

# AUTOMATIZAÇÃO DE SERVIÇOS DE INFORMAÇÃO TELEFÓNICA

Gonçalo Marques  
INESC

Isabel Trancoso  
INESC/IST

Luís Oliveira  
INESC/IST

António Serralheiro  
INESC/IST

INESC

R. Alves Redol, 9, 1000 Lisboa

## SUMÁRIO

Este artigo aborda a automatização dos serviços de informação telefónica, com especial ênfase nos serviços de consulta directa e inversa de listas telefónicas AudioSIT<sup>1</sup>. Começa por discutir formas de diálogo homem-máquina utilizando o terminal telefónico, prosseguindo com uma descrição detalhada dos vários protótipos desenvolvidos no INESC e concluindo com algumas considerações breves sobre a integração de reconhecimento automático de fala e de síntese de fala a partir de texto neste tipo de serviços.

## 1. INTRODUÇÃO

A vulgarização dos serviços telefónicos de valor acrescentado associado ao desenvolvimento do processamento automático de chamadas (Automated Call Processing) e à consultada automática a bases de dados tem chamado a atenção para processos de interacção utilizando simplesmente o terminal telefónico.

As tecnologias da fala são fundamentais para este tipo de interacção nomeadamente o reconhecimento de fala e a síntese de fala a partir de texto [3]. O estado da arte nestes dois campos não nos permite ainda falar com o computador em linguagem natural e a sua “voz” também não é ainda muito natural. No entanto, para um grande leque de aplicações, é possível impôr certas restrições, de modo a poder usar efectivamente o que está já disponível. Assim, muitos sistemas funcionam na base de um diálogo mais ou menos restrito no qual todas as mensagens são geradas pelo computador, concatenando palavras previamente gravadas e as únicas respostas aceites por este pouco excedem o conjunto dos dígitos. Outros sistemas adoptam uma abordagem ainda mais simples

que, em vez de reconhecimento de dígitos falados, usa detecção multi-frequência para distinguir quais os dígitos premidos no teclado telefónico.

Para muitos dos serviços telefónicos de valor acrescentado actualmente em vigor, a imposição destas restrições não constitui um factor impeditivo, sendo extremamente competitivas de um ponto de vista de preço, relativamente à utilização de operadoras humanas e permitindo o seu rápido alargamento a outros tipos de bases de dados. A utilização deste tipo de sistemas audiotexto tem assim crescido exponencialmente, em paralelo com o progresso em reconhecimento e síntese de fala.

Os serviços automatizados de informação telefónica permitem ao utente navegar numa base de dados de informação mais ou menos complexa, através de um menú hierarquicamente estruturado de opções. Como exemplos, podem citar-se os serviços de informação sobre comboios, aviões, museus, exposições, farmácias de serviço, cinemas, teatros, restaurantes, instruções de preenchimento de formulários, consulta de listas telefónicas, etc. Em muitos serviços, o utente não tem apenas o papel passivo de obter informação, gerando ele próprio informação que ficará contida na base de dados. O caso mais típico desta classe de sistemas é o das encomendas via telefone, indicando o código do produto e o número do cartão de crédito, por exemplo. Em qualquer destes serviços a validação do acesso à base de dados pode ser feita mediante a introdução de números de identificação pessoais. Outros serviços mais complexos implicam a geração e/ou o reencaminhamento de chamadas. Está nesta classe, por exemplo, o serviço de despertar automático.

O artigo começa por introduzir formas de diálogo homem-máquina utilizando o terminal telefónico, prosseguindo com uma descrição detalhada dos vários protótipos desenvolvidos, com particular ênfase no do serviço de consulta telefónica Au-

<sup>1</sup> O INESC reserva para si todos os direitos de propriedade intelectual do serviço AudioSIT

dioSIT. Conclui com algumas considerações breves sobre a integração de reconhecimento automático de fala e de síntese de fala a partir de texto neste tipo de serviços.

