Detecção Automática de Sotaques Regionais Brasileiros: A Importância da Validação *Cross-datasets*

[1]Nathália Alves Rocha Batista, [1]Lee Luan Ling, [1]Tiago Fernandes Tavares, [1]Plínio Almeida Barbosa
[1]Universidade Estadual de Campinas, São Paulo, SP
E-mails: nathbapt@decom.fee.unicamp.br, lee@decom.fee.unicamp.br, tavares@dca.fee.unicamp.br, pabarbosa.unicampbr@gmail.com

***Resumo*— Neste artigo apresentamos uma análise sobre validação de sistemas de reconhecimento de sotaques regionais em português. Sistemas de identificação automática de sotaques tem sido, usualmente, avaliados usando uma metodologia de validação cruzada em dobras de uma base de dados. Esse procedimento parte da hipótese de que os resultados da validação cruzada generalizam para outras situações. Neste trabalho, usamos duas bases de dados gravadas independentemente para a realização de testes em um cenário *cross-dataset*. Os resultados nesse cenário, em termos de taxa de erros, são substancialmente inferiores aos encontrados na validação cruzada. Isso indica que testes em cenários *cross-dataset* são necessários para a validação adequada de sistemas de reconhecimento de sotaque.**

***Palavras-Chave*— Identificação de sotaques, Reconhecimento de fala, Sotaques Regionais Brasileiros, Forense**

***Abstract*— In this article we present an analysis on the validation of regional accent recognition systems in Portuguese. Automatic accent identification systems have usually been evaluated using a k-fold cross-validation methodology in a database. This procedure is based on the hypothesis that cross-validation results generalize to other situations. In this work, we use two independently recorded databases to perform tests in a cross-dataset scenario. The results in this scenario, in terms of error rates, are substantially lower than those found in cross-validation. This indicates that tests in cross-dataset scenarios are required for proper validation of accent recognition systems.**

***Keywords*— Regional Accent Identification, Speech Recognition, Brazilian Regional Accents, Forensic**

I. Introdução

A fala de um indivíduo traz informações que permitem inferir seu gênero, idade, emoção, ritmo e o sotaque. O sotaque engloba tanto características fonológicas, isto é, maneira como os fones (sons) se organizam dentro da língua, quanto fonéticas, isto é, aspectos acústicos e fisiológicos dos sons da fala referente à produção, articulação e variedades. Ambas são fortemente determinadas por aspectos geográficos (região de origem), sociais e étnicos do falante [1] [2].

O reconhecimento de sotaques regionais permite a adaptação automática de modelos acústicos de reconhecimento de fala, tornando-os potencialmente mais robustos às variações fonéticas do locutor [3]. Também, pode ter aplicações em situações forenses, uma vez que um sistema automático pode ser capaz de detectar variações da fala que são imperceptíveis ao ouvido humano [4]. Por fim, a identificação de sotaques pode auxiliar na personalização de sistemas *Text to Speech*, permitindo construir modelos mais realísticos que os que usam fala genérica.

Estudos anteriores tiveram como objetivo identificar automaticamente idiomas e dialetos [5], [6], [7], [8], usando técnicas de modelagem do espaço acústico GMM-UBM (*Gaussian Mixture Model-Universal Background Model*) e *iVector*. As principais línguas estudadas na literatura são inglês americano, inglês britânico, francês, mandarim, árabe, chinês, indiano e suas variações.

A maior parte desses estudos sobre reconhecimento de sotaques realizam avaliações usando um procedimento de validação cruzada em dobras de uma base de dados. Como mostra a Seção II, esse procedimento é comumente aplicado em bases de dados foneticamente balanceadas e gravadas em condições de baixo ruído. Tais cuidados visam garantir que os algoritmos adaptativos ou de aprendizado modelem características da voz, tais como, as variações de sotaque, e não características espúrias. Apesar disso, não há estudos demonstrando experimentalmente sua relevância.

Neste artigo apresentamos um procedimento de validação *cross-datasets* com duas bases de dados em português brasileiro em dois sistemas de classificação automática de sotaques regionais. Ambos os sistemas foram, primeiramente, testados em um cenário *closed-set* que replica os procedimentos mostrados na literatura, resultando em indicadores de acerto semelhantes aos reportados na literatura para outros idiomas. Após, foram aplicados em cenário *cross-datasets*. Nesse experimento, os indicadores de acerto caem substancialmente em relação aos anteriores, para ambos os sistemas.

