画像情報特論 (5)

- ディジタル圧縮とメディア表現 (2)

音声・オーディオ、SMIL、グラフィクス

情報ネットワーク専攻　甲藤二郎

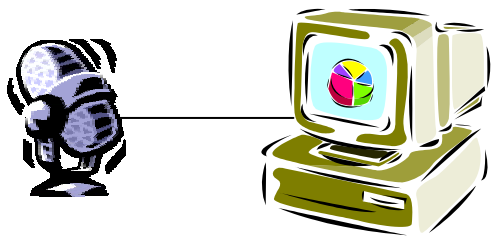
E-Mail: katto@waseda.jp

音声・オーディオ圧縮の原理

ディジタルオーディオ

- キャプチャ＆圧縮

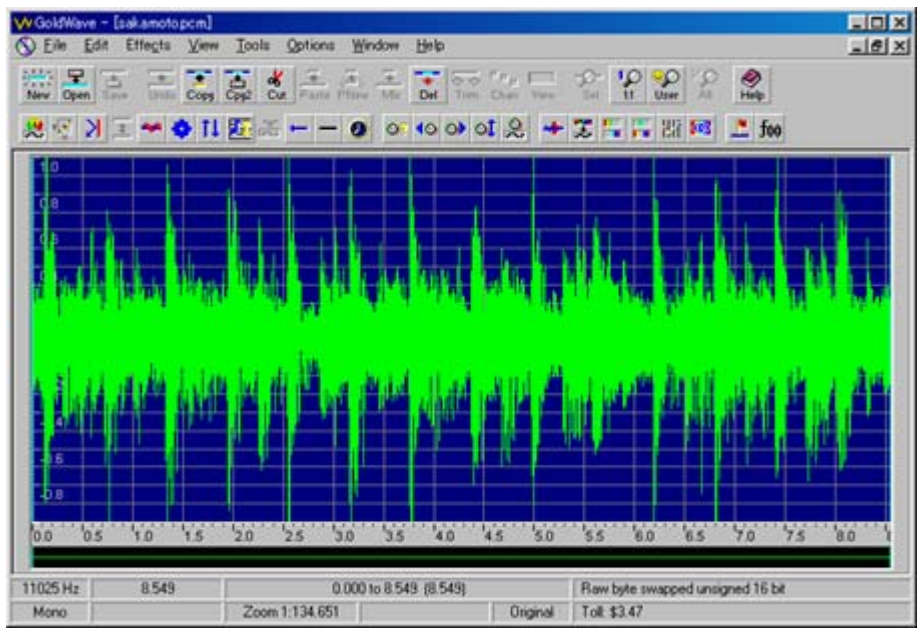
マイク　　サウンドキャプチャ



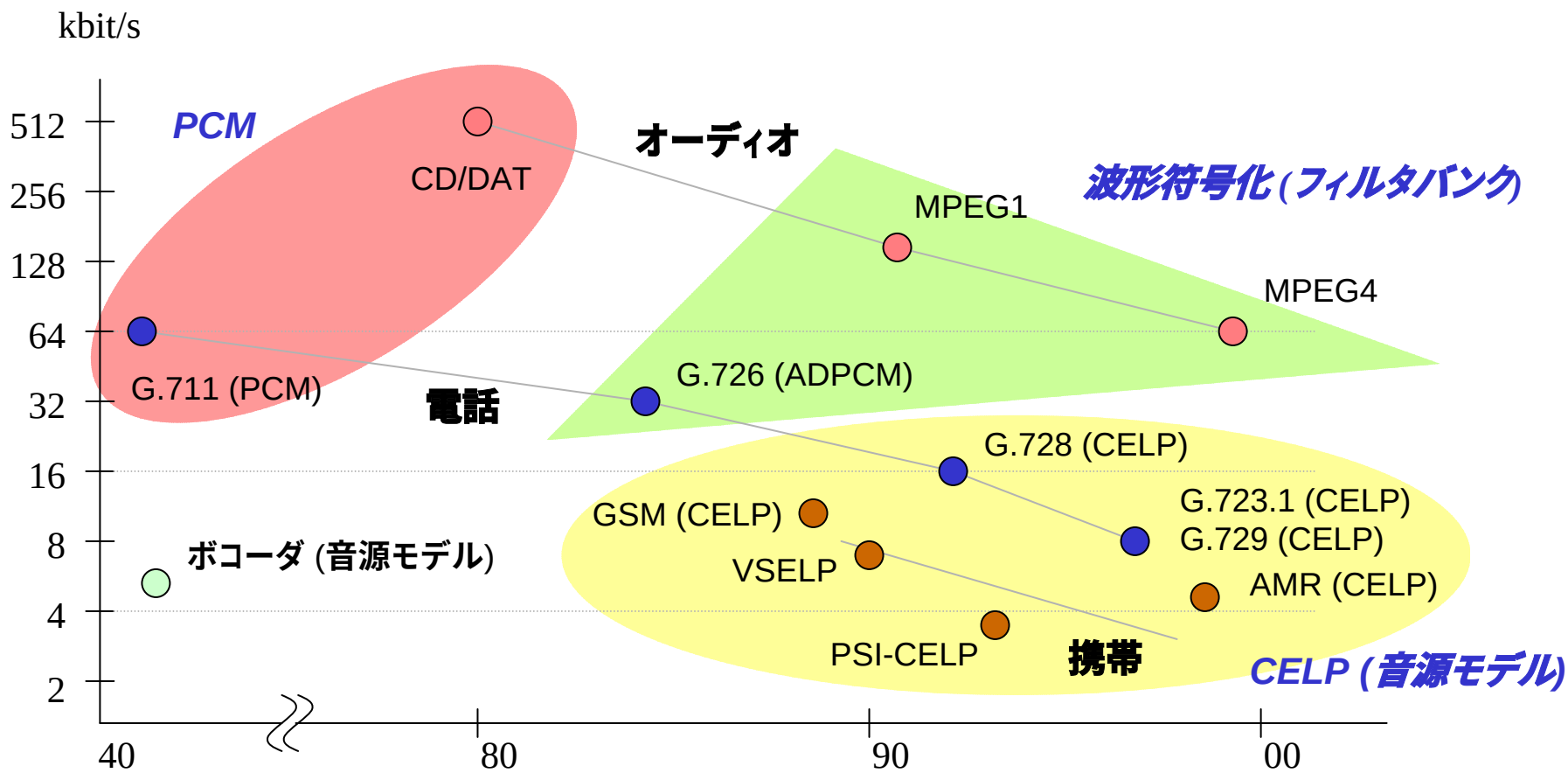
典型的なサンプリングレート

音声:
8 kHz 、8 ビット

オーディオ:
22.5, 44.1, 48 kHz、16 ビット



音声・オーディオ符号化の歴史



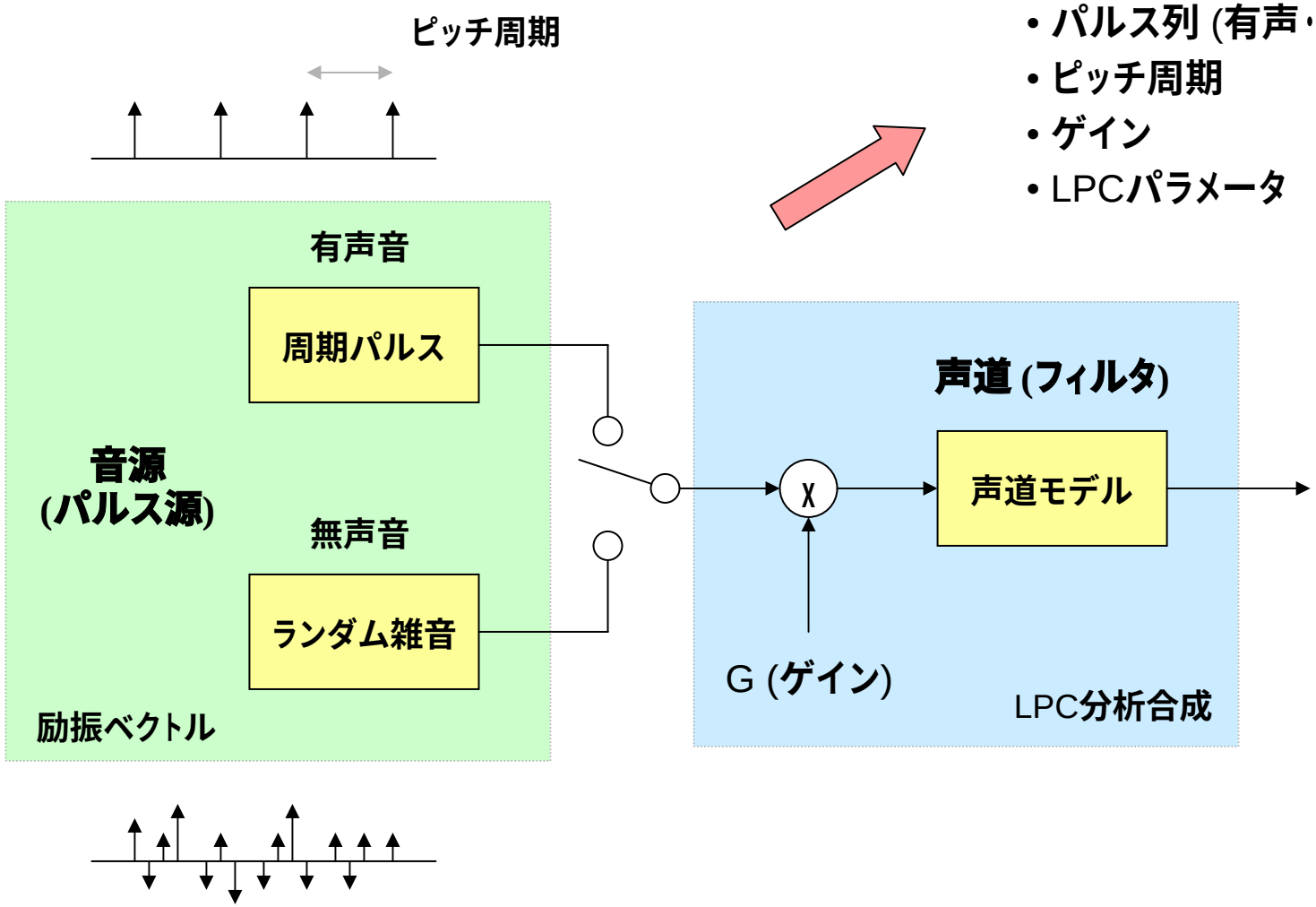
守谷：“音声符号化”

音声符号化 (1)

- 音声合成モデル

以下のパラメータを推定 (予測) して送信する

- パルス列 (有声・無声)
- ピッチ周期
- ゲイン
- LPCパラメータ

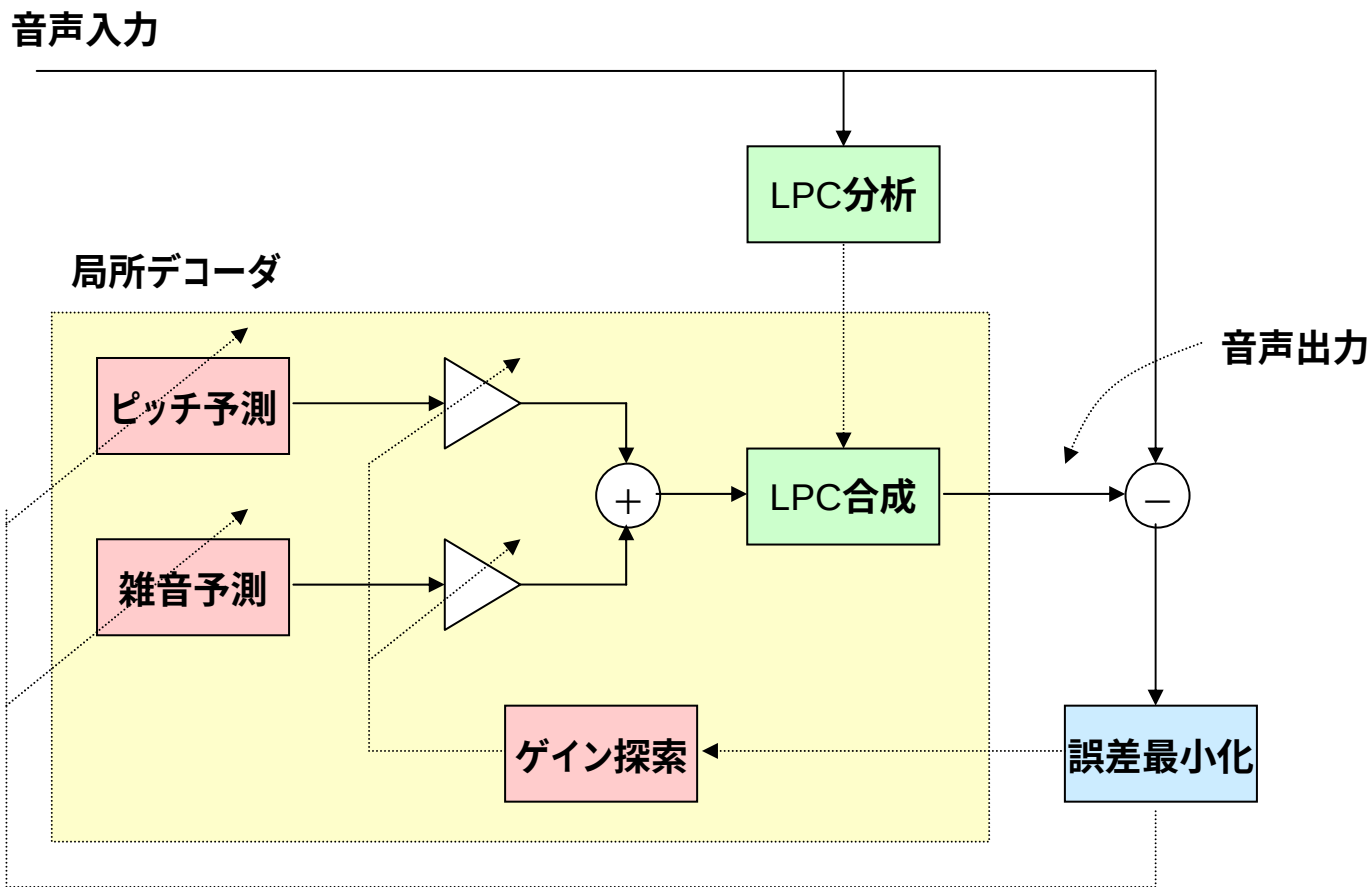


L.Rabiner et al: “Fundamentals of Speech Recognition”

音声符号化 (2)

CELP: Code Excitation Linear Prediction

- **CELP**



守谷：“音声符号化”

音声符号化 (3)

LPC: Linear Prediction Coding

- **LPC 分析 (線形予測分析)：声道モデル**

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + G \cdot u(n)$$

過去の k 個のサンプル値から線形予測

(注) 通常、画像のモデルでは雑音と扱う

$s(n)$: **音声サンプル**
a_k: **LPC係数**
p: **LPC分析次数**
G: **励振ゲイン**
$u(n)$: **正規化励振項**

予測誤差二乗平均の最小化

$$\frac{\partial e(n)}{\partial a_k} = 0$$

$$\sum_{k=1}^{p} r_n(i-k)\,\hat{a}_k = r_n(i)$$

自己相関法 (Durbin**のアルゴリズム**)

$r(k)$: **自己相関係数**
$\hat{a}_k$: **推定LPC係数**

音声符号化 (4)

- **ベクトル量子化：音源パルス列**

励振ベクトルとゲインの探索：

$$d = \|x - gAc\| \rightarrow \min$$

となる励振ベクトルとゲインを探索

さまざまな探索手法 ...

