Loquendo TTS Director: The next generation prompt-authoring suite for creating, editing and checking prompts

Davide Bonardo with the collaboration of Simon Parr

1. Overview

The release of **Loquendo TTS SDK 7.4** has made available a series of innovative new features to prompt designers, foremost of which is a landmark release of TTS Director – the complete authoring suite for writing prompts, now remodelled and upgraded.

TTS Director enables application designers to create their prompts both simply and rapidly, and to get the very best results from Loquendo speech synthesis technology.

Following the painstaking work of our systems engineers, significant advances have been made to the Loquendo TTS speech engine, while, thanks to an extension of the JAVA API set, our software developers have successfully ported these advances to the more user-friendly interface of our prompt-authoring suite. All this means that the latest version of TTS Director is rich with key new features, the most significant of which will be described below. In essence, these new features offer users the following benefits:

- Enhanced speech fluency and clarity.
- Increased choice and flexibility – different reading styles, selection of audio sampling rates, unit selection, etc.
- Simplified interface – streamlined menu, word-wrap, colour-coded proprietary tags, highlighted phoneme keys for each language, etc.
- Simplified SSML creation – colour-coded SSML tags, syntax error check, etc.

In the following pages we will look at each of the new features in turn, offering a brief description of the benefits offered and enabling you to get the absolute best out of TTS Director.

2. Loquendo TTS Director: the new features

The **User-Driven Unit Selection Tool** gives you more control over how Loquendo TTS constructs phrases from the speech database. If the reading of a prompt by the text to speech engine is not exactly as you want it in terms of intonation, emphasis or pronunciation, simply launch the User-Driven Unit Selection (UDUS) tool within TTS Director: the UDUS allows you to highlight the relevant word or words, and the speech engine then re-generates that part of the prompt, offering you an alternative acoustic realization. This process can be repeated many times over, giving you plenty of alternatives to choose from to get just the result you are looking for.
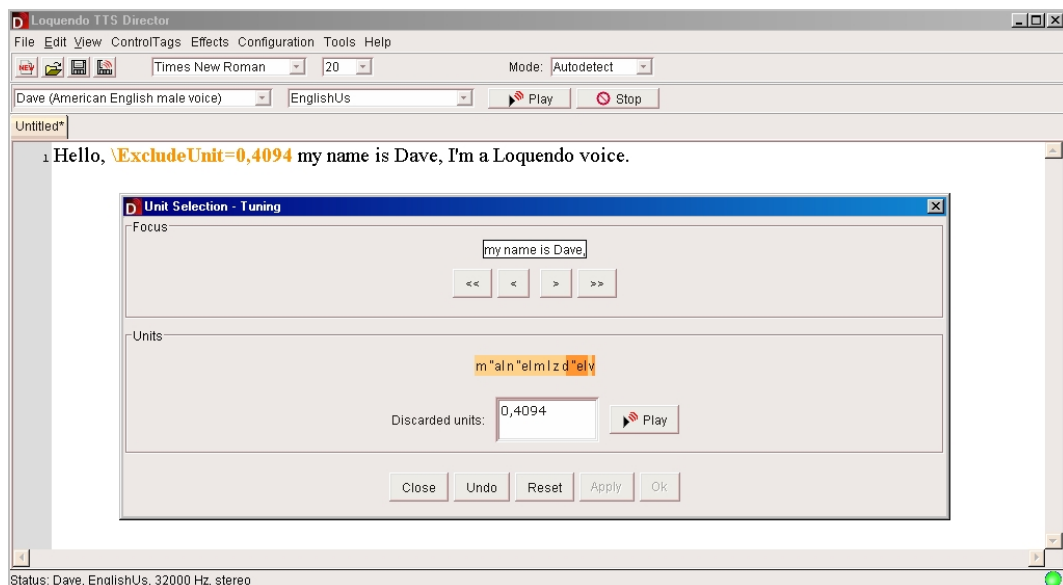

Figure 1: User-Driven Unit Selection Tool

1

As shown in Figure 1, the 'Focus' section enables you to move backwards and forwards through the text of your prompt, selecting the part you wish to work on - which will then be displayed phonetically in the 'Units' section. By clicking on the phonetic unit or units you wish to modify, the speech engine automatically offers you an alternative reading - which you can then listen to by simply clicking 'Play'. The process may be repeated many times, allowing you to choose the version which suits your application best. Clicking 'Apply' will select the unit selection currently displayed and insert it into your text, in the TTS Director editing window, with the appropriate control tag: **\ExcludeUnit=0,4094**

The inclusion of this tag ensures that the phrase in question will always be read in the way you have chosen. Note that the control tag is colour-coded to differentiate it from the text of the prompt, for ease of legibility.

