

# Modelagem estocástica do ritmo na fala

Antonio Galves

Denise Duarte

## 1 Introdução: modelagem estocástica e reconhecimento de padrões

Suponhamos que  $N$  falantes de  $L$  línguas diferentes gravem  $M$  frases, sendo  $N \geq L$  e  $M > N$ . A questão é saber se é possível apesar disso agrupar esses  $M$  registros sonoros em grupos correspondendo às línguas de seus locutores. O problema só aparecerá em toda sua beleza, se supusermos que as  $N$  línguas dos locutores nos sejam totalmente desconhecidas, de forma a não podermos utilizar o conhecimento das palavras ou da sintaxe para nos ajudar na classificação. Também estamos supondo que as vozes dos locutores sejam suficientemente neutras e semelhantes de forma a tampouco podermos utilizar seus traços distintivos, como timbre, para nos ajudar na classificação. O que estamos dizendo é que essa classificação só pode se basear em características intrínsecas de regularidade do sinal acústico em cada língua.

Quem alguma vez visualizou um exemplo de função de onda produzida pela fala, sabe como o sinal acústico pode variar de uma frase a outra, mesmo que as duas sejam produzidas numa mesma língua e pelo mesmo locutor. Essa diversidade de aparência condena ao fracasso qualquer tentativa de classificação baseada num critério ingênuo de busca de semelhanças visíveis na forma do sinal.

O que fazer então? Este texto tem como objetivo defender uma abordagem ao problema que em vez de buscar semelhanças superficiais entre as amostras, tenta identificar a "máquina" responsável pelas diversas amostras. Essas "máquinas" são objetos matemáticos de um tipo particular. Na literatura especializadas eles são chamados de *Processos Estocásticos*.

Nossa classificação começa pela identificação do tipo de processo estocástico capaz de produzir aquelas amostras. Cada língua será identificada por parâmetros específicos que deverão ser estimados estatisticamente. Uma vez feito esse trabalho de estimação, as amostras serão agrupadas tendo como critério a proximidade dos parâmetros dos processos que as produziram.

O trabalho de identificar o processo capaz de produzir um certo conjunto de amostras é chamado de *Modelagem Estocástica*. Ele é a base do trabalho de classificação que na literatura estatística é muitas vezes chamado de *reconhecimento de padrões*.

O nosso objetivo é ilustrar essa abordagem do problema da classificação de classes através do estudo de um problema científico atual e relevante, a saber a identificação de correlatos acústicos do ritmo nas línguas naturais. Trata-se um problema que permaneceu aberto por mais de meio século e que apenas no últimos anos pode ser atacado com êxito.

Além do problema linguístico aqui tratado, é fácil entender que a abordagem proposta pode ser aplicada a uma grande variedade de situações, onde se trata de identificar regularidades estatísticas em sinais complexos. Um primeiro exemplo é a análise de dados genéticos, onde a modelagem estocástica vem sido utilizada com grande sucesso. Outros exemplos que estão na ordem do dia são a identificação dos tipos de internautas a partir de registros de suas excursões pela *grande teia mundial*, ou a identificação de tipos de consumidores feitas por redes comerciais, bancos e empresas de crédito a partir dos registros de suas movimentações e compras.

## 2 A conjectura das classes rítmicas e a revolução de Ramus, Nespór e Mehler

Lloyd James ( anos 40) e Abercrombie ( anos 50) conjecturam que as línguas se agrupam em classes rítmicas da seguinte maneira

1. línguas acentuais: Holandês, Inglês, Polonês, ...
2. línguas silábicas: Catalão, Espanhol, Francês, Italiano, ...
3. línguas moraicas: Japonês, ...

A pergunta é que características do sinal acústico de fala identificam essas classes. Até recentemente ninguém havia encontrado correlatos acústicos que dessem base empírica à intuição dos linguistas. Isso acontece pela primeira vez com um artigo de 1999, assinado por um jovem pos-doutorando Franck Ramus e co-assinado por seus orientadores Marina Nespór e Jacques Mehler.

Em Ramus, Nespór e Mehler (1999) é proposta uma abordagem que dá evidências acústicas e estatísticas para a hipótese da existência das classes rítmicas.

Isso foi conseguido através da seguinte experiência: foram escolhidas para a pesquisa 5 sentenças, em 8 línguas diferentes, que foram pronunciadas por 4 falantes de cada língua, resultando num total de 160 sentenças. Essas sentenças foram extraídas de um corpus de sentenças com várias línguas.

As frases são declarativas simples, inicialmente escritas em francês e posteriormente traduzidas para as demais línguas. As leituras foram feitas por locutores nativos do sexo feminino.

Os intervalos vocálicos são definidos como as seqüências de vogais consecutivas não interrompidas por consoantes e os intervalos consonantais, ou ainda intervalos intervocálicos, são definidos como as seqüências de consoantes consecutivas não interrompidas por vogais. Sendo assim, o sinal acústico é totalmente segmentado em intervalos vocálicos, consonantais e pausas. A segmentação de todos os intervalos foi feita com

base na observação do gráfico do sinal acústico, do seu espectrograma e na percepção auditiva (ver Figura 1).

