



音声音響処理特論



- NTTと自己紹介とデジタル標準化
- 基本技術
- 音声符号化の標準
- 音楽符号化の標準
- ロスレス符号化

NTTコミュニケーション科学基礎研究所

守谷 健弘

自己紹介



- 1978年 計数工学科
卒論：磁束量子を用いた論理回路のアナログシミュレーション
- 1980年 計数工学科
修論：音楽を素材とした人間の情報処理機構の研究
- 1980年 基礎第四研究室
- 1988年 ヒューマンインタフェース研究所
- 1989年
学位：周波数領域のベクトル量子化を用いた中帯域音声符号化
- 1989年 AT&Tベル研究所
- 1990年 ヒューマンインタフェース研究所
- 1999年 サイバースペース研究所
- 2004年 コミュニケーション科学基礎研究所

NTT持株会社

NTT東日本

NTT西日本

NTTコミュニケーションズ

NTTデータ

NTTドコモ

- ・NTTビジネスアソシエ
- ・NTTエレクトロニクス
- ・NTTアドバンステクノロジー
- ・NTTロジスコ
- ・NTT都市開発
- ・NTTファシリティーズ
- ・NTTリース
- ・NTTアド
- ・NTTコムウェア
- ・NTTソフトウェア
- ・NTTアフティ
- ・日本カーソリューションズ
- ・NTTアイティ
- ・情報通信総合研究所
- ・NTT出版
- ・NTTラーニングシステムズ
- ・サイバー・ラボ
- ・NTT CAPITAL(U.K.) Ltd.
- ・NTTトラベルサービス
- ・クリニカルサポート
- ・NTTヒューマンソリューションズ
- ・日本情報通信
- ・NTTレゾナント
- ・NTTクラリティ

サイバーコミュニケーション総合研究所

- サイバーソリューション研究所
- サイバースペース研究所

情報流通基盤総合研究所

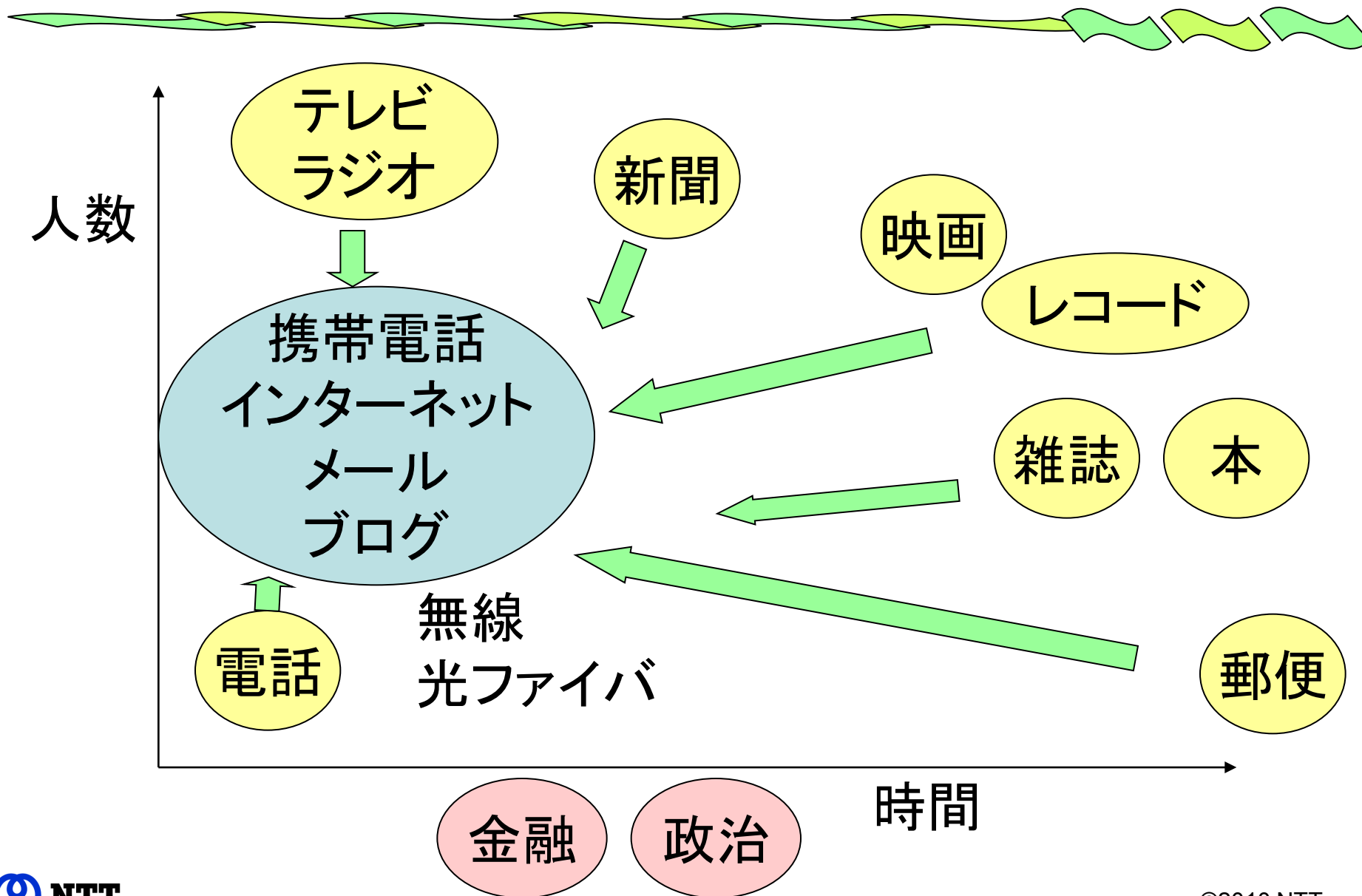
- サービスインテグレーション基盤研究所
- 情報流通プラットフォーム研究所
- ネットワークサービスシステム研究所
- アクセスサービスシステム研究所
- 環境エネルギー研究所

先端技術総合研究所

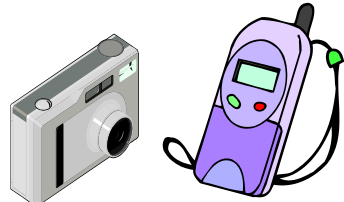
- 未来ねっと研究所
- マイクロシステムインテグレーション研究所
- フォトニクス研究所
- コミュニケーション科学基礎研究所**
- 物性科学基礎研究所

12研究所・6ロケーション・2500人

ICT (Information and Communication Technology) による変化



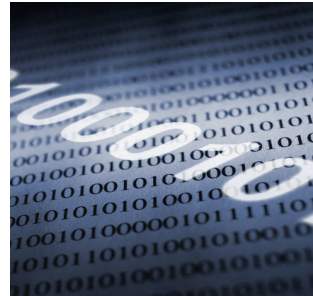
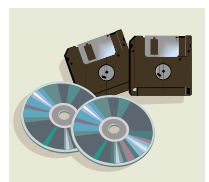
デジタル情報の落とし穴



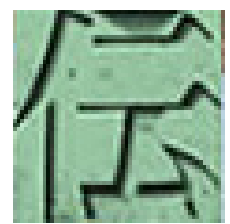
高性能・高品質

永久保存

もとはただの0と1

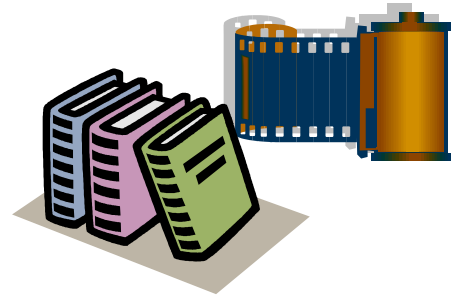


厳密な規格が必要



旧音響研究棟
(武蔵野)

石に刻むのが一番長持ち
1000年



紙やフィルムの
ほうがまだまし
100年

国際標準の例

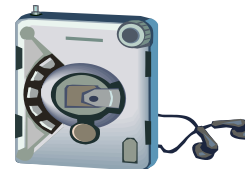
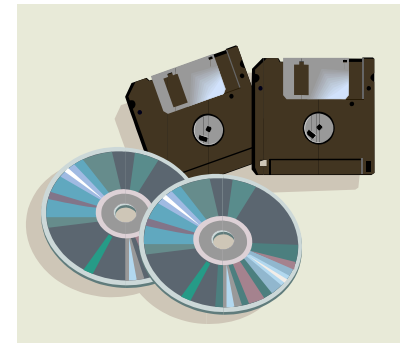
- ITU-T