By carrying out several User-Driven Unit Selections and by listening carefully to the results, you may find that the UDUS tool not only resolves the irregularity you were unhappy with, but also provides an alternative emphasis or intonation that better suits your application needs.

The **Reading Style** feature allows prompt designers to activate/deactivate predefined reading styles. *User*-defined reading styles can also be created, specifying Style in a configuration file (<style name>.ycf).

By selecting Reading Style from the Control Tags menu in TTS Director, with one simple click of the mouse, users can command the speech synthesis to read a phrase, or an entire text, in a pre-defined style by assigning values to a number of configuration parameters, e.g. to read all numbers as dates/time/phone numbers; to spell out each word; to insert pauses at every phrase boundary/only at punctuation; to read words in capitals letter by letter; to read SMS/email abbreviations according to the SMS pronunciation lexicon; to interpret/ignore secondary stress in SAMPA transcriptions, plus many more.

In this way, prompt designers can control a large number of parameters with one mouse click, activating or deactivating them as required. This means that long and complex prompts can be read exactly as they were intended, without having to manually adjust each parameter individually for each new prompt or set of prompts.

TTS Director can be shipped with a number of pre-loaded, pre-defined Reading Styles, but VUI designers can also create their own user-defined reading styles, which can then be activated and deactivated as required in just the same way.

During text-to-speech conversion, Reading Style, just like Voice and Language, can be loaded and set on the Reader by means of APIs (ttsNewStyle and ttsSetStyle) or escape sequence (\style=<style name>). If it is set via the escape sequence, the change applies only to the single prompt.

Examples of instructions that can be contained in a Reading Style are: spell out each word; insert pauses at every phrase boundary/only at punctuation; pronounce all numbers as ordinal numbers, etc., expressed in the following syntax:

"SpellingLevel"="spelling "
"ProsodicPauses"="Automatic"
"ProsodicPauses="Punctuation"
"DefaultNumberType"="MasculineOrdinal"

A Reading Style can also be configured to apply to lexicons: User Pronunciation Lexicons can be combined with the other speech synthesis parameters described above, and included together in a user-defined Reading Style to be loaded and unloaded as and when required. Users can thus adjust a series of parameters *and* activate/deactivate lexicons with one mouse click. (To learn more about the creation of Lexicons, see section 6 on the Lexicon Manager).

The Reading Style feature gives users full control over a wide range of TTS reading settings, precisely defining the interpretation of texts by Loquendo TTS, simplifying the creation of lengthy prompts and guaranteeing their correct pronunciation every time.

