東京大学 工学部 計数工学科/物理工学科

応用音響学：

スペクトルパターンベクトルと距離尺度

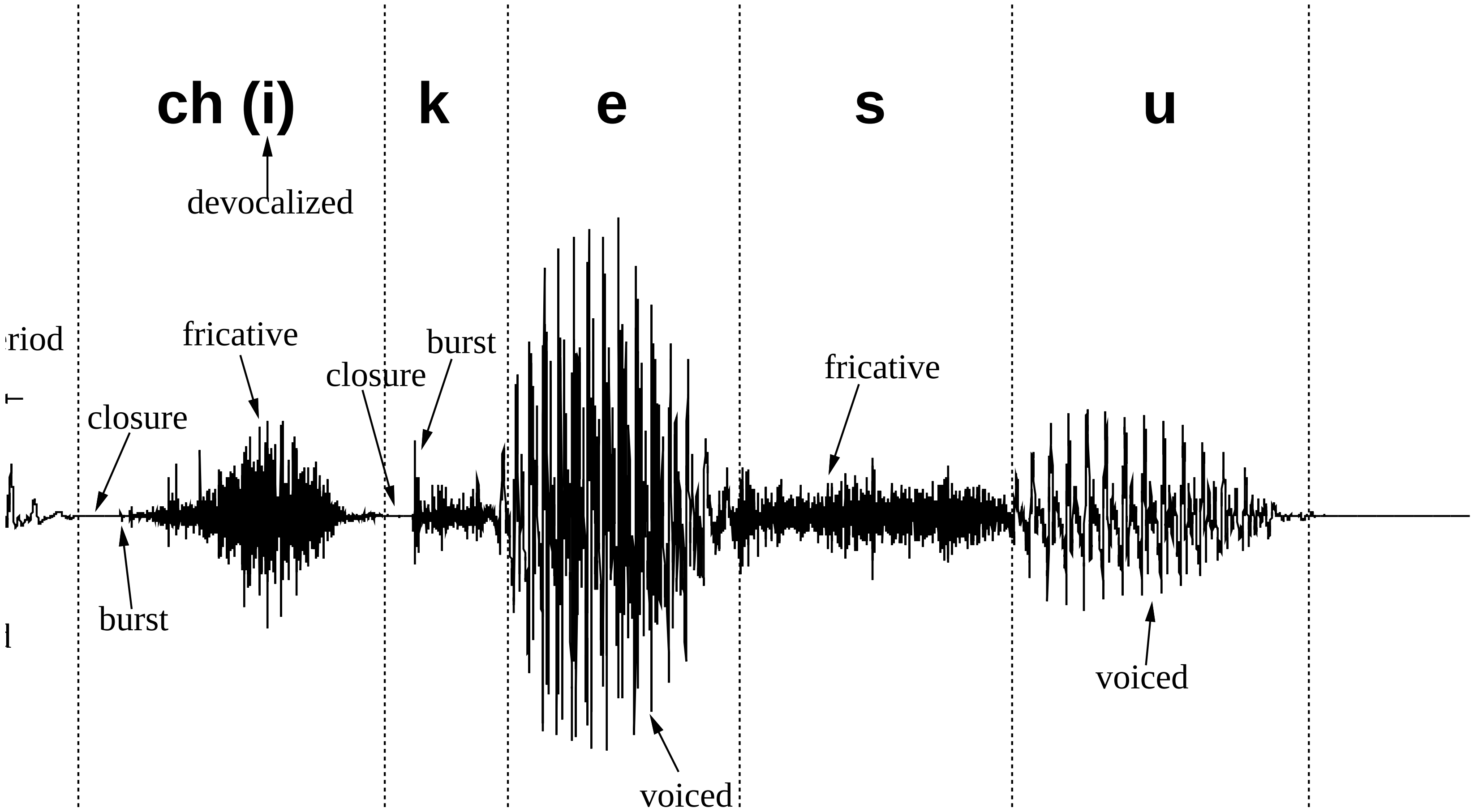
嵯峨山 茂樹 <sagayama@hil.t.u-tokyo.ac.jp>

東京大学 工学部 計数工学科 http://hil.t.u-tokyo.ac.jp/

参考文献

- 北 研二・中村 哲・永田 昌明「音声言語処理」森北出版
- 中川 聖一「パターン情報処理」丸善
- 古井 貞熙「音声情報処理」森北出版
- 谷萩 隆嗣「音声と画像のディジタル信号処理」コロナ社
- 中川 聖一「確率モデルによる音声認識」コロナ社

