# Using an Adaptive Voice User Interface to Gain Efficiencies in Automated Calls

Daniel O'Sullivan

Contact Solutions

Inventing Real Customer Service™

# Abstract


Traditional speech applications are "static" and make no dynamic adjustments for the real-time behavior of individual callers. As a result, all callers are handled in the same way, regardless of their ability to use the automated system. With an adaptive approach, automated interactive voice response (IVR) systems continuously monitor individual caller behavior during each call to the system.

Specifically, speech and/or dual-tone multi-frequency (DTMF) responses (such as entering account numbers, entering a PIN, making selections from menus, etc.) are monitored for speed and accuracy in real-time, node-by-node in the call script.

When initially run, the software learns how effectively callers navigate each of the nodes in the application, and continues to monitor how callers behave in the IVR for the first several hundred passes through each node. This monitoring process allows the adaptive system to capture the specific information required to adapt WPM and/or content for individual callers, based on how they interact with the system, within specific calls.

This personalizes the call experience as it happens, creating a friendlier, more responsive and productive caller experience. Analysis of adaptive business-to-consumer (B2C) retail, financial, travel, medical insurance and government applications, using 95 percent confidence intervals, indicated:

- Improvements in IVR utilization of approximately 17.24-20.44 percent.
- Reduction of first-attempt caller input errors of about 1.02-1.75 percent (relative reduction ranging from 4.7-8.0 percent).
- Increase in average handle rate (AHR) of about .5-3 percent.
- Reductions in average handle time (AHT) of about 6-16 percent.

# Background

## Although user-personalization has demonstrated success for Web-based interactions, it has yet to be fully leveraged in the handling of automated calls.

Advances in speech technology such as automated speech recognition (ASR), natural language understanding (NLU), caller-directed dialogues and Web profiles are excellent enabling technologies. But, these technologies are only part of the solution. Used in conjunction with a well-designed voice user interface (VUI), they represent a significant improvement over earlier user personalization. However, none of these technologies incorporate the individuality and in-call behavior of callers themselves.

Callers to speech applications have their own unique set of aural, speech, hand-eye coordination (as used in DTMF keypad entry) and comprehension skills. Additionally, there are a variety of environmental variables to factor into the equation, such as background noise, poor mobile phone signals and caller distraction. Callers often opt for call center agents when they get frustrated with automated speech systems, because agents are able to adjust to environmental variables in real time, and handle the dynamics of human conversation intuitively and with ease.

To the extent that speech applications can monitor and adjust to how individual callers behave during a call, a proportionate number of automated calls can be more efficient and productive. Although a well-designed call script with optimal structure and content, intentional pauses, grammar tuning and context forming are excellent design principles, the system falls short if it does not consider the real-time behavior of callers, just as a call center operator would under the same circumstances.

This paper describes how a real-time adaptive VUI can be used in conjunction with best-practice design principles to provide optimal use and efficiency gains for automated speech applications.

# The Dynamics of Human Conversation

Research indicates that speakers of English typically produce 130 - 200 words per minute (WPM). This wide WPM range applies to 90 percent of the English-speaking population.

| 130 - 145 WPM | 145 - 175 WPM | OVER 175 WPM |
|---|---|---|
| FOR COMPLEX MATERIAL | FOR AVERAGE COMPLEXITY | FOR SIMPLE MATERIAL |

Listeners can be lost to boredom or complexity or fully engaged in a conversation based on the speaker's ability to deliver all types of material at the optimal rate for each listener. Good communicators are aware of this and continuously monitor their audience. They periodically adjust their conversational pace, verbal content and emphasis to get the message across effectively and efficiently. They make these adjustments in an instinctive, fluid and natural way, thereby quickly "tuning in" to establish optimal harmony with the listener to keep them fully engaged in the dialogue. Traditional speech applications are "static" and make no dynamic adjustments for the real-time behavior of individual callers.