d: **ひずみ**

x: **目標ベクトル (入力音声)**

A: **LPC係数行列**

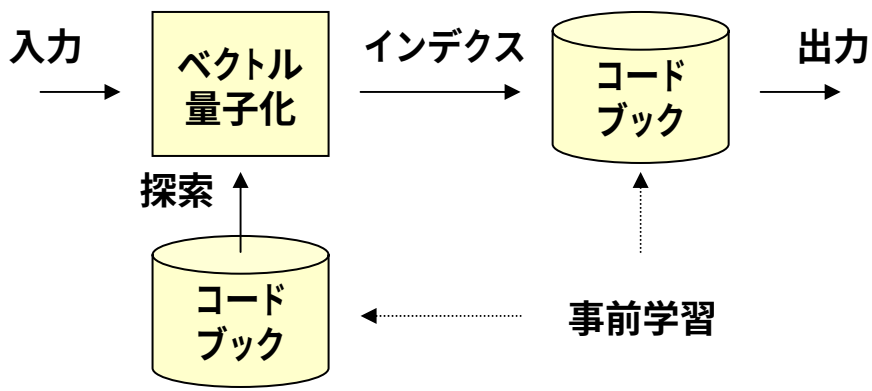
g: **ゲイン**

c: **励振ベクトル (パルス列)**

励振ベクトル → ベクトル量子化

ゲイン → スカラー量子化

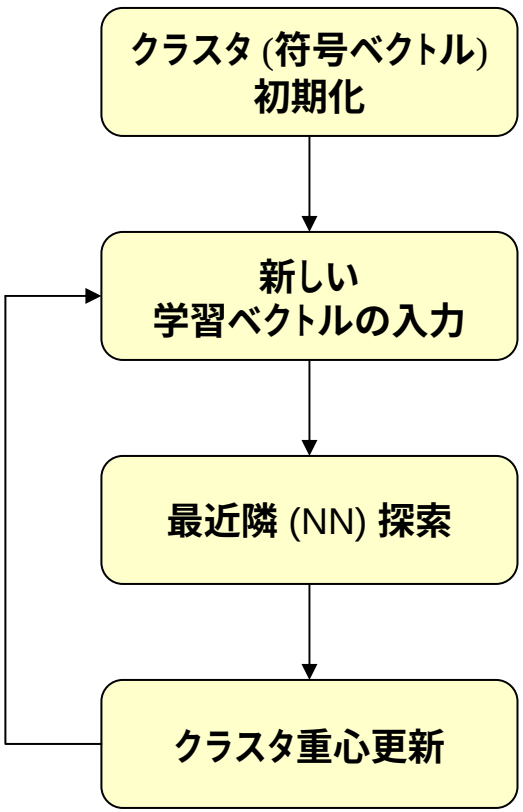
(声道パラメータ → ベクトル量子化)



音声符号化 (5)

- **ベクトル量子化：コードブックの学習 (1)**

K-**平均アルゴリズム** (**一般化** Lloyd **アルゴリズム**)



$\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_M$ (M: **符号ベクトル数**)

$\boldsymbol{r}_k \ (k = 1, 2, \ldots, L)$ (L: **学習ベクトル数**)

$\left\| \boldsymbol{c}_i - \boldsymbol{r}_k \right\|^2 \rightarrow \min$ (**ユークリッド距離の場合**)

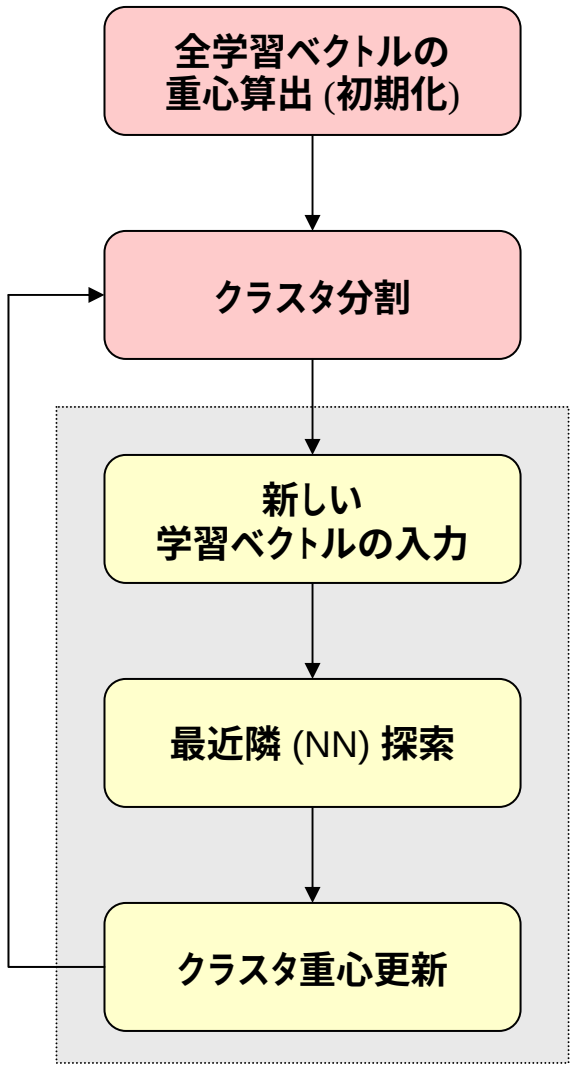
$$\boldsymbol{c}_i = \sum_{\boldsymbol{r}_k \in C_i} \boldsymbol{r}_k$$

欠点：**最終結果が初期ベクトルに依存**

音声符号化 (6)

- **ベクトル量子化：コードブックの学習 (2)**

LBG アルゴリズム

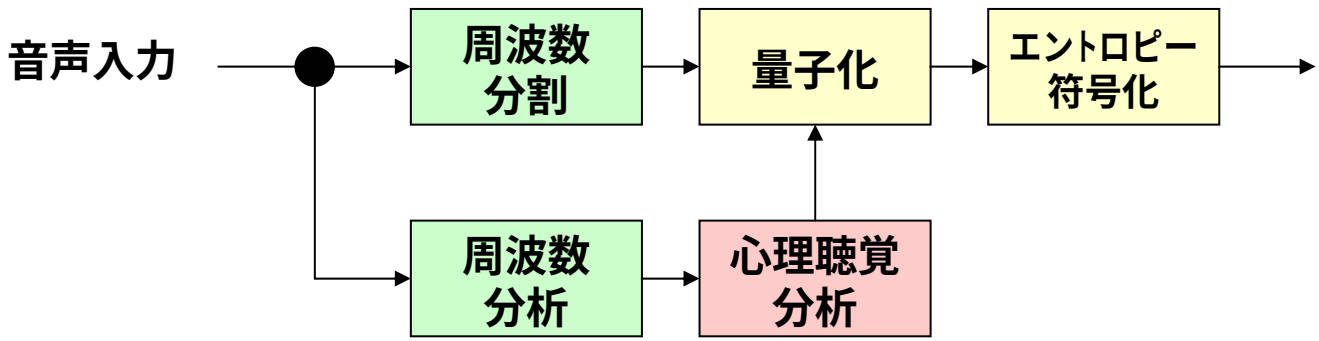


$$\begin{cases} \boldsymbol{c}_n^+ = \boldsymbol{c}_n(1+\varepsilon) \\ \boldsymbol{c}_n^- = \boldsymbol{c}_n(1-\varepsilon) \end{cases}$$

k-平均アルゴリズム

オーディオ符号化 (1)

- **オーディオ符号化の基本**

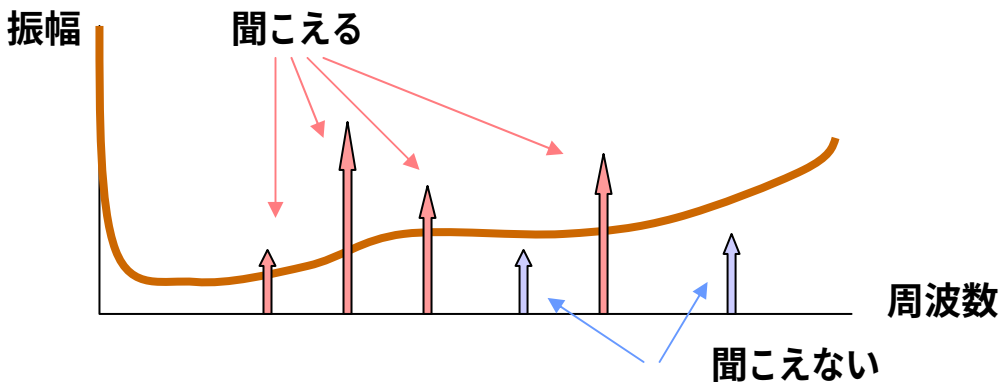


- **周波数分割、周波数分析**：FFT、**サブバンド分割** (QMF)、MDCT
- **心理聴覚分析**：**絶対閾値とマスキング**
- **量子化、エントロピー符号化**：**スカラー量子化とハフマン符号**

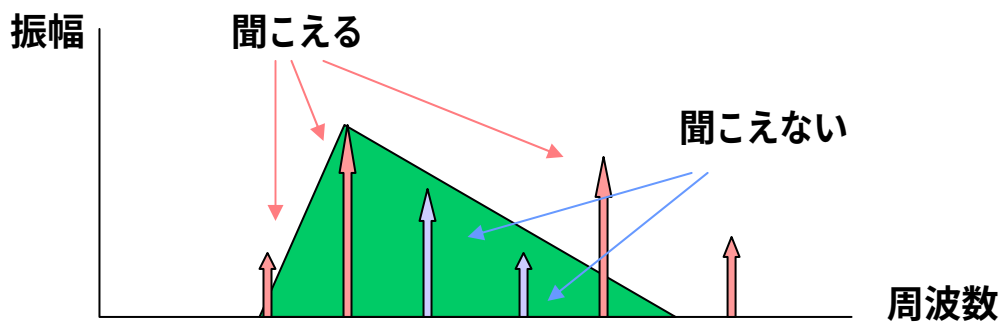
オーディオ符号化 (2)