## 2. FORMAS DE DIÁLOGO UTILIZANDO O TERMINAL TELEFÓNICO

A grande maioria dos sistemas de audiotexto implica não só a reprodução de palavras e frases previamente gravadas, mas também a leitura oral de valores numéricos. Esta pode assumir variadíssimas formas desde a leitura de dígitos isolados usada também no serviço de consulta, à leitura de valores monetários (e.g. 123\$50 - cento e vinte e três escudos e cinquenta centavos), datas (e.g. 28/2/94 - vinte e oito de Fevereiro de mil novecentos e noventa e quatro), horários (21h30m - vinte e uma horas e trinta minutos), etc. Sempre que a leitura de numerais não deva ser feita dígito a dígito, há que efectuar um pré-processamento do valor em questão de modo a produzir uma cadeia de palavras.

As mensagens faladas pelo sistema resultam tipicamente da concatenação da informação que este foi buscar à base de dados, contendo ou não valores numéricos, com frases portadoras (exemplo: “O saldo da conta bancária n.º ... é de ...”). A concatenação deve ser feita tendo em conta a coarticulação e a entoação tipicamente utilizadas na leitura do tipo de informação em questão. Isto implica por vezes a gravação da mesma palavra em contextos diferentes e algum processamento adicional, pelo que a adopção ou não destas medidas resulta muitas vezes de um compromisso entre a memória disponível e a naturalidade exigida para a mensagem sintetizada.

A entrada de dados impõe sérias restrições ao diálogo dado que se dispõe apenas de um teclado com 10 a 12 teclas. Daí a maior parte das aplicações oferecer apenas os 10 dígitos como opções normais, reservando tipicamente as teclas “\*” e “#” para funções especiais (e.g., menú de instruções, saída de emergência, terminador, etc.). De modo a permitir a entrada de palavras através deste reduzido leque de teclas, cada dígito é habitualmente associado a um conjunto de letras (e.g., 1:ABC, 2:CDE, etc). Mesmo ignorando as vogais acentuadas e o c cedilhado é necessário gerar os 23 símbolos do alfabeto português usando apenas as 10 teclas numéricas. Este problema é resolvido num teclado convencional quer pela utilização simultânea de duas teclas (Control e Shift, por exemplo), quer pela utilização de uma ou mais teclas que mudam o valor das seguintes (p.e., CAPS LOCK). No caso do terminal telefónico, a primeira hipótese é claramente inviável e a segunda limitaria ainda mais o número de teclas disponíveis. Outra alternativa seria a utilização de variações temporais (um toque rápido para a primeira, letra mais longo para a segunda e ainda mais longo para a

terceira) o que teria o inconveniente de exigir uma grande disciplina por parte do utilizador e de estar fortemente sujeito a erros.

A solução encontrada consistiu em marcar os dígitos das teclas que contenham as letras da palavra pretendida e deixar ao sistema a tarefa de identificar a palavra com o auxílio de um dicionário, rejeitando as combinações de letras impossíveis [2]. No entanto este processo poderá produzir uma resposta ambígua, isto é, o número de palavras correspondente a uma dada sequência de teclas é frequentemente superior a um. Suponhamos, para exemplificar, que se utilizava apenas um léxico de nomes próprios, e que as letras “A” e “B” eram associadas à mesma tecla, o mesmo acontecendo com “J” e “L”. Neste caso, “JOÃO” corresponderá à mesma sequência de teclas que “LOBO”, exigindo assim a utilização de processos de desambiguação. Desta forma, transfere-se a possibilidade de erros por parte do utilizador para um grau de incerteza no diálogo com o sistema.

Este tipo de associação tecla-letras tem sido utilizado noutros países como auxiliar de memorização de certos números de telefone. Por exemplo, em qualquer telefone dos Estados Unidos da América, a marcação de 1-800-PORTUGAL liga-nos à delegação do Instituto do Comércio e Turismo Português (ICEP) em Nova Iorque. O facto de não existir uma norma internacional para esta associação levou-nos a tentar encontrar uma distribuição que minimize a ambiguidade mencionada para as palavras da Língua Portuguesa. Este estudo levou-nos a uma combinação optimizada para um dicionário de nomes próprios, mantendo como critério de base a sequencialidade das letras do alfabeto, mas permitindo um número variável de letras por tecla.