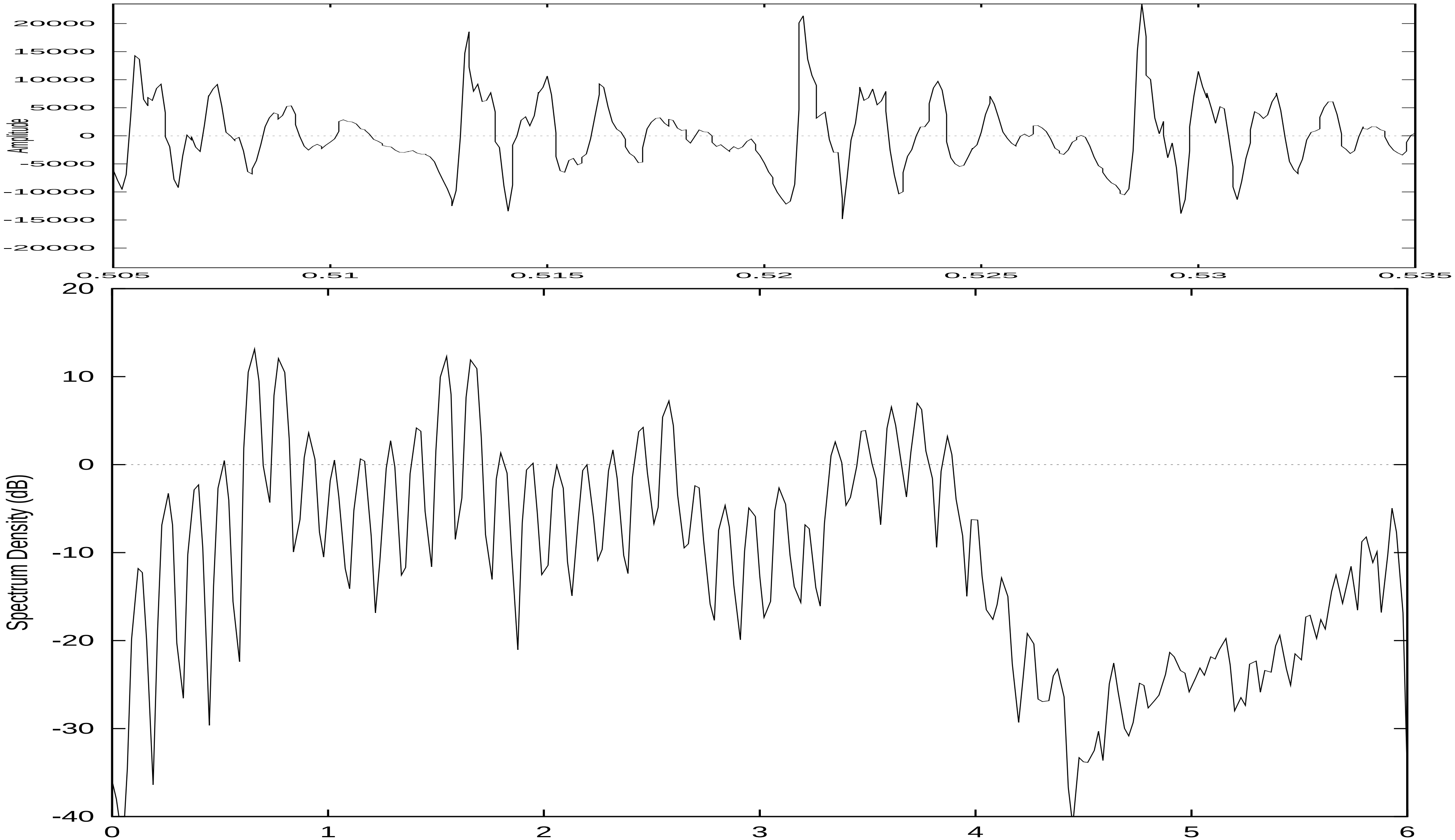
音声波形の例 (男声 単語)