- 携帯電話
- インターネット電話
- TV電話
- FAX



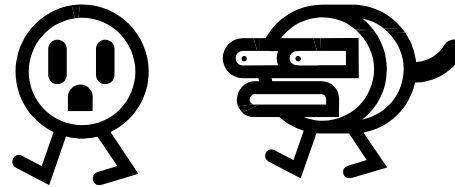
- ISO/IEC JPEG・MPEG

- デジタルカメラ
- デジタル放送・DVD
- 携帯音楽プレーヤー



国際標準の利点

- 相互接続性
- 仕様の公開
 - 長期保守性
 - 特許権者の確定
- 最高レベルの技術の結集
- 量産効果



市場拡張の循環を生む

音楽符号化技術の特許紛争例

MicrosoftとLucentのMP3特許侵害訴訟，地裁がMS不利の評決を覆す

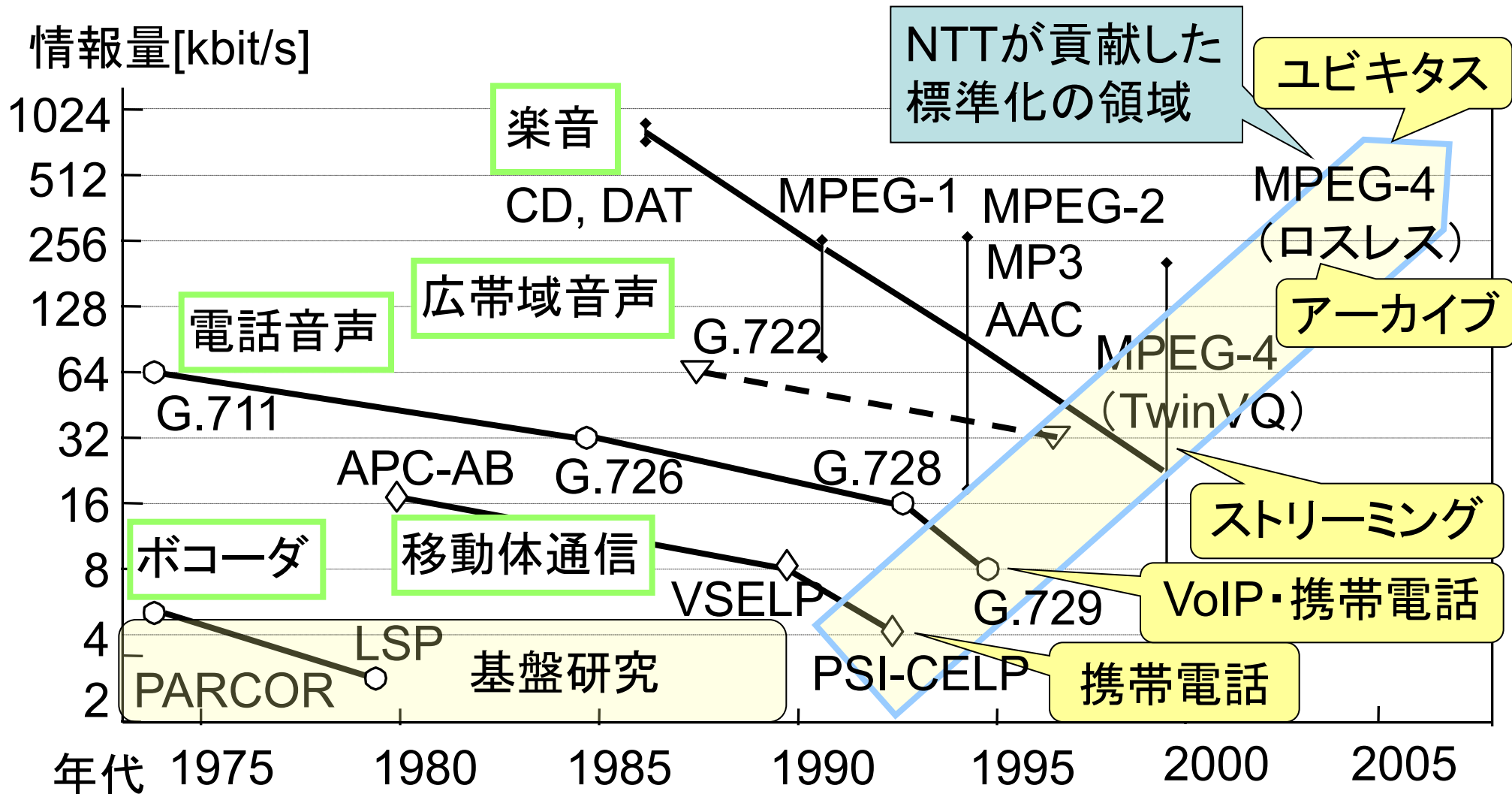
この訴訟は、2003年に米Lucent Technologies(現Alcatel-Lucent)が米Dellと米Gatewayを訴えたことに端を発する。Lucentは、2社が採用しているWindows OSに、同社の特許に触れる技術が使われていると主張した。Microsoftは顧客であるDellおよびGatewayを救援しようと、Lucentを相手取り提訴。これを受けLucentが反訴し、両社間の係争に発展した([米InfoWorldの報道](#))。

陪審団は2007年1月29日、米国特許番号5,341,457(457特許)と引用特許番号39,080(080特許)の特許2件について陪審裁判を開始した。2月22日にMicrosoftの特許侵害などを認め、約15億ドルの支払いを命じた(関連記事:[音声圧縮技術を巡る特許侵害訴訟、Microsoftに約15億ドルの支払い命令](#))。

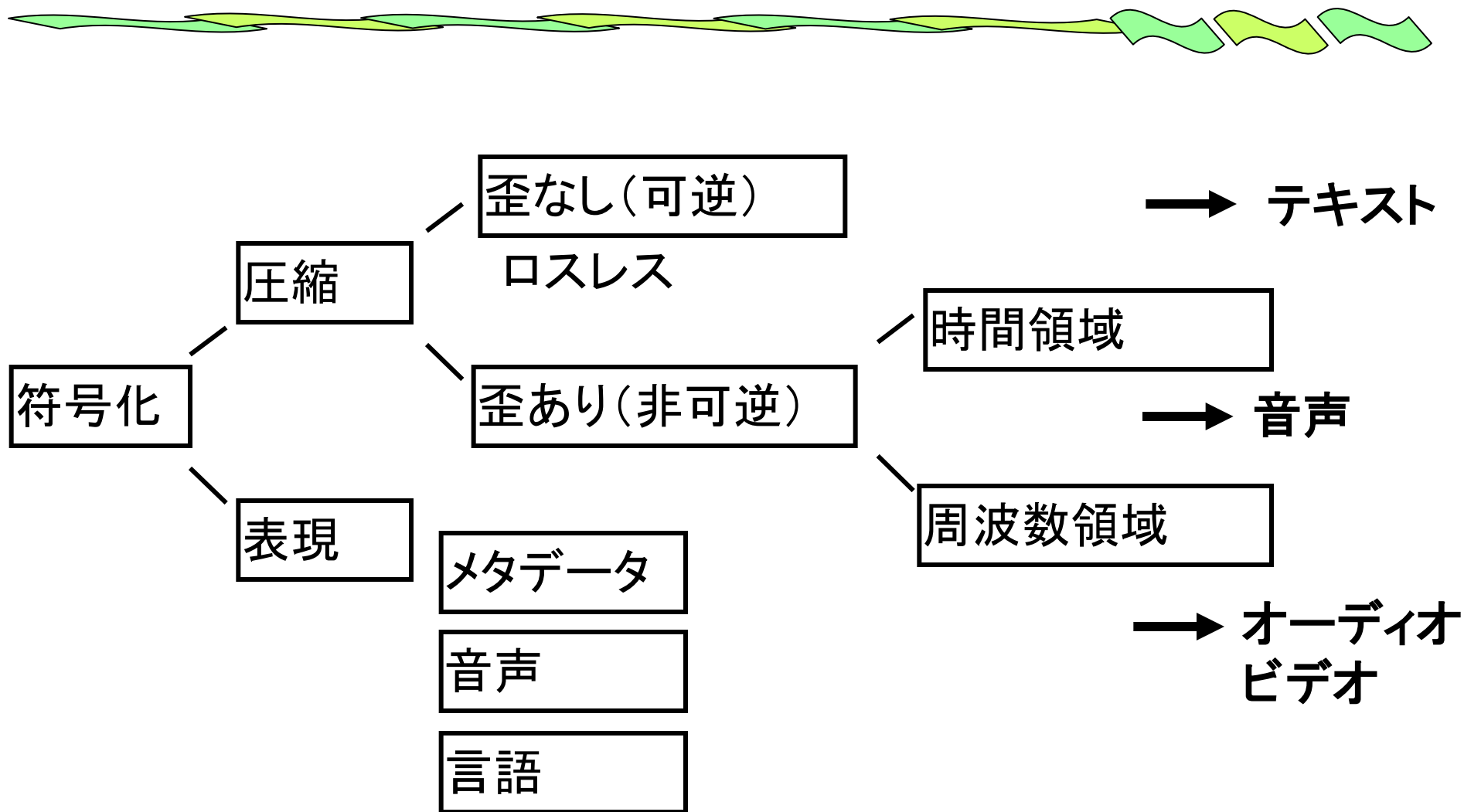
これに対しMicrosoftは、「当社はMP3技術のライセンス供与を行っているドイツのFraunhofer Institute for Integrated Circuitsから、適切にライセンスを取得したと確信している。当社がFraunhoferに支払った1600万ドルという金額を考えると、評決の賠償額はあまりに法外だ」とし、カリフォルニア州南地区連邦地裁に判断を仰いだ。

その結果、同地裁のRudi M. Brewster判事はMicrosoftが457特許を侵害していないことと、Fraunhoferが080特許の共同所有者であると認定し、評決と異なる判断をした。

