Desenvolvimento de algoritmo de análise automática da curva de frequência por meio de convoluções gaussianas do histograma de alturas¹⁵

André Ricardo de Souza; Maressa Vieira, Daniel Peres; Marcus Martins; Waldemar Ferreira Netto

Introdução

A convolução de um histograma por meio da função normal de Gauss é um dos inúmeros métodos de prospecção de dados que visam a encontrar possíveis pontos de concentração dos dados em uma amostra (clusters) e, portanto, possíveis alvos recorrentes do sistema que produziu aquela amostra. A convolução suaviza o “ruído” presente no histograma da amostra, permitindo identificar picos de probabilidade correspondentes a certos valores da variável aleatória por meio da estimativa da função densidade de probabilidade (probability density function, doravante PDF). No caso particular da análise automática da curva de frequência em amostras de fala, a PDF nos permite identificar alturas preferenciais empregadas pelo falante, o que pode contribuir para um aperfeiçoamento dos modelos teóricos da entoação.

Este trabalho apresenta o desenvolvimento e a implementação de um método de análise de amostras de fala baseado no

¹⁵ COLÓQUIO BRASILEIRO DE PROSÓDIA DA FALA, 4. UFAL, Maceió, 2013.

conceito de convolução gaussiana, acoplado ao algoritmo ExProsodia (Ferreira Netto 2006).

O método de convolução empregado neste trabalho é conhecido como “janela de Parzen” (Parzen 1962). Segundo o método de Parzen, a PDF é estimada por meio de uma função nuclear — comumente a função normal de Gauss — estabelecida em cada ponto da escala a que pertencem os dados da amostra. Cada um destes pontos é um candidato a centroide desta amostra. Empregando-se a função normal de Gauss, o cálculo da probabilidade de cada resultado da amostra pertencer ao centroide c é feito pela fórmula:

$$N(x \mid c, s) \frac{1}{\sqrt{2\pi}s} \times e^{\frac{-(x-c)^2}{2s^2}}$$

Nessa fórmula, x é o valor da variável aleatória correspondente àquele resultado, c é o valor do centroide na escala e s é o parâmetro de suavização, que corresponde aproximadamente à largura da janela do histograma. Depois de calculadas as probabilidades de cada resultado da amostra pertencer a cada candidato a centroide, calcula-se a média dessas probabilidades para cada um desses possíveis centroides, que será o valor da PDF para cada ponto da escala. O resultado desta função depende, portanto, do estabelecimento de uma escala à qual pertencem os valores da amostra e do parâmetro suavizador s , correspondente ao desvio padrão do kernel gaussiano aplicado.

O programa ExProsodia parte da hipótese de que alguns fatos prosódicos têm restrições mecânico-fisiológicas e outros decorrem das necessidades expressivas dos falantes (Xu e Wang 1997). Segundo Ferreira Netto (2006; 2008) e Peres e seus colegas (2009; 2011), a produção da fala exige esforço para sustentar a voz com uma frequência relativamente estável, definida aqui como tom médio ideal (TM) de F_0 , que se repete

nos momentos $Z(t)$ mensurados de F_0 . A supressão desse esforço desencadeia uma declinação pontual que exige a retomada da tensão inicial. A sustentação (S) é consequência do esforço que se acrescenta a cada um dos momentos da fala, incluindo-se o inicial, para compensar a declinação pontual de finalização (F). A componente F associa-se ao fato de que se trata do tom alvo da declinação pontual, estabelecida por um intervalo ideal decrescente de 7 st do TM obtido até o momento $Z(t)$. TM é a tendência central dos valores válidos de F_0 calculada como a média aritmética acumulada no tempo. Os valores válidos mensurados são os momentos de F_0 que cumprem as restrições de altura, intensidade e duração. A série temporal se configura aditivamente como $Z(t)=S(t)+F(t)+E(t)$. A seleção das unidades $Z(t)$ — ora chamadas de UBI (Unit of Base of Intonation) — para análise é feita pelo aplicativo ExProsodia® (Ferreira Netto 2010). O aplicativo faz a análise automática de porções da curva de frequência estabelecida por autocorrelação pelo software Speech Filing System® (Huckvale et al. 1985). Três parâmetros são considerados para essa definição: frequência maior do que 50 Hz e menor do que 700 Hz; intensidade suficiente para ser percebida e, garantidos os critérios anteriores, duração maior do que 20 ms. Esses valores podem ser modificados pelo usuário.

