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

Daniel O'Sullivan

Interactive Digital, 250 Community Drive, Smithtown, NY 11787
dan@interactive-digital.com

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 knowledge, experience, navigation skills and willingness to use the automated system. 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. When run for the first time, the software listens and learns how effectively callers navigate each of the nodes in the application, and continues monitoring for the first several hundred passes through each node to get the specific information needed later to make intelligent decisions as to when and how to adjust the WPM and/or content for a caller of a particular skill level. This personalizes the call experience as it happens, creating a friendlier, more responsive and productive caller experience. For sample B2C Retail, Financial, Travel, Medical Insurance and Government applications, analysis using 95% confidence intervals indicated improvements in IVR Utilization of about 17.24-20.44%, a reduction of First-Attempt Caller Input Errors of about 1.02-1.75% (relative reduction ranging from 4.7-8.0%), an increase in Average Handle Rate of about .5-3% and reductions in Average Handle Time of about 6-16% when incorporating adaptive functionality.

Keywords: IVR Utilization, Customer Satisfaction, Customer Service, IT Efficiency, IVR Efficiency, Adaptive Audio, IVR Retention, IVR Containment, IVR Call Duration, IVR Calls, IVR Software, Call Center Cost, Telecom Cost, Phone Bill, Telephone Cost, Handle Rate, Handle Time, AHT, AHR, Completion Rate, Voice User Interface, VUI, Call Experience, Personalized Call, Adaptive Interface, Interactive Voice Response, Voice Application, Voice System, Automated Call, Speech Technology, Speech Recognition

1 Introduction

1.1 Background

Although user-personalization has had demonstrated success for web-based interactions [1, 2], 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 are only part of the solution. Used in conjunction with a well-designed voice user interface (VUI), they represent a significant improvement over earlier technologies. However, one important design factor that has until recently been overlooked is the individuality and in-call behavior of the telephone callers themselves.

Every caller to a speech application has his or her own individual set of aural, speech, hand-eye coordination (as used in DTMF keypad entry) and material comprehension skills. Add to this environmental variables such as background noise, poor mobile phone signals and caller distraction, and it becomes clear that each call to the IVR system is truly a unique interaction.

This is one of the main reasons human operators are so good at handling any type of call – they can handle the dynamics of human conversation intuitively and with ease. Callers know this all too well and will often opt for an agent the first time the speech system fails to be productive for them.

To the extent that the speech application can monitor and adjust to the behavior of a particular caller during the 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 the caller just as a human would under the same circumstances.

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

1.2 The Dynamics of Human Conversation

Research [3] shows that speakers of English typically produce 130 - 200 words per minute (WPM). This wide WPM range applies to 90% of the English-speaking population.

- For complex material, a rate of 130 - 145 WPM may be required.
- For material of average complexity, 145 - 175 WPM can be optimal.
- For simple material, many listeners can accommodate over 175 WPM.

Listeners can be lost to boredom, lost to 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 and keeping 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 consequences in terms of customer satisfaction and the costs associated with the handling of automated calls.

Table 1. Human factors – compromising has its consequences

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 towards 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 skill levels, knowledge and aural, vocal and hand-eye coordination talents, 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
- Unfamiliar callers are given a reason to become repeat users
- Salvaging the automated call for callers using mobile or public phones.
- Distracted callers get the flexibility they need

From the preceding, it is clear that having the automated speech system listen to how callers behave during the call and adjusting the audio responses of the system accordingly promotes 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 principals of good communication.

1.3 Purpose of this Paper

The purpose of this paper is to (1) provide a description of an approach to adaptively modifying audio playback to overcome the limitations of nonadaptive applications and (2) to describe case studies demonstrating the effectiveness of the approach.

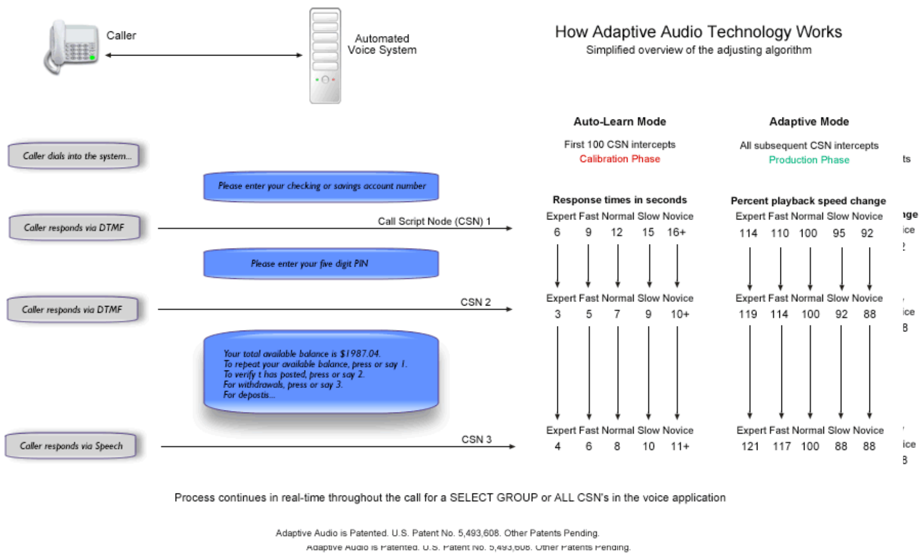
2 An Approach to Adaptive Modification of Audio in IVRs