O artigo está organizado conforme a seguir. A Seção II descreve os trabalhos relacionados à pesquisa. Seção III mostra a descrição matemática dos sistemas utilizados. Seção IV descreve a metodologia dos experimentos. Seção V mostra os resultados e discussões, e, por fim, a Seção VI conclui este trabalho.

II. Trabalhos Relacionados

Na literatura as abordagens linguísticas como análise prosódica, da fonotaxe e acústica são usadas em sistemas LID (*Language Identification*) e DID (*Dialect Identification*) [7] [11] [12] [13]. Essas abordagens são também aplicáveis aos sistemas de reconhecimento de sotaque regionais (RAI) [14].

Em [12], a proposta se baseia na identificação das características prosódicas representadas por durações de sílabas, pitch e contornos de energia como entrada do classificador SVM (*Support Vector Machine*). Essa proposta foi avaliada em uma base de dados gravada com 10 locutores pronunciando a sentença "EVARO ANNAM THINNARU. NENU EVARINI CHUDALEDHU " por cinco vezes. Os falantes são de origem de três regiões do Sudeste da India. O sistema alcançou um índice de acerto de 72%, em um cenário *closed-set*.

Em Hanani et al, o GMM-UBM, GMM-SVM (*Support Vector Machine-Gaussian Mixture Models*) e *iVector* foram usados para identificar 4 sotaques regionais árabe palestino, em uma análise acústica. Os resultados mostram que *iVector* tem melhor índice de acerto que GMM-UBM e GMM-SVM, alcançando índices acima de 80%. Em outra pesquisa [10], um sistema automático de reconhecimento de sotaque é aplicado para identificar os sotaques franceses pronunciados em 4 diferentes regiões da Suíça, usando GMM-UBM e *iVector*. Foi observado melhoria relativa de $15,3\%$ na precisão geral de identificação dos sotaques do *iVector* sobre o sistema baseado em GMM.

Najafian et al [15] propõem um sistema *iVector* e outro com fusão de fonotaxe para identificação de sotaques regionais do inglês britânico. O sistema *iVector* é implementado usando os subconjuntos de treinamento ABI *corpus* [16], que representa dados de 13 diferentes sotaques regionais com padrão do inglês britânico do sul. O sistema alcança uma precisão de 76% em *3-Fold Cross Validation*, considerando também um cenário *closed-set*. Testes em cenários desse tipo não retratam casos em situações reais de sistemas de identificação.

Como mostra essa discussão, o procedimento de teste em cenário *closed-set* tem sido usado sistematicamente para a avaliação de sistemas de reconhecimento de sotaque. Isso revela a hipótese subjacente de que esse procedimento indica que os resultados encontrados extrapolam para aplicações reais. Porém, até o limite do conhecimento dos autores, essa hipótese ainda não foi testada experimentalmente. Esse teste foi realizado neste trabalho, como mostram as seções a seguir.

III. Descrição dos Sistemas de Identificação de Sotaques

Utilizamos os sistemas para classificação automática dos sotaques regionais *Gaussian Mixture Model-Universal Background Model* (GMM-UBM) e vetor de identidade (*iVector*) descritos a seguir.

A. Gaussian Mixture Model-Universal Background Model (GMM-UBM)

Nessa abordagem um *GMM-UBM* é primeiramente treinado para representar a distribuição independente dos dados acústicos de todos os sotaques. Os parâmetros do GMM de cada sotaque são então estimados usando o algoritmo de *Maximization-Expectation* (EM). Posteriormente, um GMM para cada sotaque é derivado pela adaptação *Maximum A-Posteriori* do GMM-UBM, conforme ilustrado na Figura 1. Nessa abordagem, as amostras de voz são representadas por um conjunto de vetores de características espectrais e o reconhecimento é baseado na estimativa de máxima verossimilhança.