Speech Waveform /uchikesu/

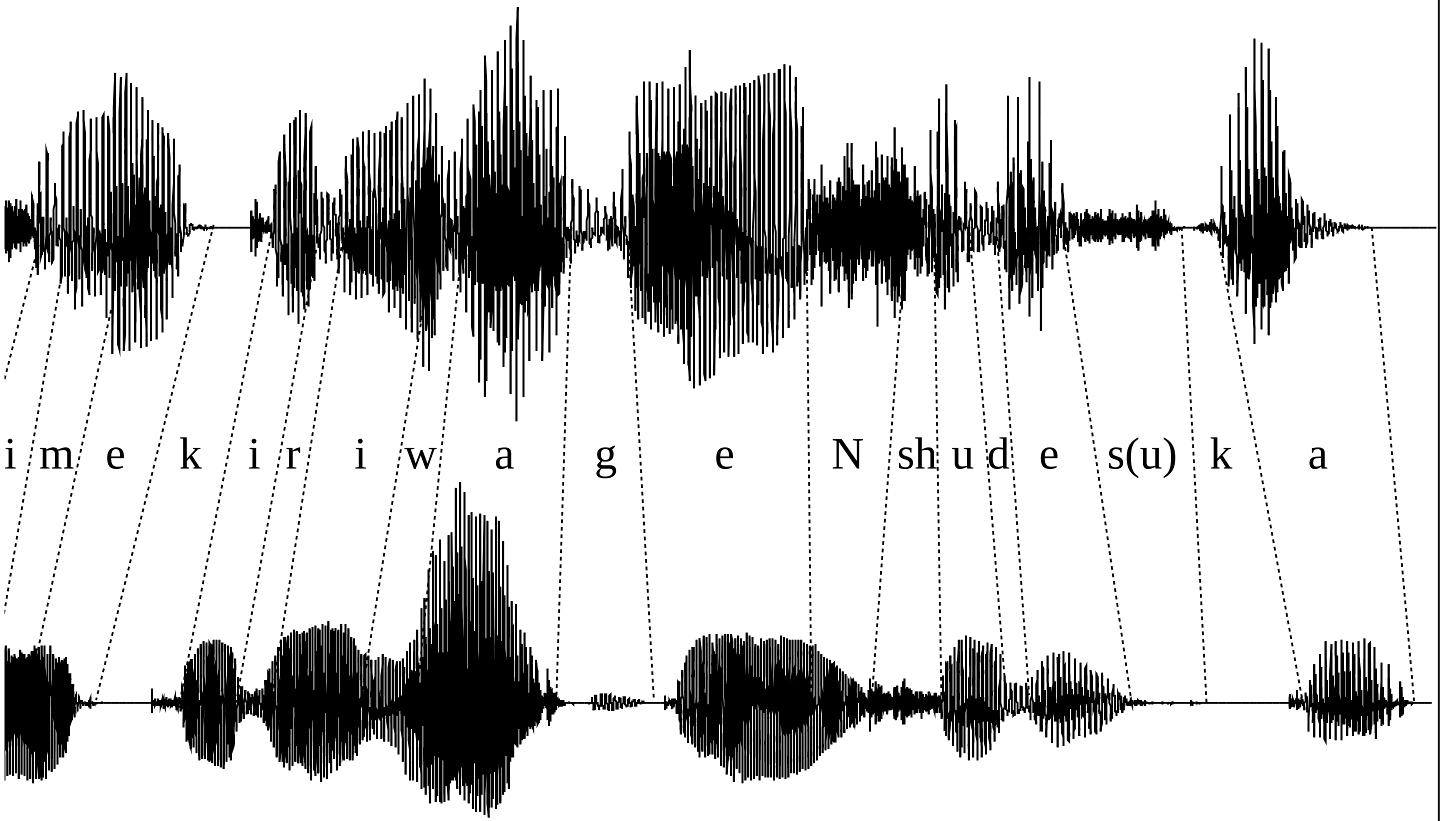
図1. 音声波形の例「打ち消す」

音声の短時間スペクトル(periodogram)の例



短時間音声波形とスペクトルの例 (男声 /a/)