As a result, all callers are handled in the same way regardless of their knowledge, experience, navigation skills and willingness to use the automated system. Specifically, all audio content is delivered to the caller at the same WPM rate regardless of their demonstrated behavior during the call. These applications do not listen for signs that the listener understands what is being said and is comfortable with the pace and content of the dialogue. Without "tuning in" to a caller's behavior during the call, real IT efficiencies are lost. As shown in Table 1, this can lead to negative outcomes in terms of customer satisfaction and the costs associated with handling automated calls.

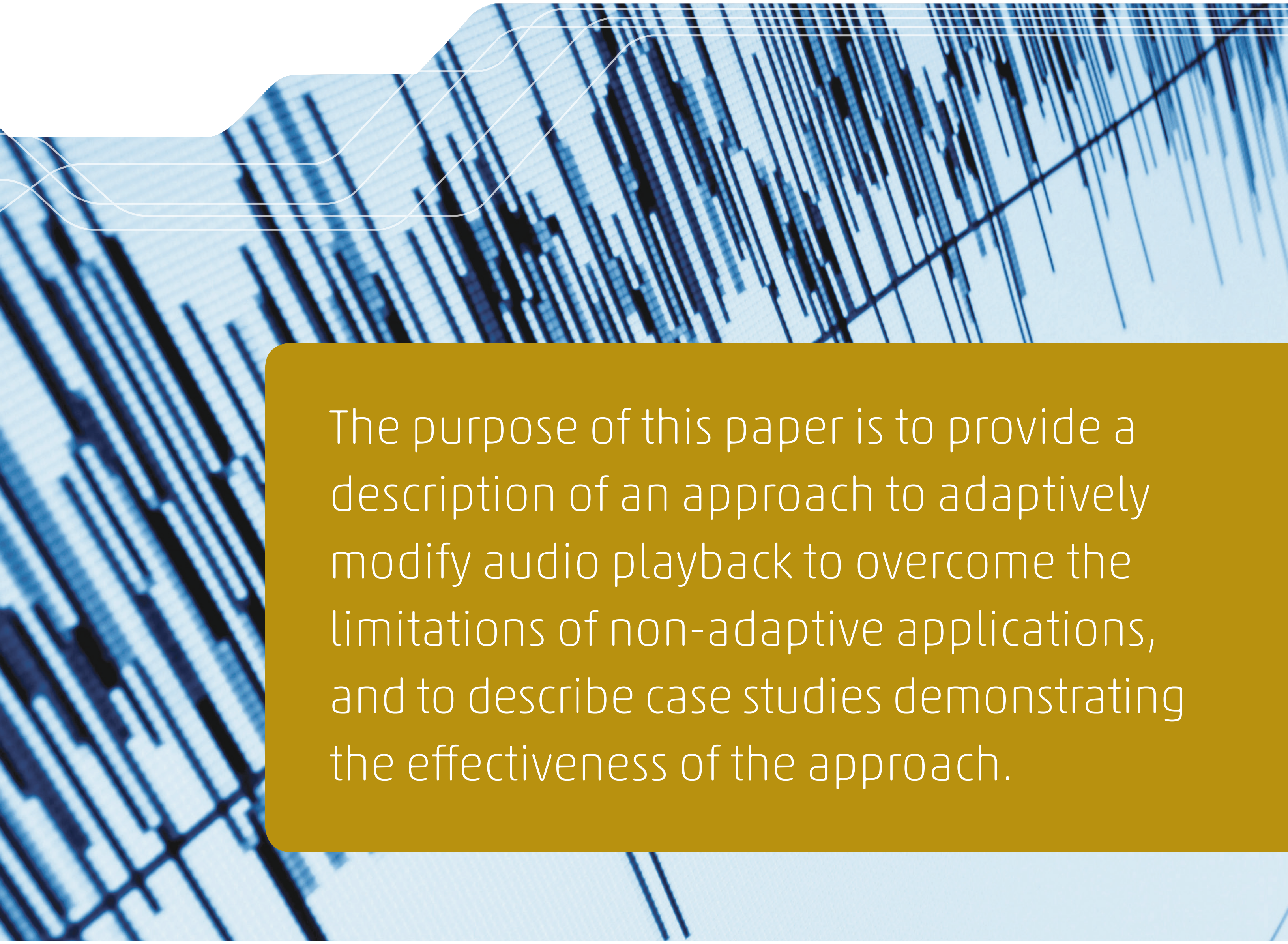
TABLE 1

| HUMAN FACTOR | COMPROMISE | CONSEQUENCE |
| --- | --- | --- |
| **Caller skills:** Individual conversational and speech/keypad navigation skills. | Automated calls are not tuned to each individual caller's skill level. | Call is longer, less productive and less efficient than it should be. |
| **Caller attitude:** Level of general like or dislike for automated voice systems. | All callers handled in the same way regardless of acceptance for automated voice systems. | Marginal users are more inclined to opt for an agent. |
| **Environment:** Callers may be using a home, mobile, car, public or office phone. | No real-time adjustment for real world dynamics. | Even accepting and power users can get frustrated. |
| **Attention:** Callers (expert and novice alike) can easily be distracted during a call. | No allowances for the caller's circumstances. | Caller knows an agent will understand. |
| **Familiarity:** All callers are not equally familiar with the voice application. The trend is toward automating increased amounts of information. | No provisions for less familiar/ unfamiliar callers. | Caller knows they can ask the agent anything. |

When callers interact with a system that is tailored to their specific set of skills, knowledge and aural, vocal and hand-eye coordination, efficiencies are gained in the handling of automated calls including:

- Highly skilled power users can more quickly navigate the application flow.
- Greater tolerance for callers having difficulty hearing/understanding the audio.
- More flexibility for distracted callers.
- Increased user preference for automated systems.
- Higher success rates for callers using mobile or public phones.

Experience and data indicate that giving automated speech systems the capability to adjust audio responses of the system to how callers behave during calls would facilitate improved communication. Individual callers have their own natural conversational rhythm, a dialogue pace at which they are most comfortable and productive. This form of "tuning in" to the individual caller during the call emulates the principles of good communication demonstrated by skilled call center agents.



The purpose of this paper is to provide a description of an approach to adaptively modify audio playback to overcome the limitations of non-adaptive applications, and to describe case studies demonstrating the effectiveness of the approach.

# An Approach to Adaptive Modification of Audio in IVRs

With an adaptive approach, automated IVR systems continuously monitor individual caller behavior during each call to the system. Specifically, speech and/or DTMF responses (such as entering account numbers, entering a PIN, making selections from menus, etc.) are monitored for speed and accuracy in real-time, node-by-node, in the call script.