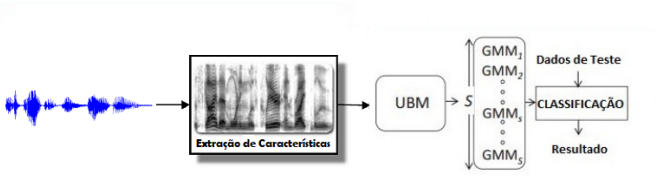

Fig. 1. Modelo GMM-UBM convencional - Um GMM-UBM é treinado em um conjunto de dados de um grande número de sotaques (amostras de voz). Os modelos GMM específicos de cada sotaques são então adaptados do UBM usando a estimação máxima a posteriori (MAP).

Dado o GMM-UBM e o vetor de características para cada sotaque, é calculado o alinhamento probabilístico do vetor de treinamento X em relação as componentes das misturas do UBM. A distribuição das características de cada sotaque é modelada como a soma ponderada de N funções de densidade de probabilidade gaussinadas, conforme a Expressão (1).

$$p(x|\lambda) = \Sigma_{i=1}^{N} w_i \mathcal{N}(x|\mu_i, \Sigma_i) \tag{1}$$

em que w_i é o peso dado para i-ésima mistura. μ_i e Σ_i são as médias e matriz de convariância da Gaussiana.

A densidade de probabilidade $\mathcal{N}(x|\mu_i, \Sigma_i)$ para $i = 1, ..., N$ é dada por:

$$\mathcal{N}(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{M}{2}} |\Sigma_i|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\} \tag{2}$$

Do treinamento de um UBM independente, um GMM de cada sotaque é adaptado a partir dos parâmentos do modelo λ_{UBM} pela *Maximum-A-Posteriori* (MAP). As estatísticas são calculadas com base no vetor de treinamento e atualizadas no UBM da i-ésima mistura durante a etapa de *Expectation-Maximization*.

O modelo GMM-UBM é discriminativo e baseado em pontuações, de forma que as características correlacionadas dos sotaques são agrupadas na mesma categoria. Assim, durante o processo de reconhecimento, dado um novo modelo de teste, um modelo será selecionado de modo a fornecer a máxima verossimilhança entre eles.

B. Identity-Vector (iVector)

Recentemente Dehak et al. [18] propuseram uma técnica de modelagem de variabilidade total (TV) usando o espaço de supervetores GMM, que se baseia em JFA (*Joint Factor Analysis*), na qual toda a variabilidade intra-classe e inter-classe são representas em um único vetor de pequena dimensão. Essa matriz *Total Variability Space* contém os autovetores com os maiores autovalores da matriz de covariância T.

Os supervetores GMM são obtidos pela concatenação dos vetores das médias $\mu = [\mu_1^T \mu_2^T \mu_3^T ... \mu_C^T]$ dos UBM λ_ω

resultantes da adaptação MAP. O supervetor GMM M é representado matematicamente na Expressão 3.

$$M = m + T\omega \tag{3}$$

em que m é o vetor dos valores das médias do UBM $\lambda_\Omega = (\omega_i, \mu_i, \Sigma_i)$, T é a matriz retângular de variabilidade entre conteúdo da locução (amostra de voz referente ao sotaque) e o canal, w é um vetor aleatório independente com distribuição normal $N(0, I)$, em que a média dessa distribuição corresponde ao *iVector*.

IV. Experimentos

A. Base de Dados

Poucas bases de dados em português brasileiro estão disponíveis para estudos em reconhecimento de locutor e da fala. Diante disso, neste estudo confeccionamos uma base de dados chamada *braccent*, a partir da leitura de 16 frases foneticamente balanceadas e avaliadas por um profissional em linguística do português. As frases foram gravadas por diversos locutores voluntários em diferentes partes do Brasil, abragendo as 7 classificações de sotaques regionais. As gravações foram realizadas em ambiente não-controlado e com microfone de *headset*, notebook e computador. A Tabela I mostra a quantidade de gravações da base de dados organizada por sexo e sotaque.

TABELA I
DESCRIÇÃO DA BASE DE DADOS BRACCENT.