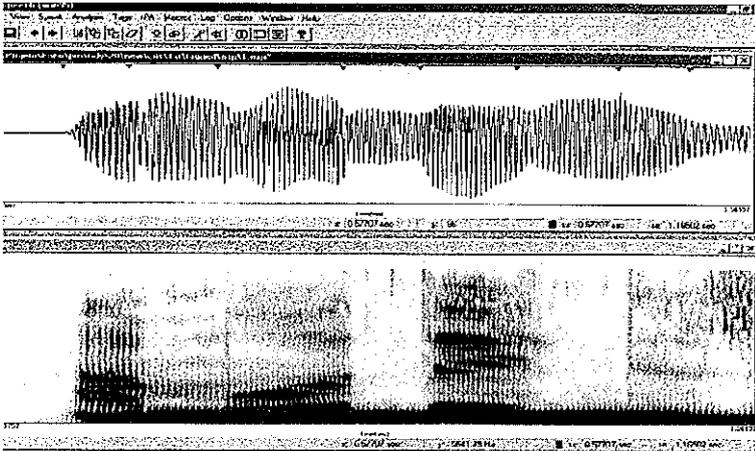


Figura 1: sinal acústico ampliado com as marcações dos intervalos

Para cada língua foram calculados os valores das seguintes estatísticas: porcentagem de tempo ocupada por intervalos vocálicos (notada como  $\%V$ ) e o desvio padrão (notado como  $\Delta C$ ) dos comprimentos dos intervalos consonantais .

A definição formal é a seguinte. Sejam  $c_i$  e  $v_i$  respectivamente os comprimentos do  $i$ -ésimo intervalo vocálico e consonantal e seja  $n$  o número desses intervalos ( para simplificar a notação estamos supondo que o número de intervalos vocálicos e consonantais é o mesmo). Então

$$\%V = \frac{\sum_{i=1}^n v_i}{\sum_{i=1}^n v_i + \sum_{i=1}^n c_i}, \quad (1)$$

$$\Delta C = \sqrt{\frac{\sum_{i=1}^n (c_i - \bar{c})^2}{n - 1}}, \quad (2)$$

sendo

$$\bar{c} = \frac{\sum_{i=1}^n c_i}{n}.$$

A Figura 2 apresenta as médias dos valores de  $\%V$  e  $\Delta C$  nas oito línguas analisadas em Ramus et al. (1999): Holandês (dut), Polonês (pol), Inglês (eng), Francês (fre), Catalão (cat), Italiano (ita), Espanhol (spa) e Japonês (jap).

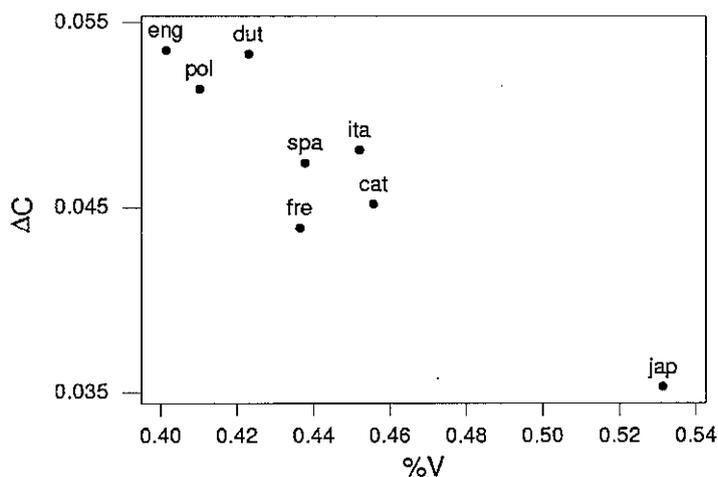


Figura 2: Distribuição das línguas no plano ( $\%V, \Delta C$ ) (Ramus et al. 1999.)

Três fatos são claramente sugeridos pela Figura 2.

1) Os pontos do gráfico parecem estar distribuídos em três aglomerados distintos. Estes aglomerados correspondem à hipótese de três classes rítmicas distintas.

2) Há uma alta correlação linear entre  $\Delta C$  e  $\%V$ . O coeficiente de correlação linear é de  $-0,93$ .

3)  $\Delta C$  é uma função decrescente de  $\%V$ .

## 2.1 Uma base estatística para o resultado de Ramus, Nespor e Mehler

Podemos dizer que os valores de  $\Delta C$  e  $\%V$  encontrados para as línguas estudadas por Ramus et al. definem de fato três aglomerados distintos? A resposta a essa questão passa pela construção de um modelo probabilístico para as classes rítmicas. Isso foi feito em Duarte et al. (2001) e resumido aqui.

Inicialmente constatou-se que havia evidências estatísticas de que os comprimentos dos intervalos consonantais eram independentes. Segue disso que os comprimentos sucessivos de intervalos consonantais podem ser modelados como uma sequência de variáveis aleatórias independentes.