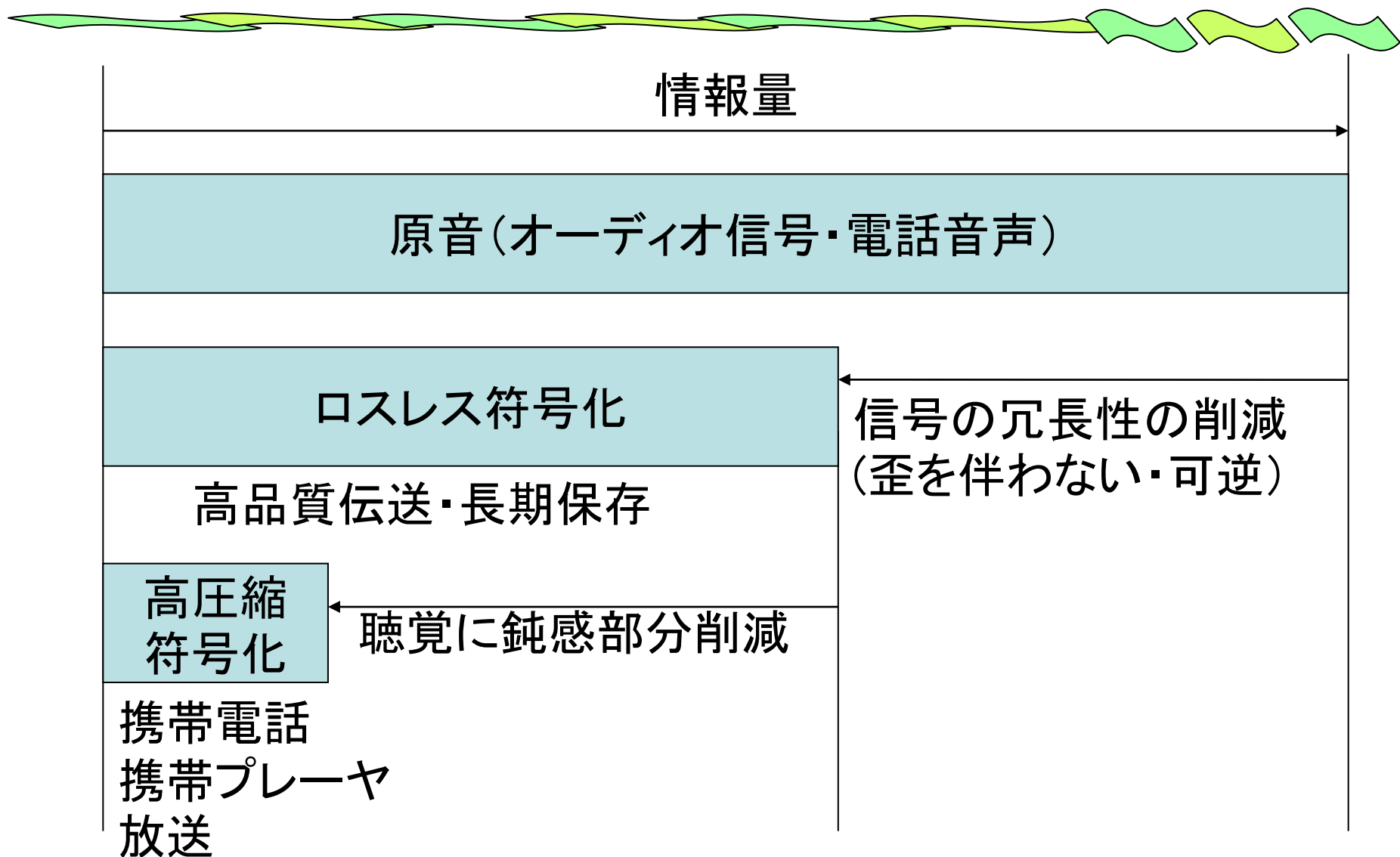
音声・音響符号化の標準化



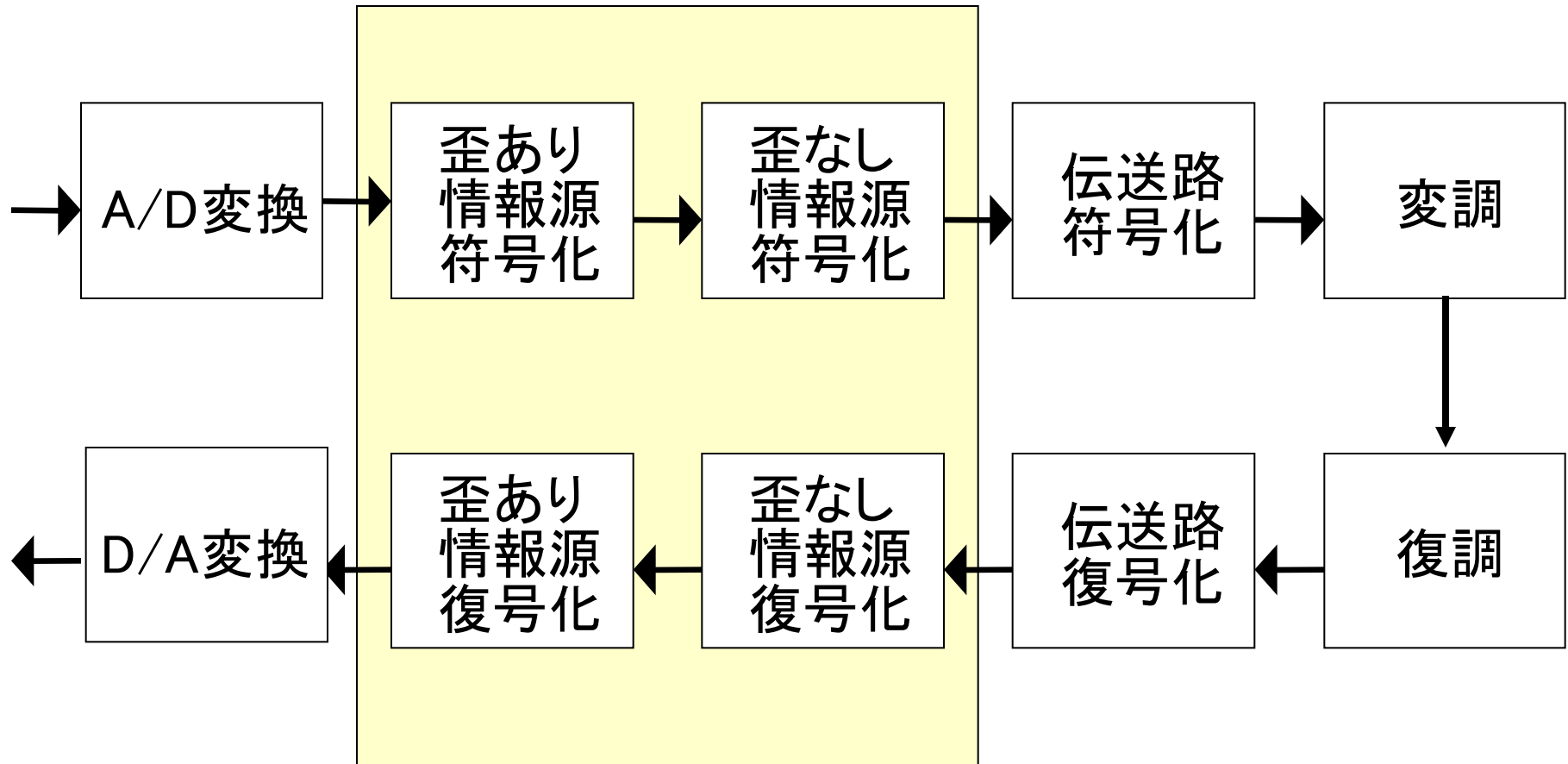
符号化の分類



符号化と情報量



情報圧縮の枠組み



歪削減の基本技術



- 予測符号化と変換符号化

- 予測パラメータ
- DCTとMDCT
- 適応ビット配分
- 変換利得と予測利得

- ベクトル量子化

- 共役ベクトル量子化
- 励振ベクトルとCELP

- 誤り制御との連携

