Checking Under the Hood of your ASR Engine

by Joanne Appleton for UX Magazine
http://uxmag.com/articles/checking-under-the-hood-of-your-asr-engine

If you put your foot on the accelerator and listen to your car rev before finally kicking into gear, you can usually tell when it's time to tune-up your engine. The same is true for an automated speech recognition (ASR) engine.

If calls into a speech application start sputtering their way into an interaction with a customer service rep because words become "unrecognizable," it is probably time to tune-up your ASR software.

Like the diagnostic tool that a mechanic plugs into your dashboard that tells them almost everything about your car's performance, a speech tuner is a powerful software tool that allows ASR users to evaluate their speech application and get feedback on how it is performing. There are other tuning tools available out there but this article focuses on the LumenVox Speech Tuner as an example of one possible tool with which to tune a speech application.

The Engines of Today

Speech recognition engines developed by today's speech technology leaders are designed to be robust enough to recognize natural language with a fair amount of tuning flexibility already prebuilt into the software. Yet, many of you reading this can undoubtedly say there has been at least one instance when you've given up trying to talk to a machine and said "operator" or pressed the zero key so many times the system gives up and routes you to customer service.

We then likely decide that, in future, it's best not to go through the automated attendant to contact that organization. It was too much trouble, and speaking to a live customer service representative seemed the only way to satisfy the original purpose of your call. You might even tell your friends about it so they don't get stuck in the system, too. To the organization, this equates to increased labor costs as more calls are routed to human attendants.

"In most cases, the fix can be simple," says Axel An, a senior core technology engineer in the speech industry who helped develop one of the first speech tuners nearly a decade ago. "Both businesses and customers could benefit from tuning. It can reduce the call length and the number of people handling calls. And that can translate directly to a monetary return."

From an efficiency standpoint, the car engine and the speech engine are the same; a regularly tuned-up car will get good gas mileage and release good emissions. A regularly tuned-up speech engine will allow higher call volumes and less human work. Although the speech recognition industry estimates that about 40-50% of total development and deployment time should be spent on the tuning process, speech tuners are greatly under-utilized.

What separates tuning from the normal process of testing an application is that tuning generally relies on actual production data collected from call logs. It provides a statistical method—rather than an anecdotal one—for evaluating the speech engine, making tuning easier and more precise.

The Tuner Interface

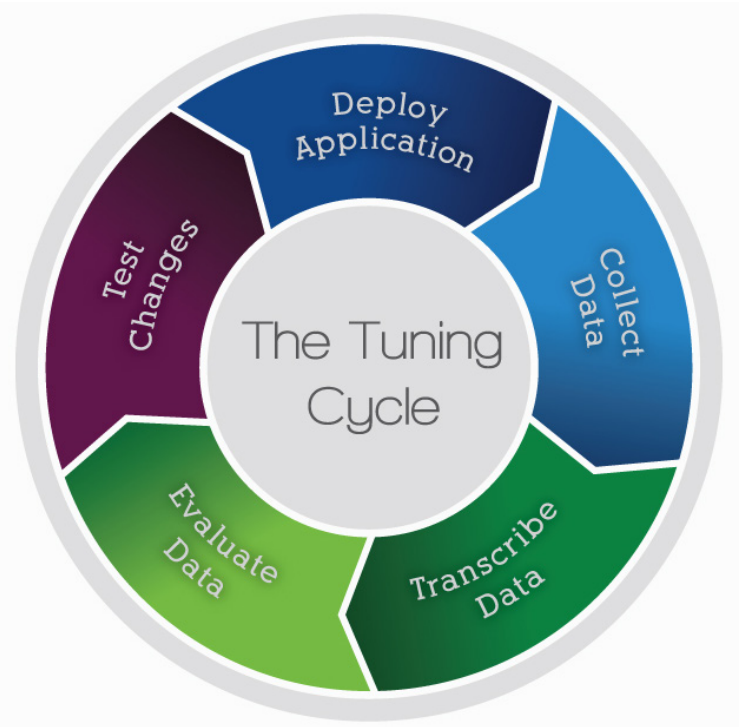
Our latest and greatest Speech Tuner has a user interface that is surprisingly easy to use. Six, colorful icons make up most of a tool bar to the left of the screen when the tuner is started up.

