

SISTEMA DE RECONHECIMENTO DE FALA COM RUÍDO PARA AUTOMÓVEIS

André G. Chiovato, Francisco J. Fraga, Fernando L. Attia, Giovanne F. Amaral
Nilson J. S. Martins, Rodrigo B. Brito, Carlos A. Ynoguti

INATEL – Instituto Nacional de Telecomunicações

Av. João de Camargo, 510 – Bairro Inatel

CEP. 37540-000 Sta. Rita do Sapucaí, MG – Brasil

Tel.: (035) 3471-9354 Fax: (035) 3471-9314

<http://www.inatel.br>

fraga@inatel.br

Resumo - Este trabalho mostra o desenvolvimento de um sistema de reconhecimento automático de fala a ser usado no interior de um automóvel em pleno trânsito, sujeito a diversos tipos de ruídos inerentes a esta aplicação, tais como: ruído de tráfego, ruído do motor e dos pneus, do vento e da chuva. O artigo está dividido em duas partes: descrição da montagem de uma Base de Dados de Fala gravada em situações reais; e descrição de um projeto prático, apresentado em uma feira tecnológica com o intuito de averiguar a viabilidade técnica e a aceitação comercial do referido sistema.

Abstract - This article shows the development of an automatic speech recognition system for cars in traffic, which are exposed to several noise types: traffic noise, engine noise, noise from tires, wind and rain. The article is divided in two parts: description of a Speech Data Base recorded under real conditions and description of a practical project, shown in a technology fair with the aim of verifying the technical feasibility and commercial acceptance of the referred system.

I. INTRODUÇÃO

Até mesmo os melhores sistemas de Reconhecimento Automático de Fala – *RAF* [1], sofrem uma substancial degradação de seu desempenho quando trabalham com sinais de fala corrompidos por ruído. Na situação específica de reconhecimento de fala no interior de um automóvel em trânsito, a relação sinal/ruído (em dB) chega a ser negativa [2].

O reconhecimento de nomes (agenda) para discagem em telefones celulares (usando sistema de viva-voz para automóveis) e o uso de comandos de voz para controle de funções de automóveis (faróis, vidros, travas, etc.) são algumas das diversas aplicações existentes que estão sendo desenvolvidas por alguns alunos de Iniciação Científica pertencentes ao Grupo de

Pesquisa em Processamento Digital de Sinais do INATEL. Em muitas aplicações, os sistemas de *RAF* podem ser extremamente atraentes, como por exemplo na situação em que o usuário está dirigindo o seu automóvel, e conseqüentemente, seus olhos e mãos estão inteiramente ocupados nesta tarefa. Vejamos algumas justificativas que comprovam sua aplicabilidade nas seguintes situações: Em congestionamentos de tráfego, tão comuns na vida moderna, usar a voz parece ser uma boa forma de aproveitar o tempo, cada vez mais escasso. Ou em viagens longas, especialmente durante a noite, onde o controle de voz sobre os faróis nas ultrapassagens (comandos do tipo “farol alto” e “farol baixo”), pode diminuir a sonolência do motorista. Nestas circunstâncias, por que não utilizar um Sistema de Reconhecimento de Fala em Ambiente Ruidoso – *RAFAR* [3], como interface para navegar na Internet, discar para o telefone de um amigo ou cliente, ou comandar via voz o ar condicionado e/ou outras funções (vidros, faróis, buzina, toca-fitas ou CD, etc.) do próprio carro?

Neste artigo descreveremos os métodos utilizados e os resultados parciais obtidos em dois projetos distintos, porém interligados: “*Otimização de Sistemas de Reconhecimento Automático de Comandos de Voz baseados em Modelos Ocultos de Markov*”, projeto de Iniciação Científica¹ orientado pelo Prof. Dr. Francisco Fraga; e “*Sistema de Reconhecimento Automático de Comandos de Fala para Automóveis*”, projeto prático apresentado na 20^a Feira Tecnológica do Inatel – *FETIN’2001*. Sendo assim, tanto a metodologia quanto os resultados são apresentados em duas partes. A primeira (relacionada com o projeto de Iniciação Científica) trata da confecção de uma Base de Dados de

¹O projeto está sendo desenvolvido por dois alunos, um deles com bolsa da *FINATEL* – Fundação Instituto Nacional de Telecomunicações e outro com bolsa da *FAPEMIG* – Fundação de Amparo à Pesquisa do Estado de Minas Gerais.