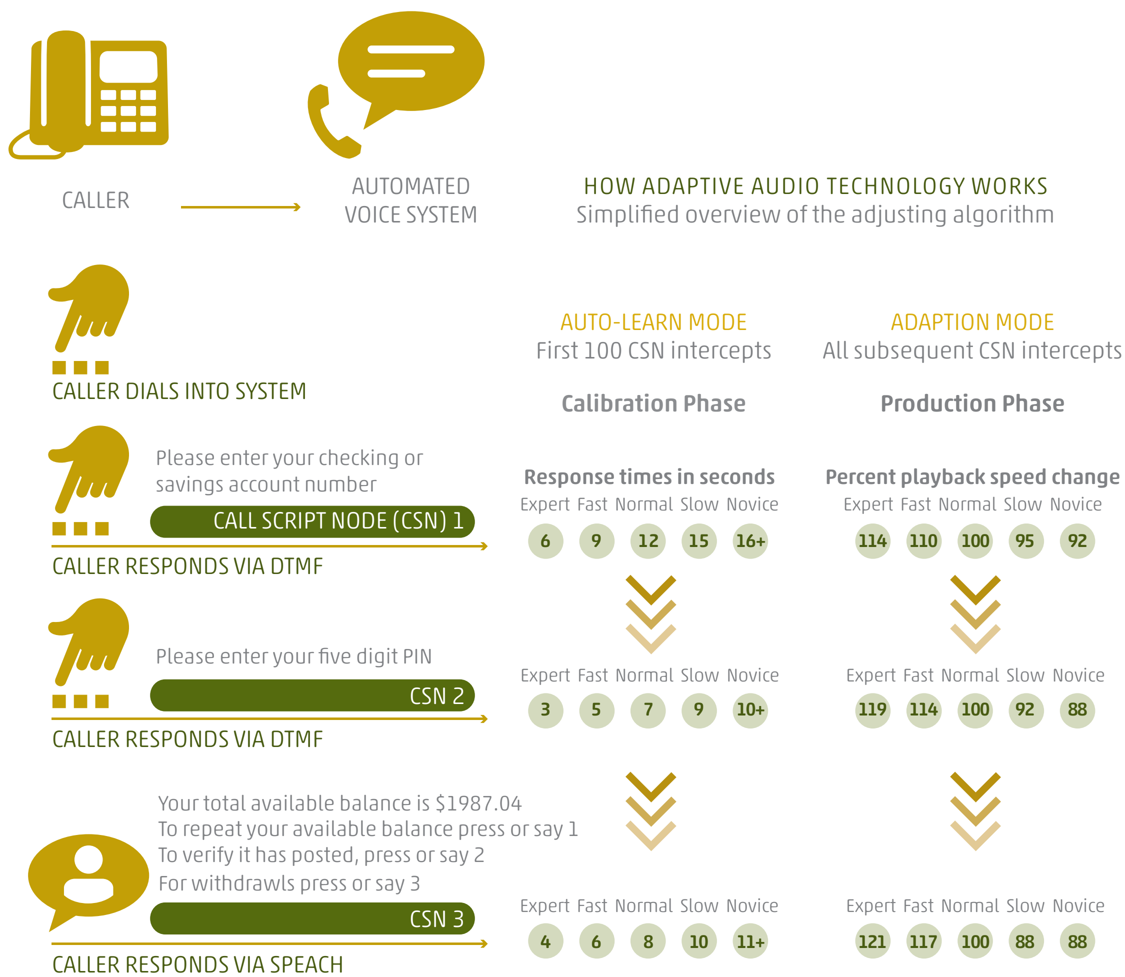
By definition, a call script node (CSN) is a unique point in the application call script at which caller input is requested. This could be as simple as a single DTMF response to a simple menu, or as complex as a string of spoken digits representing a member account number or PIN. An IVR turn is defined as a CSN and caller response/timeout pair.

When run for the first time, Adaptive software listens and learns how effectively callers navigate each of the nodes in the application, and continues to monitor caller behavior for the first several hundred passes through each node. This is the first step in capturing the specific information required to adjust the WPM and/or voice message content for the unique skill level of each individual caller, within each call.

After acquiring a sufficient calibration sample, the system automatically switches to adaptive mode. The software uses the previously stored behavioral information to automatically adjust the speaking rate and voice message content of the system to suit the skills and exhibited behavior of each individual caller in real time. This personalizes the call experience as it happens, creating a friendlier, more responsive and productive caller experience.

For example, after detecting slower response times, the system slows the playback rate of system speech to allow less experienced callers extra time to understand and respond to system prompts. Figure 1 illustrates how the basic speed adjustment part of the algorithm works.

FIG. 1. DYNAMIC ADAPTATION AND PERSONALIZATION THROUGHOUT THE CALL




System administrators can set minimum, intermediate and maximum playback speeds as well as caller response times to trigger a change in playback speed or message content. This process works autonomously for every caller without the need for automatic number identification (ANI), customer databases or Web-based profiles.

There is no need for prior knowledge about callers, and the process works in real time. This is important because even power users do not behave the same way when distracted and ANI-based profiles are not helpful when callers use a different phone.