With an adaptive approach (for example, see www.interactive-digital.com), 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, the software listens and learns how effectively callers navigate each of the nodes in the application, and continues monitoring for the first several hundred passes through each node to get the specific information needed later to make intelligent decisions as to when and how to adjust the WPM and/or content for a caller of a particular skill level.

After acquiring a sufficient calibration sample, the system automatically switches to adaptive mode. Here 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



Adaptive Audio is Patented. U.S. Patent No. 5,493,608. Other Patents Pending.

System administrators can set minimum, intermediate, and maximum playback speeds, as well as the caller response times that will 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.

3 Case Studies

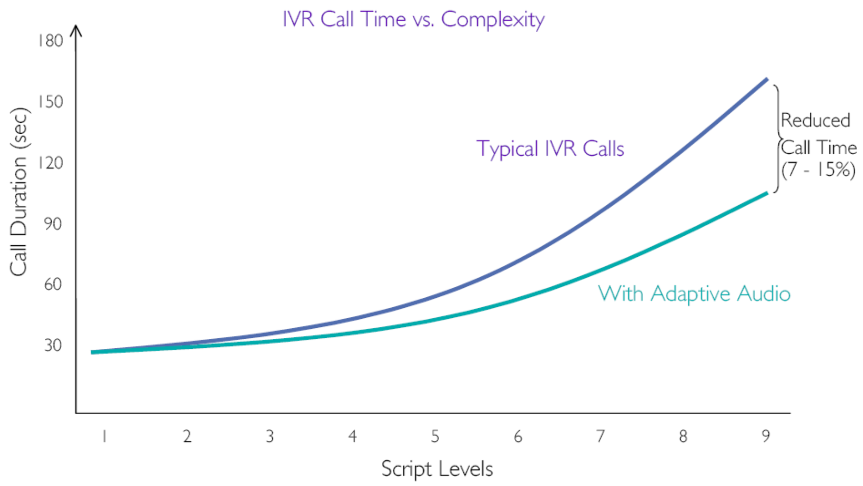
3.1 Operational Efficiency Gains in Production Metrics

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.

As a rough estimate, a voice application should have a minimum average automated call length of 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.

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 illustrate production results gathered from various types of voice applications using adaptive technology. These data indicate improvements in both the Average Handle Time (AHT) and Average Handle Rate (AHR).

For example, Travel 1 (a rail travel application) experienced a drop of 27 seconds (16%) in AHT, while the AHR simultaneously increased by 1.6%. 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.

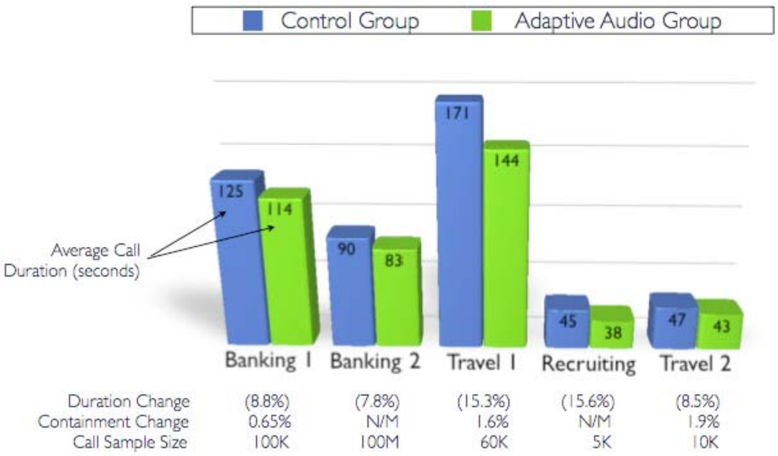
These results indicate that power users are more likely to stay with the automated system provided it moves at their comfortable speaking rate – even though that rate is 10-21 percent faster than it would be without adaptation. This is consistent with the hypothesis that tuning into the natural rhythm 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% ($t(4) = 6.28$, $p = .003$). A 95% confidence interval constructed for the percentage reduction in AHT ranged from about 6-16%.