Metodologia

Para a realização do experimento consideraram-se as variáveis: (a) gênero — masculino e feminino — e (b) tipo de fala — leitura, colérica, neutra e triste — combinadas entre si. Os dados foram coletados na internet em sites que disponibilizam podcasts (Webcombrasil 2013; A Voz... 2013; Podcast 2013). Também foram coletadas gravações de vídeos no site (Youtube 2013). Para a análise das variáveis gênero e manifestação emocional, foram selecionados 80 arquivos sonoros, distribuídos em grupos de 10. Os grupos compreendiam: leitura

masculina (HL), leitura feminina (ML), fala masculina colérica (HC), fala feminina colérica (MC), fala masculina neutra (HN), fala feminina neutra (MN), fala masculina triste (HT) e fala feminina triste (MT). A avaliação das falas como colérica, neutra e triste decorreu de interpretação semântica feita pelos membros da equipe de pesquisa. Não houve restrições quanto à qualidade da gravação.

Resultados

Na Tabela 1 abaixo, são apresentados os resultados relativos obtidos com a análise das convoluções gaussianas. A análise feita mostrou que a maioria significativa dos intervalos utilizados pelos locutores permaneceu entre 1% e 2% no que diz respeito à variação média das frequências utilizadas nas UBIs, sendo que variações de frequência entre UBIs de até 2%, representam mais das metade das variações verificadas (ver Figura 1).

	HL	ML	HC	MC	HT	MT	HN	MN
]-.01	0,01	0,09	0,20	0,33	0,05	0,18	0,12	0,20
[.01-.02[	0,28	0,42	0,43	0,40	0,30	0,37	0,40	0,39
[.02-.03[	0,33	0,25	0,17	0,14	0,29	0,22	0,23	0,21
[.03-.04[	0,16	0,09	0,09	0,06	0,14	0,10	0,12	0,10
[.04-.05[	0,11	0,07	0,04	0,02	0,09	0,04	0,06	0,04
[.05-	0,11	0,08	0,07	0,05	0,14	0,08	0,06	0,06

Tabela 1 Dados intervalares (linhas) referentes às frequências das variáveis de sexo e emoção dispostas nas colunas

Figura 1. Curva da média acumulada das frequências da Tabela 1 referente aos intervalos obtidos com o cálculo das convoluções gaussianas.

Em relação às categorias de ocorrência, foi possível verificar que, comparando os coeficientes de variação (cv) de cada categoria, os intervalos de frequências com variação entre 1% e 2%, têm uma dispersão muito reduzida ($cv=,015$, $z<0,01$) e que os intervalos de frequência com variação abaixo de 1% são os que têm a maior dispersão ($cv=0,68$, $z<0,01$).

Em relação às emoções, foi possível verificar que, a partir de um teste z aplicado às frequências em relação às das demais emoções, todas as categorias frequência de HL e de HT apresentaram diferenças significativas em relação ao valores das demais categorias ($z<0,01$ em todos os casos). Também foi possível verificar que MC teve um aumento significativo da categoria de intervalos menores do que 1% ($z<0,01$), e um redução, também significativa dos intervalos maiores do que 2% ($z<0,01$ em todos os casos). Essa diferença pode ser observada na Fig. 2 abaixo.

Figura 2. Histograma ocorrência (eixo vertical) das categorias dos intervalos de frequências, agrupados por tipo d das e manifestação emocional (eixo horizontal). A descrição das categorias vai apresentada por código de cores na legenda à esquerda.

Discussão

A análise das variações das emoções por meio da aplicação do teste de convoluções gaussianas do histograma de alturas apontou diferenças significativas entre os intervalos dos picos de probabilidades. De maneira geral, a tendência desses intervalos acentuou-se de forma significativa nos intervalos entre 1% e 2%. Esse fato se deu em praticamente todas as manifestações emocionais. Nas leituras masculinas e das manifestação de tristeza masculina, houve um aumento significativo na ocorrência dos intervalos entre 2% e 3%. A manifestação feminina da cólera, entretanto, mostrou um comportamento invertido, estabelecendo intervalos menores do que 2%. Esse fato é compreensível pois, na manifestação de cólera há o aumento significativo do TM, e, portanto, uma tensão extrema das pregas vocais, impossibilitando variações maiores.