

Nuance Dragon SDK Server Edition

Speech-enable background transcription workflow.

Dragon SDK Server Edition (DSS) provides a set of tools and interfaces to help software developers successfully build and integrate automatic speech recognition into transcription workflow systems.

DSS is specifically designed for integration into solutions that require speech recognition to run in the background, without any interactive dictation or self-editing by users. Applications can include for example, court reporting and legal document transcription and other broadcast or business document transcription. Back-end speech recognition enables the batch processing of recorded speech from a range of devices, including digital recorders, smartphones, tablets, and telephone dictation systems.

Although DSS is not a transcription workflow itself, it can be integrated within existing systems to automate transcription, increasing transcription productivity and reducing transcription costs. Instead of manually transcribing the original dictation, the transcriptionist simply edits a draft transcription, correcting any errors in the text.

Speech-enable background transcription workflow

Enhanced Dragon runtime engine
Configured for back-end transcription tasks, this engine does

not require the presence of a sound card, enabling it to run on a server system.

Amazingly fast and accurate transcription

DSS comes with a collection of base acoustic models that deliver amazingly accurate recognition results for trained user profiles. Dragon supports an unlimited number of speakers with different voices and accents, by using personal profiles that store specific speaker information, and delivers accurate results the more it is used. DSS supports audio from multiple recorded dictation sources, such as dictation recorded over the telephone or audio files from digital recorders. DSS can determine the sound-to-noise quality of an audio file and disable utterance detection to increase accuracy for small audio files.

New in DSS is an update with the latest speech recognition engine and new acoustic models making it even more accurate than ever (up to 15% more accuracy than Dragon v12 “out of the box” and up to 26% over v11).

Custom vocabulary

DSS comes with a set of vocabularies, which include a word list with information about all the words a program can recognize and a language model that contains usage information about those

If you are a developer of documentation workflow systems, benefit by integrating DSS into your application to:

- Add value by enabling speech recognition and automatic transcription running in the background
- Keep your own interface look and feel and entire workflow, and create custom vocabulary
- Enable batch processing of recorded speech from a range of recording devices

What's new in DSS:

- Faster and more accurate, using latest speech engine
 - .m4a transcription file support
 - Canadian English support
 - Updated for Windows 10 support
-

words. DSS uses the vocabulary to recognize words correctly based not only on the sound of the words, but on their context. Developers can create custom vocabularies to tailor the nomenclature and words to specific industries, companies, departments or job roles. Legal vocabularies are available for US English, UK English, French and German, but medical vocabularies are not included. DSS supports English, including variants for US, UK, Australian, Southeast Asian, Indian and—new in DSS—Canadian, German, French, Italian, Spanish and Dutch.

Interface snapshot save and align

The DSS interface has the ability to align audio playback with the final, corrected transcript using timestamps provided by Dragon. In this use case, also known as “bouncing ball” playback, the app highlights each word in a transcript as it plays back the audio corresponding to the word that is highlighted – a feature of particular interest to content providers who already have both audio and text transcripts and just need the associated timestamps (i.e. audio book companies) if they would like to for example, provide a corresponding visual component.

Supported audio formats

DSS supports data formats from 8 kHz (for telephony) to 99 kHz in PCM and MS-ADPCM and IMA-ADPCM formats.

DSS supports the following audio file types in both mono and stereo: WAVEPCM, MS-ADPCM, IMA-ADPCM, a-law, mu-law, VOX, .mp3, .wav, .wma, dss, .ds2 (including encrypted files) and (new in DSC) .m4a.

Developer environments

Dragon SDK Server Edition (DSS) provides developers with the tools that are needed to integrate Dragon’s transcription capabilities into a transcription workflow application. Among these tools are the Dragon APIs used to access Dragon’s ActiveX controls when programming in the DSS environment. ActiveX is supported by many compiled and interpreted (script) languages such as C++, C#, Visual Basic, Python, Perl, etc. Developers can program in any language that ActiveX controls support, but the Dragon APIs are documented for Visual Basic and C++.

Comprehensive sample code and detailed documentation is available. In addition, the Dragon Developer Support Program offers:

- Direct answers to development questions from Dragon Developer Support engineers via Web form
- Access to Dragon SDK TechNotes and FAQs
- Access to minor release software updates as they become available for in-house testing and prototyping

System requirements

- RAM: Minimum 2 GB for 32-bit Windows 7, 8.1 & 10; 4 GB for 64-bit Windows 7, 8.1 & 10 and Windows Server 2008 R2 & 2012
- CPU: Intel® dual core or equivalent AMD processor. Faster processors yield faster performance.
- Free hard disk space: 8 GB if installed from the DVD, 16 GB if you copy the DVD to your hard drive and then install
- Supported operating systems: Windows 7, 8.1, 10 (32- and 64-bit); Windows Server 2008 R2 & 2012

- A DVD-ROM drive for installation (or Internet connection for product download)

Note: Does not support dictation into Electronic Medical Record (EMR) systems. For EMR support, please use Dragon Medical Practice Edition.

Application example: transcription workflow

Before back-end transcription can be implemented as part of a transcription workflow, a user and vocabulary must be created for each speaker, who should then be enrolled/trained within the system. To obtain optimal accuracy for each speaker/dictation-source combination, the application needs to adapt the factory-built acoustic models to specific characteristics of the individual speaker’s voice, accent and speaking style using speech recordings from the same dictation source (telephone, voice recorder, or microphone). DSS provides an “Enrollment from Transcription” capability (called EEnroll, for eyes-free enrollment) that uses actual dictation and the associated transcriptions for each speaker to adapt to an acoustic model. DSS also offers a Speaker Triage command-line utility that predicts the effectiveness of automatic speech recognition for an enrolled user using a set of recordings and corrected transcripts.

Step one: dictation

A speaker dictates into the digital dictation system, creating a .wav file. The transcription workflow component uses the metadata that is included with each .wav file to determine the identity of the speaker and retrieve the speaker’s user profile and the appropriate

vocabulary. The workflow component passes the .wav file, the user's profile, and the appropriate base speaker model and the base topic model to the transcription component. To further streamline workflow, recognition can also be performed from a streaming file, eliminating the need to produce an audio file.

Step two: transcribing the dictation

The transcription component transcribes the recorded text to .txt and .idx, where the .idx (index file) is the concordance containing recognized words and timestamps.

Step three: correcting the transcription

A transcriptionist compares the transcribed document with the original .wav file and fixes any errors. The transcription application's workflow component passes a

correction client all of the files it needs to update a draft transcript, including user information, .wav data, and an .idx file that contains timestamps for each utterance. After the transcriptionist has corrected the draft transcript, the speaker verifies the corrected document is accurate. The files, including the corrected transcript, are then passed back to the application's workflow component, where they are stored until it can be passed to the adaptation component.

Step four: adaptation

After the transcriptionist has corrected the draft transcript and the speaker has reviewed and accepted the corrected transcript, the corrected transcript can be used to update the speaker's user profile. Updating the speaker's profile increases Dragon's accuracy for that speaker's subsequent transcripts.

About Nuance Communications, Inc.

Nuance Communications is reinventing the relationship between people and technology. Through its voice and language offerings, the company is creating a more human conversation with the many devices, electronics, apps and services around us. Every day, millions of people and thousands of businesses experience Nuance through intelligent systems that can listen, understand, learn and adapt to your life and your work. For more information, please visit www.nuance.com.

