?function=status
Description	Allow a user to retrieve the status of it's running processes (transcription) . The user can see some detailed informations regarding its process, or list all its processes running on MSF.

6.1 Arguments :

jobid	String	Identification of the running process, this field is returned by the start process pages.

6.2 Results.

The status returns some text.

The status returns the status of the job with some additional data:

jobident	Job identifier.
processtype	Type of process.
jobstatus	Running, Queued, Ended, Finished (Ended but the result has been read), Downloading file, Canceled.
startdate	Start date and time of the process.
stopdate	Stop date and time of the process.

Additional columns may be added in the future. The separator between fields is a space.

Regarding Ended processes, once the user has access its results, it will be removed from the status display.

6.2.1   Example:

After asking for the status of all jobs (Job Id no provided), MSF will display a screen like :

```
status=0
message=

JobIdProcessNameStatusDateStartTimeStartDateStopTimeStopProgress
418.msf01 Transcription E 2009-04-15 11:54:51 2009-04-15 12:16:19?
419.msf01 Transcription E 2009-04-15 13:58:32 2009-04-15 14:20:01?
422.msf01 Transcription E 2009-04-15 15:26:57 2009-04-15 15:27:42?
```

VECSYS
Des Hommes de Parole®

Doc.: MSF_SpecificationsInterfaceRest.doc
Date : 26/03/2017

7 Transcriptions result.

function	?function=result
Description	Allow a user to retrieve the status of its running processes and retrieve the result of the transcription process. The user can see some detailed informations regarding its process, or list all its processes running on MSF.

7.1 Arguments :

jobid	String	Identification of the running process, this field is returned by the start process pages.
format	String	Gives the output format, default is xml.
	txt	Srt is the subrip format, dedicated for subtitles
	xml	Text is plain text, one paragraph per speaker
	srt	xml is the transcription standard output (see DTD in annex).

7.2 Results.

The status returns some raw text and some XML text.

The raw text is the error is returned by the transcription tools.
The XML part contains the result of the transcription process.
The process progression status may be subject to changes to ease its exploitation par users.
Transcription result example:

Here is the result of the transcription of an audio file. The error file is empty and the output file contains the result of the transcription.

```
status=0
message=

<?xml                              version="1.0"                              encoding="UTF-8"?>
<AudioDoc          name="PKCTS05_FRE_FR_0000159291_05.1"          path="/Audio/PKCTS05_FRE_FR_0000159291_05.1.wav">
<ProcList>
<Proc                              name="vrcts_part"                              version="1.0"/>
<Proc                              name="vrcts_lid"                              version="3.2"/>
<Proc                              name="vrcts_trans.fre"                         version="2.3"/>
</ProcList>
```

```
<ChannelList>
<Channel          num="1"          sigdur="304.26"          spdur="240.13"          nw="845"          tconf="0.68"/>
</ChannelList>
<SpeakerList>
<Speaker    ch="1"    dur="240.13"    gender="1"    spkid="1"    lang="fre"    lconf="0.98"    nw="845"    tconf="0.68"/>
</SpeakerList>
<SegmentList>

                            …........................
<Word          stime="299.81"          dur="0.04"          conf="0.761">          Ã          </Word>
<Word          stime="299.85"          dur="0.04"          conf="0.760">          l'          </Word>
<Word          stime="299.89"          dur="0.16"          conf="0.760">          heure          </Word>

                            …........................
</SpeechSegment>
</SegmentList>
</AudioDoc>
```

8 Logout

function	?function=logout
Description	Destroy the user's session created during the login.

8.1 Arguments :

Name	Description	Submit method

No argument is needed.

8.2 Results.

The logout returns some text.

The result of the upload process is displayed in the message field.

```
status=xxx
message=bla bla
```

9 Delete an uploaded file.

function	?function=delete
Description	Allow a user to delete an uploaded audio file and it's transcriptions.

9.1 Arguments :

Name	Description	Submit method
filename	Filename of the file to upload.	POST

9.2 Results.

The deletion returns some text.

The result of the upload process is displayed in the message field.

```
status=xxx
message=bla bla
```

10 Cancel a running transcription.

function	?function=cancel
Description	Allow a user to stop a running processes.