音声時間パターンの変動



Tempral Difference between Different Speakers
/ shimekiri wa genshu desu ka /

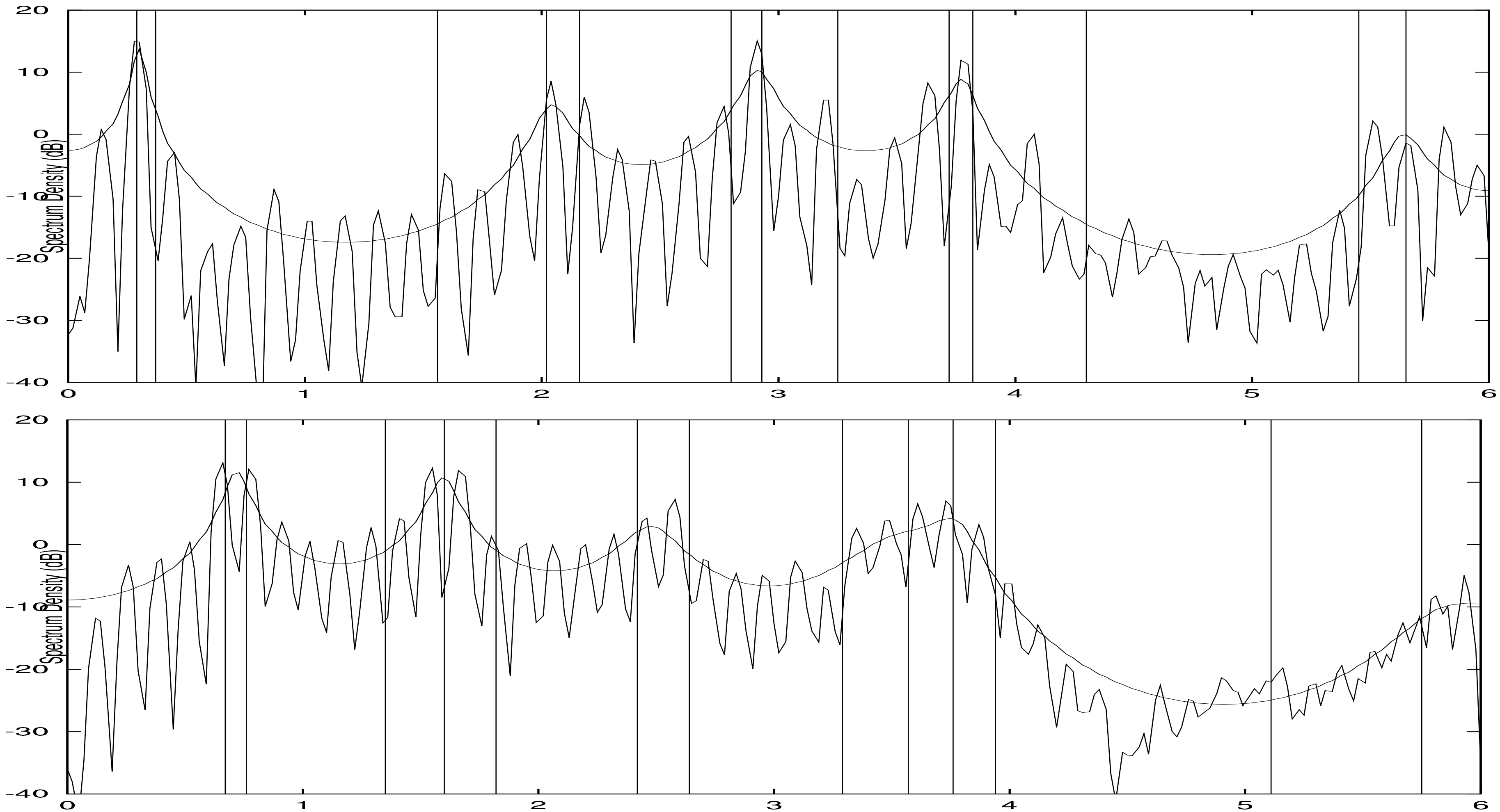
Shigeki Sagayama, ATR Interpreting Telephony Research Laboratories