- 適応化

<http://www.data-compression.com>

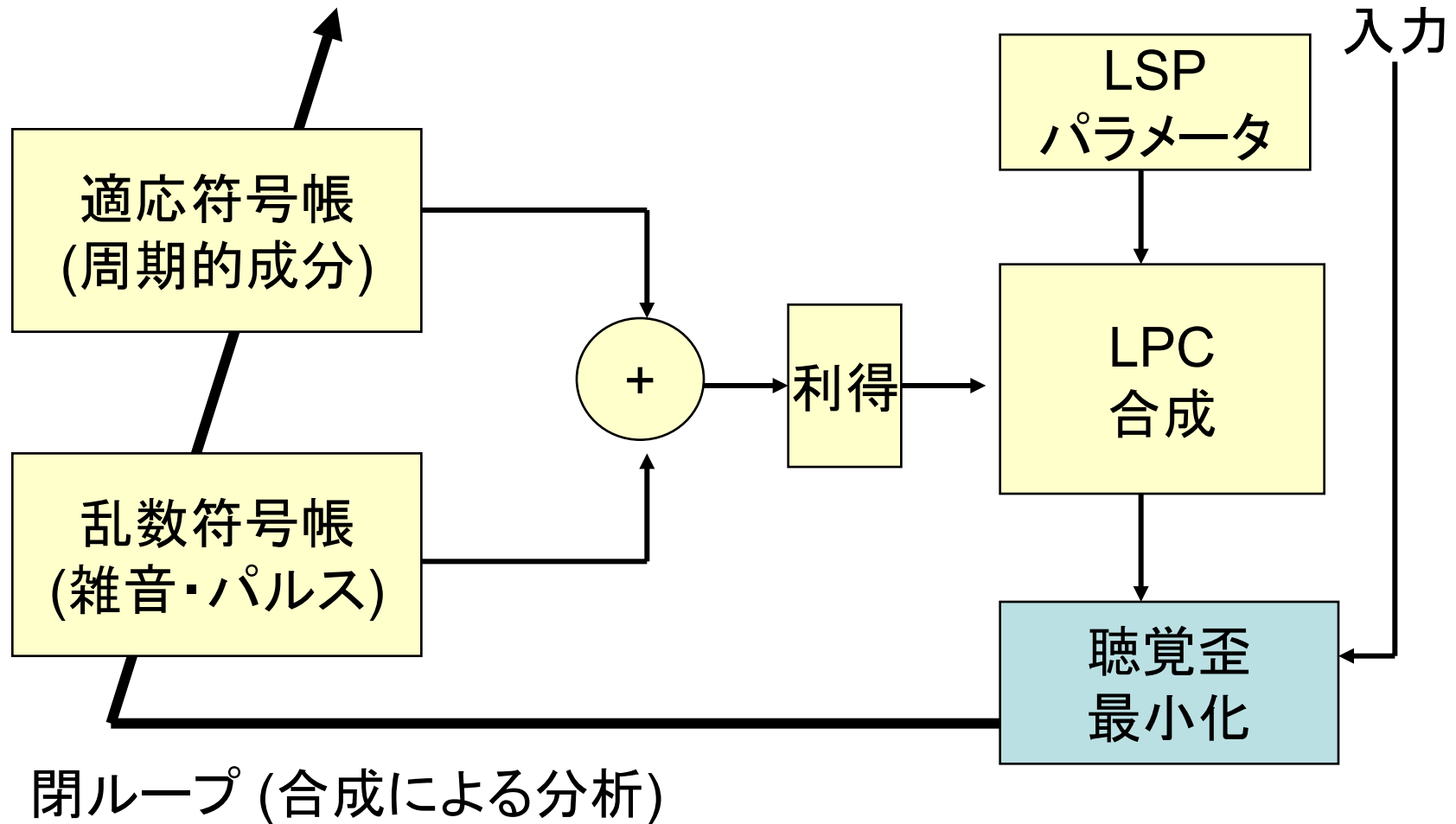
CELPの誕生



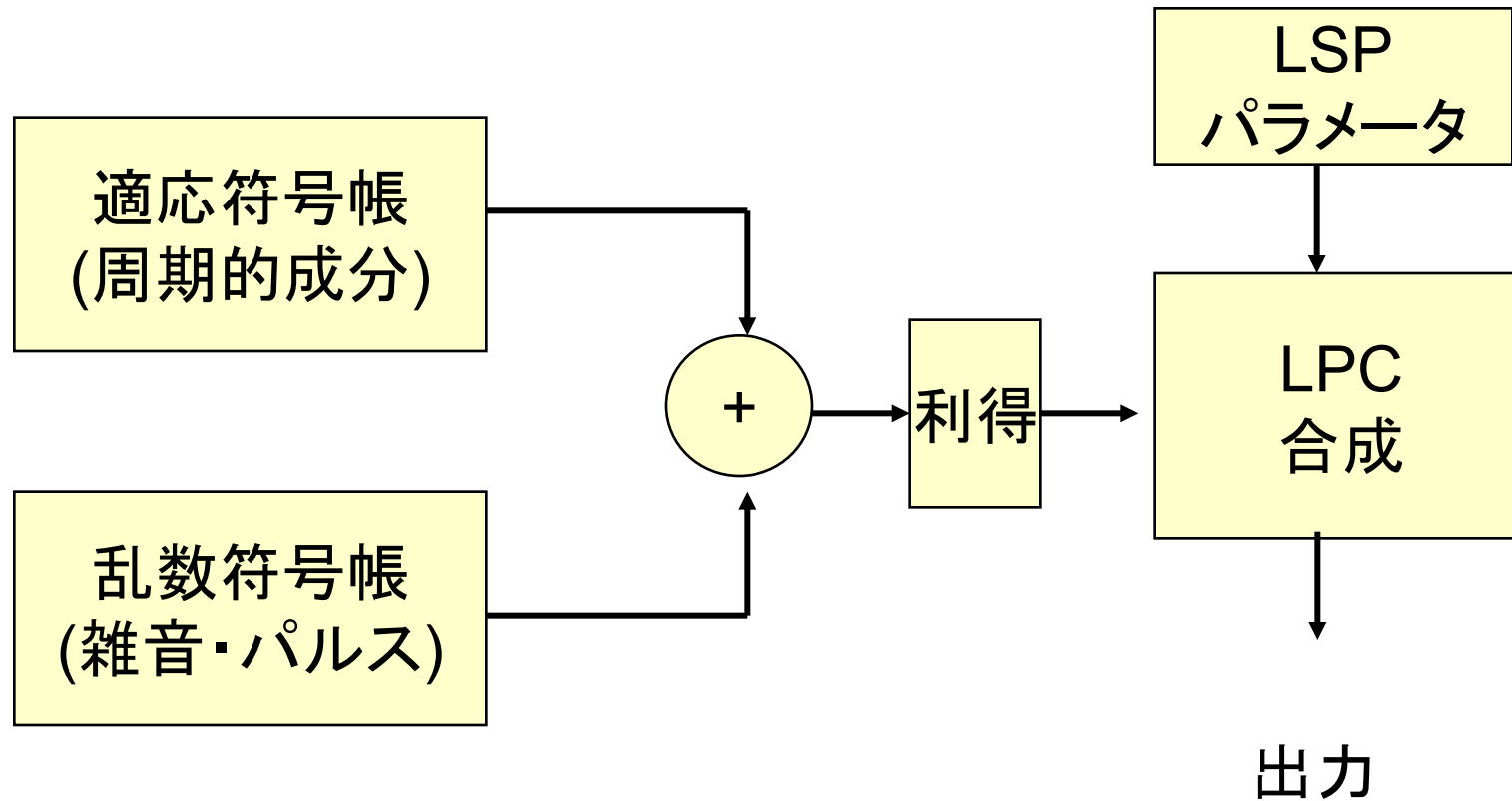
- Code Excited Linear Prediction
- ベクトル量子化と閉ループ探索 (AbS)
- 利点
 - 自然な合成音声
 - 低ビット化可能
- 問題点
 - 予測パラメータは未量子化
 - 膨大なメモリと演算量

