

【技術分類】 6 - 1 - 1 音声 / 特徴抽出技術 / 特徴パラメータ

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 1 - 1 音声認識と話者認識

【技術内容】

音声波に含まれる情報を抽出し認識する過程は、図のように通信理論(情報処理論)の問題として、確率モデルで定式化できる。情報源は、話者が文を考える文発生部に対応する。音声認識システムを音声分析部(特徴抽出部)と言語復号部に分けて考え、話者発声部と音声分析部を合わせて、一つの音響チャンネルとしてモデル化する。この音響チャンネルは、歪(雑音)のある通信路に対応する。

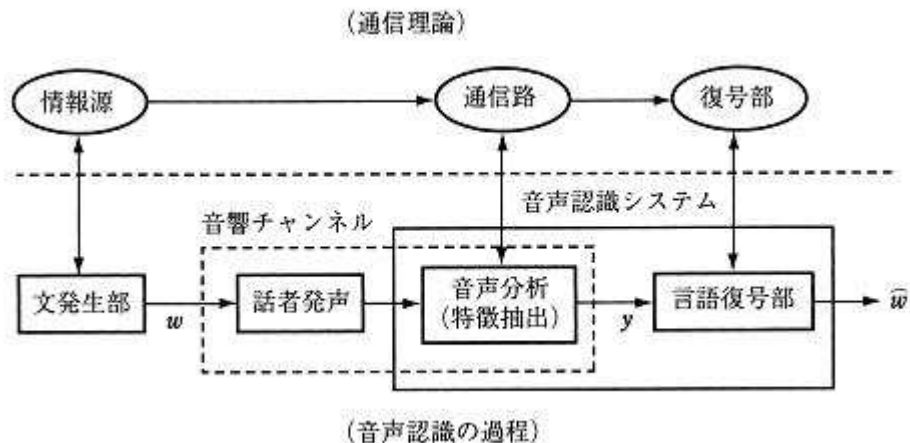
話者は文 w に従い、音声波形 s を生成する。 s には、話者の個人差、付加雑音、伝送歪などが重畳される。音声分析部は、音声波形データの分析・変換を行い、短時間スペクトルなどの時系列データ(ベクトル系列) y を出力する。言語復号部は、 y から信号文の推定値 $\hat{w}$ を出力する。 $\hat{w}$ は、事後確率 $P(w|y)$ が最大となるよう、ベイズ則により次式から求める。

$$P(\hat{w}|y) = \max_w \frac{P(y|w)P(w)}{P(y)}$$

ここで、話者認識においては信号文の内容の推定は必要なく、出力時系列データ y が、その話者が文 w を発声したときのものであろう確からしさ(尤度) $P(y|w)$ から、その声が当人のものであるか否かを判定する。そのためには、どのような特徴パラメータを用いて、どのような話者モデルを立てるかが重要である。

【図】

通信理論との対応



出典：「音声情報処理」,「電子情報通信工学シリーズ」, 1998年6月30日発行、古井貞熙著、森北出版株式会社発行、80頁 図5-1 音声認識過程の確率モデル

【出典 / 参考資料】

「音声情報処理」,「電子情報通信工学シリーズ」, 1998年6月30日発行、古井貞熙著、森北出版株式会社発行、80頁

【技術分類】 6 - 1 - 1 音声 / 特徴抽出技術 / 特徴パラメータ

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 1 - 2 スペクトル包絡とフォルマント

【技術内容】

話者認識に用いられる物理特徴としては、

- ・声の高さ、アクセント、イントネーションに関連する「ピッチ」(基本周波数)
- ・声の質に関連する「スペクトル包絡」(スペクトル概形)

