



*Appliance hardware may vary depending on appliance version

Docsoft:AV Best Practices

Using Docsoft:AV for Automatic Transcription and Captioning of the Media Files

v 1.0

Overview

This document illustrates Docsoft's Best Practices for optimizing the performance of the Docsoft:AV appliance. This overview only represents the recommended procedures and does not illustrate all environments or software applications.

The primary function of the Docsoft:AV appliance is to create automatic transcripts with time stamps or time codes that provide closed captions for media files. The Docsoft:AV appliance is designed to monitor remote network folders for new media files and automatically processes them.

Preparing Speaker Profile

The Docsoft:AV appliance utilizes an automatic speech recognition engine to process the file. The engine is using generic speaker information (sounds and vocabularies) to match sounds to phonemes, and then through a process of statistical analysis creates words and sentences. Since not all people sound the same, the appliance may misunderstand some words within a media file; however, the accuracy of the system can be enhanced through speaker training. The first step in training a speaker is to first create a speaker profile. To create speaker profile:

1. Navigate to Speaker Profiles screen;
2. Click "Create New..." button;
3. Enter the speaker profile information:
 - a. Name;
 - b. Short description;
 - c. Select language based acoustic model;
 - d. Specify whether profile is public (all users of the appliance can use it);
 - e. Click "Next" button;
4. Enter first vocabulary information:
 - a. Name;
 - b. Select base vocabulary;
 - c. Click "Finish" button.

Now that the speaker profile is created, speaker training may begin. When training conducting speaker training please keep in mind that the higher the quality of training the better the training. We recommend that a professionally produced text transcript and with its accompanying uncompressed or unaltered media file will produce the best results. The process of training is quite simple as the appliance will train itself by listening to the media file and matching it to the text transcript. To conduct speaker training:

1. Navigate to the Speaker Profiles screen;
2. Click the “Train” link for the speaker profile;
3. Specify media files/text transcript pair(s):
 - a. Browse and select the exact transcript text file;
 - b. Browse and select the media files;
 - c. Click “Add Files” button;
4. After the files have been uploaded, click “finish”.

Vocabulary training in addition to speaker training will further enhance the speech recognition engine accuracy. This process will help the system understand unknown words and more importantly will analyze each individual’s writing and speaking style. To enhance speaker profile vocabulary:

1. Navigate to Speaker Profiles screen;
2. Click “Vocabularies” link for the speaker profile;
3. Click “Populate” link for the vocabulary;
4. Select type of the training document(s);
5. Specify the training document(s):
 - a. Browse and select the text file;
 - b. Click “Add” button;
6. Click “Finish” button.

Transcribing the Files

The appliance is designed to automatically monitor predetermined locations (repositories) and process files automatically according to user specified options. To simplify this process further, a standard speaker profile and digital caption output format can be specified by the administrator during the creation of the repository or can be left to be chosen by the end user. To create file repository:

1. Navigate to the File Repositories screen;
2. Specify media files source location and repository information:
 - a. Repository name;
 - b. Source URL and credentials;
 - c. Files filter or keep it empty;
 - d. Virtual URI (if media files playing using different protocol, e.g. MMS);
 - e. Specify whether exact transcript exists for media files;
 - f. Click “Next” button;
3. Specify output location:
 - a. Target URL and credentials;
 - b. Click “Next” button;
4. Specify output result file types and speech recognition options:
 - a. Select speaker profile;
 - b. Select vocabulary;
 - c. Select one or more output types;
 - d. Choose predefined transcript style options;
 - e. Click “Finish” button.

Please consult your network administrator for information about location, username and password to access remote files.

Choosing the Media File

It is really important to remember that there are hundreds of factors that may impact the quality of the audio or video as it relates to Automatic Speech to Text Recognition. The following is a list of things to do to enhance the audio quality. These techniques can improve the transcription accuracy of the Docsoft:AV. Providing good quality media files is essential for the speaker training and transcribing processes.

The Do's:

1. Use good quality recording equipment
2. Ensure the microphone is not blocked or touching articles of clothing
3. Ensure the microphone is close to the speaker and free from outside interference
4. Record using native audio capture settings without sound improvement filters
5. Ensure speaker pronounces all words clearly
6. Eliminate as much background noise as possible such as music, applause, or loud noises

The Do Not's:

1. Do not compress audio files
2. Do not re-encode audio file from original recording
3. Do not attempt to digitally alter audio file
4. Do not try to digitally enhance audio quality

Please keep in mind that the best files for audio mining are the files that were captured with CD quality (44,100 kHz, 16 bit) and were saved in lossless or low compression formats (e.g. MP3 with 128 kbps). Altering or up-converting low quality sound to high quality parameters does not improve speech recognition accuracy.