

ANALYSIS

Configurable packages are a better fit for new markets embracing speech

In the last five years, the market backdrop for speech solutions has become substantially more favorable. The following changes have all been influential:

- The industry is maturing as exhibited through consolidation;
- The core technology has vastly improved with increased CPU performance and refined algorithms;
- Open standards (like Voice-XML) and industry protocols (like MRCP) are becoming prevalent across the solution stack;
- Best practices around speech solutions have emerged;
- Hosted and premise-based managed services have introduced new deployment and more economically palatable cost options for clients;
- Companies are achieving solid results with their speech deployments and overall awareness has improved.

However, the hefty price tag associated with speech solutions still stands in the way for many businesses. In addition, many businesses are satisfied with their existing DTMF solutions and have not reached a crisis point where they absolutely see the need to move to speech. The high average price of a speech deployment is rooted in the fact that most are still custom developed, via vendor professional services. Therefore, the bulk of investments in speech solutions have been confined to the early adopter large organizations (communications, financial services and travel & tourism) that have had the need for speech automation and the budgets to afford expensive custom application development work.

Over the years, the cost of custom application development engagements has decreased to some degree. Professional services teams have become more adept at leveraging pre-built modules and templates to expedite deployment time, and as the number of speech specialists and developers has grown, hourly rates have dropped. But custom application development still involves coding, design, tuning and integration service, none of which are cheap. In response, vendors working to bring speech mainstream began bundling reusable code assets and productizing these as configurable packages. These are gaining traction in the market as new markets embrace speech.

To date, the majority of spending on speech applications has been for custom developed applications and this will remain the case through 2012. However, the proportion of spending on packaged applications, configurable packages and pre-built modules and templates will grow over this time.

The following figure and table highlight spending on speech applications over the next five years. As shown, spending on custom developed applications will grow from \$366 million in 2007 to reach \$753 million by 2012, at a CAGR of 15.5%, pre-built modules and templates will increase from \$43 million to \$215 million at a CAGR of 37.7%, and packaged applications and configurable packages will grow from \$56 million to \$259 million, at a CAGR of 30.5%.