O próximo passo é identificar a distribuição de um único intervalo. Após várias tentativas com diversas distribuições (normal, log-normal, exponencial, etc), constatou-se que a distribuição que melhor se ajustava aos dados era a Gama.

A Figura 3 ilustra este fato. Nela é apresentado o gráfico dos percentis teóricos da Gama contra os percentis da amostra para uma língua dada.

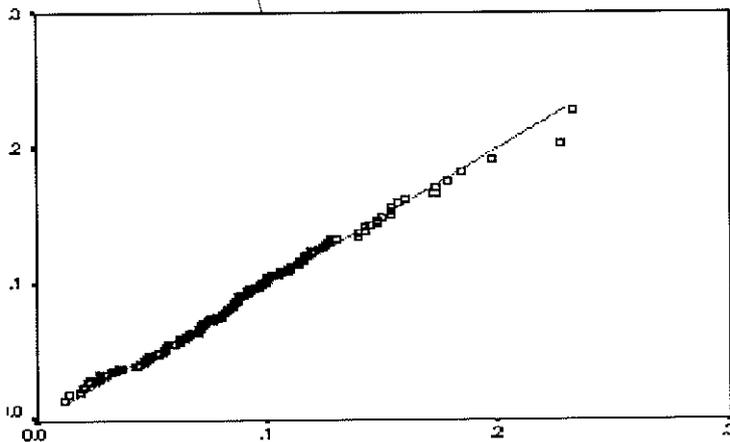


Figura 3: Gráficos de percentis teóricos da Gama contra percentis empíricos da amostra.

A Gama é uma distribuição em  $\mathbb{R}^+$  dependendo de dois parâmetros  $\alpha$  e  $\beta$  e tendo como função de densidade

$$f_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-\frac{x}{\beta}},$$

onde  $\Gamma(\alpha)$  é a famosa função Gama que é igual a  $\alpha!$  quando  $\alpha$  é um número inteiro.

A Figura 4 representa o gráfico da função de densidade  $f_{\alpha,\beta}(x)$  para diversos valores de  $\alpha$  e  $\beta$ . Na literatura estatística,  $\alpha$  é chamado de parâmetro de forma e  $\beta$  de parâmetro de escala. Os exemplos apresentados na Figura 4 ajudam a entender o porquê dessas denominações.

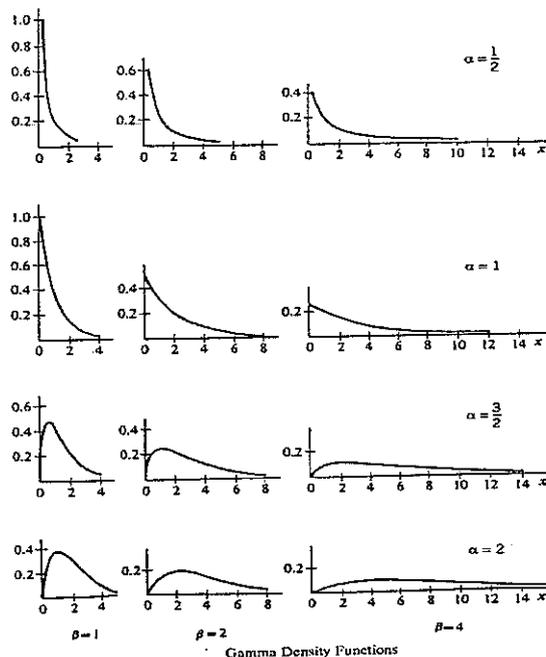


Figura 4: Exemplo da distribuição Gama com parâmetros de forma ( $\alpha$ ) e escala ( $\beta$ ) diferentes

A média e a variância da Gama são dadas, respectivamente, por

$$\mu = \alpha\beta \text{ e } \sigma^2 = \alpha\beta^2$$

Para cada uma das línguas consideradas por Ramus, Nespor e Mehler (1999), estimamos os parâmetros da distribuição Gama que melhor se ajustaram aos dados amostrais. Uma maneira habitual de estimar parâmetros é utilizar um critério de *máxima verossimilhança*. Isto significa buscar os valores dos parâmetros que maximizam a probabilidade de ocorrência da amostra (é esta probabilidade que chamamos verossimilhança). No caso da Gama, a busca dos parâmetros que maximizam a verossimilhança não pode ser feita de forma analítica. Neste caso, utilizamos um método numérico (isto foi feito com a ajuda do aplicativo S-Plus) para encontrar uma solução.