- The Summary page
- Call Browser
- Grammar Editor
- Transcriber
- Text-to-Speech
- Tester

The difficulty level of tuning a speech engine can range from rather simple to quite intricate, depending on performance indicators from the ASR application. For instance, it might be as basic as adjusting the "grammars" to include new utterances. A grammar is a vocabulary group comprised of the words or numbers that the ASR will expect callers to say, such as all the names of the employees in an organization. Tuning

can also be as involved as rewriting a prompt to sound clearer or, in some cases, redesigning the entire application to better guide callers through the flow of a call. In addition to grammar tuning, the tuner can perform transcriptions, instant parameter tuning, and version upgrade testing of any speech application.

The tuning process is essentially the same for simplistic or sophisticated improvements to a speech engine. The process is also cyclical, meaning after changes have been made to the application, the results are reviewed and more changes are made. Returning to the analogy of tuning a car, this is when the mechanic plugs the vehicle back into his diagnostic machine after changing the spark plugs and adjusting the timing to see if it is running better, or if he needs to make more adjustments. This type of feedback loop actually turns out to be an excellent way for ASR users and application developers to see the affects of tuning in real-time.



Here are the 5 basic steps to tuning an automated speech recognizer:

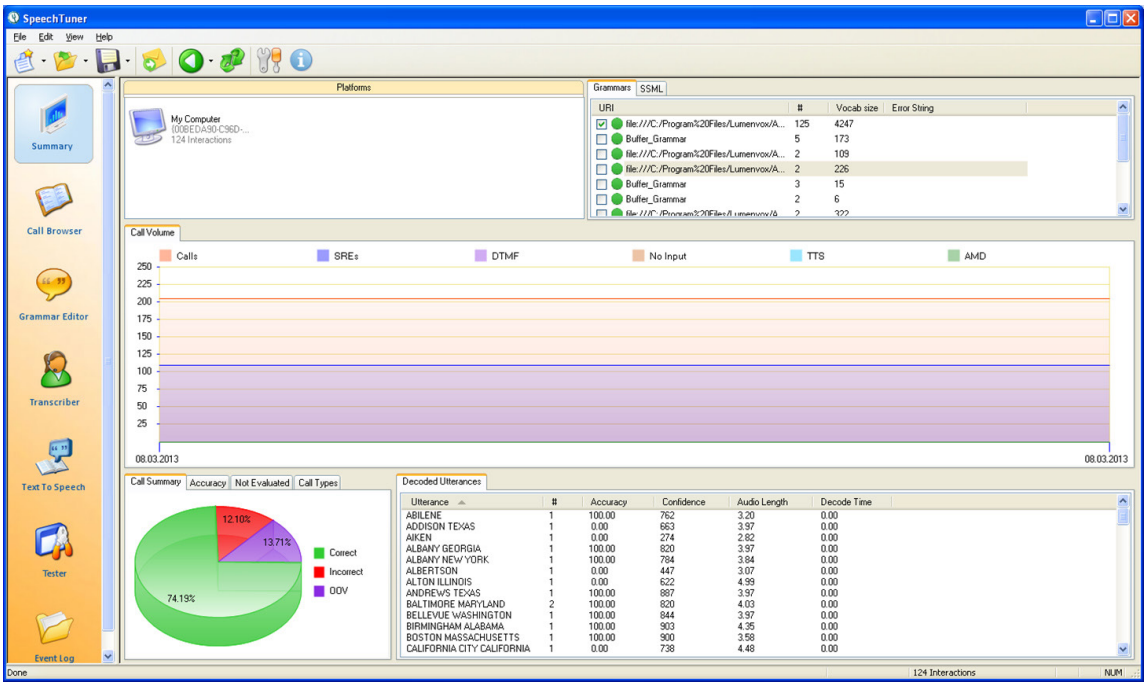
1. Deploy application
2. Collect the data
3. Transcribe the data
4. Evaluate the data
5. Test the changes (by deploying the application)

Deploy an Application

There are many places within an ASR application to make effective changes but in general grammars are the easiest and most effective place to start tuning. To begin using the LumenVox Speech Tuner you must first deploy your speech application so it can record live incoming calls and build up a dataset.