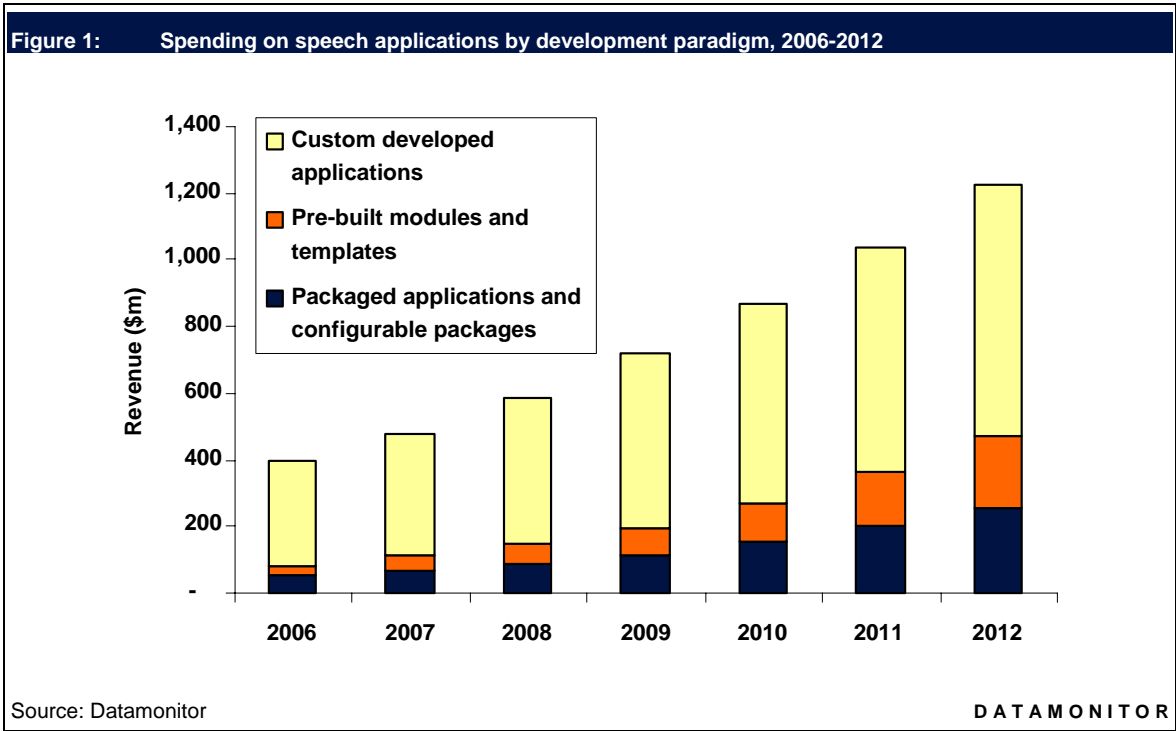
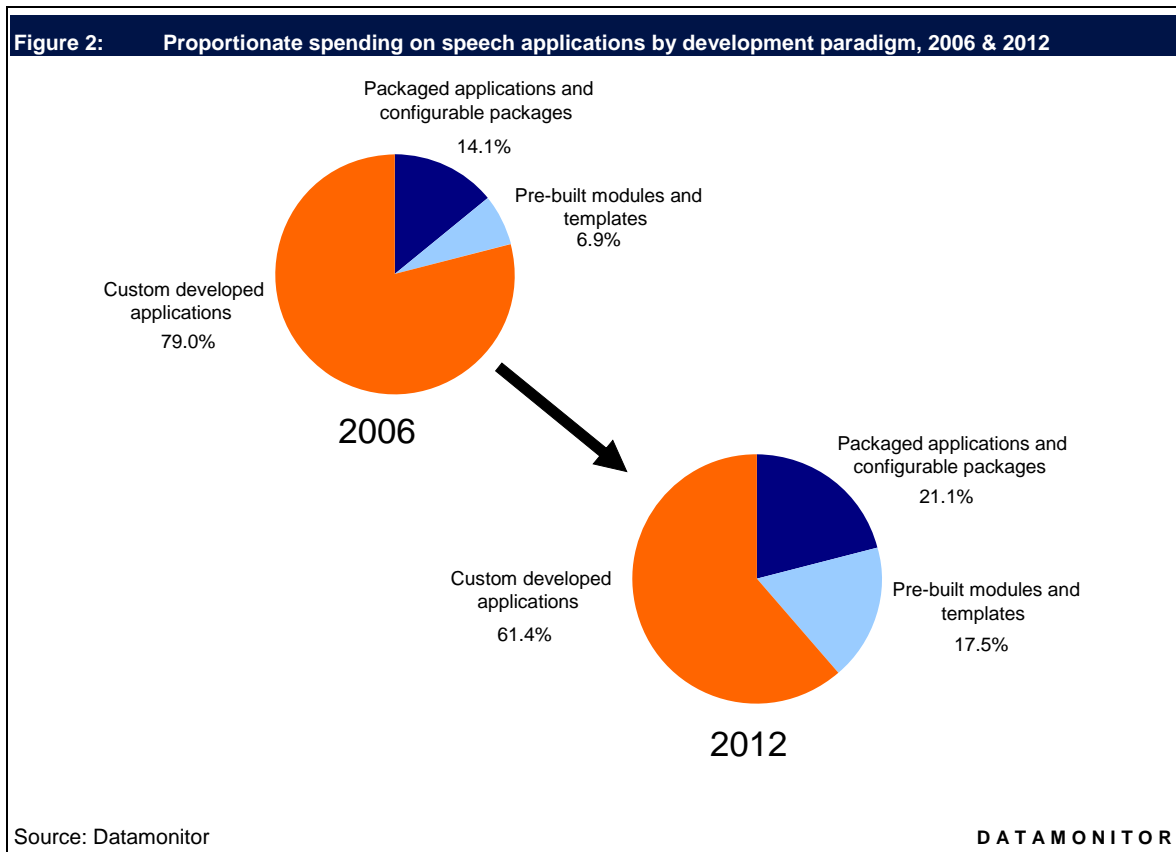