図2. 日本語「締切は厳守ですか」の男女２名の音声波形

スペクトルパラメータベクトル

- **スペクトルパラメータベクトルの例：**
  - **短時間スペクトル:** $I_N(\omega) \to \boldsymbol{s} = (s_1, s_2, \cdots, s_N)^T,\ s_i = I_N(\omega_i)$
  - **ケプストラム:** $\boldsymbol{c} = (c_1, c_2, \cdots, c_p)^T$
  - **LPC係数** $\boldsymbol{a} = (a_1, a_2, \ldots, a_p)^T$
  - **PARCOR係数** $\boldsymbol{k} = (k_1, k_2, \cdots, k_p)^T$
  - **LSP周波数** $\boldsymbol{f} = (\omega_1, \omega_2, \cdots, \omega_p)^T$

音声の短時間スペクトルとLSP周波数の例



短時間スペクトルと LSP 周波数の例 (男声 /i/ と /a/)

ベクトルの概念と基本事項 (2 次元)

- **基本事項**
  - **ベクトルの成分表示:** $\boldsymbol{a} = (a_1, a_2)$
  - **ベクトルの線形独立性** $\boldsymbol{x} = c_1\boldsymbol{a} + c_2\boldsymbol{b}$
  - **ベクトル空間 (線形空間)**
  - **ベクトルのノルム:** $||\boldsymbol{a}|| = \sqrt{a_1^2 + a_2^2} \geq 0$
  - **ベクトル間の (Euclid) 距離:**

$$d(\boldsymbol{a}, \boldsymbol{b}) = |\boldsymbol{a} - \boldsymbol{b}| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

  - **ベクトルの内積:** $\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 = |\boldsymbol{a} \cdot \boldsymbol{b}| \cos\theta$
  - **ベクトルの直交:** $\boldsymbol{a} \cdot \boldsymbol{b} = 0$**, i.e., (**$\theta = \pm\pi/2$**)**
  - **ベクトルの射影 (**a, b**は直交):**

$$\boldsymbol{x} = c_1\boldsymbol{a} + c_2\boldsymbol{b} = \frac{(\boldsymbol{x}, \boldsymbol{a})}{(\boldsymbol{a}, \boldsymbol{a})}\boldsymbol{a} + \frac{(\boldsymbol{x}, \boldsymbol{b})}{(\boldsymbol{b}, \boldsymbol{b})}\boldsymbol{b}$$

n 次元ベクトル空間

- n 次元ベクトル
  - ベクトル: $\boldsymbol{v} = (v_1, v_2, \cdots, v_n)$
  - ベクトルの線形独立性 $\boldsymbol{x} = c_1\boldsymbol{u} + c_2\boldsymbol{v}$
  - ベクトル空間 (線形空間)
  - ベクトルのノルム: $||\boldsymbol{v}|| = \sqrt{\sum_{i=1}^{n} v_i^2} \geq 0$
  - ベクトル間の(Euclid)距離: $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = \sqrt{\sum_{i=1}^{n} (u_i - v_i)^2}$
  - ベクトルの内積: $\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i$
  - ベクトルの直交: $\boldsymbol{u} \cdot \boldsymbol{v} = 0$, i.e., ($\theta = \pm\pi/2$)
  - ベクトルの射影 (v_i は直交ベクトル基底):

$$\boldsymbol{x} = \sum_{i=1}^{n} c_i \boldsymbol{v}_i = \sum_{i=1}^{n} \frac{\boldsymbol{x} \cdot \boldsymbol{v}_i}{||\boldsymbol{v}_i||^2} \boldsymbol{v}_i$$