- **心理聴覚分析**

絶対閾値：人間は絶対可聴閾値よりも大きな音しか知覚できない

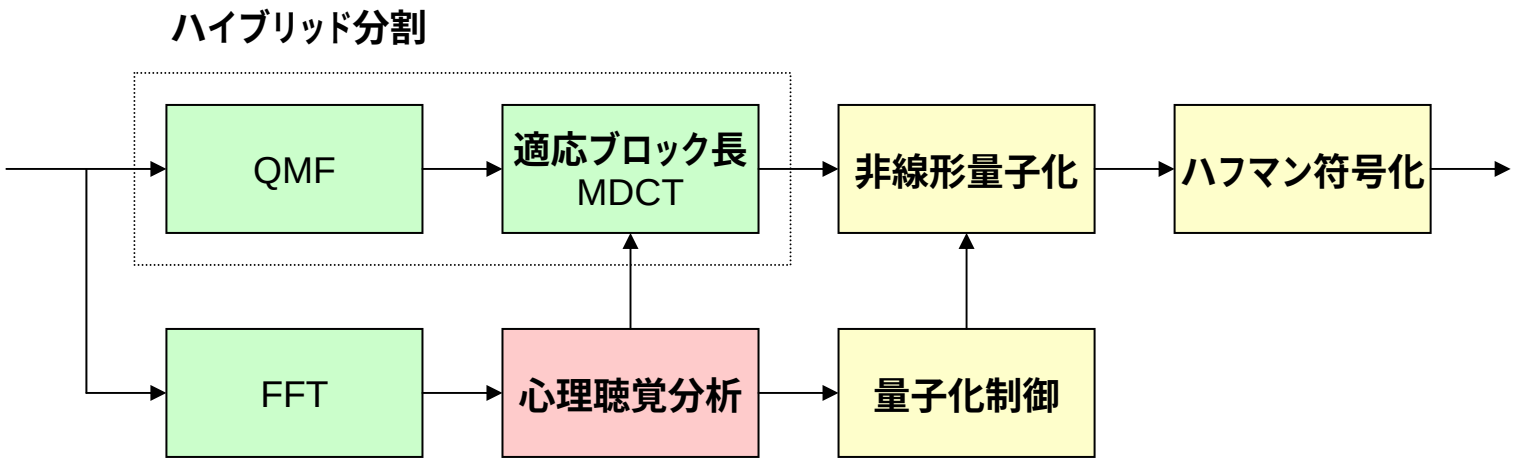


マスキング (相対閾値)：大きな音の周波数の近傍の小さな音の周波数は知覚できない

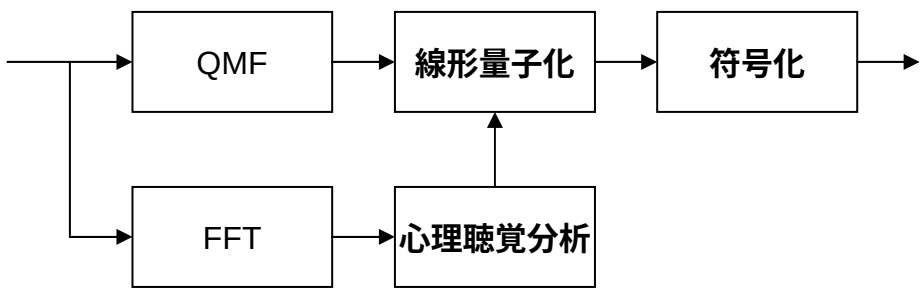


オーディオ符号化 (3)

- **MP3 (MPEG-1 Layer III)**

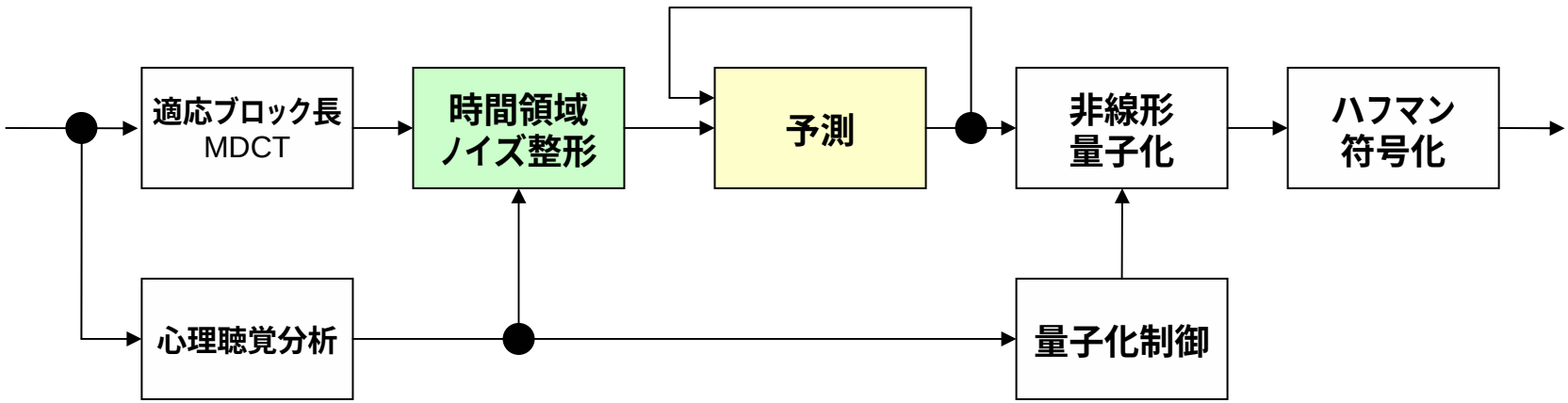


cf. Layer I, II



オーディオ符号化 (4)

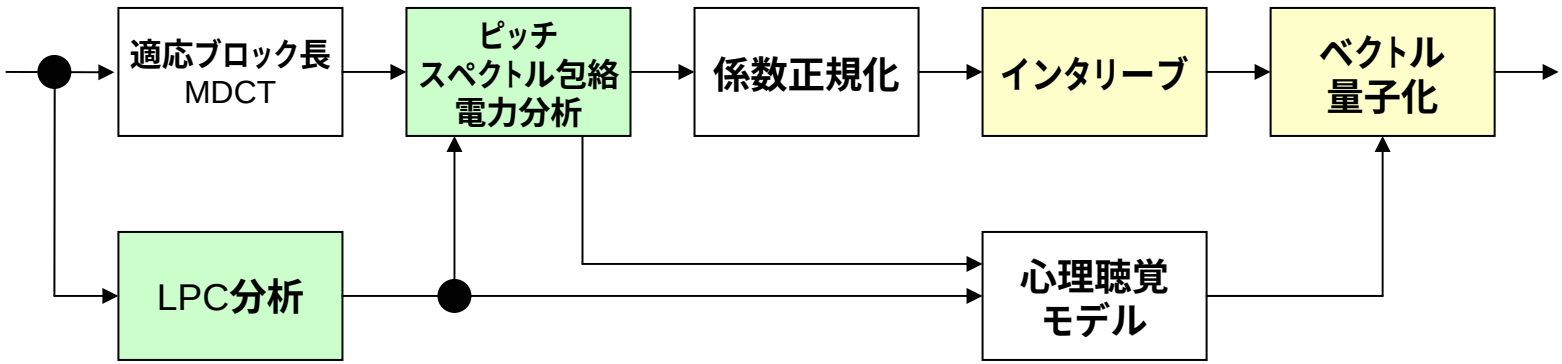
- **MPEG-2 AAC**



- **時間領域ノイズ整形** (for transient signals)：**一部の**MDCT**係数を時系列とみなして線形予測** (LPC) **分析。振幅の大きい部分に量子化雑音が集中する** (**ノイズ整形**)。
- **予測** (for stationary signals)：MDCT**係数毎に、過去**2**フレームの**MDCT**係数から予測。入力が定常的な場合に有効。**

オーディオ符号化 (5)