Table 1: Spending on speech applications by development paradigm, 2006-2012

Revenue (\$m)	2006	2007	2008	2009	2010	2011	2012	CAGR 2007-2012
Packaged applications and configurable packages	56	68	89	116	153	200	259	30.5%
Pre-built modules and templates	27	43	60	81	117	162	215	37.7%
Custom developed applications	313	366	436	521	601	675	753	15.5%
Total	396	478	585	719	872	1,037	1,227	20.7%

Source: Datamonitor DATAMONITOR



As shown in Figure 2, from 2006 to 2012 there will be sizable shifts in proportion of spend as more companies opt for packaged applications and configurable packages and pre-built modules and templates. Custom developed applications accounted for 79.0% of spending on speech applications in 2006, packaged and configurable packages 14.1% and pre-built modules and templates 6.9%. By 2012, custom developed applications will account for 61.4% of spending, packaged applications and configurable packages 21.1% and pre-built modules and templates 17.5%. This shift will be caused by increased spending on configurable packages in new markets, pre-built modules and templates by early adopters and the general evolution of the speech market that will nurture software package advancement.

Adoption of speech technology is crossing Moore's chasm and moving beyond early adopter markets (communications, financial services and travel & tourism) to pragmatist and conservative markets. In recent years, new markets including retail, healthcare, technology, entertainment, media & leisure, public sector and utilities have made significant investments in speech. Speech application requirements among many of the companies in these markets are not as stringent as those of the early adopter markets where call volumes are the greatest. Significant layers of customization are typically not needed by companies in these new markets and budgets are more prohibitive. This presents opportunities for configurable packages as they are cheaper than custom developed applications and provide adequate levels of configuration to complement specific business processes.

Pre-built modules and templates reduce early adopters' dependency on vendor professional services

Communications, financial services and travel & tourism companies – early speech adopters – are veterans when it comes to speech, having had speech deployments for several years. The TCO of a custom developed speech application is high as it requires heavy dependency on vendor professional services to make changes and upgrades to the applications over time on top of the initial costs for initial design and deployment. As we roll out the tape to 2012, early adopters will strive to become more self-sufficient in speech application development and management in efforts to reduce the TCO of the speech solution and to leverage their existing in-house staffs. There have been many unsuccessful speech applications developed in-house by some of the early adopters, regardless, these companies will continue to obtain more control and self-sufficiency when it comes to speech and this will be a growing trend over the next five years. They will shadow vendor professional services and engage in a blend of in-house development utilizing pre-built modules and templates and piece-mealed vendor professional services. This will lead to increased spending on pre-built modules and templates. Reliance on vendor professional services will not go away among early adopters but it will be reduced over the next five years.

The evolution of the speech market will help nurture software package advancement

Over the past few years there has been heavy consolidation in the IVR, contact center, enterprise applications and communications markets which has also led to heavy mergers and acquisitions (M&A) activity in the speech market as vendors move to establish long-term positions in their respective markets and other markets. This consolidation signals maturing market conditions across contact center and broader CRM and phone-based technologies. Some of the major M&A activities in the speech market are highlighted below:

- Scansoft's (Nuance) merger with Nuance and acquisitions of BeVocal, VoiceSignal and Dictaphone;
- Microsoft's acquisition of Tellme;
- Genesys' acquisition of VoiceGenie and GMK;
- Interville's acquisition of Edify and Nuasis;
- Tuvov's acquisition of Netbytel;
- HP's acquisition of Pipebeach;
- Cisco's acquisition of Audium.

