LumenVox®
Speech Understood

THE ROI OF SPEECH TUNING

Executive Summary:
Speech tuning is a process of improving speech applications after they have been deployed by reviewing how users interact with the system and testing changes. Though the process can be time-consuming, even slight gains in application performance produce a positive Return On Investment (ROI) in very short times. The LumenVox Speech Tuner can be used to accelerate this ROI by decreasing the time spent in tuning cycles, which also decreases the Total Cost of Ownership (TCO) of a speech application.

Audience:
This whitepaper is intended for sales engineers, account managers, project managers, and other decision-makers who have an interest in understanding the financial impacts of speech applications.

THE ROI
OF SPEECH TUNING

Speech tuning is a vital part of building and maintaining any successful speech automation application (for more information, see our online article *The Importance of Tuning*). Speech tuning, the process of changing an application based on data gleaned from real-world use, improves recognition accuracy and provides a host of vital metrics such as call completion rates, containment rates, and user experience scores. Most importantly, these changes affect the bottom line of anyone who builds, deploys, hosts, or purchases any kind of interactive voice response (IVR) or other application that makes use of automatic speech recognition (ASR) or text-to-speech (TTS) technology.

Because tuning has a direct relationship with how well a voice application functions, tuning an application tends to increase its success rate, providing a faster return on investment (ROI). Tuning also affects the customer experience, providing benefits which are harder to measure directly compared to ROI. Since tuning is something that must be done periodically in order to ensure the application is still performing at optimal levels, the efficiency with which an organization is able to perform tuning will be a factor in understanding the total cost of ownership (TCO) of an application.

The LumenVox Speech Tuner, a unique tool developed by LumenVox, serves two key purposes. First, it makes tuning easier and faster, allowing applications to be tuned in less time, providing a faster ROI and lowering the TCO of voice applications. Secondly, the Speech Tuner allows organizations to find and improve issues they might not otherwise be aware of, which improves the experience of users and customers of the application, increasing loyalty and satisfaction.

Tuning Basics

Many users of early speech applications will remember experiences with applications that performed poorly due to insufficient tuning. “It didn’t understand me,” or, “It pronounced that name wrong,” are common critiques users have of these applications. Speech tuning helps to improve that situation by making adjustments to the components of the speech application, including:

- Improving grammar files, which drive speech recognition. This can mean adding or removing options for users to say, choosing words that don’t sound so similar, or adding weights to make often-used options more likely to be recognized.
- Tweaking synthesized markup files, which tell a TTS provider how to pronounce words. This may include adjusting the speed or pitch of the synthesis, or spelling out tricky words phonetically to make sure they’re pronounced correctly.
- Correcting confusing prompts so users are guided to speak appropriate responses, decreasing out-of-grammar responses and increasing overall recognition rates.

In all these cases, tuning requires a few things:

1. Audio must be transcribed, meaning somebody must listen to recordings of callers and write down precisely what the callers said (the *Collecting Data for Speech Tuning* whitepaper describes this process in detail).
2. The transcriptions must be compared to the recognition results from the ASR, to generate accuracy metrics (see our whitepaper *Calculating Speech Recognition Accuracy* for details on why this can be difficult).
3. In the case of TTS, somebody must compare the synthesized audio to what should have been spoken.
4. Grammars and synthesis markup documents must be changed and tests must be run to see their effect on the system.
5. The changes are deployed and the process repeats.

Calculating an ROI for Tuning

The costs and benefits of tuning vary significantly depending on the application and the work required to tune it, but we can look at a generic case to understand how, even though it is time consuming, tuning is almost always worth the effort. Consider a typical IVR application which replaces or supplements live agents in a call center.

Using industry-standard costs for agents, telephony charges, and other costs, we can estimate the cost per minute for an agent to handle a call at **$1.045**. This compares to a typical IVR that utilizes ASR and TTS technology costs between **$0.10 and $0.25** per minute to build, operate, and maintain over a 3-5 year lifespan. This means that each minute an IVR is handling a call, the average savings is around **$0.87**.

Industry Standard Costs

[1] 0.175 is the average per minute cost based on the industry standard $0.10 and $0.25 per-minute costs.