Fala que servirá para o treinamento e teste de um sistema de *RAFAR*. A segunda parte (relacionada com o projeto prático) diz respeito à situação mercadológica de sistemas de *RAF* para automóveis, mostrando sua viabilidade técnica – comandos de voz adequados e comodidade para o usuário; e sua viabilidade comercial – aceitação do mercado, i.e., interesse por parte dos usuários de optarem por automóveis com este tipo de acessório.

II. MÉTODOS UTILIZADOS

A. Montagem da Base de Dados de Fala com Ruído

Descreveremos primeiramente a montagem de uma Base de Dados de Fala gravada no interior de um automóvel, sob diversas condições de ruído ambiente (que serão detalhadas mais adiante).

O sinal de fala captado já vem corrompido por ruídos inerentes à natureza da própria aplicação: ruído do motor e dos pneus, ruído urbano, ruído de vento e chuva, etc. Assim, tendo em vista a futura implementação de um sistema de *RAFAR* que produzisse um aumento da relação sinal/ruído antes do reconhecimento (*speech enhancement*) [4], a gravação do sinal de voz foi feita por meio de dois diferentes microfones: *Handheld Recorder Microphone – M10* e *Super Directional Desktop Microphone – M60*, da empresa *Telex Communications*, cujas ilustrações podem ser vistas nas Figuras 1 e 2.



Fig.1 - M10



Fig.2 - M60



Fig.3 - TCD-D100

Os dois microfones foram posicionados estrategicamente sobre o painel do carro, de modo a realizar uma captação homogênea e estereofônica (dois canais) do sinal de fala proferido pelo motorista. A gravação dos sinais foi feita por meio de um gravador digital (*DAT – Digital Audio Tape*) portátil, modelo *TCD-D100* da Sony, conforme ilustra a Figura 3. A Figura 4 mostra o cenário de gravação com os microfones direcionados para o motorista, conectados ao *DAT* através de 2 cabos de áudio blindados.

Após as gravações no interior do automóvel com o uso do *DAT*, os sinais de fala corrompidos por ruído foram transferidos para o computador (tipo *desktop*). Foram então separados em arquivos e nomeados, com a ajuda do software *Wave Studio* da *Creative Sound Blaster*.

A nomeação dos arquivos foi feita segundo uma sistemática criada pelo grupo, na qual primeiramente aparece o comando (o vocabulário completo dos comandos será listado a seguir) e sua respectiva ordem de locução (1ª, 2ª ou 3ª vez pronunciada), seguida do estilo de percurso (avenida, paralelepípedo, rodovia) e por último a situação do vidro (aberto, semi-aberto ou fechado). Segue um exemplo:

TemperaturaInternal_Av_Ab.wav

ou seja, foi gravado o comando “temperatura interna”, pela primeira vez, na avenida, com os vidros abertos.