Heavy consolidation is changing the landscape of the voice business market as illustrated in Figure 4. By analyzing the Herfindahl-Hirschman Index (HHI) concentration measure for the global supply-side voice business market (platforms, enabling software, applications, services and hosting), data suggests that the market's concentration level has increased steadily for the past few years. This gauge in addition to the elapsed time indicates a maturing market. The HHI concentration measure is equal to the sum of the squares of the market shares of all vendors. The following is the standard equation for HHI measurement:

$$HHI = \sum_{i=1}^f S_i^2$$

Where f = number of firms participating in the industry | S_i = each firm's market share | i = firm in a given industry

The HHI measure can range between zero – when thousands of firms are competing and have close to 0% market share – and 10,000 when a single firm accounts for 100% of the market, indicating a monopoly. The US Department of Justice uses the HHI for evaluating mergers and provides the following guidelines to classify market concentration levels:

- HHI measure < 1,000 is an unconcentrated market;
- 1,000 < HHI measure < 1,800 is a moderately concentrated market;
- 1,800 < HHI measure is a highly concentrated market.

Figure 3 illustrates the increase in HHI concentration levels from 2004 to 2007 for the global supply-side voice business market. As shown, the market was more fragmented in 2004 where the HHI measure was less than 700 – this is classified as an unconcentrated market. Over the next three years the HHI measure for the voice business market rose to just above 1,000, crossing over to the moderately concentrated market category. This rise in the HHI was primarily due to mergers and acquisitions followed by organic vendor market share growth. Over the next five years more M&A activity will occur in the applications, services and hosted areas of the voice business value chain paralleling the shift in value within the value chain, towards applications and services and away from platforms and enabling software. This will cause the HHI measure to further increase as vendors' increase market share.