Simplifying SSML and control tag use. For prompt designers using **SSML**, the new-look TTS Director has just made their life much easier: SSML tags are now colour-coded, so that even with long sections of mark-up language, SSML commands can be differentiated from prompt texts at a glance, making the editing and refining of prompts both simpler and faster. Completed SSML can also be checked for errors, using the SSML Syntax Check, greatly speeding up the trouble shooting process.
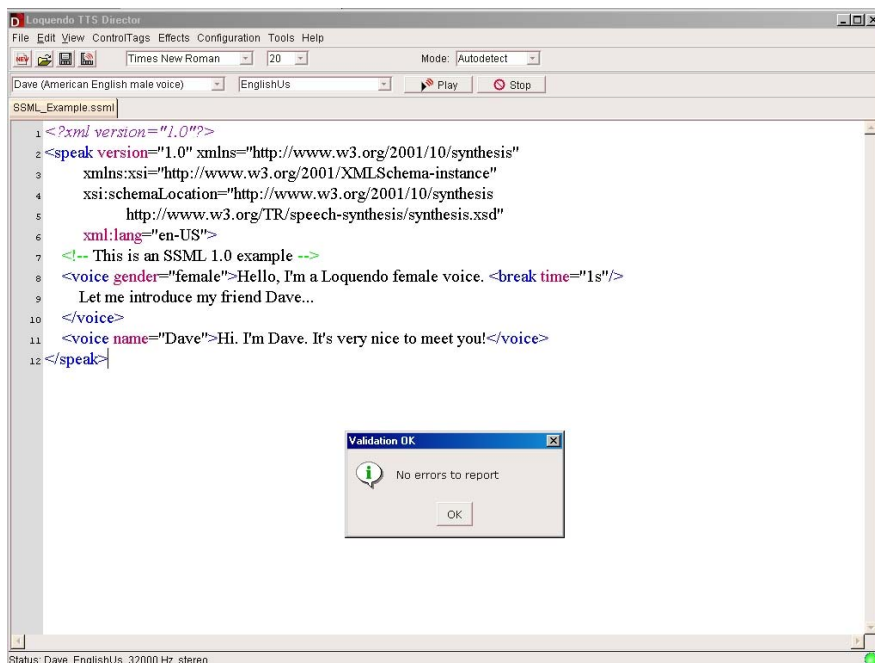
Figure 2: SSML Highlighting and SSML Validation

Similarly, the Loquendo control tags can also be highlighted; this distinguishes between commands and text and makes the reading of text easier.

The **Graphic Equalizer** provided with the new version of TTS Director has been enhanced and expanded to give users more control over their audio output: 30 bands of EQ (1/3 octave bands from 30 Hz to 25 kHz), with the option to cut or boost the gain of each frequency by up to 18dB. Several preset EQ settings are available, e.g. low fidelity, voice presence, allowing users to select a predefined frequency curve with one mouse click. A 10 band EQ can also be selected if required, for less demanding audio set-ups.

The Graphic Equalizer is easily accessed from the audio menu: *Configuration/Graphic Equalizer*
and appears as an on-screen, virtual EQ unit. Adjustments can be made simply and quickly by selecting individual frequencies with the mouse and dragging the faders up and down as required, to cut or boost. Click on Apply to hear your changes, or Reset to cancel them.
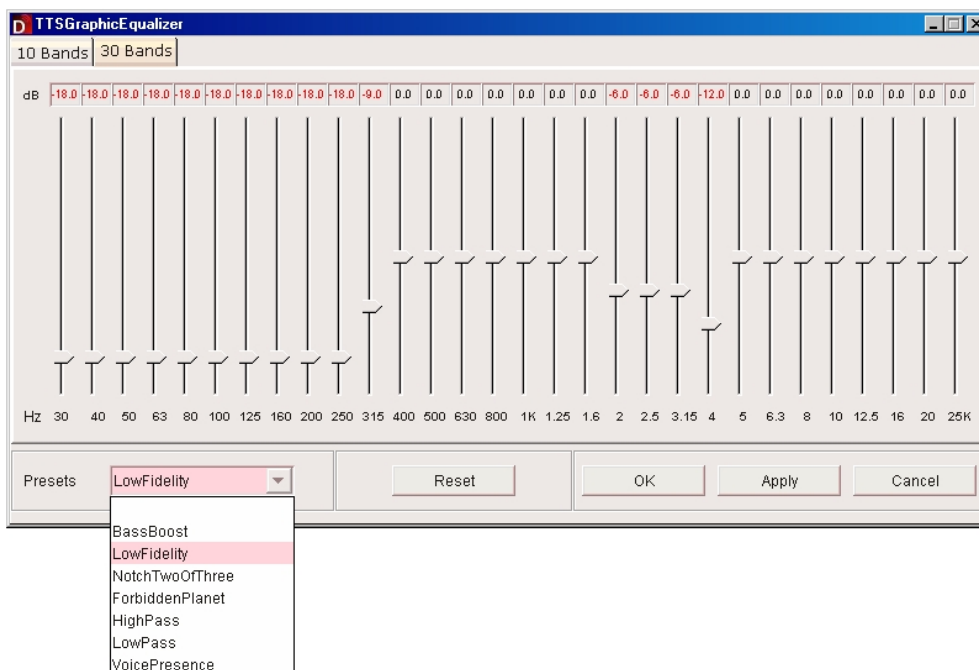