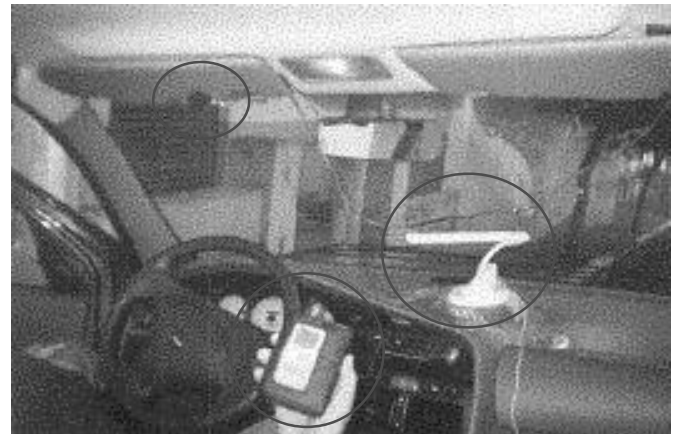


Fig.4 - Lay-out do interior do carro

TABELA I
Vocabulário Completo da Base de Dados de Fala

Comandos para telefone celular	Comandos para controle do carro
1. Bloquear	1. Farol Alto
2. Desbloquear	
3. Menu	2. Farol Baixo
3.1. Agenda	
3.1.1. Nomes	3. Pisca – Alerta
3.1.2. Número	
3.1.3. Polícia	4. Travas
3.1.4. Bombeiros	
3.1.5. Emergência	5. Capô
3.2. Chamadas	
3.2.1. Atender	6. Porta – Malas
3.2.2. Receber Mensagem	
3.3. Personalizar	7. Ventilação
3.3.1. Nome (do usuário)	
3.3.2. Perfil original	8. Alarme
3.3.3. Campanha	
3.3.3.1. Aumentar	9. Tanque
3.3.3.2. Diminuir	
3.4. Assistente Pessoal	10. Combustível
3.4.1. Calendário	
3.4.2. Atividades	11. Óleo
3.4.3. Despertador	
3.4.4. Calculadora	12. Abaixar Vidros
3.4.5. Administrador de contatos	13. Levantar Vidros
3. Funções Avançadas	14. Temperatura
3.1. Memória (tamanho)	
3.2. Conferência Telefônica	14.1. Interna
3.3. Rediscagem	14.2. Ambiente
4. Serviços Internet	14.3. Motor
4.1. Lazer	
4.1.1. Teatro	
4.1.2. Restaurante	
4.1.3. Shows	
4.1.4. Cinema	
4.2. Trânsito	
4.2.1. Livre	
4.2.2. Congestionado	
5. Desligar	

A Tabela I apresenta o vocabulário completo usado na confecção da Base de Dados de Fala, em formato de *Menu*.

B. Apresentação do Projeto Prático na FETIN'2001

Em segundo plano, este artigo também trata do efeito que o protótipo experimental de um sistema de *RAF*, apresentado como projeto prático na 20^a Feira Tecnológica do Inatel – *FETIN'2001*, causou sobre o público em geral. Tal projeto serviu para testar e analisar a situação mercadológica de *RAF* em automóveis, i.e., sua viabilidade técnica (comandos adequados, tempo de resposta, comodidade de uso e

segurança) e sua aceitação comercial (opcional atraente para os consumidores).

Foi utilizado um carro da marca *Fiat*, modelo *Marea SX 1.8 16v 4c - 4 portas*. O sistema de *RAF* apresentado ao público durante a feira não foi treinado com a Base de Dados de Fala descrita anteriormente. A justificativa deste fato radica na inexistência de um prazo razoável para que o grupo pudesse treinar e implementar seu próprio sistema de reconhecimento de fala, baseado na ferramenta computacional *HTK – Hidden Markov Model Toolkit*, versão 3.0; uma vez que a feira tecnológica acontece sempre no mês de Outubro e a montagem da base de dados foi concluída somente em Agosto de 2001. Com isso, foi necessário criar um novo conjunto funcional, apenas de caráter expositivo, dividido em duas partes [5], para que diversas funções do automóvel pudessem ser controladas pela voz.

A primeira parte consistiu em adaptar um programa de reconhecimento de fala (desenvolvido na Unicamp em linguagem *C++* pelo Prof. Dr. Carlos Ynoguti na sua tese de Doutorado [6]) de maneira que sua resposta fosse capaz de acionar um *driver* de controle externo ligado à *porta paralela* do *PC*. Este *driver* foi constituído de um *PLD – Programmable Logic Device* (modelo *EPM7128SLC84-15* da Altera Corporation) e de uma placa de circuito impresso de face simples com relés de contato seco, do tipo *NA* e *NF* (que acionavam os botões, alavancas e chaves correspondentes à cada função do automóvel).

A segunda parte, mais simples, é formada de um microfone (modelo *headphone*) unidirecional, conectado ao computador que continha o programa em *C++*; além de um sensor de toque fixado ao volante para detecção da presença do motorista (que acionava o sistema de reconhecimento através da mão esquerda, pois logicamente a outra estaria ocupada com o controle do câmbio do carro).

Este sistema de acionamento por toque foi monitorado pela *porta serial* do micro através do pino 3 (*nível alto* – motorista desejando comandar o veículo por meio da voz; *nível baixo* para o caso contrário).

TABELA II
Comandos de Fala Usados no Projeto da Fetin'2001

Abaixar Vidro Motorista	Apagar Farol
Abaixar Vidro Passageiro	Travar Portas
Abaixar Vidros Traseiros	Destruir Portas
Farol Alto	Buzina
Farol Baixo	Limpador de Pára-brisas
Levantar Vidro Motorista	Iluminação Interna
Levantar Vidro Passageiro	Ventilação Interna
Levantar Vidros Traseiros	Pisca Alerta

Na Tabela II são apresentados os comandos de fala que formam o vocabulário utilizado no projeto apresentado na FETIN'2001.