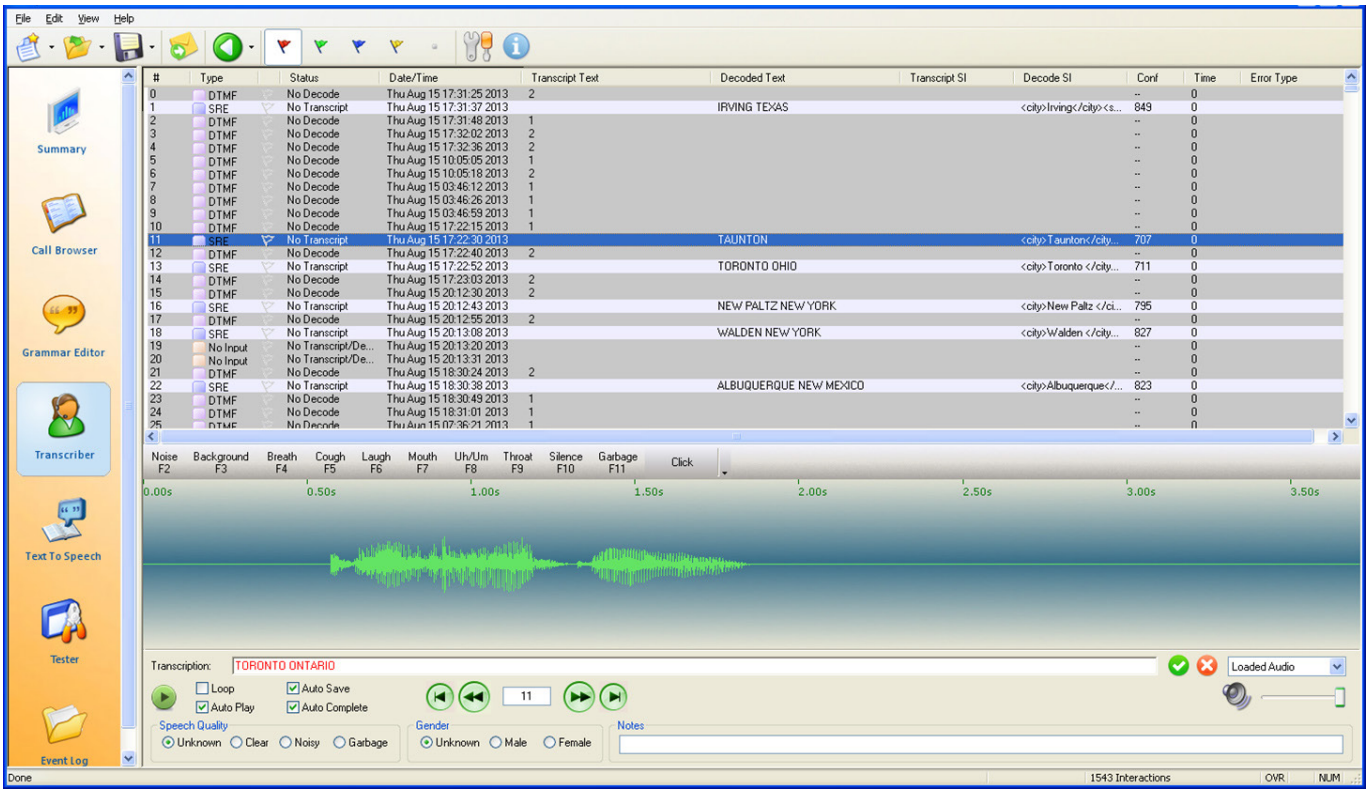
Collect the Data

The data is then imported into the Speech Tuner application in the form of a recording file, or utterance file that contains a list of each call into the speech application. Once the logs are loaded into the Speech Tuner they appear in the summary screen separated by grammar type. For example, one grammar might be a company's directory of employee names and extensions; another might be made up of city and state names. It depends on the type business for which the auto attendant is taking calls.