# Case Studies

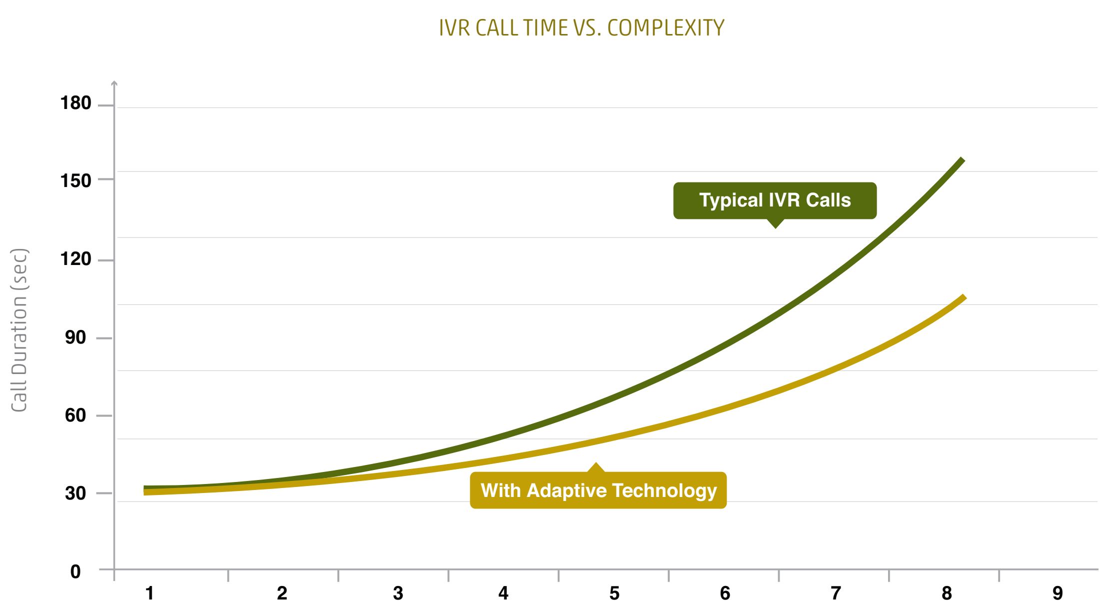
## Operational Efficiency Gains in Production Metrics

The benefits of an adaptive VUI vary based on the design, content and average call duration of the speech application. Applications must provide a sufficient amount of caller interaction to make the technology worthwhile.

Voice applications should have a minimum average automated call length of approximately 40 seconds, with at least two discrete (menu choices, yes/no options or similar) or one multiple value (account number, PIN or similar) caller responses to make adaptive technology worth implementing.

In general, the more levels of scripting and the higher the average automated call duration, the greater the benefit to users. Based on production metrics gathered at various sites, Figure 2 illustrates the relationships among automated call script length, script levels and the effectiveness of the adaptive process.

FIG. 2. ADAPTIVE CALLS ARE MORE EFFICIENT

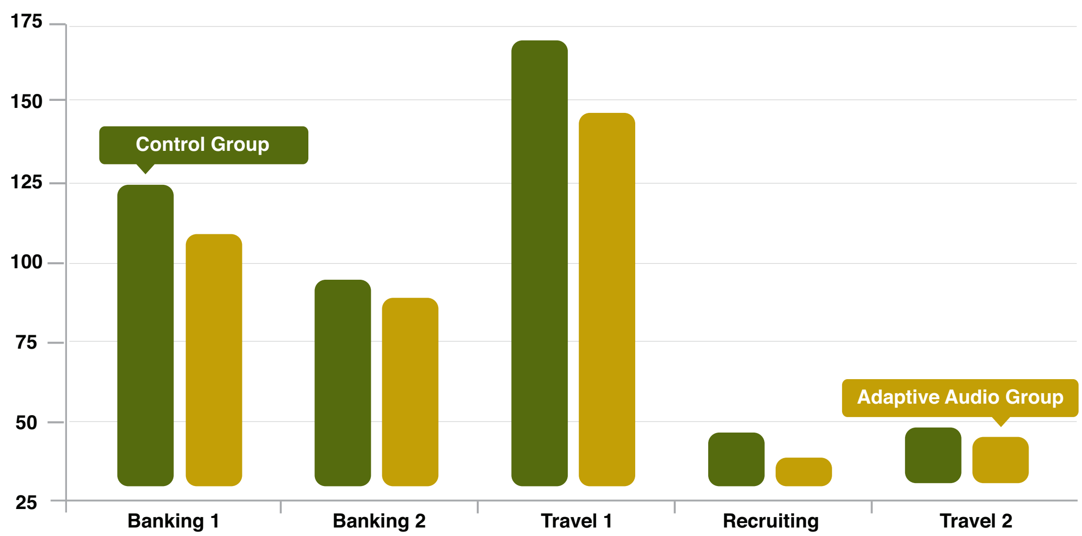


The data in Figure 3 illustrates production results gathered from various types of voice applications using adaptive technology. This data indicates improvements in both the AHT and AHR.