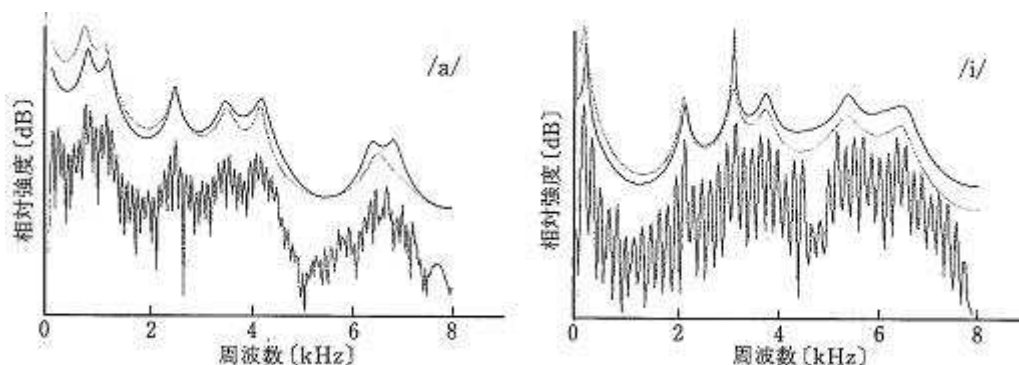
などがある。他に、発声レベル(声の大きさ)、発声速度なども個人差を有するが、抽出は難しい。ピッチは有用な特徴であるが、正確な測定が難しく、他人にも真似しやすく、個人内変動が大きいといった問題がある。

スペクトル包絡には図のように優勢な周波数成分が複数あり、周波数の低い側から順に第 1、第 2、第 3...フォルマントと呼ばれる。第 1、第 2 フォルマントが主に母音に関係するが、声道の形が急に变化できないため、スペクトル包絡は時間的に複雑に変化する。フォルマント周波数は声道の共振周波数に対応しているため、スペクトル包絡を真似ることは極めて難しく、そのため個人の識別に有効である。

高域の成分からは、発声時期による変動や雑音に対して安定した情報が得られる。計算機能力が低い時代には、5kHz 付近の話者間/話者内分散比(F 比)や、第 3、第 4 フォルマントなどに注目して話者を識別することがなされていたが、近年ではスペクトル包絡の全体の形が用いられることが多い。

【図】

音声スペクトルの微細構造と包絡の例



出典：「改訂 音声」,「日本音響学会編 音響工学講座7」,1995年2月20日発行、中田和男著、株式会社コロナ社発行、30頁 図4・1 日本語母音の周波数スペクトルの一例

【出典 / 参考資料】

・「音声情報処理」,「電子情報通信工学シリーズ」,1998年6月30日発行、古井貞熙著、森北出版株式会社発行、11頁

・「HMM による話者認識」,「電子情報通信学会技術研究報告 音声 SP95-111 pp.17-24」,1996年1月発行、松井知子著、電子情報通信学会発行

・「改訂 音声」,「日本音響学会編 音響工学講座7」,1995年2月20日発行、中田和男著、株式会社コロナ社発行、30頁

【技術分類】 6 - 1 - 1 - 3 音声 / 特徴抽出技術 / 特徴パラメータ / ケプストラム

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 1 - 3 - 1 ケプストラムの抽出

【技術内容】

音声のスペクトル包絡はケプストラムで表現されることが多い。ケプストラム(cepstrum)とは、対数スペクトラムを逆フーリエ変換したものであり、spectrum のつづりを逆にした造語である。同様にその変数は、frequency を逆にしてケフレンシー(quefrequency)と呼ばれる。ケフレンシーの次元は、周波数の逆変換であるから時間となる。音声波形は、音源波形と声道共振系の各インパルス応答の畳み込みで表される。各フーリエ変換を $G(\lambda)$ 、 $H(\lambda)$ と表すと、次の関係が得られる。