Transcribe the Data

Now that you have a dataset to work with, the next step in the tuning process is to transcribe the data. Start by selecting a grammar set from the list “Grammars” in the summary screen and click the Transcriber tab on the toolbar (pictured above). The list of calls will show in the top half of this screen and a waveform of the sound quality of each call along a time scale is shown in the section below. Select each call one by one, press play and listen to the audio.



#	Type	Status	Date/Time	Transcript Text	Decoded Text	Transcript SI	Decode SI	Conf	Time	Error Type
0	DTMF	No Decode	Thu Aug 15 17:31:25 2013	2				--	0	
1	SRE	No Transcript	Thu Aug 15 17:31:37 2013		IRVING TEXAS		<city>Irving</city><s...	849	0	
2	DTMF	No Decode	Thu Aug 15 17:31:48 2013	1				--	0	
3	DTMF	No Decode	Thu Aug 15 17:32:02 2013	2				--	0	
4	DTMF	No Decode	Thu Aug 15 17:32:36 2013	2				--	0	
5	DTMF	No Decode	Thu Aug 15 10:05:05 2013	1				--	0	
6	DTMF	No Decode	Thu Aug 15 10:05:18 2013	2				--	0	
7	DTMF	No Decode	Thu Aug 15 03:46:12 2013	1				--	0	
8	DTMF	No Decode	Thu Aug 15 03:46:26 2013	1				--	0	
9	DTMF	No Decode	Thu Aug 15 03:46:59 2013	1				--	0	
10	DTMF	No Decode	Thu Aug 15 17:22:15 2013	1				--	0	
11	SRE	No Transcript	Thu Aug 15 17:22:30 2013		TAUNTON		<city>Taunton</city...	707	0	
12	DTMF	No Decode	Thu Aug 15 17:22:40 2013	2				--	0	
13	SRE	No Transcript	Thu Aug 15 17:22:52 2013		TORONTO OHIO		<city>Toronto </city...	711	0	
14	DTMF	No Decode	Thu Aug 15 17:23:03 2013	2				--	0	
15	DTMF	No Decode	Thu Aug 15 20:12:30 2013	2				--	0	
16	SRE	No Transcript	Thu Aug 15 20:12:43 2013		NEW PALTZ NEW YORK		<city>New Paltz </ci...	795	0	
17	DTMF	No Decode	Thu Aug 15 20:12:55 2013	2				--	0	
18	SRE	No Transcript	Thu Aug 15 20:13:08 2013		WALDEN NEW YORK		<city>Walden </city...	827	0	
19	No Input	No Transcript/De...	Thu Aug 15 20:13:20 2013					--	0	
20	No Input	No Transcript/De...	Thu Aug 15 20:13:31 2013					--	0	
21	DTMF	No Decode	Thu Aug 15 18:30:24 2013	2				--	0	
22	SRE	No Transcript	Thu Aug 15 18:30:38 2013		ALBUQUERQUE NEW MEXICO		<city>Albuquerque</...	823	0	
23	DTMF	No Decode	Thu Aug 15 18:30:49 2013	1				--	0	
24	DTMF	No Decode	Thu Aug 15 18:31:01 2013	1				--	0	
25	DTMF	No Decode	Thu Aug 15 07:36:21 2013	1				--	0	

