



UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO

CENTRO DE INFORMÁTICA

2012.2

SISTEMAS DE RECONHECIMENTO DE LOCUTOR INDEPENDENTE DE TEXTO

PROPOSTA DE TRABALHO DE GRADUAÇÃO



Aluno: Hector Natan Batista Pinheiro (hnbp@cin.ufpe.br)

Orientador: Tsang Ing Ren (tir@cin.ufpe.br)

Recife, janeiro de 2013

Contexto

Com o surgimento da telefonia e da Internet, a utilização da fala como modalidade de interação homem-máquina cresceu e, conseqüentemente, cresceu também a necessidade de se gerenciar uma quantidade cada vez maior de sinais de voz, presentes em áudios e aplicações multimídias.

Enquanto a área de reconhecimento de fala está concentrada em extrair mensagens linguísticas de uma dada locução, a de reconhecimento de locutor está concentrada em extrair a identidade da pessoa que produziu a locução.

Aplicações de reconhecimento de locutor são bem abrangentes e incluem: facilitar ou controlar o acesso a um computador [1,2]; autenticação por telefone para ligações distantes ou acesso a bancos [3]; máquinas que respondem de forma inteligente à requisições personalizadas pelo usuário [4]; e indexação e categorização de áudio a partir do locutor (*speech skimming*) [5,6].

Dependendo da aplicação, a área de reconhecimento de locutor é dividida em duas tarefas específicas: identificação e verificação. Em identificação de locutor, o objetivo é determinar qual pessoa de um grupo fixo de locutores produziu um determinado sinal de voz. Essa tarefa também é conhecida como identificação de locutor em um conjunto fechado. Já em verificação de locutor, o objetivo é determinar, a partir de uma locução, se uma pessoa é ou não quem ela afirma ser. Essa tarefa é geralmente conhecida como identificação de locutor em um conjunto aberto, uma vez que ela requer que o sistema determine se a locução de entrada foi produzida por um determinado locutor específico (cadastrado no sistema) e não por qualquer outro locutor qualquer (possivelmente desconhecido pelo sistema), os “locutores impostores”. Verificação é a base para a maioria das aplicações de reconhecimento de locutor e é a mais comercialmente viável.

Essas tarefas podem também ser divididas com respeito às restrições impostas pelas locuções de treino e teste do sistema, ou então ao ambiente em que essas locuções são produzidas [7]. Em sistemas texto-dependentes, as locuções de treino e teste devem possuir a mesma palavra ou fala. Já em sistemas independentes de texto, nenhuma restrição é imposta às locuções. Nesse caso, o

sistema deve fazer o reconhecimento a partir de qualquer locução produzida pelos usuários. Entre sistemas dependentes e independentes de texto, os sistemas dependentes de vocabulário restringem as locuções a possuírem um vocabulário limitado, como dígitos, letras ou frases específicas.

Objetivo geral

Estudar técnicas e algoritmos que visem à produção de um sistema de verificação de locutor independente de texto que consiga competir diretamente com as técnicas do estado da arte.

Objetivos específicos

- Estudo das atuais técnicas e algoritmos utilizados para o reconhecimento de locutor independente da fala [8-13];
- Conduzir experimentos com uma nova técnica para reconhecimento de locutor utilizando uma combinação entre Gaussian Mixture Models e Type-2 Fuzzy [14].

Cronograma

O desenvolvimento deste trabalho está dividido em atividades a serem realizadas, segundo o cronograma presente na tabela abaixo:

Atividades	Mês			
	Dezembro/12	Janeiro/13	Fevereiro/13	Março/13
Levantamento do estado da arte e definição do escopo do trabalho				
Implementação das técnicas escolhidas				
Condução dos experimentos				
Elaboração do relatório final				
Preparação e defesa do trabalho				

Possíveis Avaliadores

A priori, o possível avaliador para o trabalho de graduação é o professor:

George Darmiton da Cunha Cavalcanti (gdcc@cin.ufpe.br)

Caso haja alguma indisponibilidade para que o mesmo avalie, outra possibilidade é o professor:

Carlos Alexandre Barros Mello (cabm@cin.ufpe.br)

Referências

- [1] J. Naik, G. Doddington: Evaluation of a high performance speaker verification system for access Control. Proc. ICASSP, pp. 2392 – 2395, 1987.
- [2] A. Higgins, L. Bahler, J. Porter: Speaker verification using randomized phrase prompting. Digital Signal Proc. 1, pp. 89 – 106, 1991.
- [3] J. Naik, L. Netsch, G. Doddington: Speaker verification over long distance telephone lines. Proc. ICASSP, pp. 524 – 527, 1989.
- [4] C. Schmandt, B. Arons: A conversational telephone messaging system, IEEE Trans. Consumer Electron. 30(3), xxi – xxiv, 1984.
- [5] L. Wilcox, F. Chen, D. Kimber, V. Balasubramanian: Segmentation of speech using speaker identification. Proc. ICASSP, pp. l.161 – l.164, 1994.
- [6] B. M. Arons: *Interactively Skimming Recorded Speech*, Ph.D. Thesis (MIT Press, Cambridge 1994).
- [7] G. Doddington: Speaker recognition – identifying people by their voices. Proc. IEEE 73(11), pp. 1651 – 1664, 1985.
- [8] T. Kinnunen and H. Li: An overview of text-independent speaker recognition: From features to supervectors. Speech Commun., vol. 52, no. 1, pp. 12–40, 2010.
- [9] H. Hermansky, N. Morgan, A. Bayya, P. Kohn: RASTA-PLP speech analysis technique. Proc. ICASSP, pp. l.121 - l.124, 1992.
- [10] F. Bimbot, J. -F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, D. A. Reynolds: A tutorial on text-independent speaker verification, EURASIP J. Appl. Signal Process. 4, 430 – 451, 2004.

- [11] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn: Speaker verification using adapted gaussian mixture models. *Digital Signal Process.*, vol. 10, no. 1, pp. 19–41, 2000.
- [12] J. L. Gauvain, C. -H. Lee: Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Trans. Speech Audio Process* 2(2), pp. 291 – 298, 1994.
- [13] W. M. Campbell, D. E. Sturim, and D. A. Reynolds: Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 308–311, 2006.
- [14] Jia Zeng, Lei Xie, Zhi-Qiang Liu: Type-2 fuzzy Gaussian mixture models, *Pattern Recognition*, 41, 12, pp. 3636-3643, 2008.

Assinaturas

Hector Natan Batista Pinheiro
Orientando

Tsang Ing Ren
Orientador

Recife, Janeiro de 2013.