AbS: Analysis by Synthesis

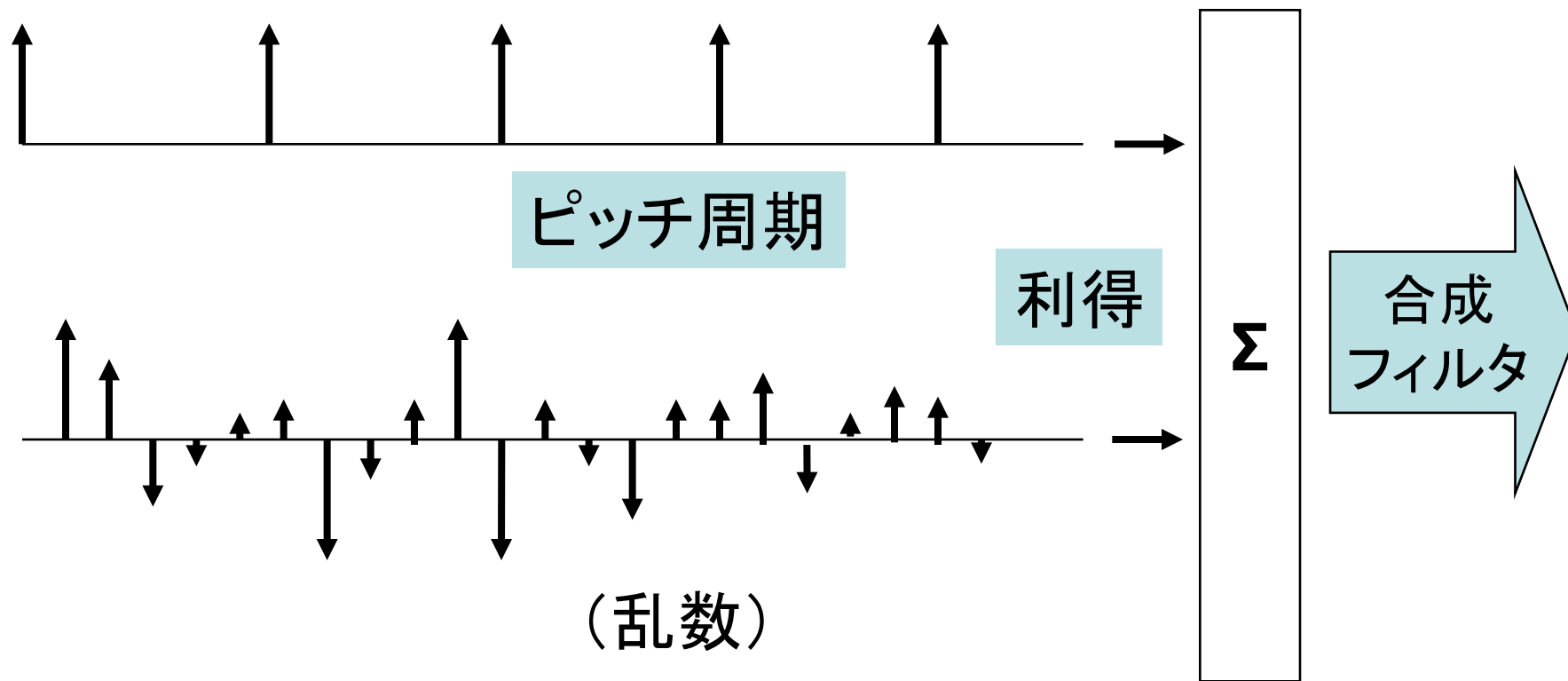
音声符号化 CELP



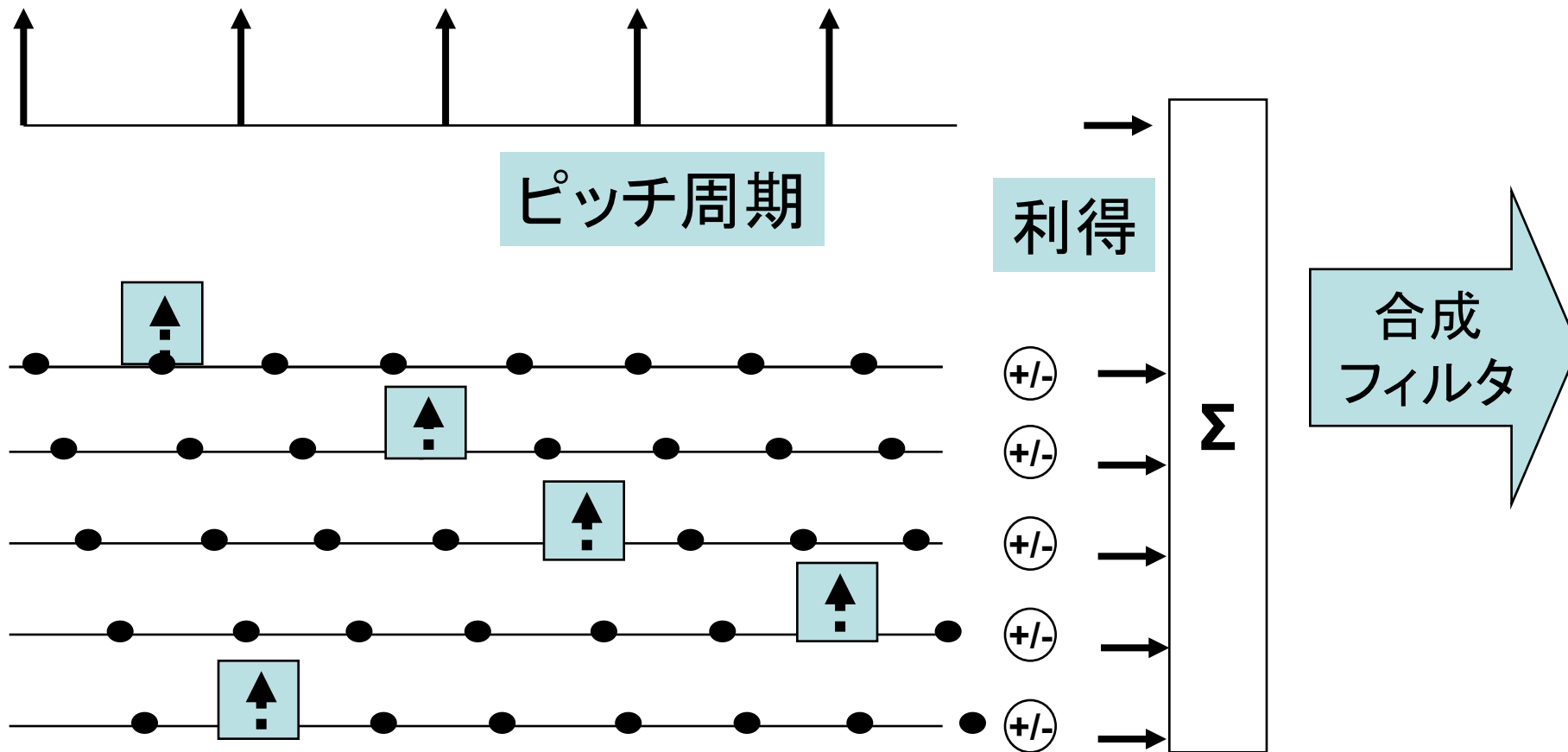
合成モデル



ボコーダの合成モデル



代数 (ACELP) の合成モデル

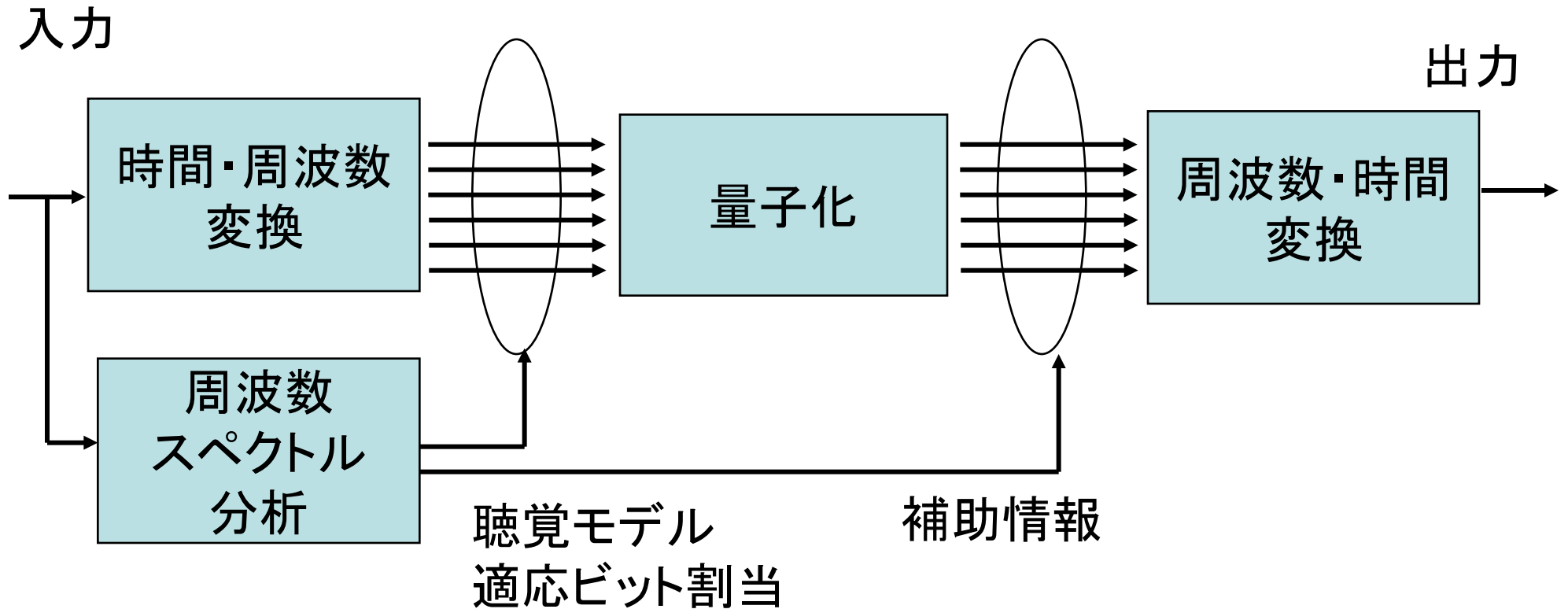


単位パルスの位置の選択

周波数領域の符号化

- 帯域分割符号化
- 直交変換符号化
- 時間分解能と周波数分解能
- 適応情報割り当て(重み付け)
- マスキング特性の利用
- QMF (Quadrature Mirror Filter)
- DFT・DCT・MDCT

変換符号化



適応ビット配分

- 量子化するサンプル数 N 、エネルギー f_i
- 1ビットの情報で量子化歪は $1/4$
- ビット数の総和一定 $m = \sum b_i$
- 量子化歪の総和 $d = \sum f_i 4^{-b_i}$
- 最適ビット配分 $b_i = m/N + (1/2)\log_2 f_i$
- 各サンプルの量子化歪が均一 \Leftrightarrow 歪最小