Sotaques	Número de Gravações	Feminino	Masculino
Nortista	20	1	2
Baiano	72	2	4
Fluminense	44	3	1
Mineiro	97	3	4
Carioca	65	3	2
Nordestino	199	7	11
Sulista	682	26	27

A segunda base de dados utilizada foi desenvolvida em trabalhos anteriores de Ynoguti et al. [26], em que cada um dos 71 locutores (ambos os sexos) pronunciaram 40 frases, foneticamente balanceadas, e com duração de 3s cada. As amostras dessa base foram capturadas em ambiente silencioso e com microfone direcional de boa qualidade. Essa base contém apenas amostras de sotaques referente à 5 classes, ou seja, não contém amostras do sotaque nortista e carioca.

O desbalanceamento entre os números de amostras para cada sotaque e gênero, em ambas as bases de dados, é compatível com o encontrado em outros trabalhos na literatura. Para este trabalho, todas as amostras de áudio das duas bases de dados foram re-amostradas para $16kHz$.

B. Procedimento experimental

Seguindo um procedimento comum na análise de fala [21] [22], as amostras de voz foram pré-enfatizadas com $H(z) = 1 - 0,97z^{-1}$. Após, o vetor de características acústicas foi extraído usando uma janela de Hamming em quadros de 25 ms, sobrepostos de tal forma que o início de dois quadros consecutivos difere de 10 ms. Para cada quadro, são calculados 19 MFCCs (*Mel Frequency Cepstral Coefficients*), além de suas derivadas de primeira e segunda ordem, totalizando em um vetor de características de dimensão 60.

Os experimentos foram conduzidos tanto para o sistema com *iVector* quanto para o GMM-UBM em dois cenários diferentes. O primeiro deles, chamado de *closed set*, consistiu de uma avaliação em validação cruzada de 10 dobras (*10-Fold Cross-Validation*) usando isoladamente as bases *braccent* e *Ynoguti*. O segundo cenário, chamado de *cross-datasets* consistiu em ajustar (treinar) os parâmetros do sistema de classificação usando uma das bases de dados e, após, testá-lo na outra base. Nos testes desse cenário, só foram considerados os sotaques que existem simultaneamente nas duas bases de dados, isto é, os sotaques carioca e nortista não foram utilizados.

A quantidade de gaussianas dos sistemas GMM-UBM e o tamanho da matriz T do *iVector* foram alterados durante os experimentos. Isso permitiu verificar como os resultados variam com a alteração destes parâmetros. O algoritmo *K-Means* foi utilizado para inicialização dos vetores de média para os dois sistemas.

C. Medidas de Desempenho

1) Equal Error Rate (EER): O EER é o valor em que FAR (*False Accept Rate*) e FRR (*False Reject Rate*) são iguais. Quanto menor o valor, melhor é o desempenho do sistema.

2) Recognition Rate (RR): Mensura a quantidade de amostras que foram corretamente reconhecidas pelo sistema. Também conhecido como *Sensitivity* ou *Recall* ou *Probability of Detection*, isto é, é o número de positivos verdadeiros dividido pelo número de positivos verdadeiros mais o número de falsos negativos.

$$RR = \frac{\text{TP}}{\text{TP + FN}} \tag{4}$$

3) F1-Score: É a média harmônica entre *Precision* e *Recall*. *Precision* é definido como o número de positivos verdadeiros (sotaques reconhecidos corretamente) dividido pelo número de positivos verdadeiros mais o número de falsos positivos.

$$F1_{Score} = \frac{\text{2TP}}{\text{2TP + FP + FN}} \tag{5}$$

V. Resultados e Discussão

A. Closed set

Nesse cenário, os sistemas GMM-UBM e *iVector* foram treinados e testados em cada uma das bases *braccent* e *Ynoguti*.

A Tabela II mostra os resultados obtidos ao treinar um sistema GMM-UBM independente de gênero com diferentes quantidades de componentes gaussianas na base de dados *braccent*. Os testes foram realizados nessa mesma base, usando *Stratified K-Fold Cross Validation*. Foram verificados os valores de taxa de erro igualitário, taxa de detecção e identificação, F1-Score, de acordo com a variação da quantidade de componentes gaussiana. Quanto maior o F1 Score, melhor é o desempenho do sistema.