正規直交ベクトル空間

正規直交ベクトル空間

- 正規直交基底ベクトル$\{\boldsymbol{u}_i\}$: $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- 任意のベクトル $\boldsymbol{x}$ は、$\boldsymbol{x} = c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_n\boldsymbol{u}_n$ と表せる。
- 部分和(m個, $m < n$)での最良近似

$$\boldsymbol{x} = \sum_{i=1}^{m} c_i\boldsymbol{u}_i + \varepsilon$$

とすると誤差のノルム$||\varepsilon||^2$が最小になる条件は、

$$c_i = \frac{\boldsymbol{x} \cdot \boldsymbol{u}_i}{||\boldsymbol{u}_i||^2}$$

パターン空間と距離尺度

- **パターンベクトル(特徴ベクトル):** $\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{pmatrix}$
- **距離の公理**
  - $d(\boldsymbol{x}, \boldsymbol{x}) = 0$
  - $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x}) > 0$
  - $d(\boldsymbol{x}, \boldsymbol{z}) + d(\boldsymbol{z}, \boldsymbol{y}) \geq d(\boldsymbol{x}, \boldsymbol{y})$ **(三角不等式)**
- **距離尺度の例**
  - **Euclid 距離:**

$$D_E(f, g) = ||\boldsymbol{v}_f - \boldsymbol{v}_g||^2 = \sum_{i=1}^{p} (v_i^{(f)} - v_i^{(g)})^2$$

  - **Mahalanobis 距離:**

$$D_M(f, g) = (\boldsymbol{v}_f - \boldsymbol{v}_g)^T \boldsymbol{V}^{-1} (\boldsymbol{v}_f - \boldsymbol{v}_g) = \sum_{i=1}^{p} (v_i^{(f)} - v_i^{(g)})^2$$

$\boldsymbol{V}$ は自己共分散行列。(正値行列)

音声スペクトル間距離

- 瞬時スペクトル間の距離尺度

$$\text{距離} = d(\ \text{スペクトル}\, f(\omega),\ \text{スペクトル}\, g(\omega)\ )$$

- 距離尺度の例
  - 板倉・齋藤距離
  - COSH 尺度
  - WLR 尺度(重みつき尤度比)
  - PWLR 尺度(パワー重みつき尤度比)
  - LPC ケプストラム間ユークリッド距離
  - LPC 重みつきケプストラム間ユークリッド距離

パターン空間

1. パターンベクトル $\boldsymbol{x} = (x_1, x_2, \cdots, x_p)^T$
2. p次元ベクトル空間
3. パターン間距離尺度: $\boldsymbol{x}, \boldsymbol{y}$間

- ユークリッド(Euclid)距離

$$d(\boldsymbol{x}, \boldsymbol{y}) = \| \boldsymbol{x} - \boldsymbol{y} \|^2 = \sum_{i=1}^{p} |x_i - y_i|^2$$

- マハラノビス(Mahalanobis)距離

$$d(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{V}^{-1} (\boldsymbol{x} - \boldsymbol{y})$$

$\boldsymbol{V}$: x, yの共分散行列。

正規分布

確率変数: x
平均: μ
分散: σ^2
1次元正規分布:

$$\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma}}$$

対数表現:

$$L(\mu,\sigma^2;x) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma - \underbrace{\frac{(x-\mu)^2}{2\sigma}}_{\text{Mahalanobis 距離}}$$

n 次元正規分布

確率変数ベクトル: $\boldsymbol{x}$
平均ベクトル: $\boldsymbol{\mu}$
分散行列: Σ
n **次元正規分布の確率密度分布:**

$$\mathcal{N}(x;\mu,\sigma^2) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

対数表現(対数尤度):

$$L(\boldsymbol{\mu},\boldsymbol{\Sigma};\boldsymbol{x}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \underbrace{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}_{\text{Mahalanobis 距離}}$$