$$X(\lambda) = G(\lambda) H(\lambda)$$

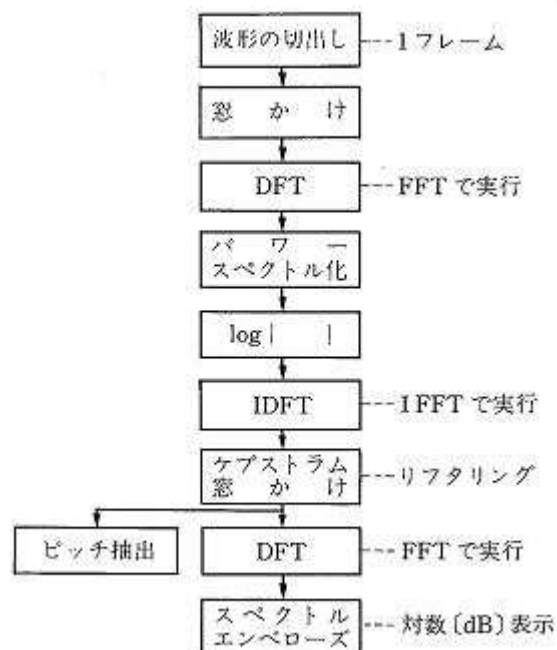
対数をとることにより積を和に分解し、フーリエ逆変換を取ると、ケプストラム $c(\tau)$ が得られる。

$$c(\tau) = F^{-1} \log |X(\omega)| = F^{-1} \log |G(\omega)| + F^{-1} \log |H(\omega)|$$

スペクトル包絡のピーク同士の周波数間隔が広ければ、それを波形とみなしたときのスペクトルは低周波成分しか持たない。つまり逆変換された次元で、ケプストラムの低域成分はスペクトル包絡の全体的な形に、高域成分は音源パルス列に相当する。そこで適当な周波数分離により、音源成分と包絡成分とに分離できる。包絡成分の抽出は、個人の声道の音響特性を抽出していることに相当する。

【図】

基本周期とスペクトル包絡の抽出の手順



出典：「改訂 音声」、「日本音響学会編 音響工学講座 7」、1995 年 2 月 20 日発行、中田和男著、株式会社コロナ社発行、31 頁 図 4・2 ケプストラム分析の手順

【出典 / 参考資料】

・「音声情報処理」、「電子情報通信工学シリーズ」、1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、25 頁