DCTとMDCT



DCT

$$\cos((2n+1)m\pi / (2N))$$

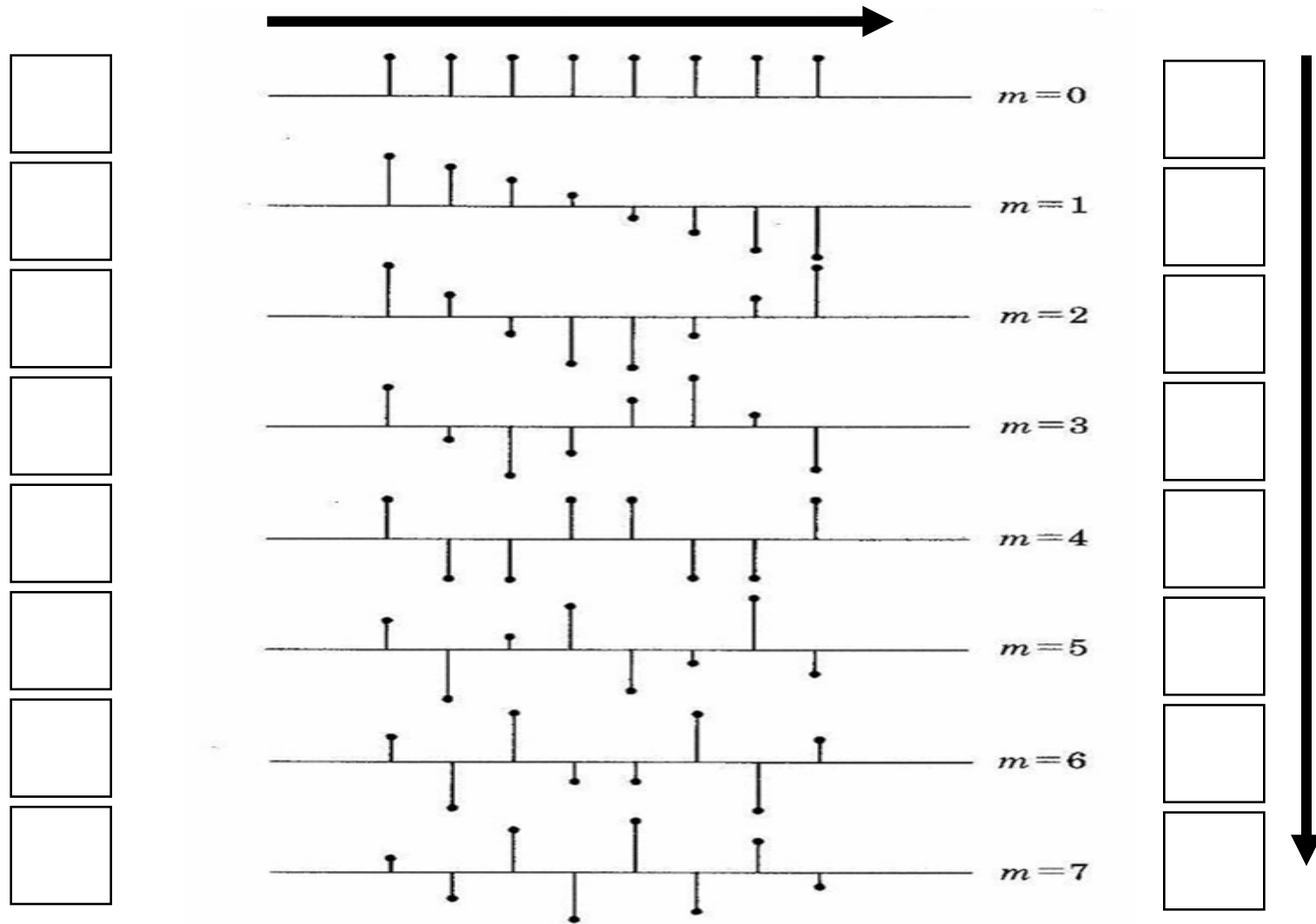
$$n=0, \dots, N-1, m=0, \dots, N-1$$

MDCT

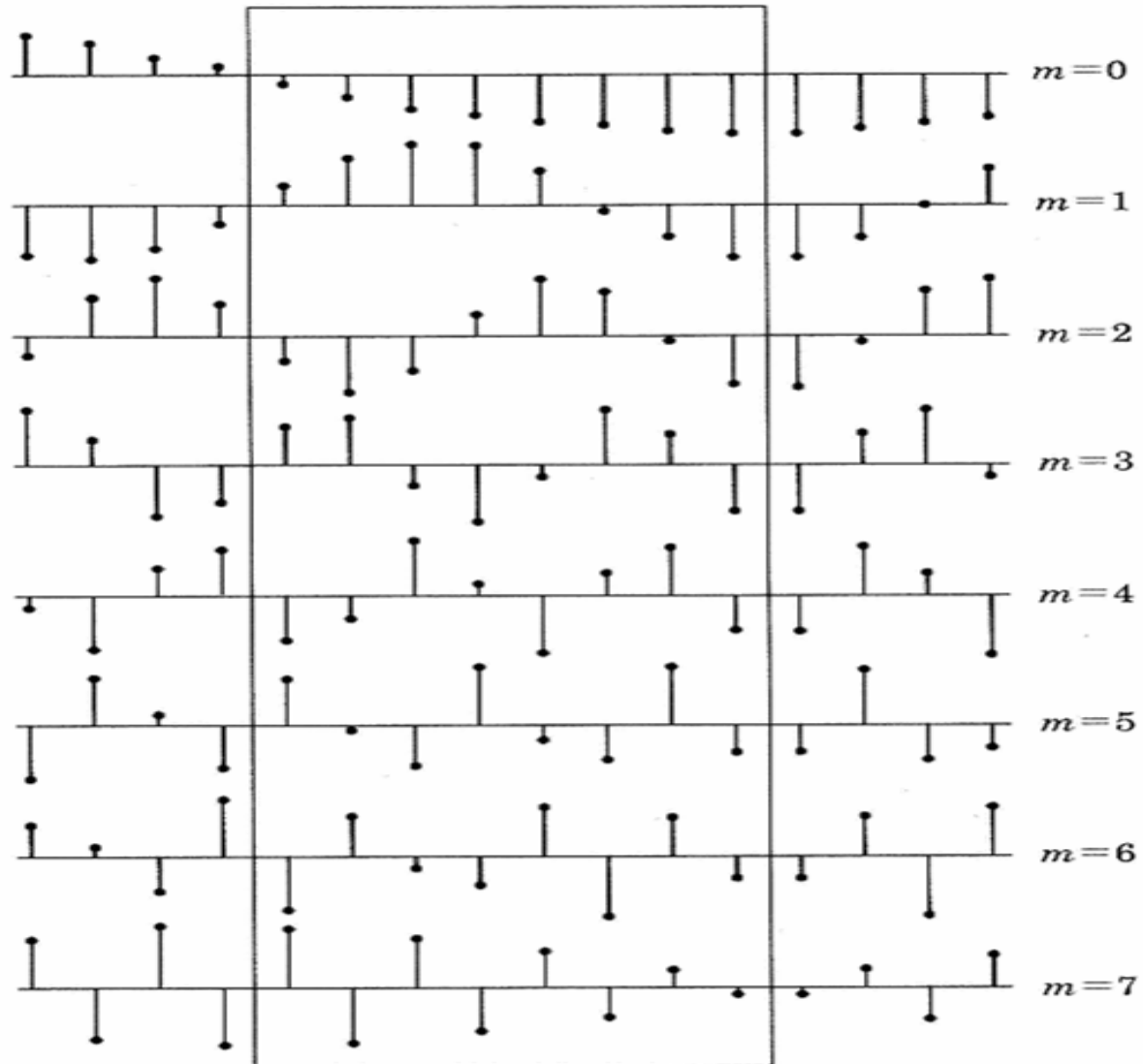
$$\cos((2n+1+N)(2m+1)\pi / (4N))$$

$$n=0, \dots, 2N-1, m=0, \dots, N-1$$

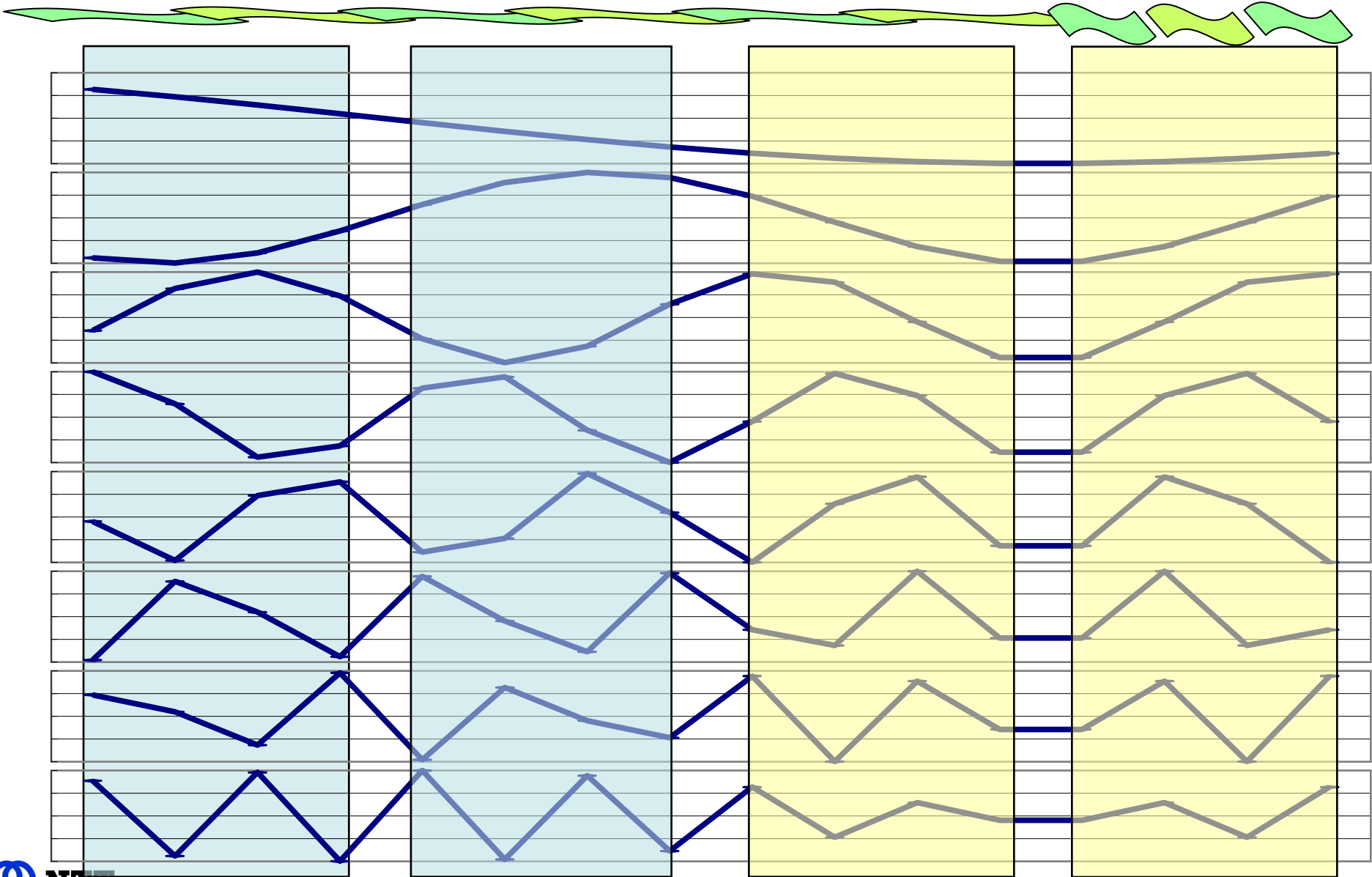