Como medida de ambiguidade, adoptou-se a percentagem de palavras com descodificação ambígua (que resulta em mais do que uma palavra do dicionário) em relação ao número total de palavras do dicionário. Na aproximação inicial começou-se por permitir que cada uma das 10 teclas tivesse entre 1 e 4 letras. Esta restrição em conjunto com a sequencialidade das letras resulta em 112035 combinações possíveis. Estudou-se também a possibilidade de deixar, para além das teclas \* e #, outra tecla disponível para sinalização, atribuindo letras a apenas 9 das teclas com dígitos, o que reduz substancialmente o número de combinações possíveis. Posteriormente foram também testadas associações de 5 letras a uma tecla. A optimização da associação teclas-letras foi feita com base em 3 corpora distintos:

**Nomes** Corpus com cerca de 15000 palavras diferentes, correspondendo aos nomes de pessoas e companhias de maior frequência das lista telefónica de Lisboa e Porto (excluindo siglas e nomes estrangeiros)

ros) com 3 ou mais letras. Os acentos foram retirados e as palavras truncadas a 9 caracteres.

**PF** Corpus “Português Fundamental” recolhido pelo Centro de Linguística da Universidade de Lisboa (CLUL) em entrevistas orais realizadas em todo o País [1]. Contém as cerca de 21000 palavras mais frequentes, com 3 ou mais letras. Os acentos foram retirados e as palavras truncadas a 9 caracteres.

**Pal** Corpus misto de nomes próprios e comuns, contendo cerca de 96000 palavras diferentes.

O primeiro teste de optimização efectuado com o corpus Nomes, com base em 10 teclas, teve como resultado uma ambiguidade mínima de 3,3%. A utilização de apenas 9 teclas piorou este resultado para 4,6%. As 1000 melhores associações foram utilizadas em ensaios com os corpora PF e Pal, tendo-se obtido uma ambiguidade mínima de 2,1% (10 teclas) e 2,9% (9), para o primeiro corpus e de 9,3% (10 teclas) e 10,5% (9), para o segundo. Estes resultados mostram um nível surpreendentemente baixo de ambiguidade para dicionários de tamanho intermédio (cerca de 20.000 palavras).

Para cada um dos conjuntos de teclas adoptado (10 e 9), a melhor combinação foi seleccionada com base na média dos valores de ambiguidade obtidos para os três corpora, tendo-se obtido valores mínimos de 5,0% e 6,4%, respectivamente. Utilizando 10 teclas, a melhor combinação é a óptima para o corpus PF (2,1%), resultando numa ambiguidade adicional de 0,04% para o corpus Pal e de 0,32% para Nomes, relativamente aos valores mínimos encontrados para cada um destes corpora. Com 9 teclas, a melhor combinação é também a óptima para o corpus Nomes, fornecendo uma ambiguidade adicional de 1,11% em relação à combinação óptima para o corpus PF e de 0,06% para o corpus Pal.

Os resultados obtidos com as melhores combinações para 10 e 9 teclas foram posteriormente comparados com os obtidos utilizando as associações teclas-letras mais vulgarizadas nos telefones existentes no nosso país e no estrangeiro, tendo-se obtido valores de ambiguidade que medeiam entre 7,7 e 9,3%, para a média dos 3 corpora.