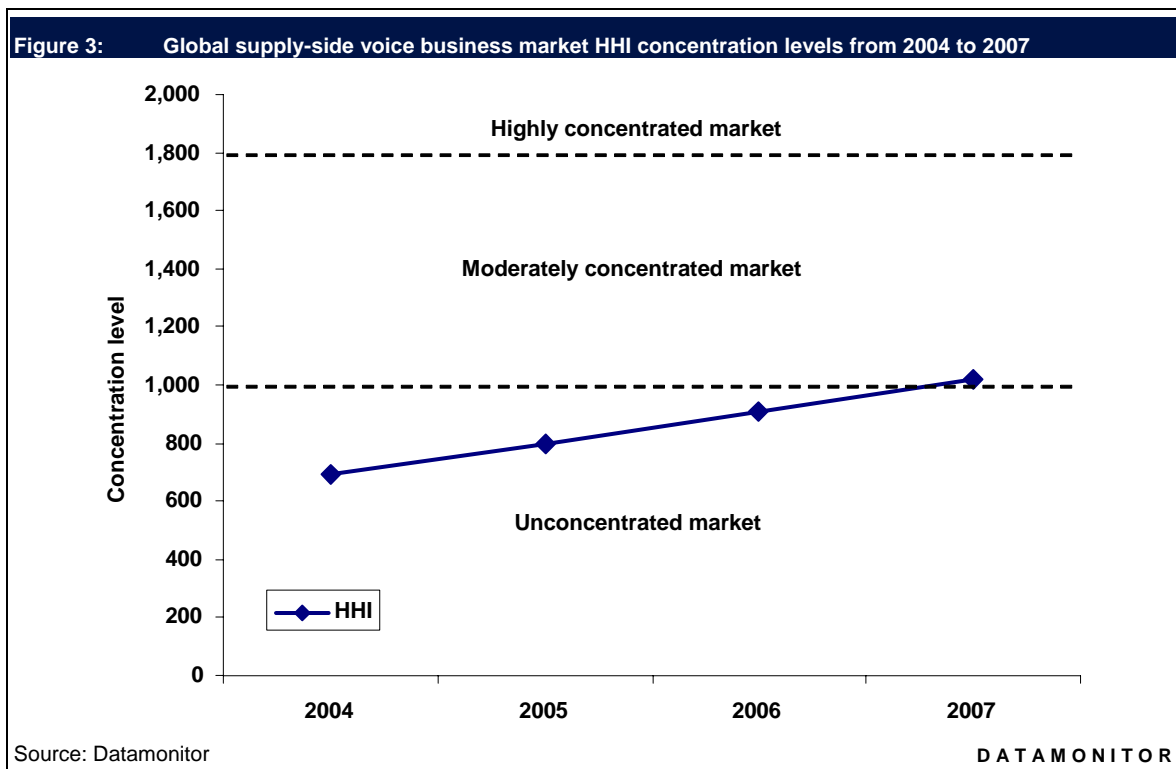
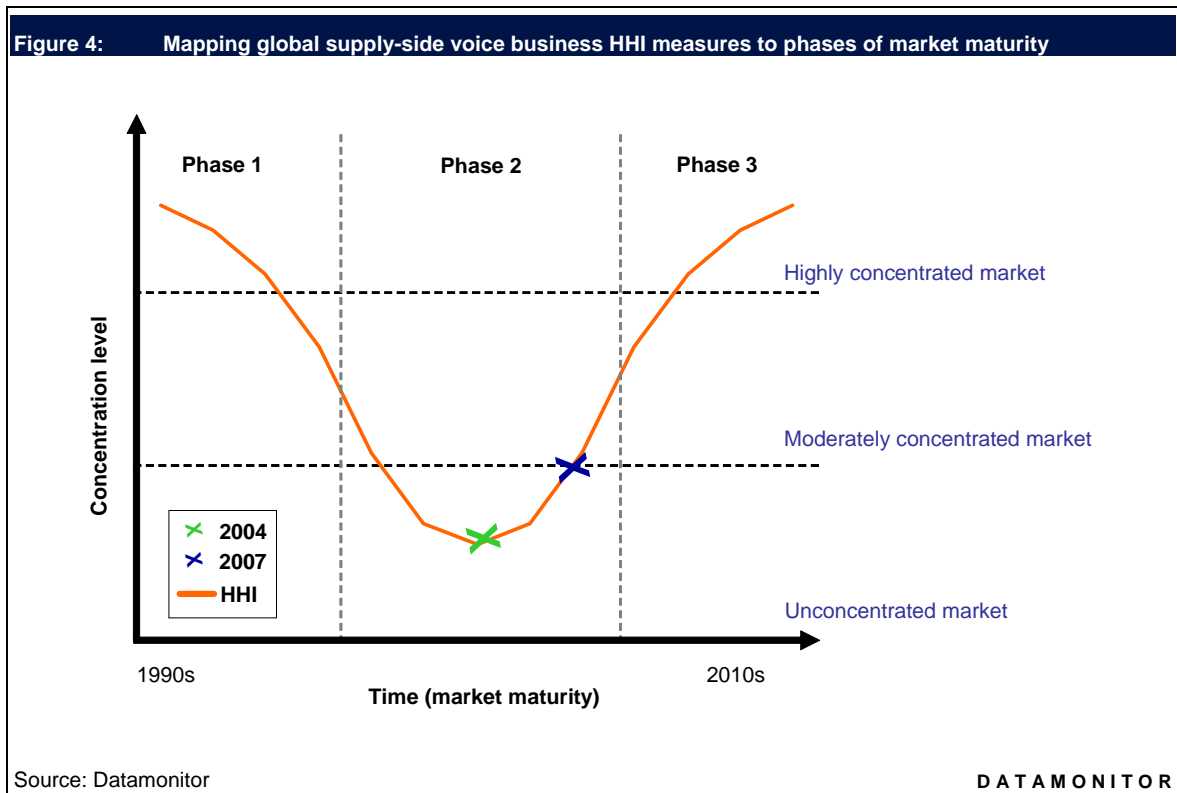


Figure 4 presents the market maturity curve and plots the HHI measures from Figure 3 along this curve. As illustrated, the global supply-side voice business market is currently in Phase 2 of market maturity and is moving towards Phase 3 over the next several years. The global supply-side voice business HHI measure for 2004 is plotted on the maturity curve in green while the HHI measure for 2007 is plotted in blue. Phases 1 through 3 are defined below.

- **Phase 1** – A nascent market with very few vendors that have majority market share. Barriers to entry are high;
- **Phase 2** – An emerging and competitive market. Barriers to entry have come down and new entrants enter the market causing fragmentation, downward pricing pressure and increased competition. Consolidation ramps up towards the latter half of this phase as commoditization occurs and differentiation becomes harder;
- **Phase 3** – A mature market. Saturation has been reached in the addressable market. Heavy M&A activity has occurred and the majority of market share is held by a handful of large enterprise software vendors and systems integrators.