Figure 3: Graphic Equalizer

3

The **Voice Flavours** feature allows prompt designers to listen back to the prompts they have created exactly as their customers will hear them in terms of audio compression, sampling rate and frequency range. This means that if the prompts you are working on are destined for use in automotive applications, for example, then selecting the associated Voice Flavour will allow you to hear them back just as your customers will hear them when using their navigation device. If your prompts are destined for use in an automated call centre, then you should listen to them using the default Voice Flavour, telephony.

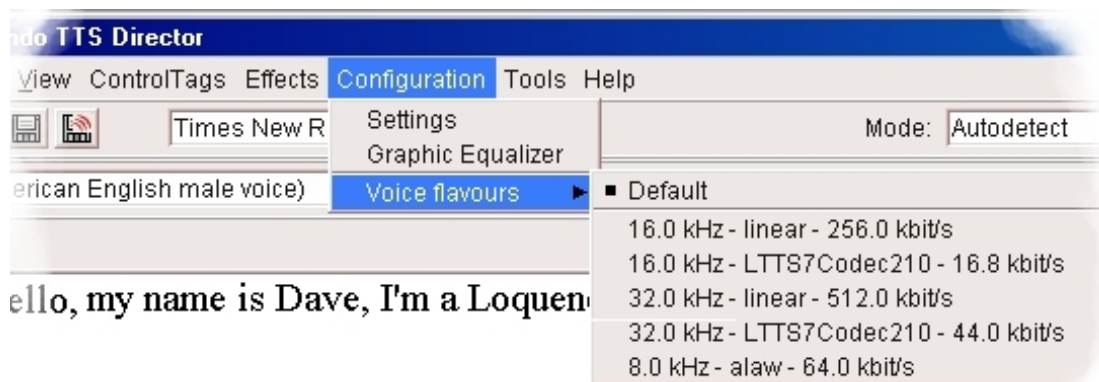This feature is easily accessed by clicking on: *Configuration/Voice Flavours*

Figure 4: Voice flavours

where a drop-down menu will appear - simply select the audio setting you require, then press Play to hear your prompts just as the end-user will hear them. The Voice Flavour feature thus enables prompt creators to adapt their messages perfectly to the device for which they are being designed.

Lexicon Manager. Similar streamlining has been carried out on the **Lexicon Manager**, the tool found within TTS Director for creating User Pronunciation Lexicons files.