### 3. SERVIÇOS DESENVOLVIDOS

De particular interesse para o INESC, dada a sua experiência passada no desenvolvimento do Sistema de Informação Telefónica para os TLP, é a automatização deste tipo de serviço recorrendo à detecção de tons multi-frequência e à síntese de mensagens pré-gravadas. São dois os tipos de serviço integrados numa aplicação deste género. No primeiro, tipicamente designado por consulta telefónica directa, o utilizador pretende saber a morada e/ou número de telefone de um dado utente, dado

o seu nome. No segundo, relativamente menos frequente e designado por consulta telefónica inversa, a informação fornecida pelo utilizador é o número de telefone do assinante do qual pretende obter o nome e/ou morada. Nesta fase de desenvolvimento, o sistema fornece apenas informação sobre números de telefone, embora o software esteja já preparado para aceder a todo o tipo de informação disponível nas listas telefónicas. Descrever-se-ão também menos detalhadamente os serviços de despertar automático, encaminhamento de chamadas e correio vocal.

### Arquitectura do sistema

Para o desenvolvimento deste tipo de sistemas, o INESC adoptou uma plataforma modular com base numa rede de PCs equipados com placas de interface telefónica que realizam a codificação e descodificação de fala, a detecção e geração de tons multi-frequência, o controlo automático de ganho e a detecção de silêncios para um conjunto de canais telefónicos. Para aplicações mais complexas, tais como a consulta oral da lista telefónica, incorpora-se na rede uma máquina de maior capacidade, tipicamente com um processador RISC usando o sistema operativo UNIX, que funcionará como servidor de informação. Cada placa pode ligar a 4, 8 ou 12 linhas telefónicas, ou a uma interface para canais E1 e T2. Pode ainda ser acoplada a placas específicas para reconhecimento de sinais gerados por marcadores decádicos e reconhecimento de fala.

O software está assim estruturado em duas partes diferentes: a primeira corre num ambiente UNIX e é responsável pelo processamento e a procura do nome introduzido. A segunda parte corre no sistema operativo DOS, e controla a placa de interface telefónica. Estas duas partes estão ligadas via rede Ethernet, e a informação necessária é também transmitida usando o protocolo de comunicação TCP/IP.

Para este tipo de aplicações, é fundamental processar várias chamadas simultaneamente, sendo necessário que os sistemas operativos utilizados tenham capacidades multi-tarefa (*multi-tasking*). Se cada pedido de informação fosse atendido sequencialmente, os atrasos causados a outros utilizadores inviabilizariam qualquer utilidade prática para este tipo de serviços. A utilização multi-tarefa não introduz quaisquer dificuldades do ponto de vista da implementação no ambiente UNIX, visto esta ser uma capacidade que lhe é inerente. No sistema operativo MS-DOS, no entanto, não existem capacidades de multi-tarefa, tendo cada função de ser executada sequencialmente. A placa de interface telefónica contorna este problema partilhando a execução do programa entre o microprocessador da placa e o CPU do PC. O sistema permite a uma dada aplicação executar uma série de fun-

ções sem que se tenha de esperar pela conclusão das mesmas. Esta característica é essencial para aplicações que tenham de gerir múltiplas linhas telefónicas, especialmente se são utilizadas funções que tenham um período de execução demorado como são, por exemplo, a síntese oral ou a gravação de mensagens, dado que, enquanto estas funções são executadas, o programa pode atender a outros pedidos concorrentes.

A informação sobre os assinantes é controlada por um sistema de gestão de base de dados INGRES e está estruturada em diversas bases, uma para cada zona do país. Por sua vez, os dados de cada assinante estão armazenados numa estrutura contendo vários campos: nome, morada, telefone, porta, zona, etc... As consultas às bases de dados são feitas através de um servidor desenvolvido para o sistema INGRES, que recebe pedidos de busca por um ou mais campos desta estrutura. No actual serviço de consulta directa só é permitido introduzir nome do assinante, e a busca é feita, por defeito, à base de dados da zona de Lisboa. No entanto, o software já está estruturado para poder não só aceder a outras zonas do país, como também permitir pesquisas por vários campos.