Estimados os valores de  $\alpha$  e  $\beta$  as variâncias correspondentes a cada língua foram estimados e o resultado é apresentado na Figura 5

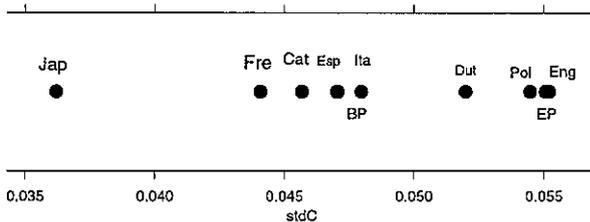


Figura 5: Desvios padrão estimados através da Gama para intervalos consonantais

A hipótese das classes rítmicas pode então ser traduzida nas seguintes hipóteses estatísticas:

1. As línguas silábicas ( Italiano, Espanhol, Francês, Catalão) têm o mesmo desvio padrão que chamaremos de  $\sigma_1$ ;
2. As línguas acentuais ( Polonês, Holandês, Inglês) também têm todas o mesmo desvio padrão  $\sigma_2$ ;
3.  $\sigma_1$ ,  $\sigma_2$  e o desvio padrão do Japonês, que chamaremos  $\sigma_3$ , são diferentes.

Para testar uma hipótese, computamos um número  $T$  com os nossos dados (a chamada "estatística de teste") que mede a discrepância entre os dados e a hipótese. Se  $T$  é muito grande, nós temos uma evidência estatística contra a hipótese. O que seria uma definição razoável de "muito grande?"

Suponhamos que nossos dados gerem  $T=3.5$ ; e que, caso a hipótese seja verdadeira, a probabilidade de obtermos um valor de  $T$  maior que 3.5 seja 0.002. Isto significa que, se a hipótese testada for verdadeira, estaríamos observando um valor extremamente raro. Ou seja, temos elementos estatísticos para acreditar que a hipótese deve ser rejeitada. Esta probabilidade é chamada de nível descritivo do teste.

Para testar o modelo, primeiro testamos (1) e (2) através de um teste da razão de verossimilhança, com um nível descritivo de 0.91, o que significa que a igualdade dos  $\sigma$ 's dentro das classes rítmicas é altamente compatível com os dados ( um valor pequeno do nível descritivo indicaria rejeição).

Finalmente testamos a hipótese de que ao menos dois dos valores  $\sigma_1$ ,  $\sigma_2$ ,  $\sigma_3$  fossem iguais entre si. Esta hipótese foi rejeitada. Por exemplo, a hipótese de que  $\sigma_1 = \sigma_2 = \sigma_3$  foi rejeitada com um nível descritivo de 0.0012.

### **3 Uma medida de sonoridade como critério para identificar classes rítmicas**

A abordagem apresentada em Ramus, Nespore e Mehler tem problemas que comprometem a continuidade do estudo para outras línguas. Primeiro, a marcação dos intervalos vocálicos e intervocálicos é feito à mão,

o que toma muito tempo, inviabilizando a análise de uma base maior de dados. Depois, a marcação manual depende de decisões difíceis de serem feitas de forma homogênea em larga escala.

Uma nova abordagem para o problema é apresentada em Galves, Garcia, Duarte e Galves (2001). Em vez de estudar durações de intervalos vocálicos e consonantais a proposta é estudar os valores de uma função que mede, em cada instante, a “sonoridade” local do sinal acústico.

O fato empírico que dá suporte à sonoridade como uma medida do ritmo linguístico é apresentado em Mehler et al (1996). Neste artigo é apresentado um experimento com bebês recém-nascidos que são capazes de distinguir entre línguas consideradas de classes rítmicas distintas, mesmo com o sinal acústico filtrado a 400 Hz. Neste intervalo de frequências, a distinção entre intervalos vocálicos e consonantais não pode ser feita com precisão, semivogais e vogais ensurdecidas se confundem com consoantes, e consoantes sonoras, como nasais, se confundem com vogais. Neste intervalo de frequências permanece simplesmente a distinção entre zonas sonoras e zonas obstruídas do sinal. Isso sugere que os bebês distinguem as classes rítmicas utilizando classificações menos refinadas do que a oposição vogais e consoantes.

### 3.1 Identificando classes rítmicas através da sonoridade

Uma medida grosseira de sonoridade possibilita automatizar a tarefa de identificar a classe rítmica de uma língua. O preço a pagar por esta abordagem é que a função sonoridade é um processo estocástico extremamente complexo e sua análise estatística exige a criação de novas ferramentas matemáticas.

Queremos definir uma função *sonoridade*  $s(t)$  que assuma valores entre zero e um. A idéia é que quanto mais sonora a região em torno do ponto  $t$ , como em vogais “típicas”, mais perto de um a sonoridade fica. Quanto menos sonora, ou “obstruída”, como em consoantes típicas, mais próximo de zero é o valor da função. Tecnicamente, definimos  $s(t)$ , a função sonoridade, onde  $t$  é o tempo assumindo valores em  $\{ku : k = 1, \dots, T\}$ , sendo  $u$  o intervalo de tempo entre duas colunas do espectrograma e  $T$  é o número de passos considerados no espectrograma

do sinal acústico. Aqui tomamos  $u = 2ms$ . Os valores do espectrograma são estimados com uma janela Gaussiana de 25ms. Consideramos somente frequências entre 0 e 800 Hz com 40 valores distanciados 20Hz um do outro. Os cálculos foram feitos através do programa de livre acesso Praat (<http://www.praat.org>).