- **Twin VQ**



- LPC分析、ピッチ・スペクトル包絡・電力分析：MDCT係数の平坦化。ベクトル量子化のコードブック削減。
- インターリーブベクトル量子化：適応量子化に替わるひずみの最小化手法。傾向の似た変換係数のグルーピング。

音声とオーディオ、ビデオの対比

- **音声符号化**

  PCM → **波形符号化** → **分析合成符号化 (音声合成モデル)**

- **オーディオ符号化、ビデオ符号化**

  PCM → **波形符号化**

  オーディオ合成モデル：楽器 (+ ボーカル)

  ビデオ合成モデル：コンピュータグラフィックス？

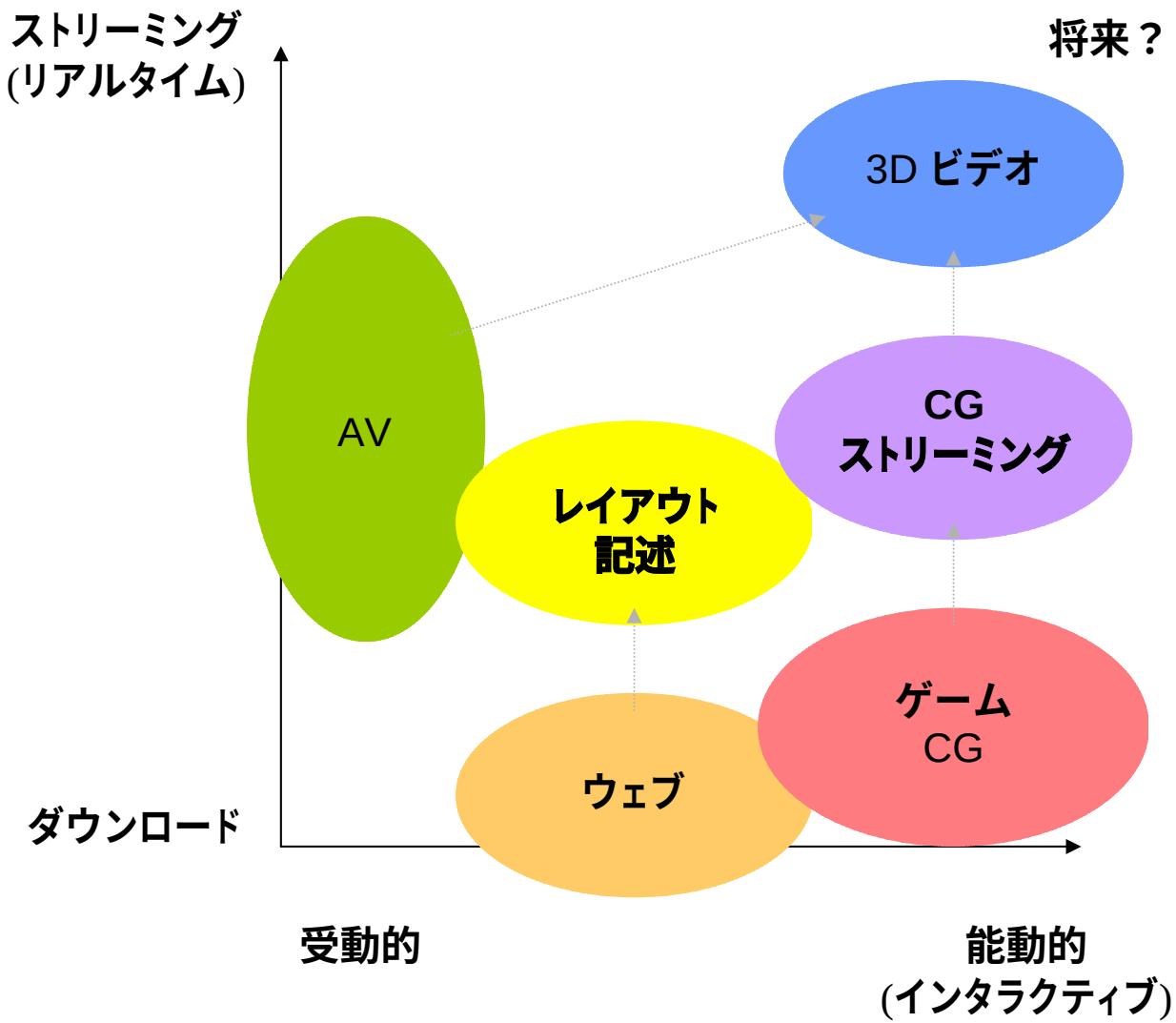
分析合成手法の試み (ブレークスルーにはなっていない)：

オーディオ符号化：音源分離

ビデオ符号化：知的符号化 (顔画像アニメーション)

SMIL

コンテンツの進化


ストリーミング
(リアルタイム)
将来？
3D ビデオ
AV
CG
ストリーミング
レイアウト
記述
ゲーム
CG
ウェブ
ダウンロード
受動的
能動的
(インタラクティブ)

SMIL

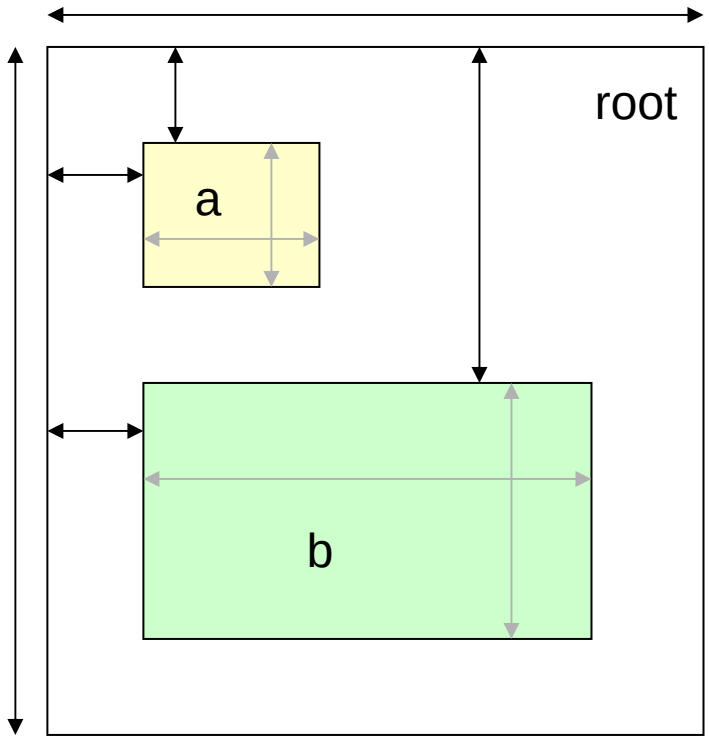
＊ Synchronized Multimedia Integration Language

・ストリーミングのためのレイアウト記述言語

```
<smil>
<head>
    <layout>
        レイアウト記述
    </layout>
</head>
<body>
    <par>
        メディア記述
    </par>
</body>
</smil>
```

** XML ベース ... HTML に慣れていれば習得は簡単*

レイアウト記述




```
<root-layout width="500" height="400"/>
<region id="a" top="50" left="50"
                    width="100" height="80" />
<region id="b" top="200" left="50"
                    width="400" height="200" />
```

表示画面

レイアウト記述

メディア記述

ストリーミング

```
<par>
   <video region="b" src="rtsp://www.foo.ac.jp/guide.sdp" />
   <seq>
      <img region="a" src="http://www.foo.ac.jp/point0.jpg" dur="10s" />
      <img region="a" src="http://www.foo.ac.jp/point1.jpg" dur="10s" />
      <img region="a" src="http://www.foo.ac.jp/point2.jpg" dur="10s" />
   </seq>
</par>
```