### **Consulta Telefónica directa e inversa**

De um ponto de vista de automatização, a consulta inversa é de implementação mais fácil, visto que a cada número de telefone indicado pelo utilizador corresponde um único utente. A resposta produzida pelo sistema pode ou conter o nome e a morada desejados ou limitar-se a informar que o número de telefone digitado não está registado na base de dados. Já o mesmo não se passa com a consulta directa. Em primeiro lugar porque a um mesmo nome, nomeadamente se incluir apenas nomes de baptismo/apelidos bastante vulgares (e.g. Manuel Silva) podem corresponder dezenas de assinantes, e em segundo lugar devido à ambiguidade introduzida pelo teclado do terminal telefónico.

No sistema desenvolvido (AudioSIT), a ambiguidade começa por ser parcialmente removida recorrendo ao corpus Nomes acima descrito, o qual garante uma cobertura superior a 90% de todos os nomes de assinantes do país, em termos de palavras isoladas e de cerca de 85% em termos de nomes completos. Embora a associação teclas-letras adoptada foi a seleccionada para um conjunto de 10 teclas utilizando a medida de ambiguidade mínima mencionada, a adaptação do programa a outra forma de associação é imediata.

Após este processamento inicial, o sistema faz uma busca do nome ou nomes correspondentes às teclas introduzidas na base de dados. A consulta é feita via rede (usando o protocolo TCP/IP) ao servidor da base de dados. O resultado da busca pode assumir três formas distintas:

no caso de não existirem assinantes com o nome introduzido, uma mensagem com esta informação é transmitida ao utilizador; no caso de haver demasiados assinantes com o mesmo nome (mais que 100 na actual aplicação), é pedido ao utilizador para introduzir mais nomes; finalmente, quando existe um número reduzido de assinantes (inferior a 5), os nomes completos e os respectivos números de telefone são reproduzidos oralmente. Dado que se trata apenas de um protótipo, a informação pré-gravada inclui apenas os nomes de pessoas que ocorrem mais de 100 vezes na base de dados. Sempre que um ou mais nomes do assinante procurado não constarem nesta informação pré-gravada, estes são abreviados para a sua letra inicial. Qualquer dos limites máximo e mínimo para o número de assinantes é programável pelo gestor da aplicação.

A base de dados põe alguns problemas específicos. Em primeiro lugar, o número de grafias alternativas para um mesmo nome pode ser bastante elevado (> 40). Algumas são aceitáveis, decorrendo, por exemplo, da substituição de “s” por “z” e de “e” por “i” (ou vice-versa), da omissão de “h”, etc.. Estas grafias alternativas estão por vezes tão vulgarizadas que se torna difícil determinar qual a original (e.g. Sousa e Souza). Mais frequente porém é a omissão de acentos e cedilhas, o que por vezes traz problemas de ambiguidade (Franca e França), e a troca de caracteres contíguos ocorrida durante a introdução do nome. Por conseguinte, a um mesmo nome gravado pode corresponder várias grafias diferentes. Na determinação desta tabela de correspondências foi fundamental o trabalho efectuado no projecto Europeu LRE Onomastica, no qual o INESC conta com a colaboração do CLUL, tendo a empresa TLP como parceiro associado. De facto, o estudo da pronúncia dos nomes próprios efectuado neste projecto permite associar facilmente todas as entradas da base de dados a que corresponde uma mesma transcrição fonética.

Um outro dos problemas da pronúncia de nomes próprios diz respeito às empresas. Um grande número de empresas adopta palavras do léxico comum para a sua designação (e.g. Instituto Superior Técnico), ou ainda nomes de localidades e apelidos. Em qualquer dos casos, as palavras mais frequentemente utilizadas estão também contidas na lista das palavras pré-gravadas. Muitas vezes, porém, estas empresas são designadas pelas siglas correspondentes. A dificuldade em desambiguar nomes de empresas constituídos por uma única “palavra” com um número muito reduzido de caracteres torna preferencial a introdução do nome completo da firma, em vez da respectiva sigla, ao pretender aceder ao respectivo número de telefone. Outras empresas adoptam acrónimos para a sua designação, ocorrendo, por conseguinte, muito poucas vezes na base de dados. A versão actual do sis-

