

# LOQUENDO TTS DIRECTOR TUTORIAL

Davide Bonardo  
Morena Danieli

## Welcome to the Loquendo TTS Director Tutorial!

The main aim of this document is to introduce the Tutorial on Loquendo TTS Director. The tutorial will be available with the next few editions of the Loquendo Newsletter, and also on the Loquendo Web site. The outline of the Loquendo TTS Director Tutorial is as follows (in this document we will cover the Introduction):

### Introduction

- Basic Requirements
- A brief overview of Loquendo TTS
  - Functionalities and controls
  - New features of Loquendo TTS Version 7
- Why Loquendo TTS Director?

### TTS Director in action

- Environment configuration
- Basic functionalities

### Control tags, inline commands

- Mixer capabilities
- Multilanguage capabilities
- Advanced prompt creation: pronunciation, prosody and reading settings

### Expressive cues

### Creating new voice personas

### Samples

## Introduction

Loquendo TTS Director (henceforth, the TTS Director) is an authoring tool, used to assist in the creation of voice output. It is designed both for the users of Loquendo speech synthesis, and for speech application developers. More generally, the TTS Director is intended to provide an environment where it is possible to approach the problem of creating rich input for text-to-speech systems. The fact you are able to annotate raw text input and control speech output parameters is of vital importance for exploiting all the potential linguistic and paralinguistic features of speech synthesis. Annotated input allows greater control over the acoustic, linguistic, and stylistic interpretation of messages. The question of input annotation can be approached at different levels: on the one hand, the speech community can now refer to standards for

annotation, such as VoiceXML based SSML. On the other hand, the producers of speech synthesis who are conscious of the issues regarding input annotation, can offer supported environments where more refined annotation can be carried out.

The TTS Director tutorial will demonstrate how to annotate texts appropriately for Loquendo TTS. Before introducing the TTS Director, we will specify the minimum system requirements needed, and we will briefly mention the main new functionalities of the upcoming Loquendo TTS version 7.

## ***Basic requirements***

The TTS Director is a multi-platform tool, written in Java. It is available for the following operating systems: Microsoft Windows 9x, ME, NT, 2000, XP, 2003 Server, Sun Microsystem, and Linux. It is compatible with Java version 1.4.2\_06, or higher (the appropriate Java version has to be present on the system in use).

When the Loquendo TTS Director setup procedure has been completed, the Loquendo TTS SDK will be available in the menu *Start → Loquendo → Loquendo TTS → TTS Director*.

## ***A brief overview of Loquendo TTS***

Loquendo TTS (henceforth LTTS) is a speech synthesis engine. It is a multilingual, multi-voice, text-to-speech system that presents high standards of acoustic quality and linguistic accuracy. Its synthesis algorithm is based on the unit selection concatenative technique.

LTTS is a software-only system: it is flexible, able to provide high quality reading performance, and, last but not least, it is characterized by efficient algorithms. Thanks to these features, the LTTS software is available for several different operation systems: Windows (from 9x to Server 2003), Sun Solaris, Linux, Pocket PC 2002/2003, CE.NET, SmartPhone 2003, WindRiver VxWorks, Windows XP Embedded and Tablet PC Edition, and SymbianOS Series 60.

The input text can be ANSI, UTF-8 or UTF-16. The output can be listened to directly on the audio board or saved in a file, or it can be sent to a suitable audio destination.

Several different levels of API and of interfaces are available: API C/C++ e Java (in the near future, also C#), ActiveX controls, SAPI 4, and SAPI 5.

## **Functionalities and Controls**

LTTS provides various controls that can be used by developers to make full use of the acoustic and linguistic features of the speech synthesis system.

The controls allow:

- the selection of the output voice,
- the selection of the language (with or without the 'Language Guesser')

- the selection of the type of prosodic analysis (for example, sentence reading, isolated word reading, spelling, interpretation of the text layout, and interpretation of the titles),
- the specification of word pronunciation (including numbers, phonetic input, specializations of the lexicon, i.e. the User Lexicon)
- the variation of prosodic parameters via the modification of pitch, speaking rate, volume, duration, emphasis, pause duration
- the mixing of the voice with music (audio mixer)
- the choice of input type, including raw text or SSML markup language, coded UNICODE, ANSI.