Transcription: TORONTO ONTARIO

Loaded Audio

☐ Loop ☑ Auto Save
☑ Auto Play ☑ Auto Complete

11

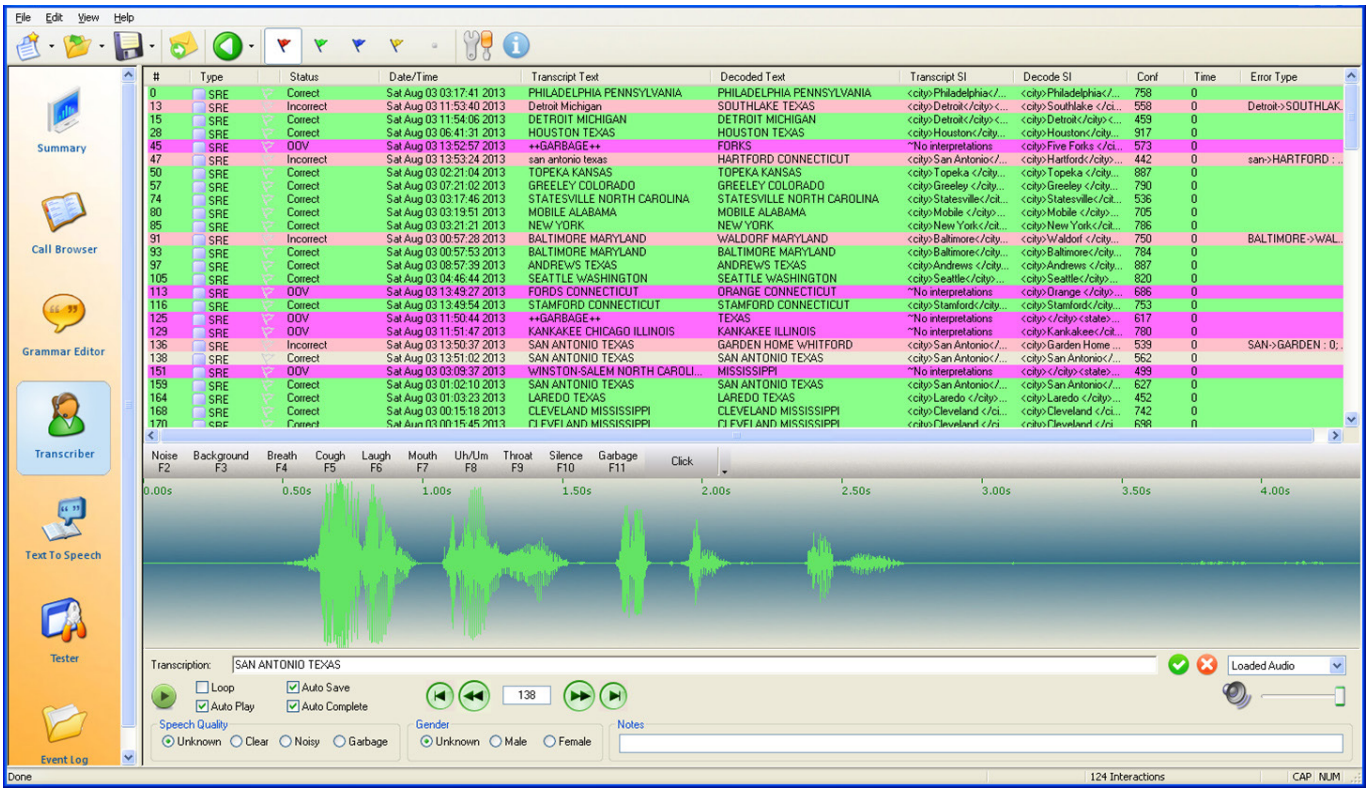
Speech Quality: ● Unknown ○ Clear ○ Noisy ○ Garbage

Gender: ● Unknown ○ Male ○ Female

Notes:

1543 Interactions

If what the caller says matches what the ASR engine interpreted them as saying, confirm it by selecting the green check mark next to the text box, or press enter to confirm and play the next call on the list. If the caller says something other than what was decoded, type in the word that the caller said and then press enter. Sometimes a caller’s response is not clear or is muddled by background noise. When this happens, select the red “X” mark or press escape to reject the call as “garbage” and play the next call.