TABELA II
DESEMPENHO DO SISTEMA GMM-UBM TREINADO E TESTADO NA BASE DE DADOS BRACCENT

Número de Gaussianas	ERR	RR	F1 Score
128	19,585%	77,4818 %	0,776758086
256	18,7539%	74,2516 %	0,739340911
512	22,6217%	67,1619 %	0,699592413
1024	31,7471%	61,2276 %	0,65264385

A Tabela II mostra que o sistema GMM-UBM tem melhor desempenho ao classificar os sotaques regionais quando modelado com 128 componentes de gaussianas. À medida que N aumenta ($N > 256$), o sistema deixa de identificar adequadamante os sotaques, chegando a uma taxa de erro de 31%. A semelhança do comportamento do F1-Score e do RR indicam que os erros estão bem distribuídos entre as classes.

Os resultados do sistema com *iVector*, mostrados na Tabela III, evidenciam um índice de erros sensivelmente maior que o do sistema com GMM-UBM. Isso pode estar associado à quantidade de dados disponíveis para treinamento do sistema *iVector*, pois a matriz de variabilidade T requer grande quantidade de dados para modelagem das características intra-classe e inter-classe [22] [25].

TABELA III
DESEMPENHO DO SISTEMA *iVector* COM 128 COMPONENTES GAUSSIANAS TREINADO E TESTADO NA BASE DE DADOS BRACCENT.

Tamanho da Matriz T	ERR	RR	F1 Score
100	54,4285%	28,3969%	0,578484446
150	56,572%	28,8813%	0,581500893
200	56,710%	29,4816%	0,583143258

As tabelas IV e V mostram os resultados dos testes na base *Ynoguti* com o sistema GMM-UBM e *iVector*, respectivamente. Dos experimentos realizados com essa base, não foi observada melhora nos sistemas GMM-UBM e *iVector* com número de gaussianas maiores que 256, por isso, reportamos apenas os resultados para 128 e 256 gaussinas.

TABELA IV
DESEMPENHO DO SISTEMA GMM-UBM TREINADO E TESTADO NA BASE DE DADOS YNOGUTI

Número de Gaussianas	ERR	RR	F1 Score
128	13.5308%	76.4806%	0,764805529
256	14.4349 %	77.4253%	0,774251811

TABELA V
DESEMPENHO DO SISTEMA *iVector* COM 256 COMPONENTES GAUSSIANAS TREINADO E TESTADO NA BASE DE DADOS YNOGUTI.

Tamanho da Matriz T	ERR	RR	F1 Score
100	28.7511%	54.0164%	0.5401642
150	29.6155%	53.3277%	0.532775647
200	29.44%	55.0274%	0.550274852

O sistema GMM-UBM alcançou índices de desempenho semelhantes aos reportados na literatura em duas bases de dados diferentes. Isso permite inferir que a implementação do sistema e as condições de validação são semelhantes às encontradas na literatura.

A seguir, são mostrados resultados referentes aos testes em cenário *cross-datasets*.

B. Cross-datasets

Os experimentos no cenário *cross-datasets* consistiram em usar tanto a base *Ynoguti* quanto a base *braccent*. Em cada experimento, uma delas foi usada para treino e a outra para teste. Os resultados são mostrados nas tabelas VI, VII, VIII e IX.

TABELA VI
DESEMPENHO DO SISTEMA GMM-UBM TREINADO NA BASE YNOGUTI E TESTADO NA BASE BRACCENT.

Número de Gaussianas	ERR	RR	F1 Score
128	39.835%	40.557%	0.1786847519
256	37.564%	43.137%	0.170052069
512	36.133%	47.368%	0.170273556
1024	35.384%	49.948%	0.174620872
2048	34.675%	52.116%	0.172506070

De acordo com os resultados, a modelagem GMM-UBM obteve seu melhor desempenho usando 2048 componentes gaussianas. Ao verificarmos se o sistema melhora o desempenho variando o tamanho da matriz T do sistema *iVector*, com 2048 gaussianas, constatamos que a melhor taxa de reconhecimento foi de 41%, porém ainda é inferior ao melhor valor com GMM-UBM.

TABELA VII
DESEMPENHO DO SISTEMA *iVector* COM 2048 COMPONENTES GAUSSIANAS TREINADO NA BASE YNOGUTI E TESTADO NA BASE BRACCENT.