For example, Travel 1 (a rail travel application) experienced a drop of 27 seconds (16 percent) in AHT, while the AHR simultaneously increased by 1.6 percent. This particular implementation of adaptive technology used only positive playback increments of 110, 114, 117, 119 and 121 percent, because the customer goal was to reduce AHT. These results indicate that power users are more likely to stay with the automated system, provided it moves at their comfortable speaking rate, even though that rate is 10-21 percent faster than without adaptation. This is consistent with the hypothesis that tuning into the natural rhythm and pace of the caller helps keep them engaged in the automated call process.

The mean reduction in AHT across the five applications in Figure 3 was a statistically significant 11.2 seconds per call ($t(4) = 2.73$, $p = .05$), with a significant mean percentage reduction of 11.3 percent ($t(4) = 6.28$, $p = .003$). A 95 percent confidence interval constructed for the percentage reduction in AHT ranged from about 6-16 percent.

FIG. 3. PRODUCTION RESULTS WITH ADAPTIVE TECHNOLOGY




| | Banking 1 | Banking 2 | Travel 1 | Recruiting | Travel 2 |
|---|---|---|---|---|---|
| Duration Change | (8.8%) | (7.8%) | (15.3%) | (15.6%) | (8.5%) |
| Containment Change | 0.65% | N/M | 1.6% | N/M | 1.9% |
| Call Sample Size | 100k | 100M | 60k | 5k | 10k |

## As a result, the relative improvement in IVR Utilization was 19.18 percent

Comparative AHR data was available for two cases (Travel 1 and Travel 2). Without adaptation, the observed rates were respectively 70.5 and 87.4 percent, with 95 percent adjusted-Wald binomial confidence intervals [6] ranging from 69.8-71.2 percent for Travel 1 and 86.8-88.0 for Travel 2.

With adaptation, the observed rates (and associated 95 percent confidence intervals) were 72.1 percent (71.7-72.5) for Travel 1 and 89.3 (88.9-89.7) for Travel 2. For both cases, the confidence intervals did not overlap, indicating statistically significant improvements in AHR ($p < .05$). Analysis of the confidence intervals indicated that the likely magnitude of improvement across the cases ranged from about .5 to three percent.

In a separate case study of the IVR for a medical insurance provider (448,646 total calls), adaptation led to an absolute reduction in first-time errors (combined no-input and no-match events) of about 1.38 percent (95 percent confidence interval ranging from 1.02-1.75 percent, relative reduction of 6.35 percent ranging from 4.7-8.0 percent) with a confidence interval ranging from 21.68 to 21.82 percent). With adaptation, the first-time error rate was 20.37 percent (15299/75120 with a 95 percent binomial confidence interval ranging from 20.08 percent to 20.66 percent). Because the binomial confidence intervals did not overlap, the difference was statistically significant ($p < .05$). Additionally, the estimated mean number of IVR turns for standard and adaptive were, respectively, 3.18 and 3.79, with respective 95 percent confidence intervals ranging from 3.18 to 3.19 and 3.74 to 3.83.

As a result, the relative improvement in IVR Utilization was 19.18 percent (ranging from 17.24 to 20.44 percent). These results indicate that callers are more likely to stick with the automated system, provided it moves at their comfortable listening rate, even if that rate is faster than it would be without adaptation. This is consistent with the hypothesis that tuning into the natural rhythm and pace of the caller helps reduce input errors and keeps them engaged in the automated call process.

# Measurable Efficiency Gains Reduced Operating Costs

For typical business-to-consumer (B2C) retail, financial, travel and government applications, savings of one to five percent in AHR and seven to 15 percent in AHT can be expected when incorporating adaptive functionality. For high-volume call centers, this translates into significant cost savings because automated speech and touch-tone calls cost about $.75 each, while agent-handled calls are about $4.25 on average.