The various controls can be activated by means of:

- Key registry or initialization file: the controls inserted via the latter can be adapted by using the appropriate API
- API
- annotated input text with LTTS control tags.

The scope of controls activated by the above listed options is varied. If the user needs to apply a synchronous control, s/he needs to use the escape-sequence mode of activation. Through control tags, normally of the type “key=value”, the escape sequences specify the appropriate action, such as a language change, the insertion of a pause, musical accompaniment to the text, or the activation of specialized lexica and plug-ins. Both plug-ins and specialized lexica are intended to better the LTTS performance in particular semantic domains, such as SMS reading, automotive environments, home banking, and so on.

A lexicon file, as well as a plug-in, can be considered a type of prepacked control, that is, sequences of instructions that are interpreted by the LTTS engine, and that are normally used for specifying particular pronunciation options for single words or expressions. The plug-ins are provided by Loquendo, but the lexica can be created by the user by using the tool provided by Loquendo.

We list below some of the more advanced, and usually highly valued, multilingual, multi-voice LTTS functionalities:

- “Phonetic mapping”, a proprietary algorithm that allows a given voice to speak any language available in LTTS, while keeping the original mother-tongue intonation and accent.
- The “Language Guesser”, along with its “Auto-guessing capabilities”, that allows LTTS to change voice and/or language automatically, by recognizing the original language of the written input.
- The “Audio-Mixer” allows the insertion, and synchronization with the text, of music and/or sounds; it also lets the user create particular effects (including separate volume controls for the music and the voice, fade-in and out, and so on...)
- The “Emotional Text-to-Speech” provides a rich repertoire of extra-linguistic yet meaningful sounds, such as laughs, coughs, hesitation sounds, crying and so on. The extra-linguistic sounds, inserted in the text, allow the creation of emotionally colored texts.

## New features of Loquendo TTS Version 7

Loquendo is continuously conducting research and development to improve its text-to-speech products. In particular, the forthcoming Version 7 of Loquendo TTS (LTTS 7) will include more efficient algorithms and new functionalities, all of which will contribute to placing LTTS 7 at the cutting edge of speech synthesis technology, while still maintaining its ease of integration, usability, and wide range of development tools.

The most interesting new features of LTTS 7 are briefly discussed below.

A **new concatenation algorithm** has been integrated into the LTTS 7 engine. With this feature, the most common speech concatenation irregularities, such as those produced by pitch and spectral discontinuities in voiced sounds, are fixed. In fact, some acoustic parameter functions are smoothed simultaneously without producing side effects. After unit selection, and before concatenation, these parameters are evaluated, and if their distances at the junction point exceed a critical threshold, a new set of values is provided that minimizes these discontinuities. This algorithm works pitch synchronously, and for each analysis frame a parametric spectral representation is considered. During concatenation, the modified spectral representation, together with the target pitch contour, are converted into speech waveforms in one step.

The **algorithm of prosodic phrasing**, once available only for the Italian language, in LTTS 7 will be provided for other natural languages (American and British English, French and German, among others). This rule-based algorithm, along with the new concatenation algorithm, improves the intelligibility and naturalness of the voice output, and provides the basis for improving the expressiveness of our speech synthesis system for moving it towards realizing truly expressive speech.

Some new functionalities for **handling the acoustic signal** will characterize LTTS 7. It will be possible to modify the voice timbre, to generate speech in stereo, to balance between right and left channels (this may be very pleasant, and useful for some families of applications), and, finally, to have output frequencies of variable values, so that the user can define them on the basis of her/his own needs.

In the new LTTS 7, the user will be able to **create her/his own voices** by using the voice databases provided with LTTS 7 and tuning the various parameters that characterize each voice. It will be possible, for example, to create a childlike Italian voice, with an English accent and intonation, that can read slowly, and can even correctly read emoticons and abbreviations as used by writers of SMSs. All these new voice functionalities may coexist, and they may be saved for use in other occasions and applications.

LTTS 7 will include **new APIs**: these will make the use of the text-to-speech system in many different environments easier and more effective, from server to embedded use.

Finally, the **internal architecture** of the speech synthesis system has been revised: the system is now completely modular, a feature that further improves the scalability of our text-to-speech.

## Why Loquendo TTS Director?

As the reader can imagine, LTTS makes available several different options and functionalities. While this is usually well accepted by users, they may be concerned about the great many controls and commands necessary to activate all the functionalities. However, LTTS is a very versatile text-to-speech system, able to read generic texts with very high (*acoustic, prosodic and intonational*) quality, without the need for tuning.