With the Lexicon Manager a user can create a **User Pronunciation Lexicon -** essentially a list of literal or phonetic transcriptions - in order to define the TTS pronunciation of foreign language words, place names, acronyms, abbreviations, etc. Each Pronunciation Lexicon can be activated/deactivated as required.
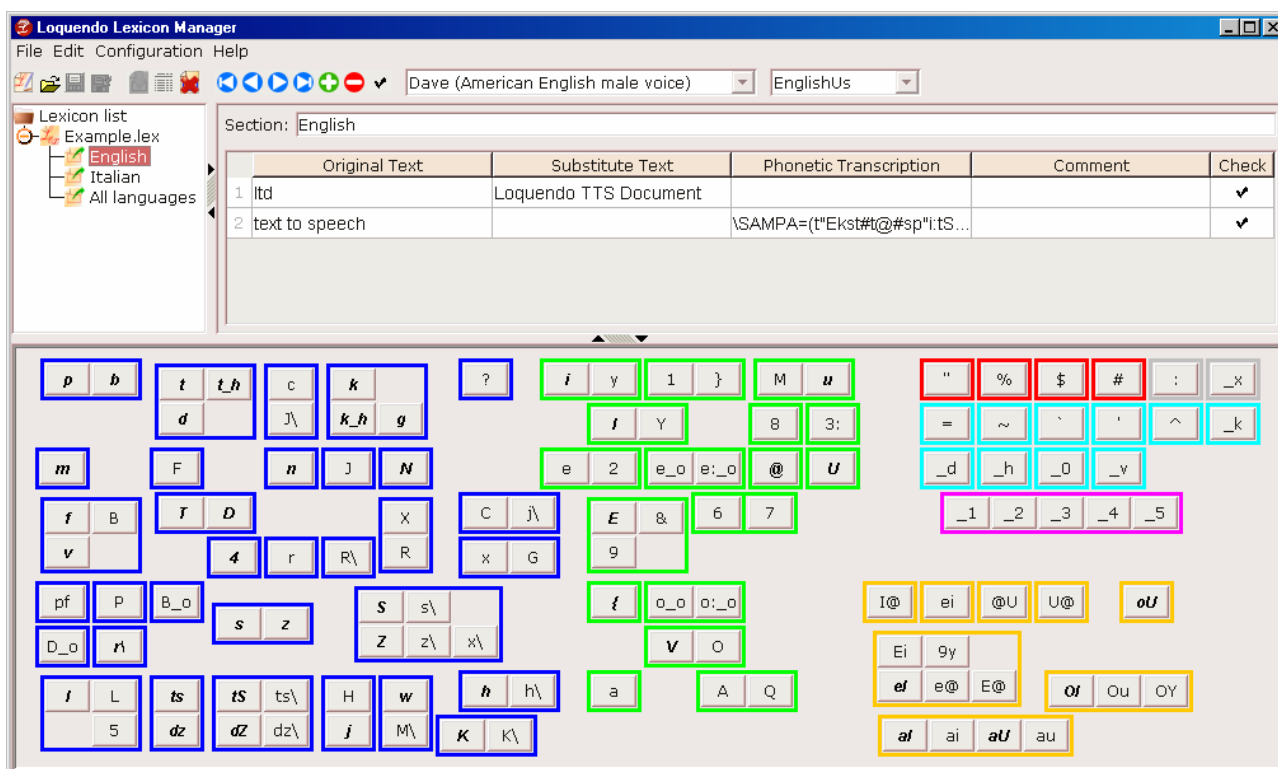
4

Figure 5: Lexicon Manager with an example of a lexicon

It is also possible to create (with a standard editor) a User Pronunciation Lexicon which contains a set of context-dependent pronunciation rules, e.g. "a single 'm' following a number is read as 'meters'", expressed in the standard syntax of regular expressions.

With the new version of TTS Director, Lexicon entries can now be checked for errors using the **Syntax Check**, which identifies any incorrect formats, phonetic transcriptions, etc. and displays an error message.
When creating or editing a user lexicon, the **Find** feature allows you to make a quick and easy search for entries that need changing or tags that need modifying, to easily handle lexicons containing many entries.
Additionally, when writing phonetic transcriptions, the virtual keyboard in the Lexicon Manager now highlights the phonemes which belong to the language you are working with; the non-highlighted phonemes can still be used, but they will be automatically mapped into the most similar phoneme native to that language.