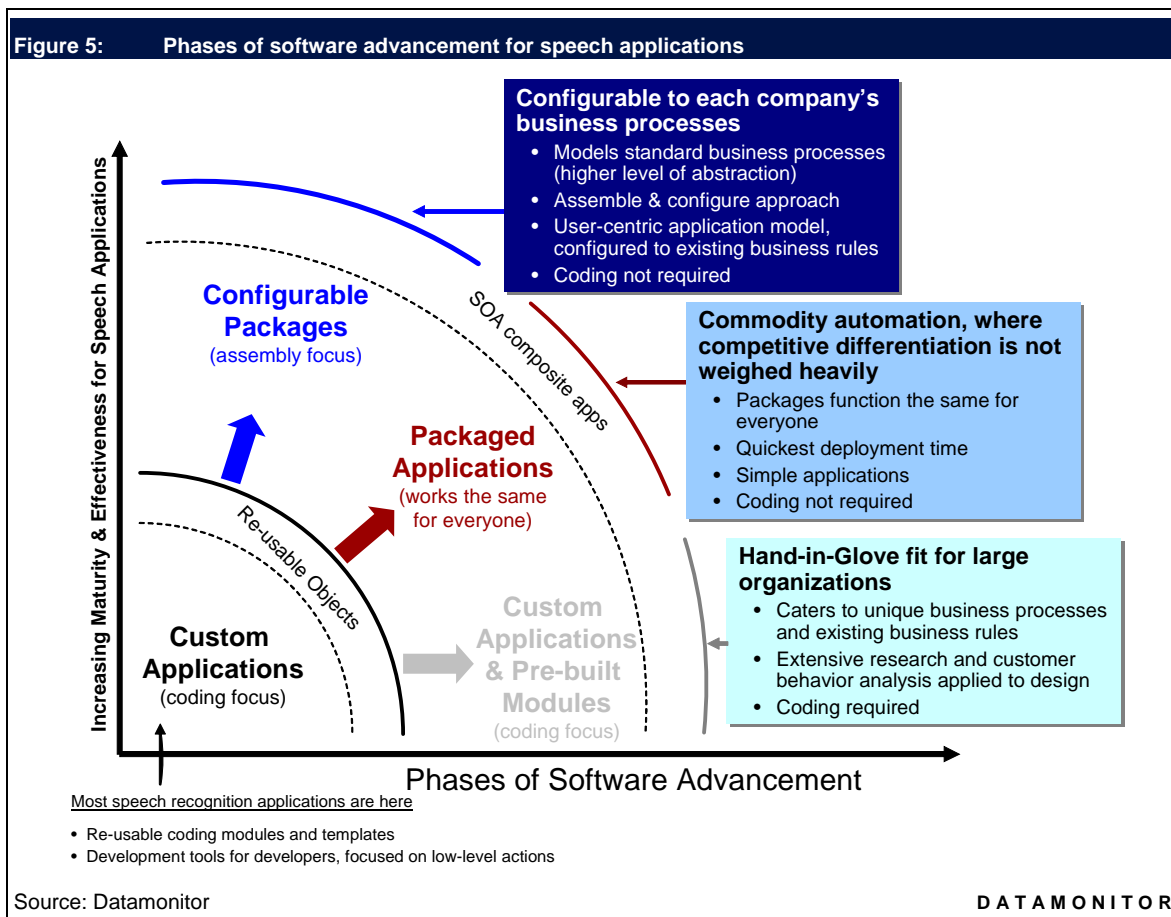


In parallel with the maturing speech market, speech applications and their effectiveness are maturing as well. In most software markets, applications begin as custom developed applications then develop into packaged type applications once enough reusable code assets emerge. In most markets, businesses have shown a strong preference for packaged over custom developed applications due to lower TCO, quicker deployment, reduced code risk, higher cost certainty and

inclusion of ongoing product updates. Businesses generally purchase two types of packaged software: packaged applications (for commodity processes) and configurable packages (for strategic processes). When companies need a basic, commodity business solution, they typically purchase off-the-shelf packaged applications. These solutions, while offering a minor degree of customization, work the same for all buyers. However, when a business process is a differentiator then companies want their solutions to complement their own, unique approach to the market. Companies have three choices here, configurable packages, pre-built modules and templates or custom developed applications.