10.1 Arguments :

Name	Description	Submit method
jobid	Identification of the running process, this field is returned by the start transcription page.	POST

10.2 Results.

The cancel returns some text.

The result of the upload process is displayed in the message field.

```
status=xxx
message=bla bla
```

11 Annex

1.1.   Mediaspeech team contact

Phone   : +33 1 69 29 87 87
Email   : support@mediaspeech.com
Forum   : https://mediaspeech.com/forum

1.2.   What about the storage of the audio file and their transcriptions?

All not uploaded results will be deleted after 1 week.

1.3.   Where should I connect to ?

https://mediaspeech.com/webservice

This page lists all functions available on MS, and allow the user to download the WSDL description file.

1.4.   What about security ?

All the communications with MS are encrypted. Only https is used.

1.5.   What are the supported audio types ?

The supported audio types are:
- All formats supported by **mplayer** should be supported.
- WAV PCM (16 KHz for BN files, 8KHz for CTS files, 16 bits.: should be used only if the original data are not compressed).

The minimum quality for the audio file is 64kbs.

1.6. Common errors

The following messages are common to all commands.

Code	Message	description

1000	Error	Generic error code
1001	invalid login	The provided username or the password are invalid.
1002	login needed	The user is not logged in.
1003	file not found, upload failed	The upload process has failed.
1004	null file size, check your local file access rights	The uploaded file size is null, may be due to your local file access rights.
1005	invalid file extension	The extension is not valid (not an audio file).
1006	invalid file name	The file name contains invalid characters.
1007	file not copied into MS content directory	MS server failed to copy the audio file.
1008	logout failed	The logout process has failed.
1009	file not deleted	The file has not been deleted !
1010	Unknown Job Id	The process has not been cancelled.
1011	No upload done	The upload has failed
1012	Function submitted in *function* not implemented	The user has submitted a function name (ex: function=toto) that is not implemented.
1013	Invalid argument	On of the argument of a function supplied to MS is invalid.
1014	Cannot cancel the job	Can't delete a job (maybe invalid job name).
1015	Cannot delete the file	Can't delete the file.
1016	Quota reached	The user has reached the limit of x available process starts.
1017	User already exists	Admin use only
1018	can't play the multimedia file	
1019	can't decrypt file	Gpg was not able to decrypt the file
1020	Maintenance in progress	The system is under maintenance, try to connect later.
1021	empty result file	
1022	system failure, please contact support : support@mediaspeech.com	
1023	text file does not exist on server	
1024	audio file does not exist on server	
1025	Your account has expired, please consider buying access rights	

VECSYS
Des Hommes de Parole®

1026	Error downloading remote multimedia file	
1027	You do not have the access rights to transcribe in the language or audio type you're asking	
1028	No language specified	
1029	Database connection problem	
1030	Blade : not enough disk space	
1031	You are out of you allowed hours	

1.7.   DTD for the XML Result file

```
<!ELEMENT AudioDoc  ( ProcList , ChannelList , SpeakerList , SegmentList )>
<!ATTLIST AudioDoc
name CDATA #REQUIRED
path CDATA #REQUIRED
>

<!ELEMENT SegmentList  ( SpeechSegment+ )>

<!ELEMENT SpeechSegment  ( Word* )>
<!ATTLIST SpeechSegment
ch CDATA #REQUIRED
sconf CDATA #IMPLIED
stime CDATA #REQUIRED
etime CDATA #REQUIRED
spkid CDATA #REQUIRED
spkname CDATA #IMPLIED
spkconf CDATA #IMPLIED
lang CDATA #REQUIRED
lconf CDATA #IMPLIED
trs CDATA #IMPLIED
>

<!ELEMENT Word  ( #PCDATA )>
<!ATTLIST Word
stime CDATA #REQUIRED
dur CDATA #REQUIRED
>

<!ELEMENT SpeakerList  ( Speaker+ )>

<!ELEMENT Speaker  EMPTY>
<!ATTLIST Speaker
ch CDATA #REQUIRED
dur CDATA #REQUIRED
gender CDATA #REQUIRED
spkid CDATA #REQUIRED
lang CDATA #REQUIRED
lconf CDATA #IMPLIED
nw CDATA #IMPLIED
tconf CDATA #IMPLIED
>

<!ELEMENT ChannelList  ( Channel+ )>
```

```
<!ELEMENT Channel  EMPTY>
<!ATTLIST Channel
num CDATA #REQUIRED
sigdur CDATA #REQUIRED
spdur CDATA #REQUIRED
nw CDATA #IMPLIED
tconf CDATA #IMPLIED
>


<!ELEMENT ProcList  ( Proc+ )>


<!ELEMENT Proc  EMPTY>
<!ATTLIST Proc
name CDATA #REQUIRED
version CDATA #REQUIRED
>
```