<par> メディア1, メディア2, … </par>　　複数メディアの「並列」再生

<seq> メディア1, メディア2, … </seq>　　複数メディアの「逐次」再生

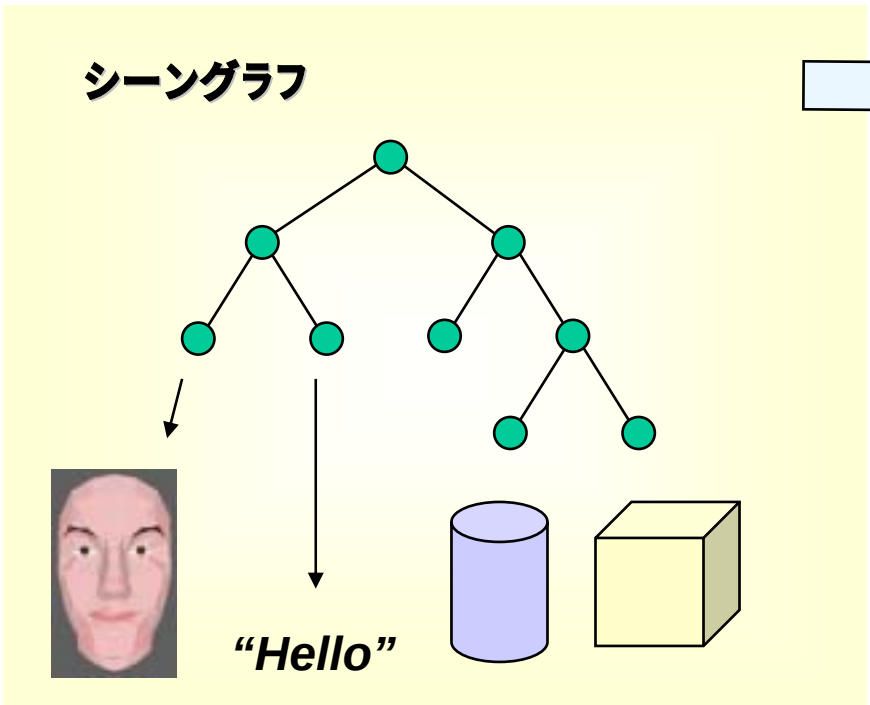
<video>, <audio>, <img>, ...　　各種メディアタグ

グラフィクス

VRML

＊ Virtual Reality Modeling Language

- 三次元CGの記述フォーマット



VRML記述

```
Transform {
    Transform {
        translation 15 10 0
        Shape {
            geometry Box 2 2 2
        }
    }
    Transform {
        translation 0 0 -1
        Shape {
            geometry Cylinder
        }
    }
}
...
```

シーン合成

VRML 2.0 のノード一覧

グループ:
Billboard
Group
Inline
LOD
Switch
Transform

形状:
Shape
Box
Cone
Cylinder
ElevationGrid
Extrusion
IndexedFaceSet
IndexedLineSet
PointSet
Sphere
Text

形状特性:
Coordinate
Color
Normal
TextureCoordinate

アピアランス:
Appearance
Material
ImageTexture
PixelTexture
MovieTexture
TextureTransform

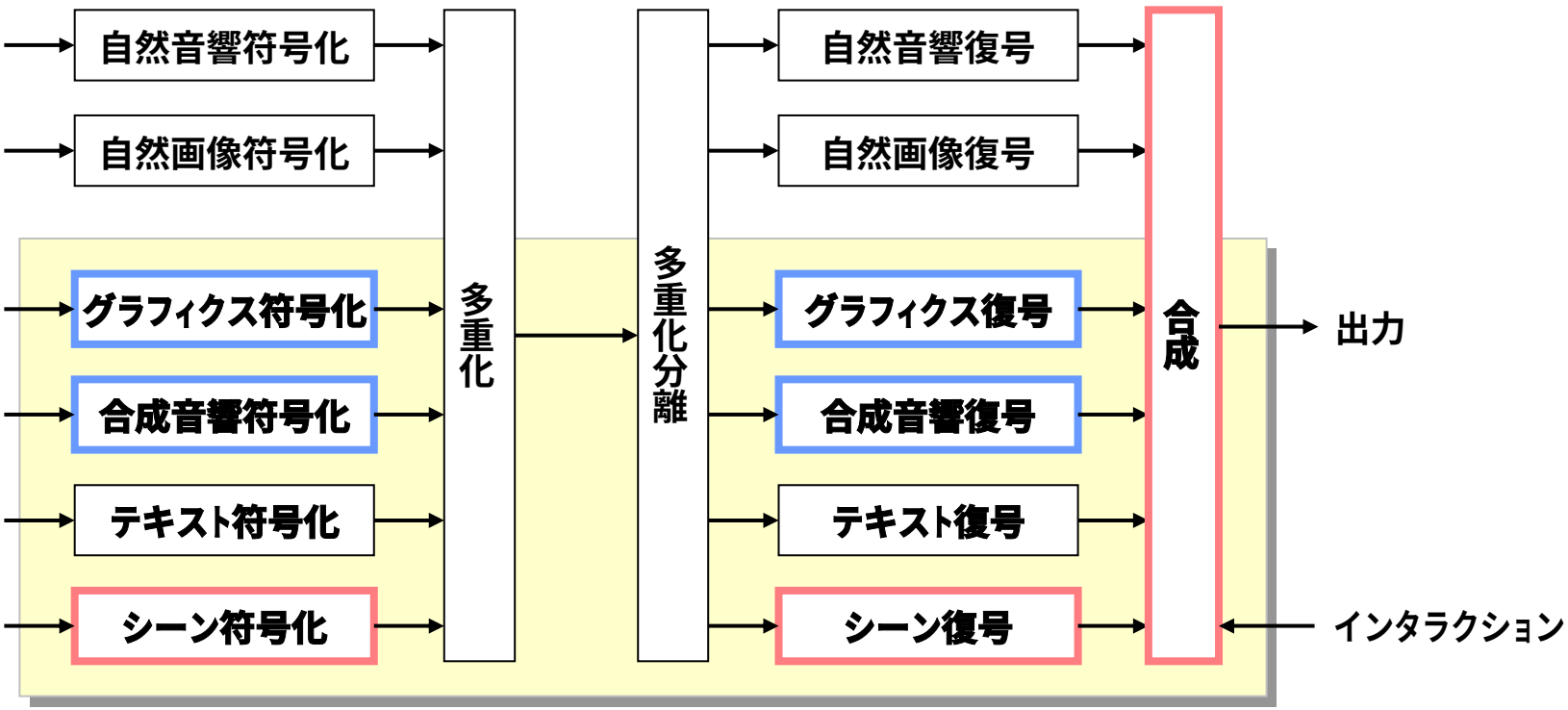
光源、視点:
DirectionalLight
PointLight
SpotLight
Viewpoint

センサ:
Anchor
Collision
CylinderSensor
PlaneSensor
ProximitySensor
SphereSensor
TimeSensor
TouchSensor
VisibilitySensor

インタポレーター:
ColorInterpolator
CoordinateInterpolator
NormalInterpolator
OrientationInterpolator
PositionInterpolator
ScalarInterpolator

その他:
AudioClip
Background
Fog
FontStyle
NavigationInfo
Script
Sound
WorldInfo

MPEG-4 Systems/SNHC

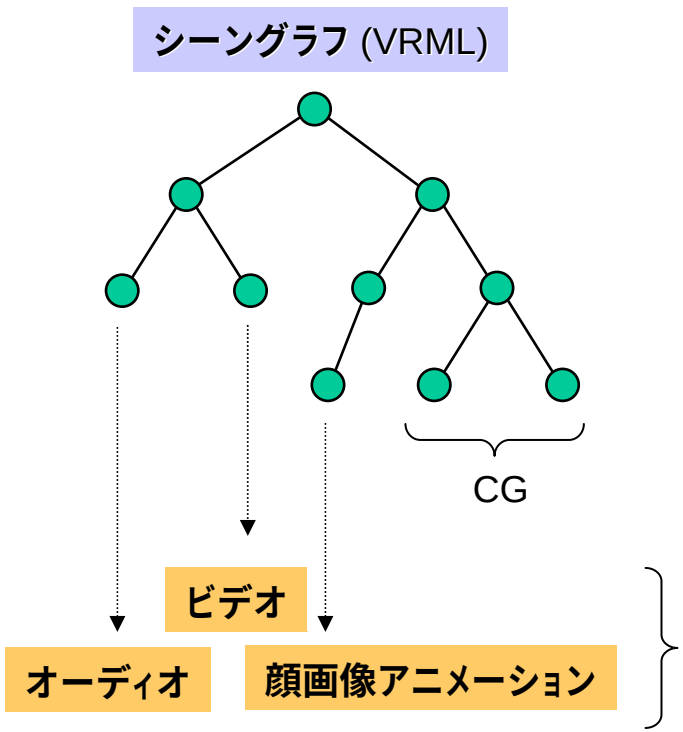


目的：従来の AV 系システムへの CG、コンピュータミュージックの取り込み

(1) シーン記述 (MPEG4 BIFS)