Nevertheless, our mission is to provide a speech synthesis system that, besides a good reading quality, allows the creation of a pleasing acoustic result in any user context. We believe that modifying voices, varying reading style, and turning on/off all variations, should be easy for the user. Giving the user the opportunity to personalize the results of her/his voice application is, we deem, a challenging but useful choice, because s/he has all the linguistic and contextual knowledge needed for tuning the text-to-speech characteristics. That is why we propose a tool able to help both users who just want to try the Loquendo synthesis, and the expert service developer who must minimize the time spent in the creation of the annotated prompts.

TTS Director is the solution we propose because it is an easy-to-use text editor. It is possible to write a text and immediately listen to the audio output produced by the text-to-speech.

In the figure below, you will find a TTS Director example screenshot.

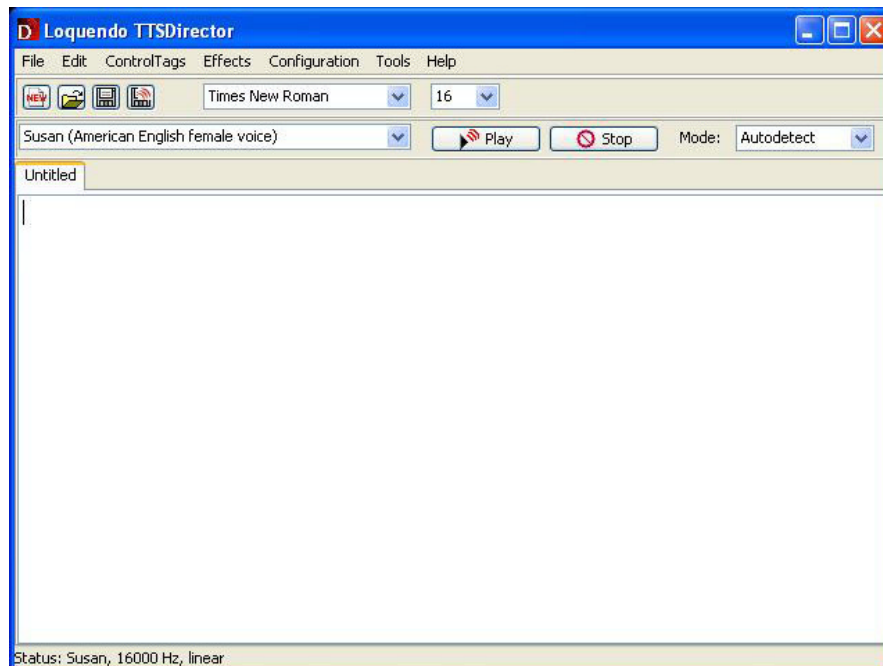


Figure 1: Opening screenshot of Loquendo TTS Director

In the central area is the edit-box; you can open several "Tabs" and manage several documents at the same time.

In the menu-box you can find the usual editor commands (to open a file, create a new file, save the document, etc...) and the special commands for the LTTS control (we will look at all these commands later).

The toolbar is divided into two parts: in the first there are three buttons for managing documents (to create a new document, open an existing one or save the current document), one button for saving the audio output as a file, and two combo boxes for choosing the font name and font size for the edit box (this doesn't affect the LTTS performance).

Just above the edit-box you can see the toolbar where an installed voice may be selected, the reading mode can be chosen, and the Play button can be clicked to start the TTS process.

In the status bar below the text area are the main characteristics of the voice selected: in this sample, the status is the voice Susan, at the frequency of 16 KHz with a linear coding.

By using the TTS Director you can easily:

- try out and demonstrate text-to-speech
- create audio-output wave files
- prepare texts for fixed prompts (useful for some types of services)
- study and test the best configurations for reading in a special context (i.e. creating a special lexica)
- make use of available control tags to control the TTS
- listen and insert into the text the "Effects", that is the expressive cues and the paralinguistic events
- create new voices, that will be immediately available on the system.

This is just a taster for what the Tutorial will be able to demonstrate. In the following issues of the Loquendo Newsletter, we will actually start to use this powerful tool and we will learn, with examples and explanations, the many secrets of Loquendo TTS, and how we can achieve the best results in any given context.