The XML transcription result fields are described below.

The Channel tag includes the following attributes:

num	channel number (1 or 2)
sigdur	signal duration in seconds
spdur	total speech duration in seconds
nw	number of hypothesized words
tconf	average transcription confidence score between 0 and 1

The Speaker tag can have up to 7 attributes:

ch	channel number (1 or 2)
dur	total speech duration in seconds
gender	hypothesized speaker gender (1 for male, 2 for female)
spkrid	an arbitrary speaker number starting from 1
spkname	name of the detected speaker
spkconf	speaker confidence score between 0 and 1
lang	language code
lconf	language confidence score between 0 and 1
nw	number of hypothesized words
tconf	average transcription confidence score between 0 and 1

The SpeechSegment tag can have up to 8 attributes:

ch	channel number (1 or 2)
sconf	speech detection confidence score between 0 and 1
stime	segment start time in seconds from the beginning of the document
etime	segment end time in seconds from the beginning of the document

spkrid	an arbitrary speaker number starting from 1
lang	language code
lconf	language confidence score between 0 and 1

The Word tag has 3 attributes:

stime	word start time in seconds from the beginning of the document
dur	word duration in seconds
conf	word confidence score between 0 and 1

The language code are:

amh	Amharic
ara	Modern standard Arabic
ara-egy	Egyptian Arabic
ara-kwt	Kuwaitian Arabic
ara-lbn	Lebanese Arabic
ara-mar	Moroccan Arabic
eng	Standard English
eng-aus	Australian English
eng-can	Canadian English
eng-gbr	UK English
eng-ind	Indian English
eng-usa	US English
eng-zaf	South Africa English
ger	standard German
fas	Farsi (Persian)
fre	standard French
fre-bel	Belgium French
fre-can	Canadian French
fre-che	Switzerland French
fre-mtq	Martinique French
fre-MAF	Middle Africa French
fre-NAF	Northern Africa French
hin	Hindi
ita	Italian
jpn	Japanese
kor	Korean
por	Portuguese
por-bra	Brazilian Portuguese
por-ago	Angola Portuguese
rus	Russian

spa	Castilian Spanish
spa-SAM	South American Spanish (including Caribbean)
tam	Tamil
vie	Vietnamese
chi	Mandarin Chinese
chi-CNT	Cantonese (Yue Chinese)

MediaSpeech® is a registered trademark of Vecsys.

12 Very simple Example of use with wget

12.1   Login:

wget --cookies=on --no-check-certificate --keep-session-cookies --save-cookies=cookie.txt --post-data 'username=UUU&password=PPP' https://mediaspeech.com/rest/?function=login -O ./res.txt

12.2   Transcription in French and high quality:

wget --cookies=on --no-check-certificate --load-cookies=cookie.txt --keep-session-cookies --save-cookies=cookie.txt

https://mediaspeech.com/rest/?function=trans\&type=bn\&filename=audio.mp3\&language=fre\&quality=2 -O ./res.txt

12.3   Status:

wget --cookies=on --no-check-certificate --load-cookies=cookie.txt --keep-session-cookies --save-cookies=cookie.txt https://mediaspeech.com/rest/?function=status\&jobid=XXX -O res.xml

12.4   Result:

wget --cookies=on --no-check-certificate --load-cookies=cookie.txt --keep-session-cookies --save-cookies=cookie.txt   https://mediaspeech.com/rest/?function=result\&jobid=XXX\&format=xml   -O ./res.xml