Configurable packages has had strong success in ERP and CRM as it is superior to custom developed applications for streamlining operational costs without sacrifice to the competitive processes that help make their business unique. Configurable packages are a more advanced phase of software delivery – providing an ‘assemble and configure’ approach. Low-level coding and reusable code modules are replaced by high-level business process abstractions.

Figure 5 illustrates the phases of software advancement and maps configurable packages in relation to packaged applications and custom developed applications.



The consolidation and maturation of the speech market is driving vendors to introduce new differentiators in and outside of the application level. Within the context of applications, vendors can no longer differentiate themselves on the availability of a configurable package, custom development capabilities or pre-built modules and templates but rather on their unique

approaches to application design, deployment and long-term value propositions for clients. The speech market is maturing at a steady rate and indicative of this is the M&A activity and the rise of packaged type applications. As the market further progresses more packaged applications, configurable packages and pre-built modules and templates will surface in the market. There will be more abundance of these packaged type applications for niche and broader markets which will lead to increase spending on packaged applications, configurable packages and pre-built modules and templates.

APPENDIX

Definitions

ASR (automatic speech recognition)

Software engine that listens to and recognizes spoken words. In most cases it processes the incoming audio to isolate words, splits these words into segments (usually phonemes or diphones), and then statistically compares these segments with a linguistic database. Depending on the word spoken, a value is returned, normally with a degree of confidence.

Contact centers

Datamonitor defines a contact center by the following features:

An Automatic Call Distributor (ACD) or Private Branch Exchange (PBX) with equivalent functionality overlaid (or soft ACD);

- 10 or more agent positions;
- Agent positions are desks from which agents make and/or receive telephone calls to and/or from internal or external customers. This is taken to imply that the call in question involves communication between the agent and the customer.

Specifically excluded from these figures are:

- Public safety centers, i.e. those centers that receive calls to the emergency services, which are counted separately and are not included as call centers;
- Air traffic control;
- Financial trading floors.

Legal interception centers, i.e. centers engaged in legal interception, where there is a law enforcement officer or other security worker listening in on a conversation in which they do not take part.

Dual tone multi-frequency (DTMF)

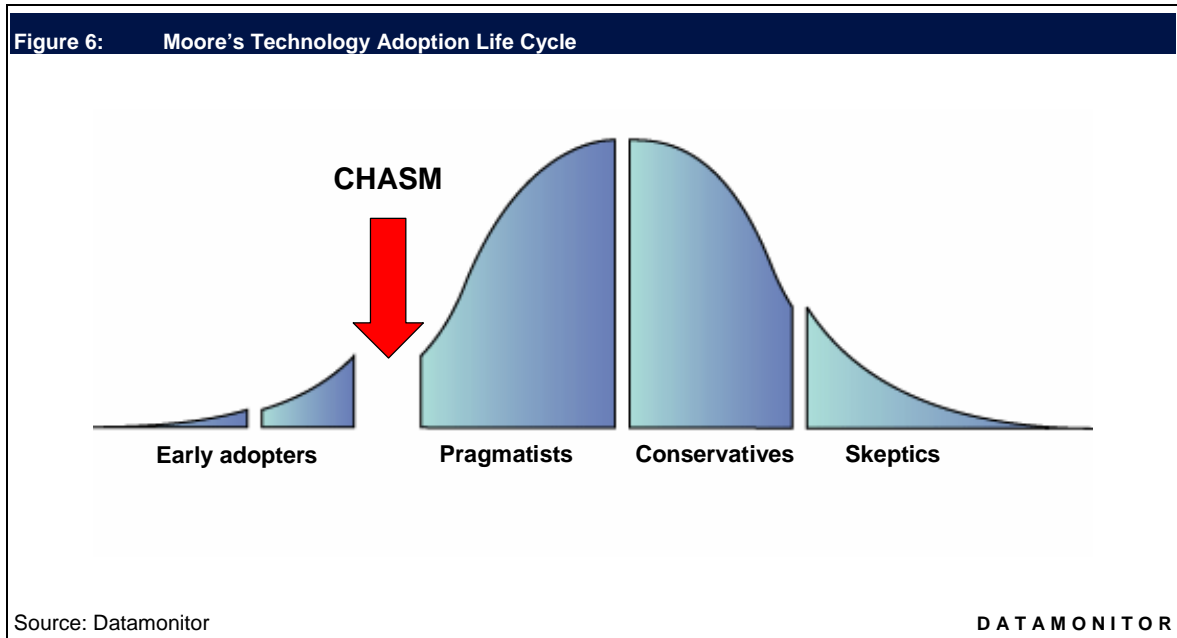
The signal to the phone company that a caller generates when he/she presses keys on a telephone's keypad. In North America and it is commonly known as touchtone phone (formerly a registered trademark of AT&T). DTMF has generally replaced loop disconnect ('pulse') dialing.