Telecom costs can range from $.02 to $.06 depending on call volume and can be as high as $.40 per minute for hosted services. As a result of shorter calls, increased automation and an interface that improves the caller experience by adjusting to their skill levels, offers the following benefits:

### Reduced customer service representative (CSR) labor costs due to increased call automation:

If self-service is quicker and easier, callers will be more likely to use it.

### Reduced CSR labor costs due to reduced caller frustration:

The average CSR talk time will be longer for callers that are frustrated with an ineffective IVR.

### Reduced telecom costs:

Shorter, automated calls result from customer experiences tailored to each individual caller's skill level. The result: lower telecom operating expenses for the call center.

Telecom costs can range from $.02 to $.06 depending on call volume, and can be as high as $.40 per minute for hosted services.

## Reduced CSR churn:

Callers invariably have a strong dislike for self-service. One reason for this is that the prompts are long, tedious, and not able to accommodate caller skill levels. Adaptation improves the caller experience and results in fewer complaints to the CSR, which makes the CSR's job more pleasant, thereby reducing CSR churn and increasing CSR productivity.

## Increased customer retention:

Previous studies have shown that the cost to acquire a new customer is approximately five times the cost of keeping that customer. For enterprises, self-service that satisfies customers also increases customer loyalty, and reduces the cost of acquiring new customers.

Table 2 shows the results for a call center handling 10 million calls per month, with an average automated call duration of 110 seconds. Typical return on investment (ROI) at a site like this will be just over 550 percent with an enterprise payback period of 1.66 months, and a total enterprise cost reduction of over $6.69 million per year.

TABLE 2. SAVINGS CALCULATIONS FOR A LARGE-SIZED CALL CENTER

| CALL CENTER VARIABLES | |
|---|---|
| Calls per month | **10,000,000** |
| Average length of talk time per call (seconds) | **280** |
| % of calls that are completed by self-service | **70%** |
| Average length of self-service call | **110 seconds** |
| Average burdened wage rate for CSRs (per hour) | **$22.00** |
| Hours worked per day | **8** |
| CSR Chrun Rate | **49%** |
| Reduction in churn rate with AA | **4.8%** |
| Cost to replace a CSR | **100% of wages** |
| Hours worked per week | **40** |
| Number of shifts? | **1** |
| CSR non-talk time (hours per day) | **2** |

| CALL CENTER VARIABLES | |
|---|---|
| % of callers that hang-up and call back again later | **3%** |
| % of calls that hang-up and call back again later | **10%** |
| Reduction in the % of callers that hang-up and call back again later | **10%** |
| Cost for Adaptive Technology | **$1,214,570** |
| Additional talk time for caller to vent | **5 seconds** |
| % of additional callers that will use self-service if it has Adaptive Audio technology | **1%** |
| Per minute toll rate | **$0.020** |
| Busy-hour calls | **85.227** |
| **REDUCTION IN SELF-SERVICE CALL LENGTH WITH ADAPTIVE TECHNOLOGY** | **8%** |

| ENTERPRISE VARIABLES | |
|---|---|
| % of callers that are lost potential customers | **0.010%** |
| Monthly lost revenue from callers that are lost potential customers | **$50** |
| % of callers that are customers that don't get issue resolved properly | **0.0500%** |
| Reduction in lost customers with Adaptive technology | **10%** |
| Cost to acquire a new customer | **$300** |
| Amount spent on advertising | **$54,208,754 per year** |
| Reduction in advertising expense | **1%** |

| OBTAIN RESULTS | |
|---|---|
| Reduction in self-service call length with Adaptive technology | **8%** |
| Call center payback period | **2.579 month** |
| Enterprise payback period | **1.666 month** |
| Call center ROI | **465.4%** |
| Enterprise ROI | **550.9%** |

| OBTAIN RESULTS | |
|---|---|
| Cost reduction in call center operating expense | **$21,410 per day**<br>**$471,017 per month**<br>**$5,652,204 per year** |
| Cost reduction in enterprise operating expense per year | **$6,691,091 per year** |
| Start number of CSRs | **1,355** |
| CSRs after adding AA | **1,272** |
| Reduction in required CSRs | **83** |
| Cost reduction of CSR labor (venting callers) | **$1,124,444 per year** |
| Cost reduction of CSR labor (additional self-service utilization) | **$2,737,778 per year** |
| Cost saving due to CSR churn reduction | **$1,481,028 per year** |
| Telephone toll expense reduction | **$308,953 per year** |
| Savings from reduction in lost customers | **$496,800 per year** |
| Cost of additional enterprise advertising | **$542,088 per year** |
| Total enterprise expense cost reduction | **$6,691,091 per year** |
| Total enterprise operation expenses | **$221,865,619 per year** |