III. RESULTADOS

A. Montagem da Base de Dados de Fala com Ruído

De forma qualitativa, o resultado mais significativo que obtivemos consistiu na confecção da Base de Dados de Fala, que permitirá a otimização dos sistemas de reconhecimento de voz em ambientes adversos, mais especificamente no interior de um automóvel. Para realizar esta otimização, é necessário fazer um treinamento adequado com elocuições gravadas em situações reais [3]; portanto a base de dados confeccionada servirá perfeitamente a este propósito. Para realizar a gravação da base de dados completa foram rodados um total de 266 quilômetros, em um carro da marca *Fiat*, modelo *Pálio 4 portas 1.6 16V*.

Com base em experimentos práticos realizados, constatou-se as seguintes configurações ótimas (Tabela III) para a gravação do sinal de fala no interior do carro (DAT) e, em seguida sua transferência e armazenamento no computador:

Assim, a Base de Dados de Fala totalizou 5.344 arquivos .wav em um universo (vocabulário) de cerca de 50 palavras isoladas, cada uma ocupando em média 120 Kbytes, alocando aproximadamente 550 Mb de memória do disco rígido. Cada palavra do vocabulário foi pronunciada 3 vezes em cada situação (condição de ruído ambiental), por 2 locutores adultos (21 anos de idade) do sexo masculino. Esta base de dados será utilizada para o desenvolvimento do protótipo de um sistema de RAFAR dependente de locutor.

TABELA III
Configurações de gravação no DAT e no computador (PC)

SET-UP DEVICE	MIC ATTENUATION	REC MODE	REC LEVEL	F _a [khz]
DAT	20 [dB]	manual	4	44,1
PC	20 [dB]	manual	1	22,050

Na Tabela IV apresentamos alguns exemplos (não exaustivos) das diversas situações nas quais foram realizadas as gravações para a montagem da base de dados.

TABELA IV
Exemplos de algumas situações de gravação

Locutor	Local	Piso	Vidros	Veloc. média
André	BR 459	asfalto ruim	fechados	80 km/h
André	BR 459	asfalto ruim	semi-abertos	80 km/h
André	BR 459	asfalto ruim	abertos	80 km/h
André	Pouso Alegre	asfalto bom	abertos	35 km/h
Rodrigo	Pouso Alegre	asfalto bom	abertos	35 km/h
Rodrigo	Pouso Alegre	asfalto bom	semi-abertos	35 km/h
André	Pouso Alegre	asfalto bom	semi-abertos	35 km/h
André	BR 381/km 794	asfalto razoável	semi-abertos	80 km/h
Rodrigo	BR 381/km 775	asfalto razoável	abertos	80 km/h
Rodrigo	BR 381/km 794	asfalto razoável	fechados	80 km/h
Rodrigo	BR 459	asfalto ruim	abertos	80 km/h
André	BR 459	asfalto ruim	abertos	80 km/h
André	Av. João de Camargo	asfalto bom	fechados	35 km/h
André	Av. João de Camargo	asfalto bom	abertos	35 km/h
Rodrigo	Rua dos Marques	paralelepípedo	abertos	35 km/h
Rodrigo	Rua dos Marques	paralelepípedo	semi-abertos	35 km/h
Rodrigo	Rua dos Marques	paralelepípedo	fechados	35 km/h

Na coluna denominada “Local”, quando aparece o nome de uma rua ou avenida, estas situam-se no município de Santa Rita do Sapucaí – MG. Igualmente, a especificação “BR459” refere-se a trechos de estrada próximos à mesma cidade. Em Pouso Alegre (MG), o trajeto percorrido foi o da avenida principal que dá acesso à catedral. Na coluna “Vidros”, a denominação “Abertos” corresponde a uma abertura de 20 cm enquanto que “Semi-Abertos” corresponde a uma abertura de 10 cm. Ambas denominações referem-se apenas aos vidros dianteiros, pois os vidros traseiros permaneceram fechados. A ventilação interna estava desligada em todas as gravações.

