Réf. : MS_WSIS_5.01_001A

MEDIASPEECH®

WEB SERVICE INTERFACE SPECIFICATION

VERSION 5.01

www.vecsys.fr

Siège social : 3, rue de la Terre de Feu • ZA de Courtaboeuf • 91940 Les Ulis • France
Adresse postale : 3, rue de la Terre de Feu • ZA de Courtaboeuf • Les Ulis • 91952 Courtaboeuf Cedex • France
Tél. : + 33 (0)1 69 29 87 87 • **Fax** : + 33 (0)1 69 07 58 58
Site Strasbourg : 37, boulevard Clemenceau • Bâtiment 6 • 67000 Strasbourg • France • + 33 (0)9 64 43 69 02
S.A. au capital de 1 011 750 Euros • 316 676 741 RCS EVRY • Code APE 6202A • Siret 316 676 741 000 • TVA Intracommunautaire FR 13316676741

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 but also audio language identification or audio partitioning.
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 MS as a WEB service based on SOAP.

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

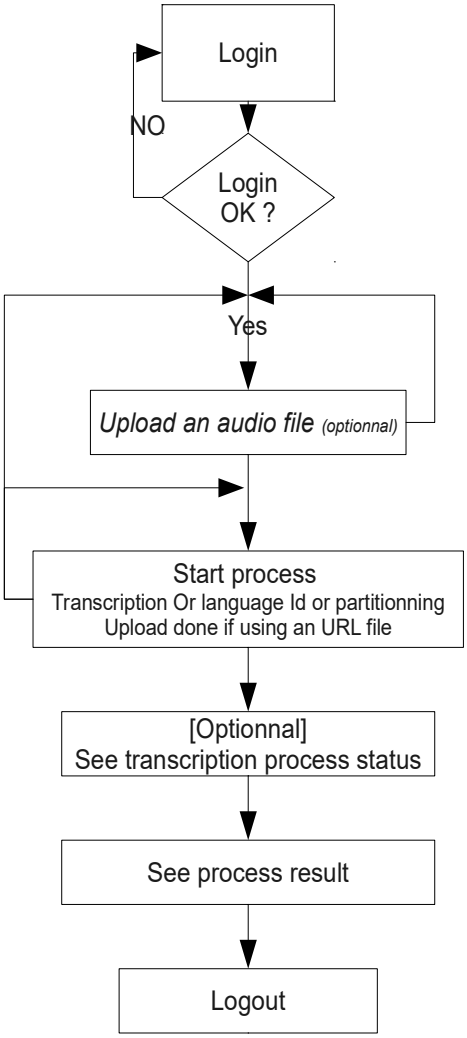
In the protocol description, the parameters in italic font with grey background are optional. In fat font is the default value.

2. Introduction to the web service 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 MS, 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 MS.
Once uploaded or downloaded by MS from an URL, the user can start the process (which can be transcription, language identification, audio partitioning, 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 MS.



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).

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE result SYSTEM "MsfResults.dtd">
<result>
<status>Error Code</status>
<message>Error Message</message>
-- additional optional data --
</result>
```

2.3. File name

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

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

MS 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.

MS 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 .

MS 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 MS sends: JobId and Status.
On error MS sends: JobId and Command and ErrorCode.

2.5. Encrypted files

You can send to MS an GPG encrypted audio file (functionality only available for the transcription process).
MS 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 MS and MF 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 MS you must provide with all functions calls the PHP session identifier provided when you login into MS.

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

In C# you will have to add the following blue lines to your code:

```
...
//webservice instanciation
WebserviceMSFService server = new WebserviceMSFService();