DCTの変換係数



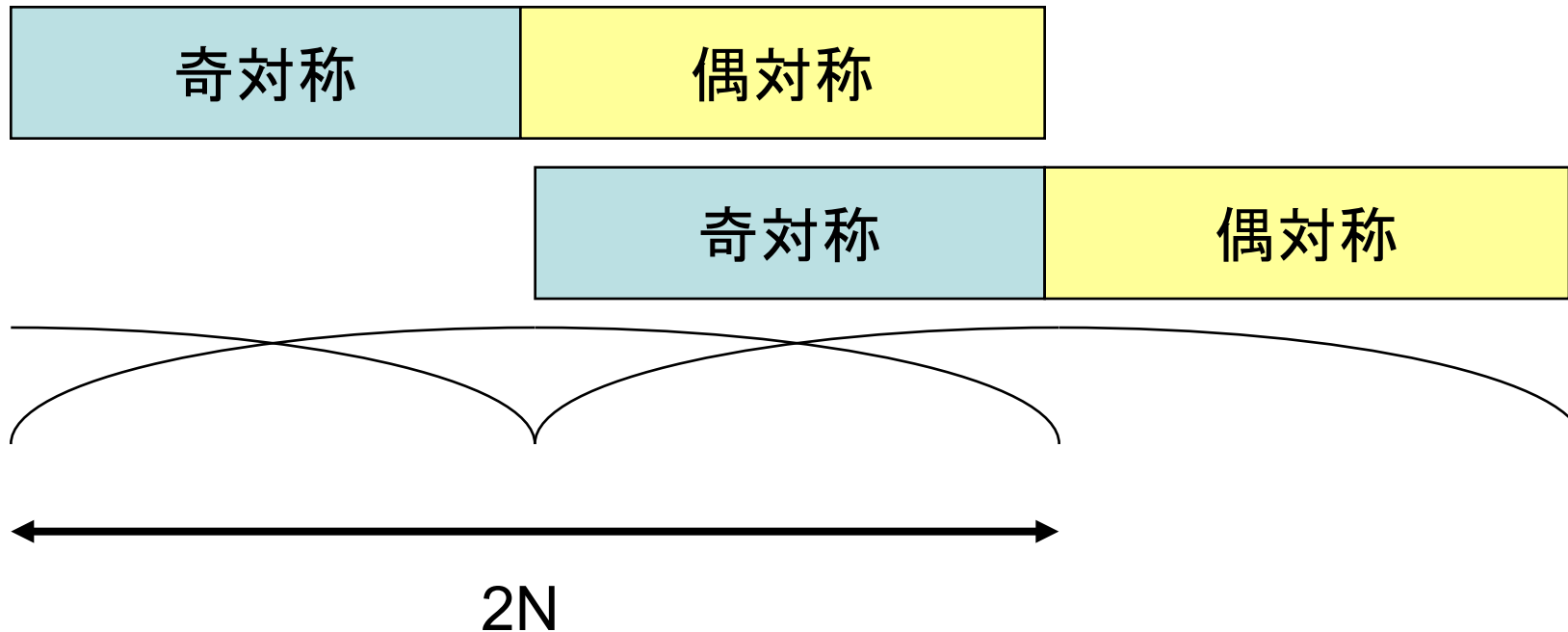
MDCTの変換係数



MDCTの変換係数



MDCTのオーバーラップ



予測符号化と変換符号化

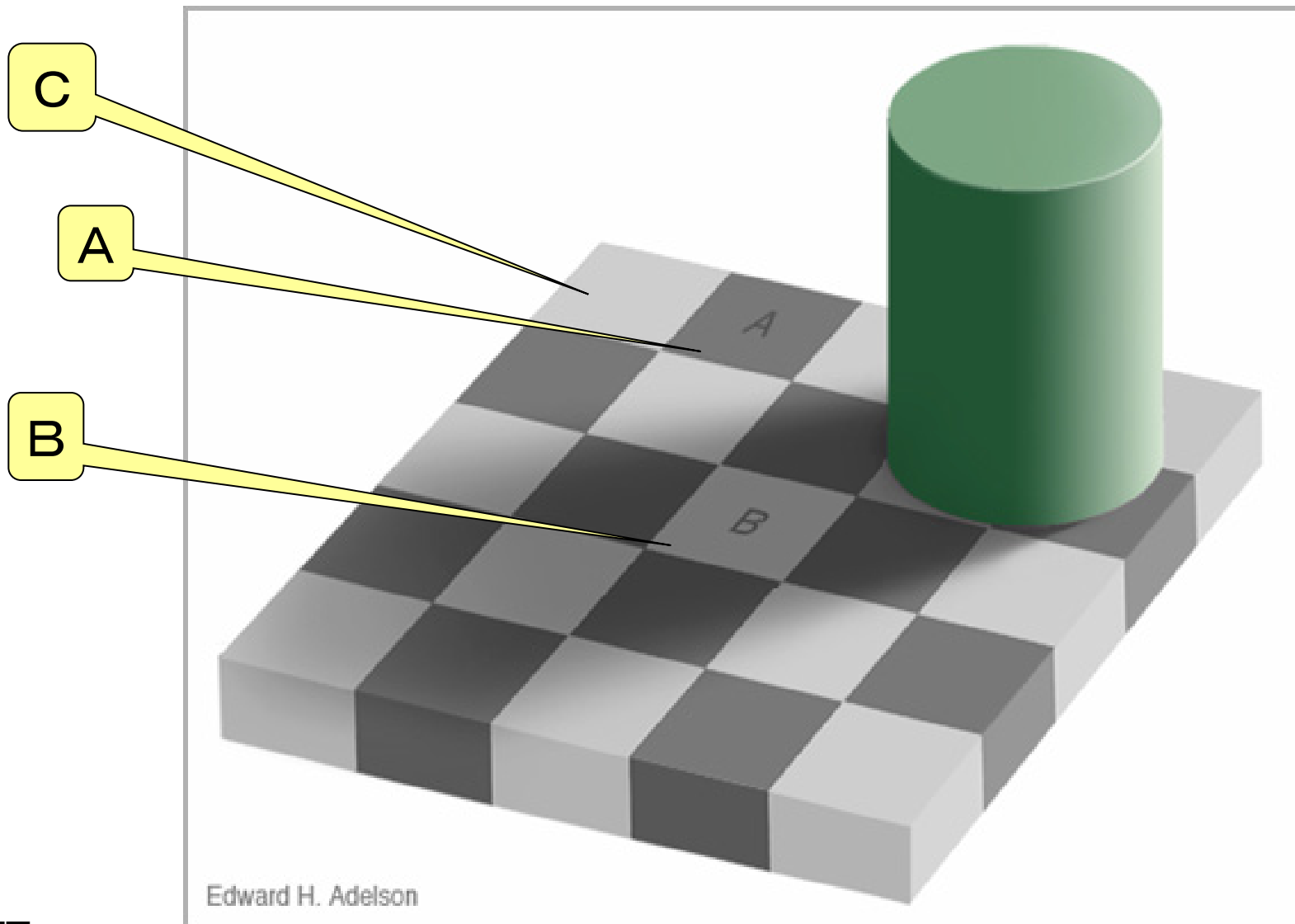
	時間領域(予測)	周波数領域(変換)
相関小さい	予測不可	スペクトル平坦
効果	予測利得	変換利得
利得	波形エネルギー	スペクトル相加平均
	予測誤差エネルギー	スペクトル相乗平均
手段	閉ループ量子化	適応割り当て 適応重み
相関大きい	予測可能	スペクトル偏り