#	Type	Status	Date/Time	Transcript Text	Decoded Text	Transcript SI	Decode SI	Conf	Time	Error Type
0	SRE	Correct	Sat Aug 03 03:17:41 2013	PHILADELPHIA PENNSYLVANIA	PHILADELPHIA PENNSYLVANIA	<city>Philadelphia</...	<city>Philadelphia</...	758	0	
13	SRE	Incorrect	Sat Aug 03 11:53:40 2013	Detroit Michigan	SOUTHLAKE TEXAS	<city>Detroit</city><...	<city>Southlake </ci...	558	0	Detroit->SOUTHLAK...
15	SRE	Correct	Sat Aug 03 11:54:06 2013	DETROIT MICHIGAN	DETROIT MICHIGAN	<city>Detroit</city><...	<city>Detroit</city><...	459	0	
28	SRE	Correct	Sat Aug 03 06:41:31 2013	HOUSTON TEXAS	HOUSTON TEXAS	<city>Houston</city...	<city>Houston</city...	917	0	
45	SRE	OOV	Sat Aug 03 13:52:57 2013	++GARBAGE++	FORKS	~No interpretations	<city>Five Forks </ci...	573	0	
47	SRE	Incorrect	Sat Aug 03 13:53:24 2013	san antonio texas	HARTFORD CONNECTICUT	<city>San Antonio</...	<city>Hartford</city>...	442	0	san->HARTFORD : ...
50	SRE	Correct	Sat Aug 03 02:21:04 2013	TOPEKA KANSAS	TOPEKA KANSAS	<city>Topeka </city...	<city>Topeka </city...	887	0	
57	SRE	Correct	Sat Aug 03 07:21:02 2013	GREELEY COLORADO	GREELEY COLORADO	<city>Greeley </city...	<city>Greeley </city...	790	0	
74	SRE	Correct	Sat Aug 03 03:17:46 2013	STATESVILLE NORTH CAROLINA	STATESVILLE NORTH CAROLINA	<city>Statesville</cit...	<city>Statesville</cit...	536	0	
80	SRE	Correct	Sat Aug 03 03:19:51 2013	MOBILE ALABAMA	MOBILE ALABAMA	<city>Mobile </city>...	<city>Mobile </city>...	705	0	
85	SRE	Correct	Sat Aug 03 03:21:21 2013	NEW YORK	NEW YORK	<city>New York</cit...	<city>New York</cit...	786	0	
91	SRE	Incorrect	Sat Aug 03 00:57:28 2013	BALTIMORE MARYLAND	WALDORF MARYLAND	<city>Baltimore</city...	<city>Waldorf </city...	750	0	BALTIMORE->WAL...
93	SRE	Correct	Sat Aug 03 00:57:53 2013	BALTIMORE MARYLAND	BALTIMORE MARYLAND	<city>Baltimore</city...	<city>Baltimore</city...	784	0	
97	SRE	Correct	Sat Aug 03 08:57:39 2013	ANDREWS TEXAS	ANDREWS TEXAS	<city>Andrews </city...	<city>Andrews </city...	887	0	
105	SRE	Correct	Sat Aug 03 04:46:44 2013	SEATTLE WASHINGTON	SEATTLE WASHINGTON	<city>Seattle</city>...	<city>Seattle</city>...	820	0	
113	SRE	OOV	Sat Aug 03 13:49:27 2013	FORDS CONNECTICUT	ORANGE CONNECTICUT	~No interpretations	<city>Orange </city>...	686	0	
116	SRE	Correct	Sat Aug 03 13:49:54 2013	STAMFORD CONNECTICUT	STAMFORD CONNECTICUT	<city>Stamford</city...	<city>Stamford</city...	753	0	
125	SRE	OOV	Sat Aug 03 11:50:44 2013	++GARBAGE++	TEXAS	~No interpretations	<city></city><state>...	617	0	
129	SRE	OOV	Sat Aug 03 11:51:47 2013	KANKAKEE CHICAGO ILLINOIS	KANKAKEE ILLINOIS	~No interpretations	<city>Kankakee</cit...	780	0	
136	SRE	Incorrect	Sat Aug 03 13:50:37 2013	SAN ANTONIO TEXAS	GARDEN HOME WHITFORD	<city>San Antonio</...	<city>Garden Home ...	539	0	SAN->GARDEN : 0; ...
138	SRE	Correct	Sat Aug 03 13:51:02 2013	SAN ANTONIO TEXAS	SAN ANTONIO TEXAS	<city>San Antonio</...	<city>San Antonio</...	562	0	
151	SRE	OOV	Sat Aug 03 03:09:37 2013	WINSTON-SALEM NORTH CAROLI...	MISSISSIPPI	~No interpretations	<city></city><state>...	499	0	
159	SRE	Correct	Sat Aug 03 01:02:10 2013	SAN ANTONIO TEXAS	SAN ANTONIO TEXAS	<city>San Antonio</...	<city>San Antonio</...	627	0	
164	SRE	Correct	Sat Aug 03 01:03:23 2013	LAREDO TEXAS	LAREDO TEXAS	<city>Laredo </city>...	<city>Laredo </city>...	452	0	
168	SRE	Correct	Sat Aug 03 00:15:18 2013	CLEVELAND MISSISSIPPI	CLEVELAND MISSISSIPPI	<city>Cleveland </ci...	<city>Cleveland </ci...	742	0	
170	SRE	Correct	Sat Aug 03 00:15:45 2013	CLEVELAND MISSISSIPPI	CLEVELAND MISSISSIPPI	<city>Cleveland </ci	<city>Cleveland </ci	698	0	