//enable the cookies
server.CookieContainer = new System.Net.CookieContainer();
...
```

3. Login

Provide a login and password to MS in order to authenticate the user.

The user MUST login into MS and handle sessions to gain access to any functionality.

The sessions are not persistent, so the same script has to login, process data then logout.

The function name is:

login

Arguments:

username	string	Provide the username of the user's account.
password	string	Provide the password of the user's account.

4. Upload a file

Allow a user to upload a local audio file into MS server.

This function should not be used, you should use the process functions with a reference to a WEB hosted file ! You will experience problems when not using medium and large sized files.

The function name is:

upload

Arguments:

filename	String	filename of the file to upload. If the file already exists on MS, the previously uploaded file will be erased.
filecontent	base64 binary	Base64binary encoded file content.

Another 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 or ftp://cluster1.mediaspeech.com. depending where is located your account.

5. Start transcription

Allow a user to start the transcription of an audio file previously uploaded or a WEB located file.

The function name is:

trans

Arguments:

filename	String	File name of the audio to process
type	String Values: cts bn ccs lcs	Specify the type of audio file. CTS: Telephonic conversation. *(if component present)* BN: Broadcast. *(if component present)* *CCS: Call Center Speech. (if component present)* *LCS :Lecture and Conference Speech (if component present)*
language	*String* *Values:* *fre* *eng* *ara* *chi* *spa* *ita* *rus*	*Specify the language of the audio file.*
channel	*integer* *Values:* *0* ***1*** *2*	*Specify the channel to decode.* *If the audio file is stereo, specify the channel 1 or 2.* *ONLY with CTS audio file.* *Use 0 for other systems*
quality	*Integer* *Values:* ***0*** *1* *2*	*used to parameter the transcription* *for MSF > v2.40* *0: normal quality.* *1: medium quality.* *2 : high quality*
corpus	*string*	*Name of the user's model, empty if none.*
duration	*Integer*	*[Reserved]Must be 0*
callback	*String*	*Url called when the status of a job change.*
gpgkey	*String*	*GPG password.*
resultinfo	*String*	*How to get the result* *ftp://login:pass@you_ftp.com/path_to/your_file.xml* *http://login:pass@you_ftp.com/path_to/your_file.xml* *Other methods are available but not no SAAS, contact Vecsys for additional information*

Result:

The result returned by this function is the standard result format plus additional information, the process identifier (JobId). **The Job id is a string with digits, dots and characters.**

The result of the transcription start process is displayed in the jobid field.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE result SYSTEM "MsfResults.dtd">
<result>
<status>Error Code</status>
<message>Error Message</message>
<jobid>xxxx.yyyyy</jobid>
</result>
```

6. Start Alignment

Allow a user to start the alignment process of an audio file.

The function name is:

alignment

Arguments:

filename	string	*Url of the* audio *f*ile to process. Just file and extension.
type	string Values : cts bn	Specify the type of audio file. CTS: Telephonic conversation. *(if component present)* BN: Broadcast. *(if component present)*
language	*String* *Values :* *fre* *eng* *ara* *chi* *spa* *rus*	*Specify the language of the audio file.* *Mandatory with BN audio file.* *If not specified MS will automatically identify the language (if optional component present).*
textfiletoalign	String	Url of the file containing the text to align
alignmode	Int	0 : special mode 1 : normal mode *(if component present)*

Result:

The result returned by this function is the standard result format plus an additional information, the process identifier (JobId). **The Job id is a string with digits, dots and characters.**

The result of the transcription start process is displayed in the jobid field.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE result SYSTEM "MsfResults.dtd">
<result>
<status>Error Code</status>
<message>Error Message</message>
<jobid>xxxx.yyyyy</jobid>
</result>
```

7. Process status

Allow a user to retrieve the status of its running processes (transcription, language identification, alignment or audio partitioning) .
The user can retrieve some detailed information regarding its process, or list all its processes running on MS.

The function name is:

status

Arguments:

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

Result:

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 data may be added in the future.

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

Examples:

After asking for the status of a jobs, MS will display a screen like:

```
<?xml                          version="1.0"                          encoding="UTF-8"?>
<!DOCTYPE                result                SYSTEM                 "MsfResults.dtd">
<result>
<status>0</status>
<message>success</message>
<joblists>
<job>
<jobident>3214.msf01</jobident>
<processtype>Transcription</processtype>
<jobstatus>E</jobstatus>
<startdate>2009-09-29                                                10:47:02</startdate>
<stopdate>2009-09-29                                                 10:49:24</stopdate>
</job>
</joblists>
</result>
```

8. Process results

Allow a user to retrieve the result of a completed process.

Depending on which system is used there a slight difference in the DTD. Ask Vecsys to be advised which system you are using.

The function name is:

result

Arguments:

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

The result is some XML formatted data which contains the result of the process (Transcription, language identification or audio partitioning).

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.

System 1 :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE result SYSTEM "MsfResults.dtd">
<result>
<status>0</status>
<message>success</message>
<resultxml>
<![CDATA[
<?xml version="1.0" encoding="UTF-8"?>
<AudioDoc name="Obama.wav" path="/Audio/1/Obama.wav.wav">
<ProcList>
<Proc name="vrbs_part" version="2.1"/>
<Proc name="vrbs_trans.fre" version="3.3"/>
</ProcList>
<ChannelList>
<Channel num="1" sigdur="94.68" spdur="89.78" nw="287" tconf="0.75"/>
</ChannelList>
<SpeakerList>
<Speaker ch="1" dur="89.78" gender="1" spkid="MS1" lang="fre" lconf="1.00" nw="287" tconf="0.75"/>
</SpeakerList>
<SegmentList>
<SpeechSegment ch="1" sconf="1.00" stime="2.04" etime="91.82" spkid="MS1" lang="fre" lconf="1.00" trs="1">
<Word stime="2.04" dur="0.73" conf="0.931"> Washington </Word>
<Word stime="3.22" dur="0.15" conf="0.177"> quand </Word>
<Word stime="3.46" dur="0.10" conf="0.601"> de </Word>
<Word stime="91.28" dur="0.27" conf="0.930"> cette </Word>
</SpeechSegment>
</SegmentList>
```

```
</AudioDoc>
]]>
</resultxml>
</result>
```

System 2 :

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE result SYSTEM "MsfResults.dtd">
<result>
<status>0</status>
<message>success</message>
<resultxml>
<![CDATA[
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE AudioDoc SYSTEM "Resultvecsys.dtd">
<AudioDoc name=" 01116podcast.mp3" path="01116podcast.mp3">
<SegmentList>
<SpeechSegment channel="1" stime="0.000" dur="1.320" score="1.00" spkrid="SPK00001" gender="male">
<Word stime="1.260" dur="0.040"> un </Word>
</SpeechSegment>
<SpeechSegment channel="1" stime="23.610" dur="14.050" score="1.00" spkrid="SPK00002" gender="male">
<Word stime="23.610" dur="0.080"> En </Word>
<Word stime="23.690" dur="0.180"> juin </Word>
</SpeechSegment>
</SegmentList>
</AudioDoc>
]]>
</resultxml>
</result>
```

The DTDs is defined in the Annex chapter.

9. Process results (Wait for completion)

Allow a user to retrieve the result of a completed process, the function waits until the result completion or the connections end up in timeout.

The function name is:

waitandgetresult

Arguments:

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

See the result function for more detail.

10. Logout

Destroy the user's session created during the login.

The function name is:

logout

No argument is needed.

11. Delete an uploaded file

Allow a user to delete an uploaded audio file and its transcriptions.

The function name is:

delete

Arguments:

filename	string	Filename of the file to upload. OR
jobid		A job identifier

If a jobid is specified, MS delete the audio file and the result file. If the filename is specified only the audio file is deleted.

12. Cancel a running transcription

Allow a user to stop a running processes (or language identification or audio partitioning) .

The function name is:

cancel

Arguments:

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

13. Example of use

On the Web site https://mediaspeech.com/webservice or http://cluster1.mediaspeech.com/webservice there example of the API. The examples are available in C#, PHP and Perl.

The following client is designed to connect to the MS Web service. It assume that you have PHP with SOAP enabled, on the computer running the script.

```
<?php

function DisplayResult($fct, $res)
{
   echo '--------------------------------<br>';
   echo '<H2>Command: <b>'.$fct." </b></H2><br>";
   if (isset($res))
   {
            foreach ($res as $line)
      {
         echo htmlentities($line)."<br>";
      }
   }
}

                                                                                    ...

   ini_set('soap.wsdl_cache_enabled', 0);

   try
   {
            // create a new SOAP client
            $client = new SoapClient(
      'https://<SERVER_ADDRESS>/webservice/server.wsdl',
                     array(
                                          'trace'=> 1,
                                          'soap_version'=> SOAP_1_1
                                 )
            );

      $retour_ws =  $client -> __call('login',array('username'=>'your_login', 'password'=>'your_pass'));
      DisplayResult('login', $retour_ws);