＊ Binary Format for Scene

VRMLのストリーミング拡張

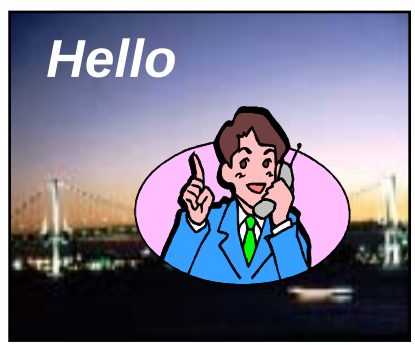


シーングラフのストリーミング

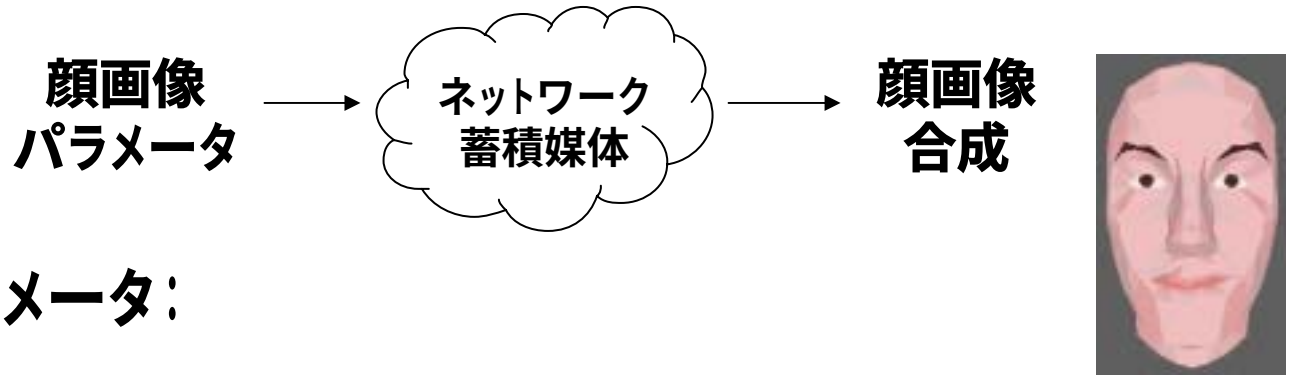
＋

従来のAVストリーミング

合成シーン



(2) 顔画像アニメーション



顔画像パラメータ:

FAP (Facial Animation Parameter)

顔の基本的な動きの表現。

FAP 初期値で基本的な顔を転送。以下は差分を転送 (ストリーミング)。

FAP を与えない場合には「ニュートラルフェイス」を使用。

FDP (Facial Definition Parameter)

FAP で与えられる一般的な顔画像のカスタマイズ。

セッション開始時に転送 (オプション)。

(3) 人体アニメーション

人体パラメータ:

BAP (Body Animation Parameter)

人体の基本的な動きの表現。

BAP 初期値で基本的な人体を転送、以下は差分を転送 (ストリーミング)。

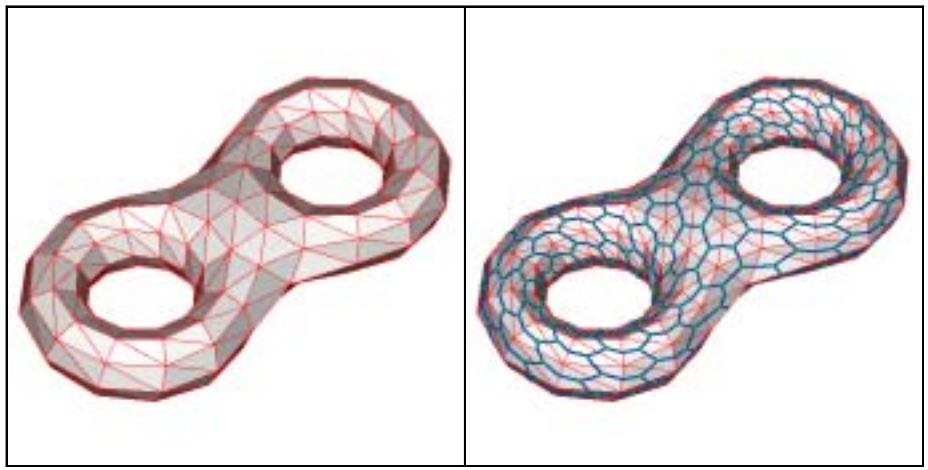
BAP を与えない場合には「デフォルト人体」を使用。

BDP (Body Definition Parameter)

BAP で与えられる一般的な人体のカスタマイズ。

セッション開始時に転送 (オプション)。

(4) 三次元メッシュ符号化



三次元メッシュ：

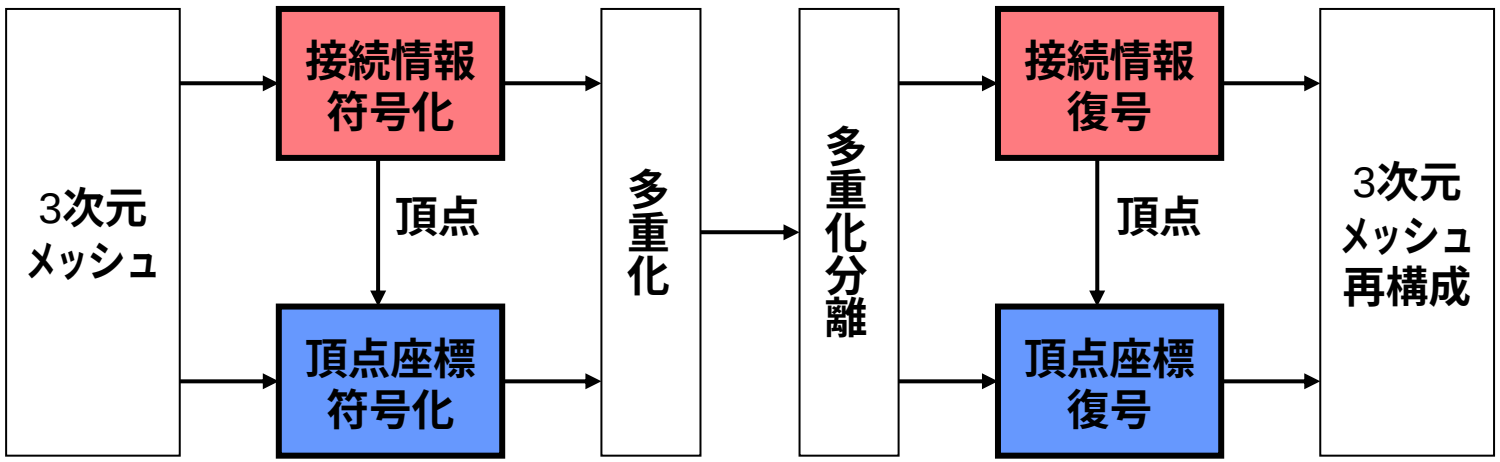
ポリゴンの頂点座標 + 頂点間の接続情報 + 各種特性情報

三次元メッシュ符号化：

上記の三次元メッシュ記述の圧縮

G.Taubin: “Geometric Compression Through Topological Surgery,” ACM Trans on Graphics.