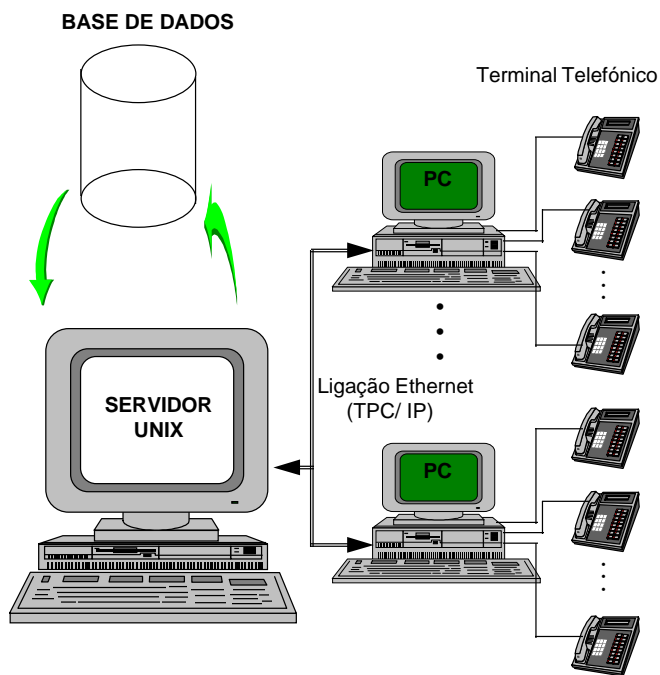


Figura 1. Arquitectura do sistema AudioSIT

tema não inclui assim acrónimos na lista de palavras pré-gravadas. Está previsto o desenvolvimento de uma nova versão que incorpore opções específicas para empresas, como sejam a leitura de acrónimos e o acesso aos números de telefone dos serviços pertencentes à mesma empresa (e.g. biblioteca). Está também prevista a normalização da interface com a base de dados para a utilização do protocolo X.500.

### Despertar Automático

Uma das aplicações desenvolvidas pelo INESC sobre a plataforma acima descrita tem como objectivo a automatização do serviço de despertar. Este sistema, ainda em fase de protótipo, oferece ao utilizador as opções de introduzir, verificar ou cancelar pedidos de despertar, através do teclado do terminal telefónico. O sistema mantém uma lista de pedidos activos e gera chamadas quer para confirmação de pedidos ou cancelamentos, quer para o despertar propriamente dito. Permite também uma integração fácil com o sistema de facturação.

### Encaminhamento de Chamadas

Uma outra das aplicações desenvolvidas tem como objectivo o encaminhamento de chamadas. Este tipo de aplicação pode ser facilmente integrado em empresas dispondo de centrais telefónicas próprias com capacidade de encaminhamento de chamadas, mas sem as facilidades do “Direct Dialing Inward”. A versão actual pede ao utente o número da extensão para onde deseja falar e efectua a transferência de chamada no caso do número

pretendido se encontrar desocupado. Está em curso o desenvolvimento de uma nova versão que não pressupõe o conhecimento do número da extensão por parte do utilizador. Neste caso, o sistema pede a este que introduza os nomes da pessoa com quem deseja falar utilizando a correspondência tecla-letras que descrevemos no serviço de consulta directa para aceder à lista de extensões telefónicas existentes localmente.

### Correio Vocal

A mesma plataforma serviu ainda para o desenvolvimento de um protótipo de correio vocal. Este tipo de serviços permite a um assinante deixar mensagens faladas que serão posteriormente ouvidas por outro assinante [4]. Concretamente, no protótipo do INESC, o assinante dispõe de várias opções, nomeadamente: gravação, audição, difusão e gestão de mensagens. A gravação, iniciando-se com a identificação do assinante, prossegue com a identificação do destinatário (número de telefone), após o que se passa à gravação da mensagem. No caso da leitura, o assinante é informado do número de mensagens que tem por ouvir, e pode optar por ouvir os primeiros  $n$  segundos de cada uma, ou por ouvir na íntegra a mensagem número 4, por exemplo. O sistema pode ainda indicar-lhe que há três mensagens por ouvir do assinante José Silva, da extensão 2298. A difusão é disponibilizada entre assinantes de um dado grupo (mensagens específicas relacionadas com um dado trabalho) ou em geral.