Interactive voice response (IVR)

A technology that analyzes a sequence of spoken and/or DTMF commands and reproduces voice prompts to the caller, the call is then routed via switch or serviced wholly within the IVR that is linked to a database. The IVR interacts with key systems, PBXs, ACDs through analog ports, digital ports and LAN/WAN connectivity.

Moore's chasm

Geoffrey Moore's theory built on the idea that the rate of diffusion in the 'Technology Adoption Life Cycle' curve is not continuous in high tech markets.



Open standards

The development of standards and standards-based platforms has challenged the proprietary siloed structure that is prevalent in traditional IVR systems. Standards offer the opportunity for platforms to be written in a standard language, thus rendering them interoperable with engines and applications developed by any other vendor, as long as the same language is used. Already in its second version, Voice-XML is more established than newer alternatives such as SALT, and is the dominant standard, with a growing sphere of deployments and developers surrounding and supporting it.

Platform

An IVR (interactive voice response) platform. Technology that analyzes a sequence of spoken and/or DTMF commands and reproduces voice prompts to the caller, the call is then routed via switch or serviced wholly within the IVR that is linked to a database. The IVR interacts with key systems, PBXs, ACDs through analog ports, digital ports and LAN/WAN connectivity. The platform segment refers to traditional IVR, standards-based platforms and browsers.

Voice-XML

Voice-XML is the World Wide Consortium's (W3C) standard markup language based on XML used for creating voice user interfaces that use advanced speech recognition (ASR) and text-to-speech (TTS) technologies. Since its commercial release in 2000, Voice-XML has emerged as the dominant open standard in IVR technology and is today's most widely deployed open standard for IVR implementations, with a growing sphere of deployments and developers surrounding and

supporting it. Although it was initially created for speech recognition solutions, platforms based on this open standard have been widely deployed for DTMF solutions as well.

Methodology

- Interviews with over 12 leading vendors in the speech application space.
- The Definitive Guide to Speech Applications (Interactive Model) – IMTC0140
- Interviews with over 10 organizations that have invested or are evaluating speech solutions.

Further reading

Packaged, Pre-built and Custom Developed Speech Applications (Strategy Focus) – BFTC1783

The Definitive Guide to Speech Applications (Interactive Model) – IMTC0140

Ask the analyst

The Technology Knowledge Center Writing team

Daniel Hong, Lead Analyst, Customer Interaction Technologies

dhong@datamonitor.com

Datamonitor consulting

We hope that the data and analysis in this brief will help you make informed and imaginative business decisions. If you have further requirements, Datamonitor's consulting team may be able to help you. For more information about Datamonitor's consulting capabilities, please contact us directly at consulting@datamonitor.com.

Disclaimer

All Rights Reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, Datamonitor plc.

The facts of this report are believed to be correct at the time of publication but cannot be guaranteed. Please note that the findings, conclusions and recommendations that Datamonitor delivers will be based on information gathered in good faith from both primary and secondary sources, whose accuracy we are not always in a position to guarantee. As such Datamonitor can accept no liability whatever for actions taken based on any information that may subsequently prove to be incorrect.