Comparative AHR data were available for two cases (Travel 1 and Travel 2). Without adaptation, the observed rates were, respectively, 70.5 and 87.4%, with 95% adjusted-Wald binomial confidence intervals [6] ranging from 69.8-71.2% for Travel 1 and 86.8-88.0% for Travel 2. With adaptation, the observed rates (and associated 95% confidence intervals) were 72.1% (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 3%.

Fig. 3. Production results with adaptive technology



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% (95% confidence interval ranging from 1.02-1.75%; relative reduction of 6.35% ranging from 4.7-8.0%). For the standard implementation, the rate of first-time errors was 21.75% (296898 errors divided by 1365172 opportunities for error, with a 95% binomial confidence interval ranging from 21.68% to 21.82%). With adaptation, the first-time error rate was 20.37% (15299/75120, with a 95% binomial confidence interval ranging from 20.08% to 20.66%). 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% confidence intervals ranging from 3.18 to 3.19 and 3.74 to 3.83. Thus, the relative improvement in IVR Utilization was 19.18% (ranging from 17.24 to 20.44%). 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 keep them engaged in the automated call process.

3.2 Measurable Efficiency Gains Reduce Operating Costs

For typical B2C Retail, Financial, Travel and Government applications, savings of 1-5 percent in AHR and 7-15 percent in AHT can be expected when incorporating adaptive functionality. 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 [4, 5]. 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, the following benefits are provided:

- Reduced Customer Service Representative (CSR) Labor Costs Due To Increased Call Automation: If self-service is a quicker-and-easier experience for the caller, they will be more likely to use it and stick with it once they have started to use it. Additionally, this results in an increase in customer satisfaction and faster response times.
- Reduced CSR Labor Costs Due To Fewer Callers Venting: The average CSR talk-time will be longer for callers that are upset and irritated with an ineffective IVR. These callers will often take time to complain, which takes the time of the CSR.
- Reduced Telecom Costs: Shorter automated calls are a product of a user experience that is tuned to each individual caller's skill level. This results in lower telecom operating expenses for the call center.
- Reduced CSR Churn: Callers invariably have a strong dislike for self-service. One of the reasons for this is that the prompts are long and tedious and not able to accommodate their skill level. Adaptive technology improves the caller experience and results in fewer complaints to the CSR. This makes the CSR's job a bit more pleasant, thereby reducing CSR churn and increasing CSR productivity.
- Reduced Advertising Costs: An enterprise that irritates callers by having a self-service that is disliked will need to spend more on advertising to compensate for this.
- Increased Customer Retention: Previous studies [7] have shown that the cost to acquire a new customer is approximately five times the cost of keeping that customer. Satisfying customers will contribute to retaining their business and decrease 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 ROI at a site like this will be just over 550%, with an enterprise payback period of 1.66 months and a total enterprise cost reduction of over $6.69M per year.

4 Conclusion

With call centers, enterprise IT departments and ASR-based hosting centers recognizing the economic benefits of automated calls versus using an agent, the trend is towards longer, more complex and information-rich speech applications. Enabling technologies such as the web-centric IVR, speech-enabled dialogs, natural language understanding and customer-directed dialogs serve to help further this trend.

With this increased caller interaction, leveraging the benefits an adaptive VUI offers several direct and indirect benefits including increased operational efficiencies, reduced operational costs, increased customer satisfaction and a very short and verifiable ROI and payback period. Thus, the use of adaptive technologies makes good business sense from the point of view of the customer and the enterprise.

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 Churn 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)	2.00	
% of callers that hang-up and call back again later.	3.00%	
% 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 to license Adaptive Audio	$1,214,570	
Additional talk time for caller to vent	5.00	Seconds
% of additional callers that will use self-service if it has Adaptive Audio technology	1.00%	
Per minute toll rate	$0.020	
Busy-hour calls	85,227	
Reduction in self-service call length with Adaptive Audio 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 Audio 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 here (see following sheets for additional details):

Reduction in Self-Service Call Length with Adaptive Audio Technology	8%	
Call Center Payback Period	2.579	month
Enterprise Payback Period	1.666	month
Call Center ROI	465.4%	
Enterprise ROI	550.9%	
Cost Reduction in Call Center Operating Expense	$21,410	per day
	$471,017	per month
	$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

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)