Este conceito de “caixa de correio” vocal pode também ser vantajosamente aplicado a um serviço de informação telefónica genérico, permitindo a um utente deixar mensagens gravadas que serão posteriormente ouvidas pelo gestor do serviço.

## 4. RECONHECIMENTO E SÍNTESE DE FALA

A substituição do teclado do telefone por um sistema de reconhecimento de fala [6] é obviamente uma das perspectivas de evolução para este tipo de sistemas. De modo a atingir este objectivo, há que tornar os actuais sistemas de reconhecimento robustos relativamente à variabilidade dos falantes (quer em termos físicos, quer de sotaque ou estilo), do ambiente, dos microtelefones, das linhas telefónicas, etc. Há também que impôr menos restrições à utilização de palavras fora do vocabulário, de frases gramaticalmente incorrectas e de outros fenómenos frequentes na fala espontânea.

Complementarmente, é fundamental investir também na área da síntese a partir de texto [5], de modo a permitir “ler” a informação contida nas bases de dados a que o sistema acede via rede telefónica, sem impôr qualquer restrição de vocabulário, permitindo assim a sua perma-

nente actualização. Há portanto que melhorar significativamente a inteligibilidade e naturalidade dos actuais sistemas, incorporando variantes de vozes e estilo adequados à aplicação, em termos de ritmo e entoação.

A capacidade de conversar naturalmente com uma máquina pressupõe a utilização de reconhecimento e de síntese de fala e também a sua integração efectiva com sistemas de processamento da linguagem natural. Pressupõe ainda um forte investimento numa área que em termos de fala está ainda muito incipiente - a das interfaces homem-máquina. Hoje em dia, o comportamento do utilizador modifica-se de modo a adaptar-se às limitações do sistema e as mensagens por este sintetizadas são propositadamente adaptadas de modo a ultrapassar a sua reduzida naturalidade e inteligibilidade. Os futuros sistemas terão que aprender a reconhecer pistas que sinalizem incerteza, repetições e correcções por parte do utilizador e serem eles próprios capazes de dar pistas prosódicas sinalizando a sua intenção ou grau de incerteza.

## 5. AGRADECIMENTOS

Os autores gostariam de agradecer aos Eng.<sup>os</sup> Manuel Mira e João Schiappa o desenvolvimento do servidor que permite comunicar com a base de dados do SIT (Serviço de Informação de Telefones). Gostariam ainda de agradecer em especial ao Prof. Luís Vidigal o incentivo e o entusiasmo com que tem apoiado o desenvolvimento do AudioSIT <sup>2</sup>.

## REFERENCES

- [1] F. Nascimento, L. Marques, L. Segura, *Português Fundamental: Métodos e Documentos*, INIC-CLUL, Lisboa, 1987.
- [2] L. C. Oliveira, I. M. Trancoso, G. C. Marques, "Composição de Palavras em Português Utilizando um Teclado Numérico", Relatório interno, INESC, 1994.
- [3] D. O'Shaughnessy, *Speech Communication – Human and Machine*, Addison-Wesley Series in Electrical Engineering: Digital Signal Processing, 1987.
- [4] M. Parker, *The Practical Guide to Voice Mail*, Osborne McGraw-Hill, 1987.
- [5] L. C. Oliveira, M. C. Viana, I. M. Trancoso, "DIXI – Portuguese Text-to-Speech System", Eurospeech'91, Génova, 1991.
- [6] A. J. Serralheiro, *Metodologias Probabilísticas no Reconhecimento de Palavras Isoladas*, Tese de Doutoramento, IST-UTL, Lisboa, 1990.

---

<sup>2</sup>O INESC reserva para si todos os direitos de propriedade intelectual do serviço AudioSIT