Tamanho da Matriz T	ERR	RR	F1 Score
100	43.034%	39.009%	0.243245543541
150	43.047%	41.589%	0.253309280164
200	42.544%	40.660%	0.262027917816

TABELA VIII
DESEMPENHO DO SISTEMA GMM-UBM TREINADO NA BASE BRACCENT E TESTADO NA BASE YNOGUTI.

Número de Gaussianas	ERR	RR	F1 Score
128	41.649%	28.280%	0.282795698925
256	41.111%	32.025%	0.320250896057
512	40.376%	39.283 %	0.392831541219
1024	39.890 %	44.803 %	0.448028673835
2048	40.108%	49.050%	0.490501792115

Os dados experimentais revelam que treinar o sistema em uma base de dados leva a resultados de identificação substancialmente inferiores quando o teste é realizado em uma base de dados diferente. Uma vez que as características referentes ao sotaque (realizações fonéticas) existem igualmente em ambas as bases de dados, essa queda de desempenho indica que

TABELA IX
DESEMPENHO DO SISTEMA *iVector* COM 2048 COMPONENTES GAUSSIANAS TREINADO NA BASE BRACCENT E TESTADO NA BASE YNOGUTI.

Tamanho da Matriz T	ERR	RR	F1 Score
100	44.086%	31.631%	0.316308243728
150	44.478%	31.756%	0.317562724014
200	43.311%	29.104%	0.291039426523

os sistemas de classificação estão modelando outras características acústicas que não se replicam entre as bases, sendo possível concluir que os sistemas estão modelando elementos espúrios de cada base de dados, e não necessariamente as características ligadas ao sotaque.

Os resultados mostrados nas tabelas VI e VII, referentes aos experimentos em que o treino é realizado na base *Ynoguti*, contradizem a hipótese de que uma base de dados com baixo ruído de gravação e foneticamente balanceada é suficiente para o ajuste de um modelo acústico de reconhecimento de sotaques. Ao contrário, os resultados encontrados ao realizar o treinamento em uma base sem controle de gravação e com maior variabilidade de canal (*braccent*) são superiores. Isso indica que as variações de canal em uma base de dados de treinamento pode ser desejável, já que adiciona variabilidade à informação usada no ajuste de parâmetros.

Isso significa que a metodologia de validação em *closed set*, comumente usada na literatura gera resultados que podem não replicar em situações reais, que são melhores simuladas em cenários *cross-datasets*. Portanto, índices de desempenho e testes mais cautelosos que a validação cruzada são necessários para demonstrar a eficácia de sistemas de detecção automática de sotaques.

VI. CONCLUSÕES

Este artigo apresenta testes de dois sistemas de detecção automática de sotaques, um baseado em modelo GMM-UBM e o outro baseado em modelo *iVector*. Os testes foram realizados em dois cenários diferentes – *closed set* e *cross-datasets*. Os testes em *closed set* reproduzem os cenários usados na literatura, e resultam em índices de reconhecimento semelhantes aos publicados para outros idiomas. Os testes em *cross-datasets*, porém, mostram uma queda sensível dos índices de reconhecimento.

Esses resultados indicam que o cenário *closed set*, comumente usado para avaliação de sistemas de reconhecimento, não é suficiente para reproduzir as condições reais de aplicação de um sistema deste tipo. Também, os resultados contradizem a ideia de que uma base de dados sem ruído e foneticamente balanceada implica em um ajuste mais adequado de parâmetros nos sistemas de reconhecimento. Por fim, os resultados mostraram que a detecção automática de sotaques no português brasileiro ainda é um problema em aberto que ainda não tem soluções que generalizam para diferentes bases de dados.

O reconhecimento automático de sotaques, portanto, ainda é um campo potencialmente prolífico para trabalhos futuros.

REFERÊNCIAS

[1] D. P. Cardoso, *Fonologia da Língua Portuguesa*. Universidade Federal De Sergipe, 2009.

[2] M. M. Alves, *As Vogais Médias em Posíção PréTônica nos nomes no dialeto de Belo Horizonte: Estudo da Variação a Luz da Teoria da Otimalidade*. Faculdade de Letras da UFMG, 2008.