Seja  $c_t(i)$  o coeficiente de Fourier da frequência  $i$  em torno do instante  $t$  do espectrograma. Definimos o espectrograma renormalizado por

$$p_t(i) = \frac{c_t(i)^2}{\sum_f c_t(f)^2}. \quad (3)$$

Isso define uma sequência de medidas de probabilidades

$$\{p_t : t = 1, \dots, T\}$$

Padrões regulares, em geral, correspondem a sequências de medidas de probabilidade próximas entre si. Esta "proximidade" pode ser medida pela entropia relativa de uma medida de probabilidade em relação a outra. A entropia relativa é definida da seguinte maneira

$$h(p_t|p_{t-j}) = \sum_f p_t(f) \log \left( \frac{p_t(f)}{p_{t-j}(f)} \right). \quad (4)$$

Definimos então  $s(t)$  como

$$s(t) = 1 - \min \left( 1, \frac{1}{27} \sum_{u=t-4}^{t+4} \sum_{j=1}^3 h(p_u|p_{u-j}) \right). \quad (5)$$

A Figura 6 abaixo mostra um exemplo de uma sentença japonesa. Na faixa superior da Figura temos o sinal acústico e na faixa seguinte temos o espectrograma correspondente. No sinal acústico são bem visíveis as regiões regulares correspondendo aos intervalos vocálicos. No espectrograma estas regiões se traduzem em zonas com padrão regular constituído por duas faixas escuras paralelas. Na terceira faixa está indicada a anotação manual dos fonemas correspondendo às diversas zonas do sinal acústicos. Finalmente, na última faixa aparece o gráfico da

função  $1 - s(t)$ . É bem clara a correspondência entre as zonas do sinal acústico ocupadas por vogais e as zonas em que  $s(t)$  é próximo de 1, ou seja, no gráfico, as zonas em que  $1 - s(t)$  é próximo de zero.

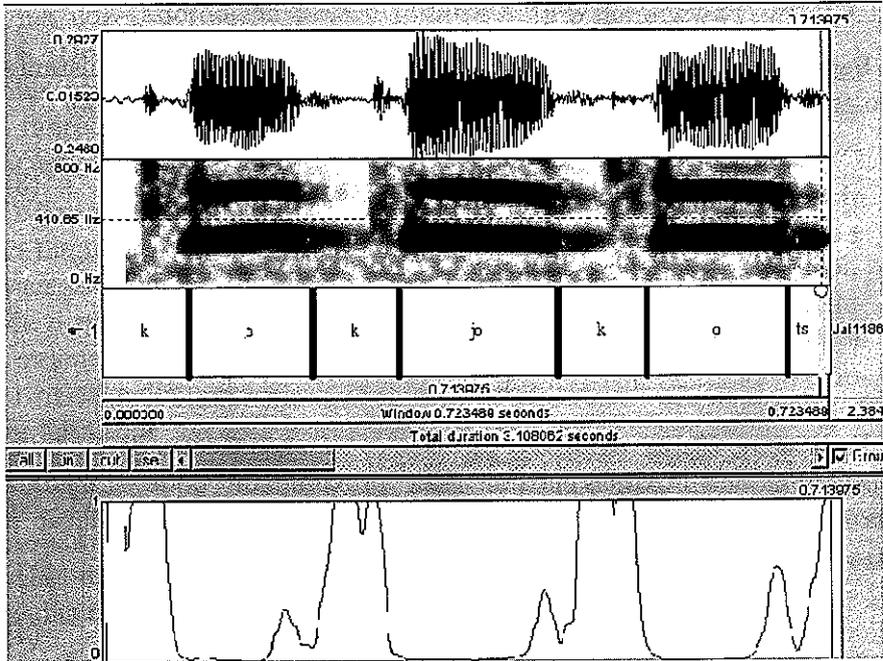


Figura 6: Exemplo de sinal acústico com espectrograma e obstruência de uma sentença do Japonês.

Vamos considerar as seguintes estatísticas da sonoridade:

1. A sonoridade empírica média, que irá fazer o papel de  $\%V$

$$\bar{S} = \frac{1}{T} \sum_{t=1}^T s(t). \quad (6)$$

2. A variação total (que irá fazer o papel de  $\Delta C$ )

$$\delta O = \frac{1}{T} \sum_{t=1}^T |s(t) - s(t-1)|. \quad (7)$$

A Figura 7 apresenta as médias de  $\bar{S}$  e  $\delta O$  para cada língua. Foram utilizados exatamente os mesmos dados acústicos considerados no artigo em Ramus, Nespor e Mehler (1999).