Assume that a newly-deployed IVR has a recognition accuracy rate of **80%**. The IVR asks 5 questions from each caller, and if there are ever 3 errors, it is transferred to an agent. This would give us a call containment rate of **94.21%**[2]. If we assume that each call that is not contained goes to an agent and takes **2.5 minutes** of the agent's time — a conservative estimate if you factor in the setup/teardown times for agents — then we know that the cost per call that is not contained is **$2.175** (this is the difference between 2 minutes of agent's time and 2 minutes of the IVR's time).

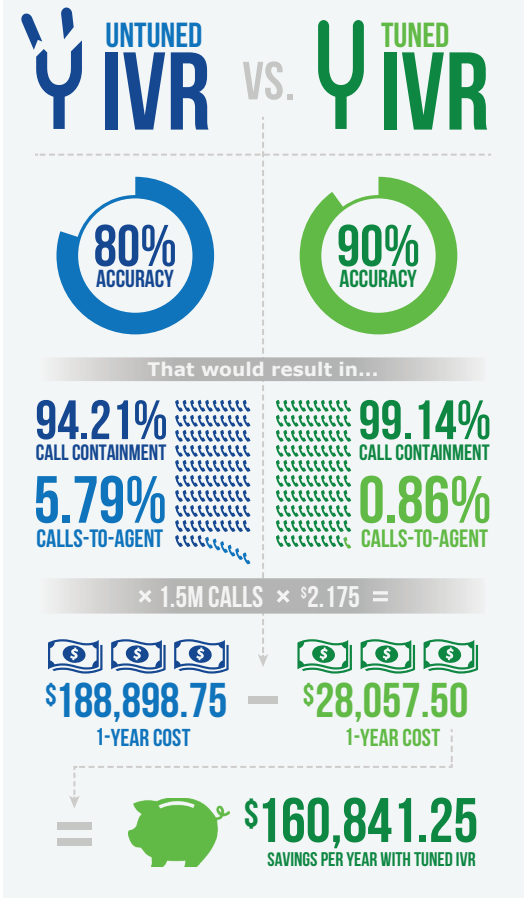
[2] This is the chance that there are 2 or fewer errors in 5 trials with a probability of success of 0.8.

Add'l Cost of Agent Calls

Assuming all agent calls last 2.5 minutes on average

Thus a medium-sized call center with 175 agents that handles **1.5 million** calls per year would experience a cost of **$188,898.75** per year due to ASR errors (1.5 million calls times a failure rate of 5.79% times $2.175 per non-contained call).

Per-Year Tuner Savings



If the tuning exercise cut the ASR error rate in half, which is normal for the first tuning cycle, then the new call containment rate would be **99.14%** — notice how a seemingly small gain in ASR accuracy greatly improves call containment. Using the same calculations as above, the call center would now only be spending **$28,057.50** per year on uncontained calls.

General industry guidelines suggest that half of an application's development time be spent on tuning. If we assume that a relatively simple IVR with just 5 dialogues represents approximately 100 hours of development time (20 hours per dialog is a reasonable approximation, though it will vary greatly depending on the complexity of a given dialog), then 50 hours would be appropriate to allocate for tuning. Considering that tuning is a highly skilled task, then we might expect to spend roughly **$300** per hour in fully-burdened costs for a speech expert to perform the tuning. The cost in tuning would thus be **$15,000**.

This gives us a net savings of **$160,841.25** per year on an investment of just **$15,000** – the tuning costs pay for themselves in less than two months of operations. Because speech automation is so much more cost effective than live agents to begin with, almost any improvement in performance translates to real savings. A similar pattern holds even for smaller applications with lower volumes.