B. Apresentação do Projeto Prático na FETIN'2001

Com relação à apresentação do trabalho ao público da FETIN, realizada nos dias 25 a 27 de Outubro de 2001 e que resultou no prêmio de 1º lugar dos projetos de Nível 5 (4º ano do curso de Engenharia de

Telecomunicações) com a maior pontuação de todos os trabalhos, a conclusão está em uma prévia comprovação da viabilidade técnica por parte dos visitantes. Segundo a Pró-Diretoria de Graduação do Inatel, 6.000 pessoas estiveram presentes ao longo do evento. Como o nosso projeto era bastante atraente e o automóvel (Fiat Marea) ficava próximo à porta de entrada da feira, centenas de pessoas puderam ver e testar o carro comandado pela voz.

Já a aceitação comercial envolve vários outros fatores de ordem não estritamente tecnológica. Trata-se, particularmente no caso da rede *Fiat*, do sistema *V.E.N.I.C.E.*, que prevê a substituição total da fiação elétrica do carro por um barramento de dados. Consequentemente, este fato provocaria ociosidade em boa parte da mão-de-obra (auto-elétrica) que os automóveis convencionais demandam nas montadoras e nas oficinas. Embora possa ocasionar um problema sindical, inicialmente tal opcional estaria disponível nos carros de categoria luxo, segundo os revendedores *Fiat*.

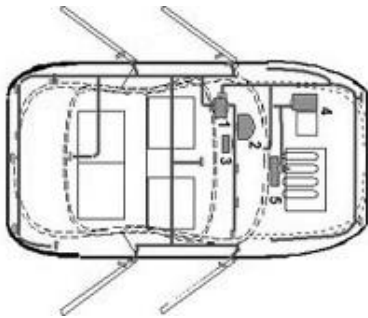


Fig.5 - Sistema V.E.N.I.C.E. (barramento de dados) da Fiat

IV. CONCLUSÕES

Pode-se afirmar que a Base de Dados de Fala obtida servirá de base para o desenvolvimento de um Sistema de Reconhecimento de Fala com Ruído em Automóveis. Porém, o grupo de pesquisa ainda não chegou efetivamente a um resultado que pudesse definir qual o tratamento do sinal de fala (*speech enhancement*) mais adequado a esta aplicação; nem tampouco está definida a configuração ideal dos diversos parâmetros do sistema de *RAFAR*. Entretanto, a apresentação do projeto prático não só motivou os integrantes do Grupo de Pesquisa em Processamento Digital de Sinais a continuar o desenvolvimento do sistema, como também provocou um aumento quantitativo e qualitativo do próprio grupo, pois alguns alunos de graduação e mestrado se dispuseram a aderir à nossa equipe, graças ao efeito que o automóvel comandado pela voz causou sobre os visitantes da *FETIN'2001*.

AGRADECIMENTOS

Agradecemos à FINATEL e à FAPEMIG pelo custeio das bolsas do projeto de iniciação científica que proporcionou a motivação para a publicação deste artigo. Agradecemos igualmente ao INATEL por fornecer as instalações e equipamentos necessários ao desenvolvimento do projeto.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] B. H. Juang; L. R., Rabiner, *Fundamentals of Speech Recognition*. Ed. Prentice-Hall, Signal Processing Series, New Jersey, 1993.
- [2] C. E. Mokbel; G. F. A. Chollet, "Automatic Word Recognition in Cars", *IEEE Trans. on Speech and Audio Processing*, vol 3, n.º 5, pp. 346-356, Sep. 1995.
- [3] J. C. Junqua; J. P. Haton, *Robustness in Automatic Speech Recognition – Fundamentals and Applications*, Kluwer Academic Publishers, Norwell, Massachusetts, 1996.
- [4] R. M. Stern; A. Acero; F. H. Liu; Y. Ohshima, "Signal processing for robust speech recognition", in *Automatic Speech and Speaker Recognition: Advanced Topics*. Eds. Norwell, MA, 1997.
- [5] Relatório do projeto "Sistema de Reconhecimento Automático de Comandos de Fala para Automóveis", *FETIN'2001*, Instituto Nacional de Telecomunicações (INATEL), outubro de 2001.
- [6] C. A. Ynoguti, *Reconhecimento de Fala contínua Usando Modelos Ocultos de Markov*, Tese de Doutorado, FEEC – UNICAMP, maio de 1999.