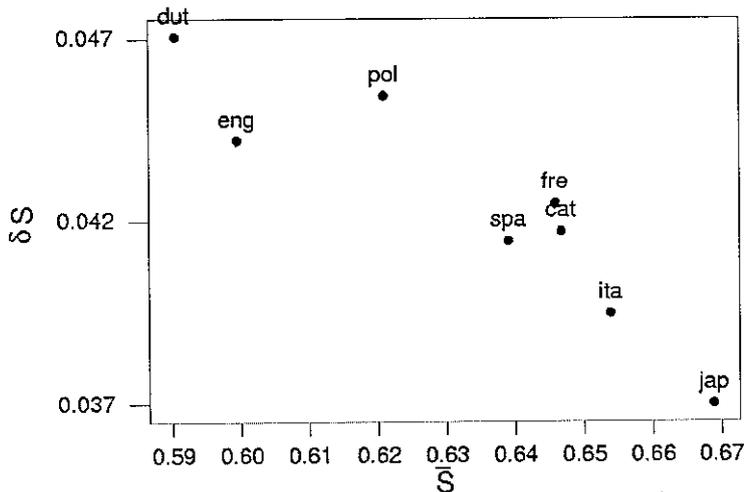


Figura 7: Distribuição das línguas no plano  $(\bar{S}, \delta O)$ .

A Figura 7 mostra um padrão parecido com o da Figura 2. A grande vantagem é que agora este resultado foi obtido sem a necessidade de uma segmentação manual prévia do sinal acústico.

A bem da verdade, a necessidade de intervenção humana na preparação das amostras ainda não foi totalmente eliminada. Com efeito, os dados considerados em Ramus, Nespor e Mehler foram escolhidos de forma a eliminar preliminarmente todos os exemplos discrepantes no que se refere à velocidade de locução.

#### 4 Questões probabilísticas na modelagem da sonoridade

Vimos que os comprimentos dos intervalos consonantais sucessivos podem ser modelados através de uma sequência de variáveis aleatórias independentes com a mesma distribuição. Já a sequência de valores da

função sonoridade são obviamente correlacionados entre si, como se pode ver pelas longas seqüências podendo ter mais de 50 passos com a função oscilando nas vizinhanças de 1.

Seqüências de variáveis aleatórias independentes de mesma distribuição constituem o exemplo mais simples de *Processo Estocástico*. Toda a teoria estatística clássica, em particular, a Lei dos Grandes Números e o Teorema Limite Central, foi desenvolvida primeiramente para este tipo de processo.

A Lei dos Grandes Números diz que para variáveis independentes, com mesma distribuição e com média e variância finitas, a média amostral converge para a média teórica. Isto deve acontecer para qualquer amostra sorteada. Tecnicamente, diz-se que a convergência se dá *quase certamente* (isto é, exceto para um conjunto de amostras de medida nula) quando o tamanho da amostra tende para infinito. Este resultado de convergência dá base ao que se chama de *inferência pontual*.

A Lei dos Grandes Números para variáveis aleatórias independentes garante que os valores de %V obtidos por Ramus, Nespor e Mehler (1999) são praticamente constantes para valores muito grandes de  $n$ .

Naturalmente, numa amostra concreta o valor de  $n$  é sempre finito. Isto significa que a média empírica não será exatamente igual à média teórica. A questão que se coloca então é que amplitude de flutuação da média amostral em torno da média teórica podemos esperar.

O Teorema Limite Central responde esta pergunta: ele diz que as flutuações estão próximas de uma distribuição Normal de média zero e variância  $\frac{\sigma^2}{n}$ , onde  $\sigma^2$  é a variância de cada uma das variáveis produzindo a amostra.

Este resultado teórico fornece a base para a definição dos intervalos de confiança. A dificuldade prática é a estimação de  $\sigma^2$ . Mas isto também pode ser feito a partir da própria amostra.

Para amostras como as produzidas pela função sonoridade, a teoria clássica para variáveis independentes não basta. Agora precisamos de uma Lei de Grandes Números e um Teorema Limite Central para seqüências de variáveis aleatórias correlacionadas.

Tais teoremas têm sido demonstrados desde os anos 50 para processos com alguma propriedade de independência assintótica. Então, se quisermos utilizar estes resultados para a função sonoridade, temos que começar por verificar que ela pode ser descrita através de um processo estocástico tendo as condições de independência assintótica requeridas pelo teorema. Mais precisamente, temos que verificar que a correlação dos valores da sonoridade em dois instantes, separados por  $t$  passos, é uma função de  $t$  que converge para zero suficientemente rápido, quando  $t$  diverge. Isto, em geral, não só para a função sonoridade, é uma tarefa estatística difícil, na fronteira da pesquisa.

Suponhamos já ter obtido evidências de que a sonoridade pode ser descrita por um processo estocástico com boas propriedades de independência assintótica. Isto não termina nosso trabalho de modelagem, precisamos também estimar a variância descrevendo as flutuações da média empírica nesse caso. Esta também é uma tarefa difícil, na fronteira da pesquisa estatística.

A função sonoridade assume valores no intervalo  $[0, 1]$ . No entanto, uma análise exploratória mostra que é possível discretizá-la, utilizando intervalos característicos e codificando o seu valor através do valor médio de cada intervalo. Para modelar esse processo podemos pensar em dois tipos de modelo: cadeias de Markov de alcance variável e cadeias de ordem infinita.