[3] C. Teixeira and I.Trancoso and A. Serralheiro *Accent identification*, Spoken Language ICSLP 96 Proceedings, October, 1996.

[4] P. Rose *Forensic Speaker Identification*, Taylor & Francis Series, 2002.

[5] A. Hanani and M. J. Russell and M. J. Carey *Palestinian Arabic regional accent recognition*. 2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), 2015.

[6] G. Friedland, O. Vinyals, Y. Huang and C. Müller *Prosodic and other Long-Term Features for Speaker Diarization*. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, NO. 5, July 2009

[7] F. Biadsy *Accent Detection of Telugu Speech Using Prosodic And Formant Features*. Phd Thesis at Columbia University, 2011.

[8] A. Hanani et al, *Human and Computer Recognition of Regional Accents And Ethnic Groups From British English Speech*, Comput. Speech Lang, January, 2013

[9] M. Benzeghiba, R. De Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet AND L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, C. Wellekens *Automatic Speech Recognition and Speech Variability*, Speech Communiation, February, 2007.

[10] A. Lazaridis and et al *Accent Identification, French Regional Accents, GMM Modelling, i-vectors, SVM*. Odyssey: The Speaker and Language Recognition Workshop, 2014.

[11] Z. Ge and Y. Tan and A. Ganapathiraju *Accent Classification with Phonetic Vowel Representation*, CoRR, June, 2017.

[12] K. Mannepalli and P. N. Sastry and V. Rajesh *Accent Detection of Telugu Speech Using Prosodic And Formant Features*, International Conference on Signal Processing and Communication Engineering Systems, January, 2015.

[13] M. A. Zissman *Language Identification Using Phoneme Recognition and Phonotactic Language Modeling*, Language identification using phoneme recognition and phonotactic language modeling, May, 1995.

[14] G. Brown *Automatic Accent Recognition Systems and the Effects of Data on Performance*. Odyssey, 2016.

[15] M. Najafian, S. Safavi, P. Weber, M. Russell *Identification of British English regional accents using fusion of i-vector and multi-accent phonotactic systems*, Odyssey, June 21-24, 2016, Bilbao, Spain

[16] S. M. D'Arcy, M. J. Russell, S. R. Browning and M. J. Tomlinson, *The Accents of the British Isles (ABI) corpus*, Proceedings Modelisations pour lIdentification des Langues, pp. 115–119, 2004.

[17] S. Jalalvand and A. Akbari and B. Nasersharif *A classifier combination approach for Farsi accents recognition*, 20th Iranian Conference on Electrical Engineering (ICEE2012), May, 2012.

[18] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, e P. Ouellet, *Front-end factor analysis for speaker verification*. IEEE TASLP, vol. 19, pp. 788-798, Maio, 2011.

[19] M. H. Bahari *Accent recognition using i-vector, Gaussian Mean Supervector and Gaussian posterior probability supervector for spontaneous telephone speech*, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, May, 2013.

[20] T. CHEN *Automatic accent identification using Gaussian mixture models*, 20th Iranian Conference on Electrical Engineering (ICEE2012), December, 2001.

[21] D. D. C. Silva *Reconhecimento de Fala Contínua para o Português Brasileiro em Sistemas Embarcados*. Universidade Federal de Campina Grande, 2011.

[22] C. J. S. de Souza *Sistemas de Verificação de Locutor Baseados em i-Vectors*. Universidade Estadual de Campinas, 2015.

[23] T. C. Silva, *Fonética e Fonologia do Português*. Editora Contexto, 2005.

[24] M. T. S. Al-Kaltakchi1, W. L. Woo1, S. S. Dlay, J. A. Chambers, *Comparison of I-vector and GMM-UBM Approaches to Speaker Identification with TIMIT and NIST 2008 Databases in Challenging Environments*, 2017 25th European Signal Processing Conference (EUSIPCO).

[25] H. Behravan, V. Hautamäki, S. M. Siniscalchi, *i-Vector Modeling of Speech Attributes for Automatic Foreign Accent Recognition*, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, no. 1, January 2016

[26] C. A. Ynoguti *Reconhecimento de Fala Contínua Utilizando Modelos Ocultos de Markov*. Unicamp - Campinas, Maio 1999.