# Conclusion

With increasing numbers of call centers, enterprise IT departments and ASR-based hosting centers recognizing the economic benefits of automated self-service, the adoption of more complex, information-rich speech applications is on the rise.

Enabling technologies such as the Web-centric IVR, speech-enabled dialogs, natural language understanding and customer-directed dialogs, are helping to drive this trend.

As a result, adaptive VUI technology offers several direct and indirect benefits, including increased operational efficiencies, reduced operational costs, increased customer satisfaction and a fast, verifiable, ROI and payback period. For that reason, the use of adaptive technologies makes good business sense for both the customer and the enterprise.

## About the Author

Daniel O'Sullivan is Principal Technologist at Contact Solutions. Previously, O'Sullivan was a member of the Technical Staff at Lucent/AT&T Bells Labs. He has optimized hundreds of millions of phone calls and saved costs for many well-known companies. O'Sullivan holds an EE degree from the Dublin Institute of Technology and a Masters in Computer Science from the Polytechnic Institute of New York University.

# References

[1] Zazelenchuk, T. W., Boling, E.: Considering User Satisfaction in Designing Web-Based Portals. EDUCAUSE Quarterly. 26(1), 35--40, (2003)

[2] Sandhu, K.: Theoretical Perspectives for E-Services Acceptance Model. In: Sobh, T. (ed.) Advances in Computer and Information Sciences and Engineering pp. 218--223, Springer Netherlands, Dordrecht (2008)

[3] Arons, B.: Techniques, perception, and applications of time-compressed speech. In: Proceedings of the 1992 American Voice Input/Output Society pp. 169-177. AVIOS, Minneapolis, MN (1992)

[4] Datamonitor: Profiting from Evolving Speech Applications - Segmenting opportunity in a growing market. Datamonitor, New York, NY (2006)

[5] Customer Contact Council: Analyzing and Defining Cost and Productivity Metrics for Contact Centers - Benchmarking Overview. CEB, Washington, DC, (2005)

[6] Sauro, J., Lewis, J. R.: Estimating Completion Rates from Small Samples Using Binomial Confidence Intervals: Comparisons and Recommendations. In: Proceedings of the Human Factors and Ergonomics Society pp. 2100-2104, Human Factors and Ergonomics Society, Santa Monica, CA (2005)

[7] Coyles, S., Gokey, T. C.: Customer Retention Is Not Enough. Journal of Consumer Marketing. 22, 101--105, (2005)

**Contact Solutions invents real customer service.** Our patented technology, optimized solutions, and culture of continuous improvement deliver a more personalized customer experience. **Guaranteed.**

Our Continuous Improvement Methodology™ has improved the customer experience for every client — with an average CX rating increase of 27.4%[1] — while lowering costs. This unique model quantifies the customer experience, enabling our clients to continuously improve performance and achieve ROI, year-after-year. Contact Solutions is one of the top-five largest hosted IVR providers in North America and the recipient of the Frost & Sullivan Product Differentiation Award.
For more, visit contactsolutions.com.

[1] The Contact Solutions CX Rating quantitatively measures customer intent, interaction and perception.

Contact Solutions

Inventing Real Customer Service™

© 2013 Contact Solutions, LLC. All rights are reserved.

Contact Solutions reserves the right to make changes in specifications and/or to discontinue any product at any time without notice or obligation and will not be liable for any consequences resulting from the use of this publication.

Printed in the U.S.A.
SDS.002.0001 MAY 2013

11950 Democracy Drive, Suite 250, Reston, VA 20190