



um SMS”, aposta o CEO da BizVox, Flávio Osso. A empresa também pretende oferecer a solução diretamente para as operadoras, para venda no modelo “white label”.

O reconhecimento de fala também pode ser usado em jogos multiplayer que simulam universos virtuais nos quais diversos jogadores interagem simultaneamente. Em vez de usar um website ou comandos via SMS, o participante fará uma ligação para um número local e dará instruções orais para o seu personagem. A BizVox planeja lançar um jogo assim. A idéia é cobrar pelo serviço uma assinatura mensal pequena, entre R\$ 3 e R\$ 5, e buscar apoio de alguma operadora de telefonia, tendo em vista que o jogo pode gerar bastante tráfego de voz.

Buscas de informações para lazer, como endereços de restaurantes e hotéis usando reconhecimento de fala, se tornarão realidade em breve. A revista “Quadro Rodas”, da editora Abril, pretende lançar o serviço logo.

Soluções para a venda de produtos através de comandos de voz também devem surgir este ano por aqui, acreditam especialistas. As maiores interessadas são empresas que vendem ingressos para shows, cinemas e teatros.

Nos EUA, já existe venda de bilhetes aéreos e até mesmo de ações da Bolsa de Valores por meio de sistemas de reconhecimento de fala.



## “O SISTEMA DE RECONHECIMENTO DE VOZ NÃO DARÁ CERTO SE A APLICAÇÃO NÃO FOR BEM ESCRITA.”

Reinaldo Bernardi, do Instituto Genius

Outra utilidade existente nos EUA e que deve chegar ao Brasil em um futuro próximo é em pesquisas de opinião pública feitas pelo telefone. Através de discadores automáticos, os institutos de pesquisa ligam para os entrevistados, que conversarão com uma máquina dotada de tecnologia de reconhecimento de fala, que transformará as respostas orais em dados.

### Tecnologia

O que está por trás desses sistemas de reconhecimento de voz é um módulo chamado “Automatic Speech Recognition” (ASR). Ele é uma espécie de banco de dados onde estão registradas as palavras. O ASR trabalha com a fonética. As diferentes pronúncias para uma mesma

**DE INTERESANTE  
DUVIDOSO ATÉ HÁ  
POUCOS ANOS,  
O SERVIÇO ACABOU  
GANHANDO  
POPULARIDADE E  
PODERÁ SER USADO  
ATÉ PARA COMANDO  
EM JOGOS**

palavra, em decorrência dos sotaques, são transcritas em fonemas dentro do ASR. A maioria dos desenvolvedores de ASR é estrangeira, como a norte-americana Nuance e a italiana Loquendo. parceira da BizVox. Os módulos de ASR são vendidos por idioma.

No Brasil, o Instituto Genius desenvolveu um ASR para o português brasileiro e, mais recentemente, um para o

espanhol latino-americano. O projeto demandou mais de um ano de trabalho de uma equipe multidisciplinar, formada por lingüistas e engenheiros, dentre outros profissionais. Foram coletadas amostras de voz de mais de mil locutores, para a versão em português. O Brasil foi dividido em cinco regiões, de acordo com os sotaques. Foram consideradas também as pequenas variações na fala, decorrentes da faixa etária e do sexo do locutor.

O sistema capta o sinal de áudio e compara o que foi dito pelo locutor com o banco de dados, para identificar as palavras. O barulho externo pode atrapalhar na identificação, embora haja filtros para reduzir esse problema. Freqüências muito diferentes da voz humana são descartadas. A taxa de acerto em laboratório gira em torno de 90%. Quando uma aplicação é lançada, costuma cair para 70%. Mas aos poucos vai sendo aperfeiçoada, com acréscimo gradual de palavras e sotaques. Esse processo de aperfeiçoamento é chamado de “fine tuning” pelos especialistas. Quando a máquina não consegue de forma alguma entender o que foi dito, o ideal é transferir a chamada para um atendente humano. Isso pode acontecer se a pessoa tiver um sotaque muito carregado, for gaga ou fanha.

O sucesso de uma aplicação, contudo, não depende apenas da qualidade do ASR, mas, principalmente, da programação do software. Cabe ao desenvolvedor definir as palavras que

## Telemar: resultados positivos no primeiro mês

O projeto teve cuidado especial para as diferenças de vocabulário nos 16 Estados onde a operadora atua

No fim de março, a Telemar adotou a tecnologia de reconhecimento de fala em seu call center. A medida foi tomada depois que pesquisas indicaram que os clientes não estavam satisfeitos com o atendimento das URAs. Agora, em vez de digitar no teclado do telefone, basta falar o que deseja para ser encaminhado a um atendente responsável. Algumas tarefas mais simples, como emissão de segunda via, desbloqueio de linha e verificação de mudança de endereço podem ser feitos somente pela máquina.