      //-----------------------------------
      if (0) // set to 1 if you want to use
      {
         $filename="/tmp/aa.wav"; //full path on your computer
         $f = file_get_contents($filename);

         $retour_ws =  $client -> __call('upload',array('filename'=> $filename,
                              'fileContent' => base64_encode($f)));
         DisplayResult('upload', $retour_ws);
      }

      //-----------------------------------
      if (0)
      {
         $retour_ws =  $client -> __call('result',array('jobid'=>'1848.msf01'));
         DisplayResult('result', $retour_ws);
      }

      //-----------------------------------
      if (0)
      {

         $retour_ws =  $client -> __call('trans',array(
                  'filename'=>'PKBN_FRE_FR_20071112_2000_FRANCE2_NEWS.wav',
                  'type'=>'bn',
                  'language'=>'fre',
                  'channel'=>'0',
                  'quality'=>'0',
                  'duration'=>'0',
                  'callback'=>'http://10.0.0.3/dev/callback.php')
                  );
         DisplayResult('trans', $retour_ws);
      }

      //-----------------------------------
      if (0)
      {

         $retour_ws =  $client -> __call('part',array(
                  'filename'=>'PKBN_FRE_FR_20071112_2000_FRANCE2_NEWS.wav',
                  'type'=>'bn',
                  'channel'=>'0',
                  'speaker'=>'0')
                  );
```

```
        DisplayResult('part', $retour_ws);
    }

    //-------------------------------------
    if (1)
    {
      //$retour_ws =  $client -> __call('status',array('jobid'=>'1848.msf01'));
      $retour_ws =  $client -> __call('status',array('jobid'=>''));
      DisplayResult('status', $retour_ws);
    }

    //-------------------------------------
    if (0)
    {
      $retour_ws =  $client -> __call('cancel',array('jobid'=>'1873.msf01'));
      DisplayResult('cancel', $retour_ws);
    }

    //-------------------------------------
    if (0)
    {
      $retour_ws =  $client -> __call('result',array('jobid'=>'1870.msf01'));
      DisplayResult('result', $retour_ws);
    }

    //-------------------------------------
    if (0)
    {
      $retour_ws =  $client -> __call('delete',array('filename'=>'aa.wav'));
      DisplayResult('delete', $retour_ws);
    }

    //-------------------------------------
    $retour_ws =  $client -> __call('logout',array());
    DisplayResult('logout', $retour_ws);

    //-------------------------------------
/*              // display soap request
                echo '<br />Requete SOAP: '.htmlspecialchars($client->__getLastRequest()).'<br />';
                echo '<br />Reponse SOAP: '.htmlspecialchars($client->__getLastResponse()).'<br />';

    // Affichage du rÃ©sultat
    print_r($retour_ws);*/
  }
  catch (Exception $e)
  {
    //TODO  exception handling
    echo $e;
  }

                                                                    ...

?>
```

14. Annex

14.1. Mediaspeech team contact

Phone : +33 1 69 29 87 87
Email : support@mediaspeech.com

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

All not uploaded results will be deleted after 1 week.

14.3. What about the language identification?

Letting the software determine the language of the audio file takes time. If you already know the language of your audio file you can save time by selecting the correct language when submitting your job.
With a BN file, it is mandatory to specify the language.

14.4. Where should I connect to ?

https://mediaspeech.com/webservice
or
http://cluster1.mediaspeech.com/webservice

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

14.5. What about security?

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

14.6. 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.

14.7. Common errors

The following messages are common to all commands.

Code	Message
1000	Error
1001	invalid login
1002	login needed
1003	file not found, upload failed
1004	null file size, check your local file access rights
1005	invalid file extension
1006	invalid file name
1007	file not copied into MS content directory
1008	logout failed
1009	file not deleted
1010	Unknown Job Id
1011	No upload done
1012	Function submitted in *function* not implemented
1013	Invalid argument
1014	Cannot cancel the job
1015	Cannot delete the file
1016	Quota reached
1017	User already exists
1018	can't play the multimedia file
1019	can't decrypt file
1020	Maintenance in progress
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
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

14.8. 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
trs	1 when the result comes from an automatic transcription, otherwise 0
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)