#### 4.1 Cadeias de Markov de alcance variável

Cadeias de Markov de alcance variável são uma classe interessante de processos, ainda de ordem finita, porém bons candidatos a modelar processos como a sonoridade. Elas são apresentadas em Bühlmann e Wyner (1999).

Nesta classe de processos estocásticos, assumindo valores em um conjunto finito, a memória da cadeia tem um alcance variável, dependendo dos valores do passado. Mais precisamente, se  $(X_n)_{n \in \mathbb{Z}}$  é uma cadeia de Markov de alcance variável, as probabilidades de transição homogêneas são dadas por

$$\mathbf{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots) = \mathbf{P}(X_t = x_t | X_{t-1} = x_{t-1}, \dots, X_{t-l} = x_{t-l}),$$

onde  $l = l(x_{t-1}, x_{t-2}, \dots)$  é uma função do passado.

Um dos maiores interesses das cadeias de Markov de alcance variável é sua parcimônia. Enquanto em modelos de cadeia de Markov de ordem finita e constante o número de parâmetros cresce exponencialmente com a ordem, em cadeias de Markov de alcance variável muitos desses parâmetros desaparecem com a perda de memória do processo para algumas histórias (passados) diminuindo o número de parâmetros a serem estimados e permitindo um melhor ajuste do modelo. Uma apresentação elementar e auto-suficiente da teoria das cadeias de Markov assumindo valores em um conjunto finito pode ser encontrada em Ferrari e Galves (1997).

Bühlmann e Wyner (1999) apresenta um algoritmo que produz estimativas com boas propriedades assintóticas, mesmo quando o modelo cresce em dimensionalidade. Eles também apresentam uma estimação consistente para o espaço de estados mínimo e propriedades de mistura dos modelos ajustados.

## 4.2 Cadeias de ordem infinita

É também natural pensarmos numa cadeia de ordem infinita para modelar a sonoridade. Mais precisamente, trabalharemos com um processo estacionário  $(X_n)_{n \in \mathbb{Z}}$ , assumindo valores em um conjunto finito  $A$  e satisfazendo as seguintes propriedades.

1. Para todo  $a_1, \dots, a_n \in A$ ,

$$\mathbf{P}(X_1 = a_1, \dots, X_n = a_n) > 0 \quad (8)$$

2. o limite

$$\lim_{m \rightarrow \infty} \frac{\mathbf{P}(X_0 = a_0 | X_j = a_j, -m \leq j \leq -1)}{\mathbf{P}(X_0 = a_0 | X_j = a_j, j \leq -1)} = \quad (9)$$

existe para todo  $a_j, j \leq -1$ .

A primeira condição garante que todas as sequências são possíveis. A segunda condição garante que a probabilidade de transição é uma função contínua do passado.

Vamos também assumir a seguinte condição de independência assintótica:

3. Existe uma sequência  $(\gamma_m)_{m \geq 1}$ , com  $\beta_m \rightarrow 0$ , tal que, para todo  $a_j, b_j \in A, j \leq -1$  com  $a_j = b_j$  para  $-1 \geq j \geq -m$ , temos a desigualdade

$$\begin{aligned} & |\mathbf{P}(X_0 = a_0 \mid X_j = a_j, j \leq -1) - \\ & \mathbf{P}(X_0 = a_0 \mid X_j = b_j, j \leq -1)| \leq \mathbf{b}_m. \end{aligned} \quad (10)$$

Necessitaremos também assumir condições de decrescimento rápido da sequência  $\beta_m$ , como por exemplo  $\beta_m = e^{-cm}$  ou  $\sum_{m \geq 1} \beta_m < +\infty$ .

Processos desse tipo são chamados de cadeias de ordem infinita, com perda de memória exponencial ou somável.

Cadeias de ordem infinita foram introduzidos na literatura probabilística pelos artigos de Onicescu e Mihoc (1935), e Doeblin e Fortet (1937) e Harris (1955). Alguns resultados recentes relacionados ao tema deste curso podem ser encontrados em Bressaud, Galves e Fernández (1999a,1999b). Uma apresentação elementar e atual do tema pode ser encontrada em Fernández, Ferrari e Galves (2001).

A análise estatística desses processos, ainda incipiente, vive um período de plena expansão. Um resultado recente de reamostragem apresentado em Collet, Duarte e Galves (2003), permite resolver a questão da estimação da variância em cadeias de ordem infinita. Assumindo para a sequência de valores discretizada da sonoridade um modelo de cadeia de ordem infinita, podemos aplicar o procedimento de reamostragem proposto em Collet, Duarte e Galves (2003) para estimar a variância da média da sonoridade das línguas e construir intervalos de confiança em torno destas médias que nos permitam compará-las. A Figura 8 abaixo mostra os intervalos de confiança para cada uma das línguas consideradas obtidos através deste procedimento de reamostragem. O que observamos é que apenas conseguimos encontrar dois grupos: um com o Japonês (Jap) e outro com Holandês (Dul) e Inglês (Eng). Isto não significa que existam apenas estes grupos, mas que, com as ferramentas que temos até agora não conseguimos identificar outros grupos através da sonoridade. Mas o trabalho de pesquisa continua e esperamos em breve poder finalizar a história começada por Ramus, Nespor e Mehler.