O projeto levou nove meses para ser desenvolvido. Houve um cuidado especial para as diferenças de vocabulário nos 16 Estados onde a operadora atua. Muitas gírias precisaram ser adicionadas ao dicionário do sistema, fornecido pela Nuance. No Ceará, por exemplo, dizer que o telefone “está no prego” significa dizer que está “com defeito”.

Levando em conta que o call center da Telemar é um dos maiores do País, com 77 milhões de chamadas atendidas no ano passado, esse pode ser entendido como um teste de fogo para a tecnologia de reconhecimento de fala. Em pouco mais de um mês de funcionamento o resultado foi positivo. A taxa de ligações encaminhadas corretamente estava acima de 70%. Isto é superior à taxa do sistema antigo, que era de 65%. “As pessoas se enrolavam para digitar as opções porque a árvore era muito grande”, lembra o diretor de serviços para clientes do varejo da Telemar, Hélivio Ferreira.

No primeiro mês houve redução de 20% no tempo de duração das chamadas e uma diminuição de 30% na quantidade de transferências de ligações. A quantidade de chamadas atendidas sem participação de atendentes humanos subiu de 16% para 23%.

serão consideradas “válidas”. Ou seja: aquelas que serão reconhecidas pelo sistema quando pronunciadas e, portanto, provocarão uma ação. “O ASR pode ser bom, mas se a aplicação não for bem escrita, não dará certo”, ressalta o gerente de pesquisa e desenvolvimento do Instituto Genius, Reinaldo Bernardi.

A linguagem mais popular para o desenvolvimento de aplicações com reconhecimento de fala é a “Voice Extensible Markup Language” (VXML). Ela é considerada uma das razões para a popularização dessa tecnologia nos últimos anos. A VXML facilitou o transporte de amostras de voz via IP entre servidores, o que barateou sensivelmente os custos de criação de aplicações de reconhecimento de fala.

### Máquina versus homem

A escolha da voz que interage com os consumidores é cuidadosa, afinal ela reflete a imagem da empresa. Muitas das frases são gravadas por uma pessoa de verdade. Outras são ditas pela máquina, através da tecnologia de text-to-speech, que transforma texto em fala. O objetivo é fazer com que a máquina pareça o mais humana possível. “O segredo está na prosódia”, revela José Carlos Cunha, sócio da LM Sistemas, que implementou reconhecimento de fala no serviço de auxílio à lista da CTBC, ainda em 1999. O executivo se refere à cadência da fala, à métrica, às pausas entre as palavras. Dependendo de como a máquina for ordenada a falar, a frase pode soar mais artificial ou mais natural.

Entretanto, por mais que tentem, as máquinas jamais substituirão completamente

## Tecnologia com mil e uma utilidades

APLICAÇÃO	O QUE A MÁQUINA FAZ?
Telefonista virtual	Encaminha chamadas para os devidos ramais de um PABX.
Auxílio à lista	Fornecer números telefônicos a partir do nome ou endereço do assinante.
Entrega de conteúdo celular	Vende jogos, ringtones e wallpapers para celular.
Call center de operadoras	Resolve pedidos simples, como emissão de segunda via, e encaminha demais chamadas para o atendente humano adequado.
Guias	Fornecer informações como endereço de hotéis e restaurantes.
Vendas	Ingressos para shows, cinema, teatro.
Jogos	Realiza a interação do participante com o universo virtual do jogo, através de comandos de voz.
Pesquisas	Liga através de discador automático, faz perguntas para pesquisa de opinião pública e registra dados com as respostas do entrevistado.

o atendimento humano, concordam especialistas. Tarefas como retenção de clientes e suporte técnico, por exemplo, continuarão a cargo de funcionários em carne e osso.

“Na prática, a fala serve para substituir a discagem das opções na URA. Ela não substitui o atendente”, afirma o presidente da Associação Brasileira de Telemarketing, Topázio Silveira Neto. Esta é uma meia verdade. Cada vez há mais solicitações que podem ser resolvidas sem a participação de um atendente humano. Um call center que adotou a tecnologia de reconhecimento de

fala, mas prefere não se identificar, informa que em poucos anos aumentou de 18% para 40% a proporção de chamadas atendidas sem a interferência de funcionários. Contudo, isso não significa necessariamente que haja demissões. Em vez disso, os atendentes podem ser remanejados para outras funções. “Não pretendemos demitir ninguém”, garante o diretor de serviços para clientes do varejo da Telemar,

Hélio Ferreira. Segundo ele, o objetivo de adotar a tecnologia de reconhecimento de fala é melhorar a qualidade do atendimento ao cliente, não reduzir custos.

O BRASIL FOI DIVIDIDO EM CINCO REGIÕES, DE ACORDO COM OS SOTAQUES, FORAM CONSIDERADAS TAMBÉM AS PEQUENAS VARIAÇÕES NA FALA