・「改訂 音声」、「日本音響学会編 音響工学講座 7」、1995 年 2 月 20 日発行、中田和男著、株式会社コロナ社発行、31 頁

【技術分類】 6 - 1 - 1 - 3 音声 / 特徴抽出技術 / 特徴パラメータ / ケプストラム

【 F I 】 G10L3/00

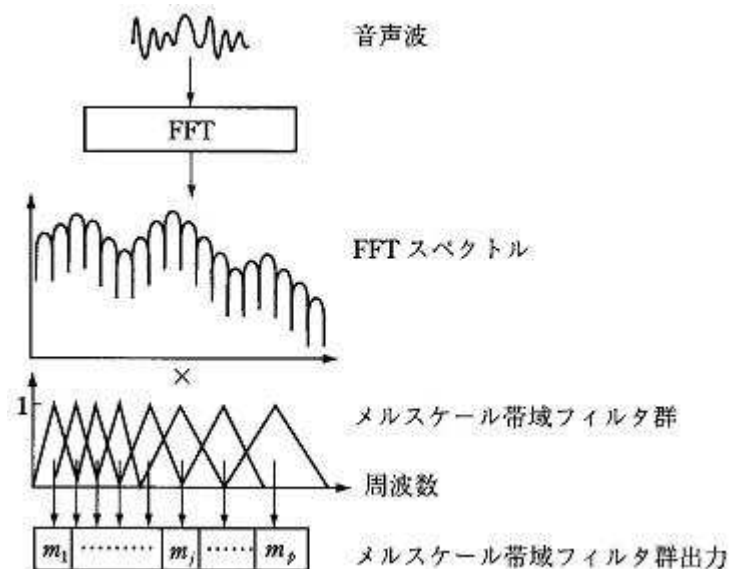
【技術名称】 6 - 1 - 1 - 3 - 2 メルケプストラム

【技術内容】

音声波のスペクトルを、人の聴覚に近い周波数間隔に切り分けてケプストラム化したものである。人の聴覚は、低い周波数では細かく、高い周波数では粗い周波数分解能を持つことが知られている。これはメル(mel)尺度と呼ばれ、対数に近い非線形特性を示す。音声を認識するためには、音声スペクトルから周波数成分ごとの時系列データを抽出する必要があるが、人の聴覚に合わせるため、各帯域フィルタを対数周波数軸上、あるいはメルスケール上に等間隔に配置して抽出する。FFT によるスペクトルを元に、メルスケールの帯域フィルタ群出力を抽出する手順を図に示す。また、全域通過フィルタの周波数軸を伸縮する方法もある。こうして抽出されたスペクトルのケプストラムはメルケプストラムと呼ばれ、種々の音声認識システムで広く用いられている。

【図】

メルケプストラム抽出手順



出典：「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、86 頁 図 5-3 FFT に基づくメルスケール帯域フィルタ群分析の手順

【出典 / 参考資料】

「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、86 頁

【技術分類】 6 - 1 - 1 - 3 音声 / 特徴抽出技術 / 特徴パラメータ / ケプストラム

【 F I 】 G10L3/00

【技術名称】 6 - 1 - 1 - 3 - 3 LPC ケプストラム

【技術内容】

線形予測(LPC)法によりモデル化した波形から計算したケプストラムである。音声信号のスペクトル密度を全極型音声生成システム関数で表し、ケプストラムを抽出する際の DFT、対数変換、IDFT をそれぞれ両側 z 変換、複素対数、両側 z 変換で置き換えて、複素ケプストラムに拡張する。系列 $x(n)$

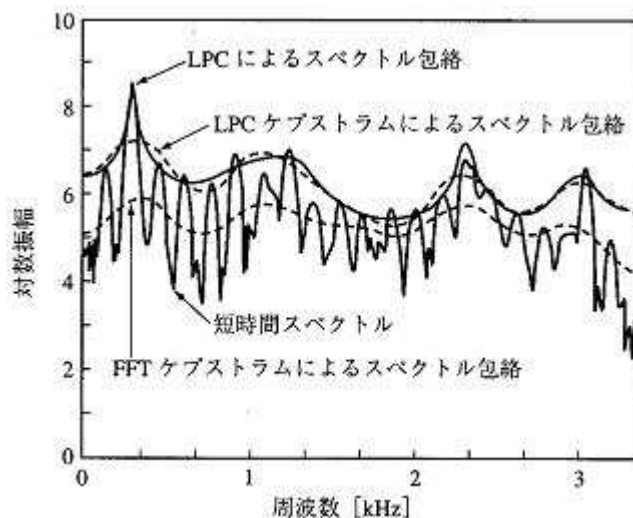
の複素ケプストラムを $\hat{c}_n$ で表すと、次の再帰式が得られる。

$$\begin{aligned}\hat{c}_1 &= -\alpha_1 \\ \hat{c}_n &= -\alpha_n - \sum_{m=1}^{n-1} \left(1 - \frac{m}{n}\right) \alpha_m \hat{c}_{n-m} \quad (1 < n \leq p) \\ \hat{c}_n &= -\sum_{m=1}^p \left(1 - \frac{m}{n}\right) \alpha_m \hat{c}_{n-m} \quad (p < n)\end{aligned}$$

図に、波形から直接計算した FFT ケプストラムと、LPC ケプストラム、LPC(最尤法)による各スペクトル包絡の比較を示す。LPC 法を用いると、ケプストラムでも最尤法でも、FFT ケプストラムによるものよりもピークが強調され、音素や個人的特徴がより強く抽出される。

【図】

スペクトル包絡の比較



出典：「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、39 頁 図 2-11 LPC ケプストラムと FFT ケプストラムによるスペクトル包絡の比較

【出典 / 参考資料】

「音声情報処理」,「電子情報通信工学シリーズ」, 1998 年 6 月 30 日発行、古井貞熙著、森北出版株式会社発行、39 頁