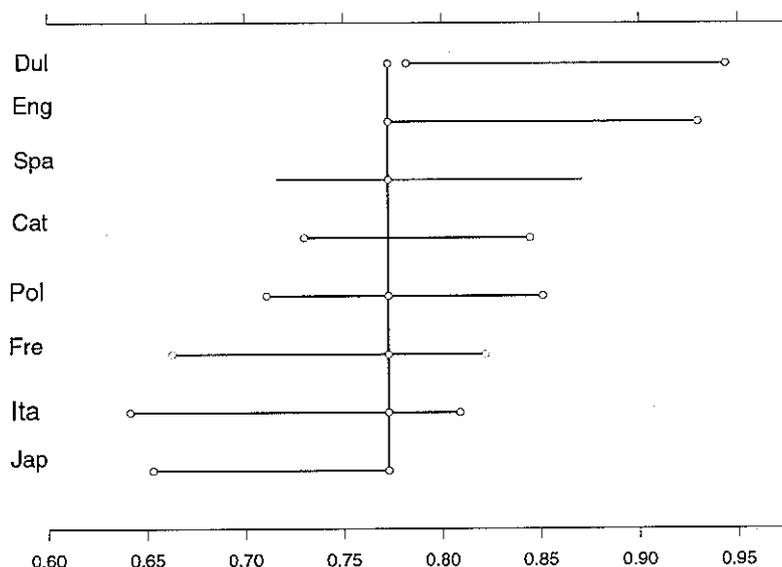


Figura 8: Intervalos de confiança de reamostragem da sonoridade discretizada para cada uma das línguas consideradas.

Uma boa parte do material deste trabalho é original e foi produzido no quadro do Projeto Temático FAPESP *Padrões rítmicos, fixação de parâmetros, e mudança lingüística*. Mais informações sobre o projeto e alguns dos artigos nos quais estas idéias foram apresentadas originalmente podem ser encontrados na página do projeto <http://www.ime.usp.br/~tycho>, com especial destaque para a página <http://www.ime.usp.br/~tycho/prosody/sonority>.

## Referências

- [1] Abercrombie, D.; 1967. *Elements of General Phonetics*. Chicago: Aldine.

- 
- [2] Bressaud, X., Fernández, R., Galves, A.; 1999a. Speed of  $\bar{d}$ -convergence for Markov approximations of chains with complete connections. a coupling approach. *Stoch. Proc. and Appl.*, 83:127–38.
- [3] Bressaud, X., Fernández, R., Galves, A.; 1999b. Decay of correlations for non Hölderian dynamics. a coupling approach. *Elect. J. Prob.*, 4. (<http://www.math.washington.edu/~ejpecp/>).
- [4] Bühlmann, P., Wyner, A.; 1999. Variable length Markov chains. *Ann. Statist.*, 27, 379-423.
- [5] Collet, P., Duarte, D., Galves, A.; 2003. Markov approximations and the bootstrap for chains of infinite order. *Manuscrito*.
- [6] Duarte, D., Galves, A., Lopes, N., Maronna, R.; 2001. The statistical analysis of acoustic correlates of speech rhythm. *Workshop on Rhythmic patterns, parameter setting and language change*, ZiF, University of Bielefeld. <http://www.physik.uni-bielefeld.de/complexity/duarte.pdf>
- [7] Fernández, R., Ferrari, P., Galves, A.; 2001. Coupling renewal and perfect simulation of chains of infinite order. *Notas da 5a. Escola Brasileira de Probabilidade*. <http://www.ime.usp.br/~galves/livro>
- [8] Ferrari, P., Galves, A. 1997. Acoplamentos e processos estocásticos. IMPA, Rio de Janeiro, Brasil. <http://www.ime.usp.br/~pablo>
- [9] Galves, A., Garcia, J., Duarte, D., Galves, Ch.; 2000. Sonority as a basis for rhythmic class discrimination. In *Proceedings of the Speech Prosody 2002 conference*. <http://www.lpl.univ-aix.fr/sp2002/papers.htm>
- [10] Harris, T.E. 1955. On chains of infinite order. *Pacific J. Math.*, 5:707–24.
- [11] Lloyd, J. (1940). *Speech signal in telephony*. London
- [12] Onicescu, O., Mihoc, G., 1935. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris*, 200:511–12.
- [13] Ramus, F., Nespore, M., Mehler, J.; 1999. Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

Antonio Galves  
Instituto de Matemática e Estatística  
USP Rua do Matão, 1010  
CEP 05508-090  
São Paulo Brasil  
galves@ime.usp.br

Denise Duarte  
Instituto de Matemática e Estatística  
Campus II - UFG Caixa Postal 131  
CEP 74001-970  
Goiania-GO  
denise@mat.ufg.br