Transcription: SAN ANTONIO TEXAS

Loaded Audio

☐ Loop ☑ Auto Save
☑ Auto Play ☑ Auto Complete

138

Speech Quality: ● Unknown ○ Clear ○ Noisy ○ Garbage

Gender: ● Unknown ○ Male ○ Female

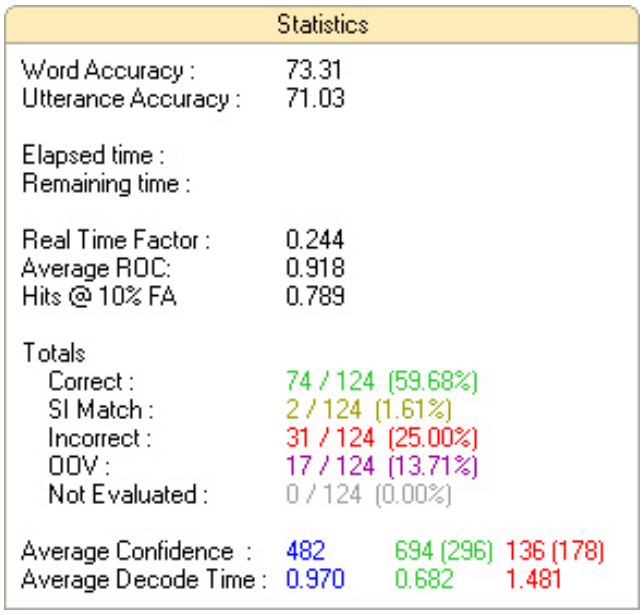
Notes:

124 Interactions

Analyze the Data

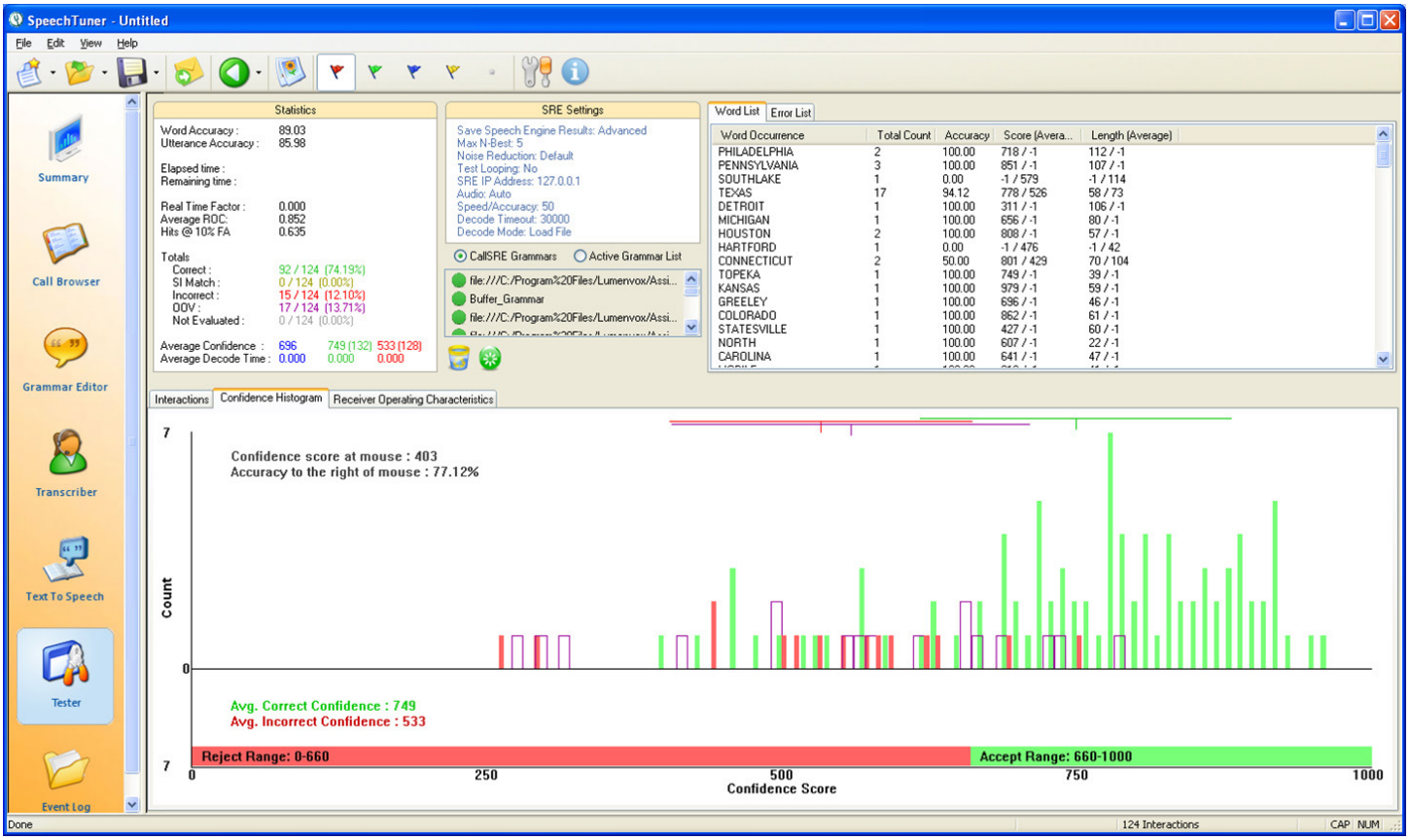
After all calls have been transcribed, you now have an aggregate dataset that can be easily analyzed. Navigate to the "Tester" screen where you can see at a glance how well, or how poorly, your ASR engine has performed. The "statistics" box displays the percentage of words the ASR engine recognized correctly, how confident it was about being correct and how many words or utterances were found "OOV," out of vocabulary or out of grammar. Evaluating the numbers will also give you an idea of how callers are using the system and help to identify problem areas.

For example, if you find most of the out of vocabulary (OOV) calls are saying the same word, you might want to add that word to the grammar so the ASR engine will recognize it in future calls. This is the point when a developer is brought into the tuning process because editing a grammar can be a highly technical task. Nevertheless, his or her input will be worth its weight in gold as this will not only increase the ASR's word accuracy levels, it will also decrease the number of calls routed to a human attendant to complete.



Test the Data

While the tester displays detailed information about the loaded dataset and provides accuracy information for the ASR engine, its main purpose is to test changes. The tester can save those changes and iterate with others until the application is optimized and ready for production.



Crossing the Finish Line

In essence, what is outlined here is "How to Tune a Speech Recognition Engine 101." There are other, more in-depth ways to quantify the accuracy of an ASR engine, and this article skimmed only the surface. A more comprehensive tune-up would measure the tuner's correct acceptance and correct rejected rates—how often the tuner accepted or rejected a user's response properly. Conversely there are false accepts and false reject rates that can be factored into calculating a speech engine's accuracy—all of which can be controlled by the speech designer.

Like cars, no two speech applications are exactly alike, so it is important to measure the accuracy of each speech application differently. But there is one thing most would agree on: it doesn't matter how well tuned your car or speech engine is, you can always do better.