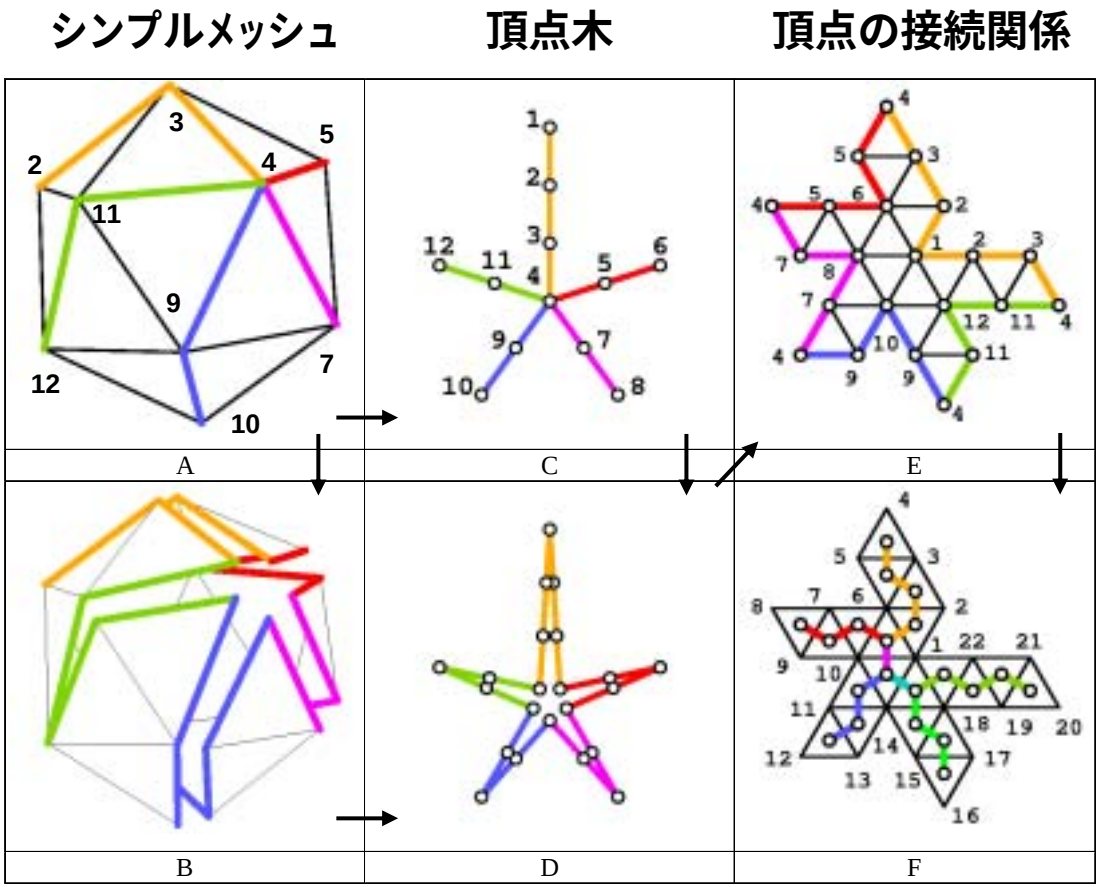
メッシュ符号化のブロック図



三段階の符号化:

1. ポリゴン頂点の接続情報 (connectivity) の符号化
2. ポリゴン頂点の三次元座標 (geometry) の符号化
3. 色、法線、テクスチャ座標などの特性 (property) の符号化

接続情報の符号化 [1]



三次元メッシュ (A)

↓

一頂点の選択と
頂点木の作成 (C)

↓

二次元平面に展開 (E)
(一番外側が選択頂点)

↓

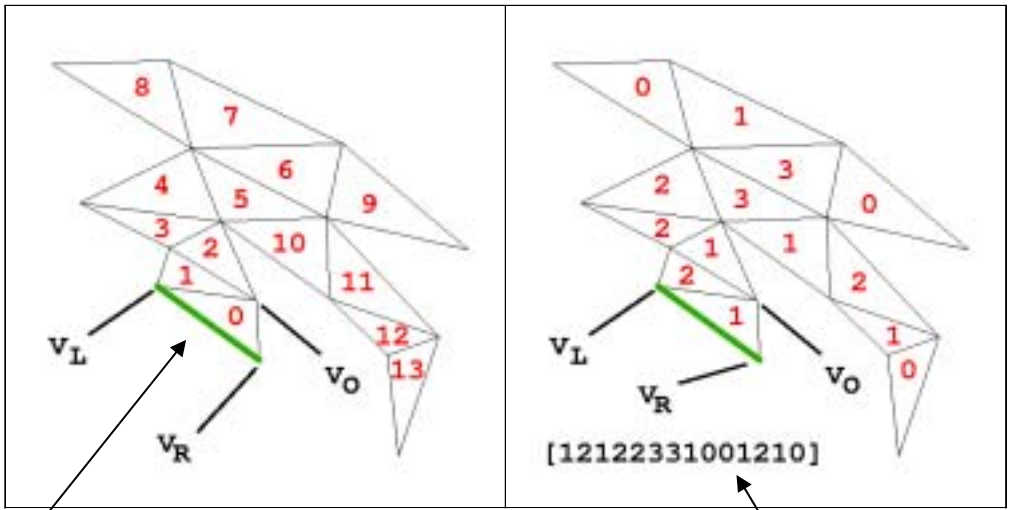
デュアルグラフ (双対木)
の作成 (F)

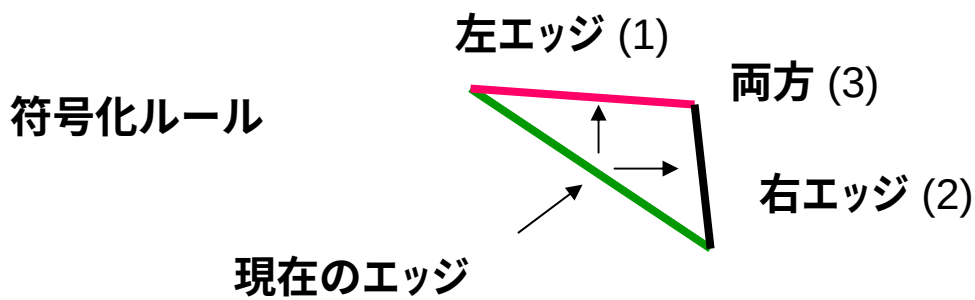
↓

双対木の符号化
(次ページ)

G.Taubin: “Geometric Compression Through Topological Surgery,” ACM Trans on Graphics.

接続情報の符号化 [2]



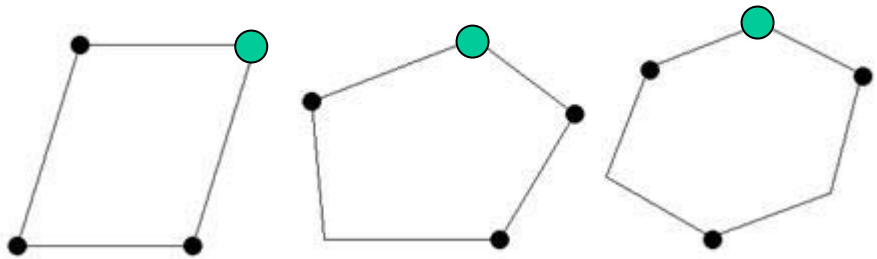


G.Taubin: “Geometric Compression Through Topological Surgery,” ACM Trans on Graphics.

頂点座標の符号化

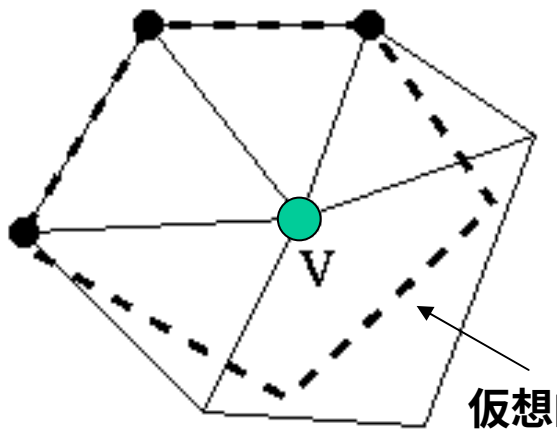
(1) ポリゴンによる予測

符号化対象の頂点を、ポリゴンを構成する頂点の一つと仮定して、座標を外挿予測。
予測誤差を符号化。



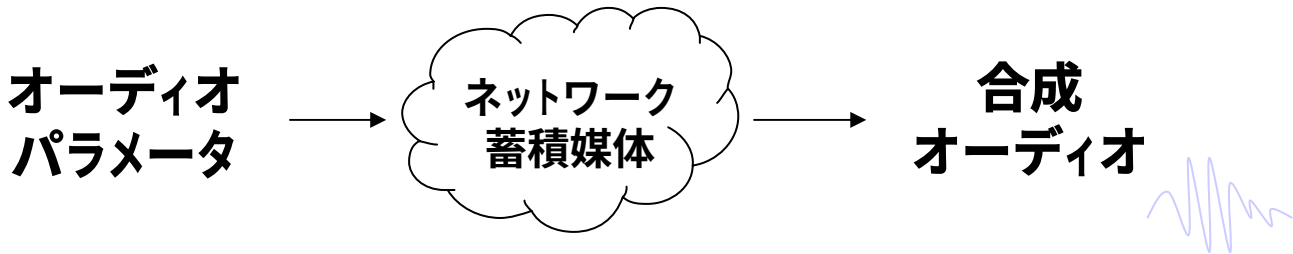
(2) 平均による予測

符号化対象の頂点を、それを囲むポリゴンの重心と仮定して、座標を内挿予測。
予測誤差を符号化。



ビデオ符号化に類似

(5) 合成オーディオ



オーディオ合成パラメータ：

SAOL (Structured Audio Orchestra Language):

楽器の特徴、信号処理方法を記述する言語 ... 音源物理モデルに相当。

SASL (Structured Audio Score Language):

楽譜情報を記述するフォーマット ... MIDI **に相当。**

SABSF (SA Bank Sample Format):

音源波形をそのまま使うフォーマット ... PCM **音源に相当。**

関連情報