1-Year Cost of Tuning

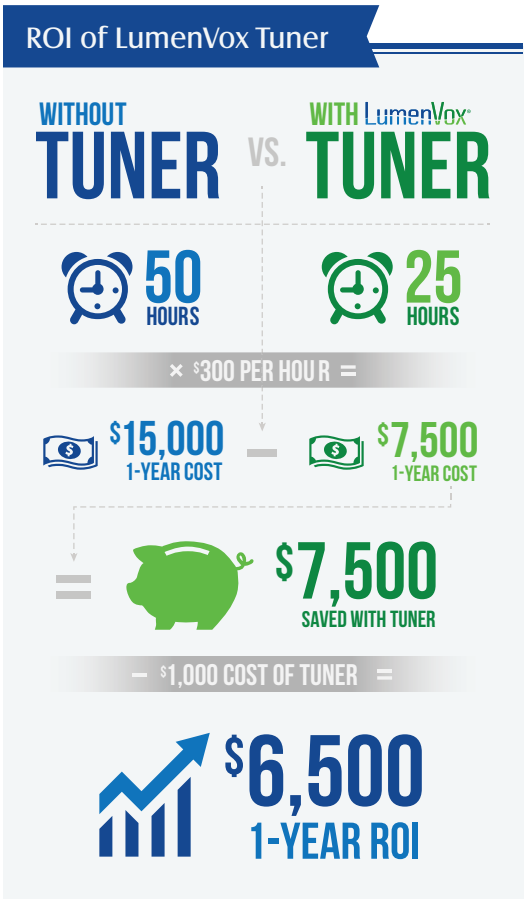
1-Year ROI of Tuning

None of the above even factors in other improvements that come from tuning, such as call completion rate (fewer people hang up in frustration), customer satisfaction from an improved user experience, etc. It is clear that tuning provides a major return on investment.

Improving the ROI Further with the LumenVox Speech Tuner

The case for tuning becomes even stronger when paired with the LumenVox Speech Tuner. The cost of a Speech Tuner license is almost always paid for and then some by a single tuning effort. We estimate that the Speech Tuner cuts down tuning times by approximately 50% compared to attempting tuning "by hand" without a specialized tool. The Tuner provides very valuable features such as:

- The ability to prepopulate transcripts with the recognized text from the ASR. Using our 80% baseline accuracy metric from earlier, this means that 80% of the time, the person doing transcriptions does not need to re-type what was said, cutting transcription time down to a fifth of what it would otherwise be.
- Grammars and SSML documents can be edited using our Tuner's Grammar Editor and SSML Editor Interactive Development Environment (IDE). This provides real-time error checking, syntax highlighting, tag completion, and other benefits that cut down grammar editing and development time by at least half.
- Changes can be made to grammars/SSML documents, ASR settings, TTS lexicons, and more and then tested immediately within the Tuner sandbox. This saves the time of having to redeploy changed documents to test environments and run tests using manual testing tools. The LumenVox 12.2 Tuner supports multi-threaded testing, as well, allowing tests to be performed orders of magnitude faster than previously.
- Advanced analytics calculate accuracy and other metrics inside the tool with a single button click, eliminating the need to export results from tests into spreadsheets or other programs for analysis (and the Tuner supports exporting this data if developers still wish to integrate it with existing workflows).



Let's look at the previous example, where 50 hours of tuning cost $15,000 but saved $160,841.25 per year. One seat of the Speech Tuner costs **$1,000** per year, but can reduce that tuning cost to **$7,500**. This is an ROI of **$6,500** by buying the Speech Tuner for just one tuning cycle. The Tuner paid for itself 6.5 times over. If the Tuner is used for multiple projects per year, the return will simply add up.

Furthermore, an important part of application development is periodic tuning of running applications in order to verify that they are working correctly and to account for any new usage periods. Thus tuning costs over the lifetime of an application must be factored into the total cost of ownership for an application.

LumenVox is a speech automation solutions company providing technology design, development, deployment, and tuning services including the LumenVox Speech Recognizer, Text-to-Speech Engine, Call Progress Analysis, Speech Tuning Services, and SLM solutions. Based on industry standards, LumenVox's core Speech Software is certified as one of the most accurate, natural sounding, and reliable solutions in the industry.

For more information, visit www.lumenvox.com

CONCLUSION

Several factors combine to make speech tuning an investment with a clearly positive return. Because speech automation enjoys such a cost-advantage over handling tasks with live agents, minor improvements in application performance translate quickly into major savings. Though speech tuning often seems expensive in an objective sense, when compared to the money it saves through improvements in performance, it generally pays for itself well within a single year. And the LumenVox Speech Tuner tool, which can cut tuning costs in half, easily enjoys a positive return on even a single tuning project of almost any size.