Client-Server Architecture. In this new version, Loquendo TTS Director has been developed with an internal client-server architecture: users can enjoy complete control over their Loquendo TTS technology by accessing TTS Director remotely from one or more other PCs. This enables configuration changes (installations of new voices and languages, or TTS engine updates) to be made on a single server to be immediately available for a variety of sites and by different application developers.
It is also possible to run Loquendo TTS Director in stand-alone mode without having to worry about the internal architecture on a single PC.
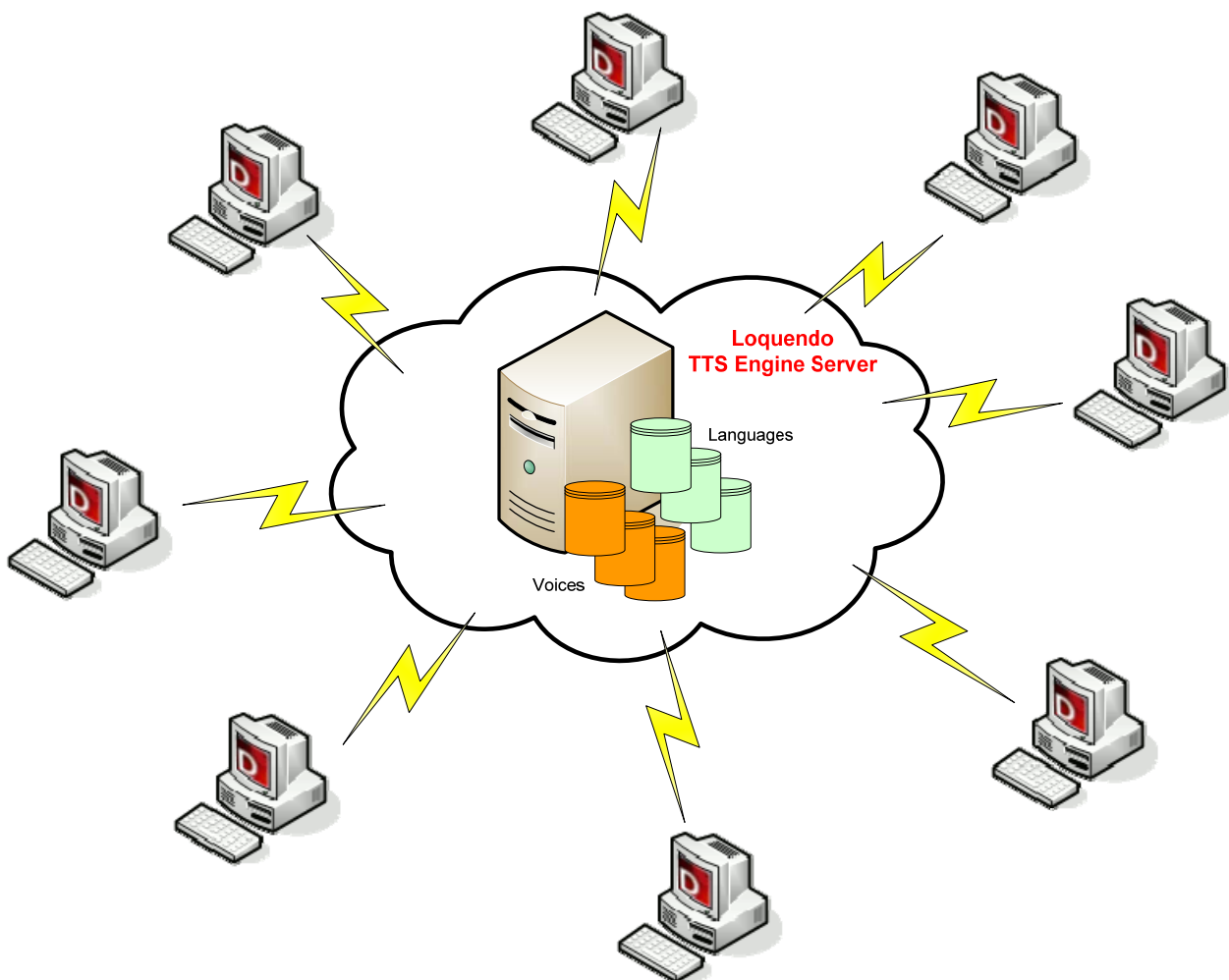
5

Figure 6: TTS Director client-server architecture

The Loquendo TTS Engine Server and Loquendo TTS Director are available both for Linux and Windows operating systems.

3. Loquendo TTS Director: a Summary of the major benefits

We hope you have found this article helpful, in which we have outlined the numerous and innovative features which Loquendo TTS Director now has to offer. We have also explained how to get started accessing these features, getting the best out of them in order to improve the reading of prompts.

This new version of TTS Director enables Loquendo's clients to keep pace with the enormous leaps forward made by Loquendo in delivering expressive speech and reading accuracy; indeed, with TTS Director prompt-designers are now able to create messages that are more fluent and natural sounding than ever possible in the past, but it also provides incredible flexibility, giving users control over a huge range of parameters for the reading of texts, ensuring pronunciation and reading settings are precisely as required.

The Author

Davide Bonardo joined the Loquendo team at the company's foundation in 2001. Within Loquendo's Technology Engineering Group, he is responsible for R&D on Text-to-Speech Synthesis interfaces, embedded TTS, development of multi-platform applications and tools, and for ensuring the compliancy of TTS with standard interfaces such as Speech Synthesis Markup Language (SSML) and Pronunciation Lexicon Specification (PLS).

6