感覚の基本特性



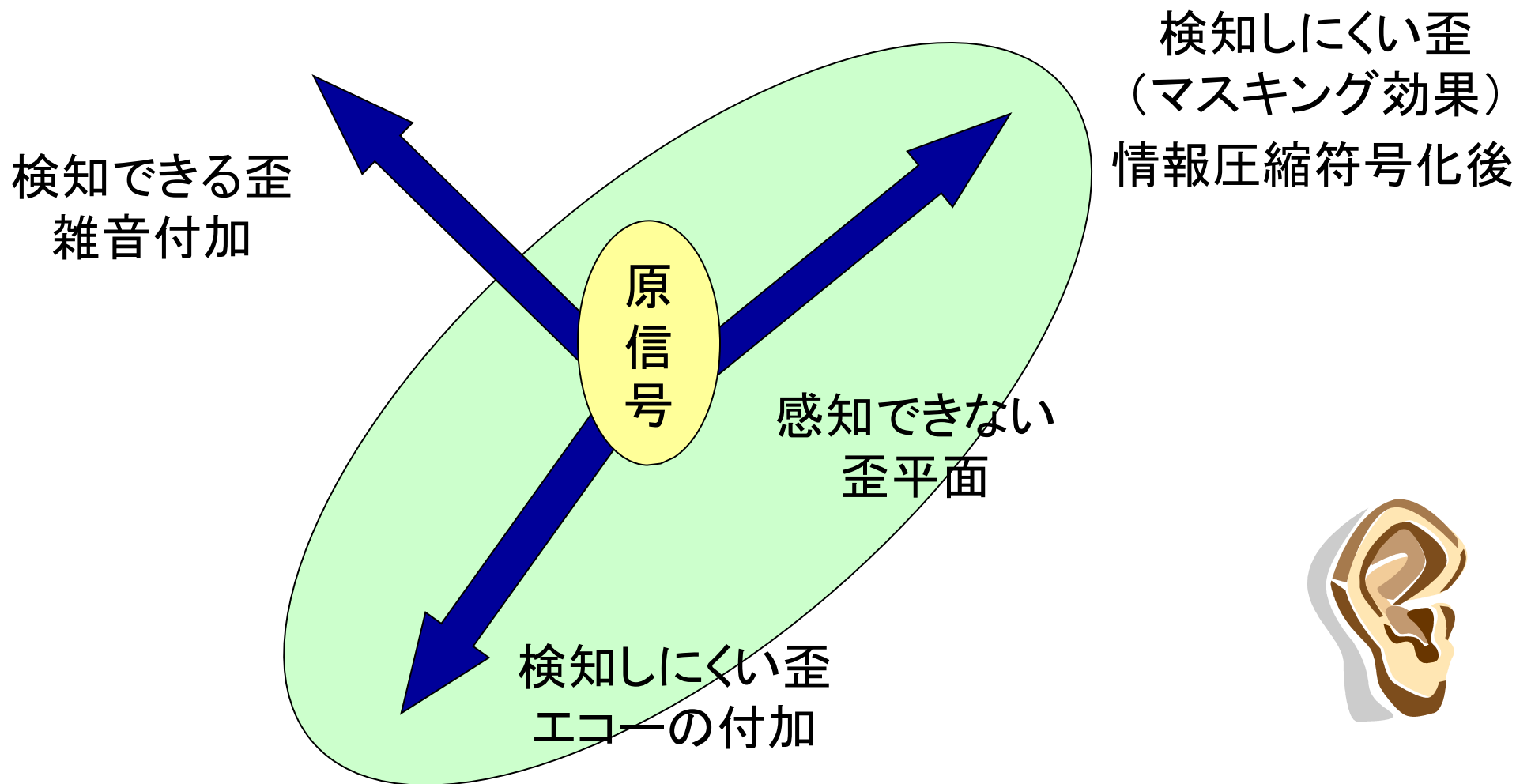
- 感覚は入力の対数に比例する
 - 音の大きさ、高さ、光の強さ、
- 感度は周波数に依存する
- いろいろな錯覚がある
- 順応現象がある
- マスキング
 - 周波数領域、時間領域

明るさとは

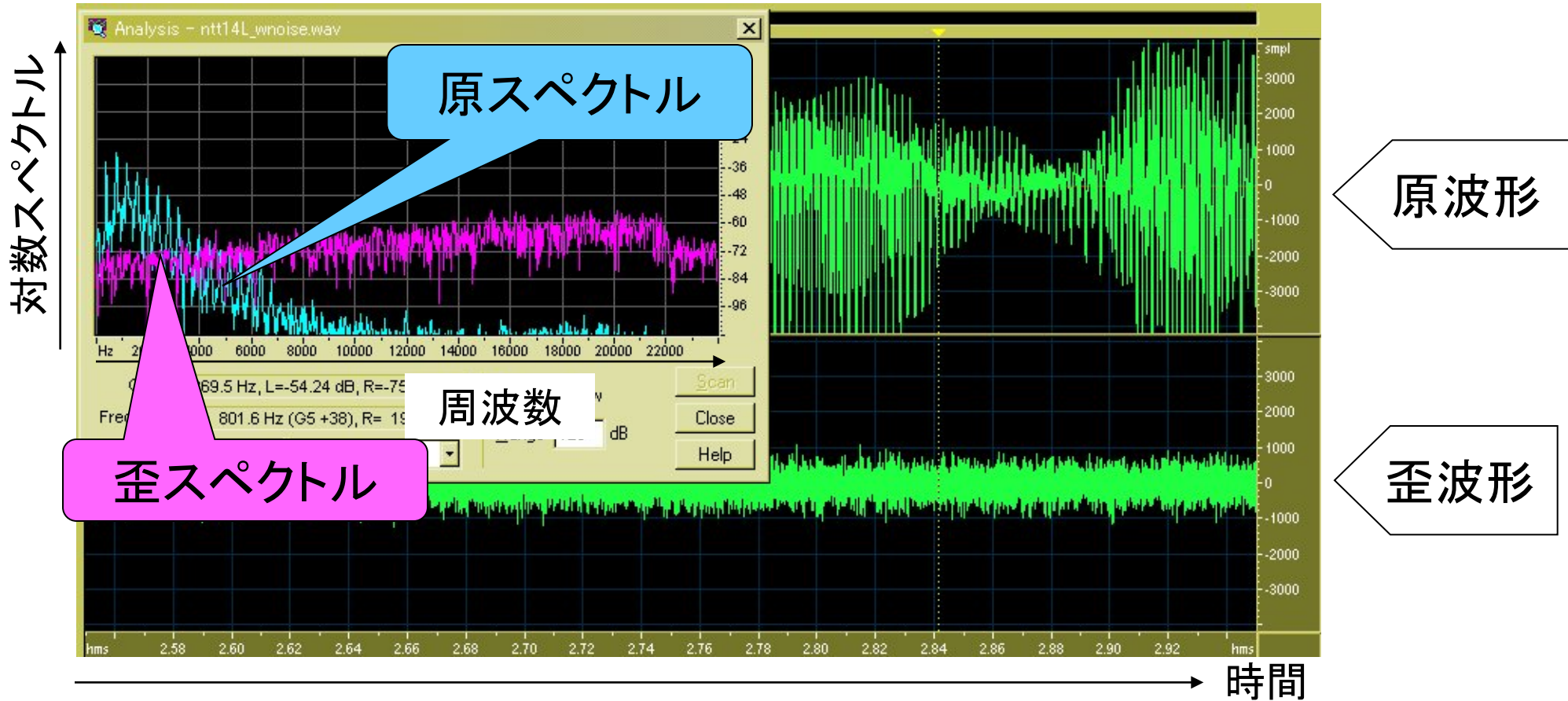


Edward H. Adelson

物理的歪と感覚的歪

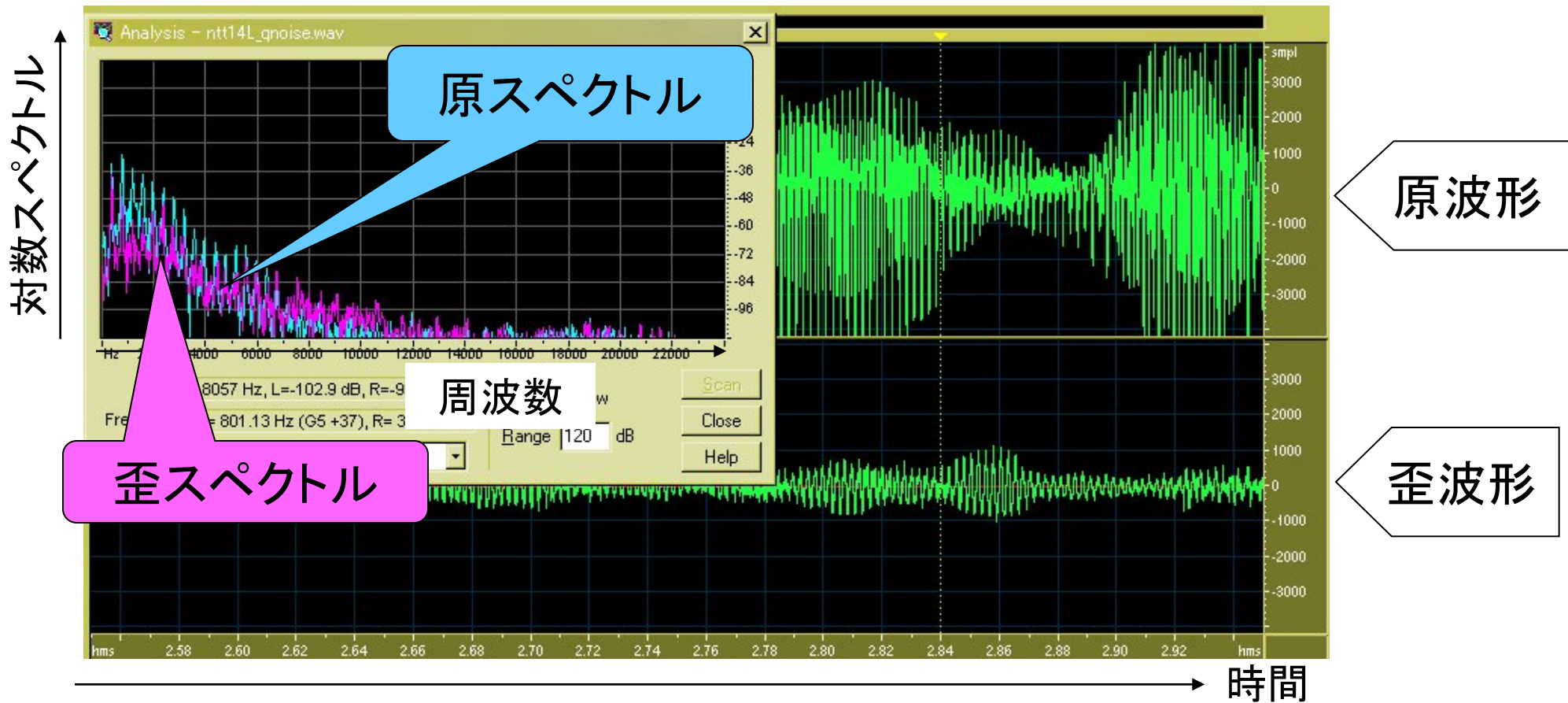


雑音付加による歪



歪が目立つ

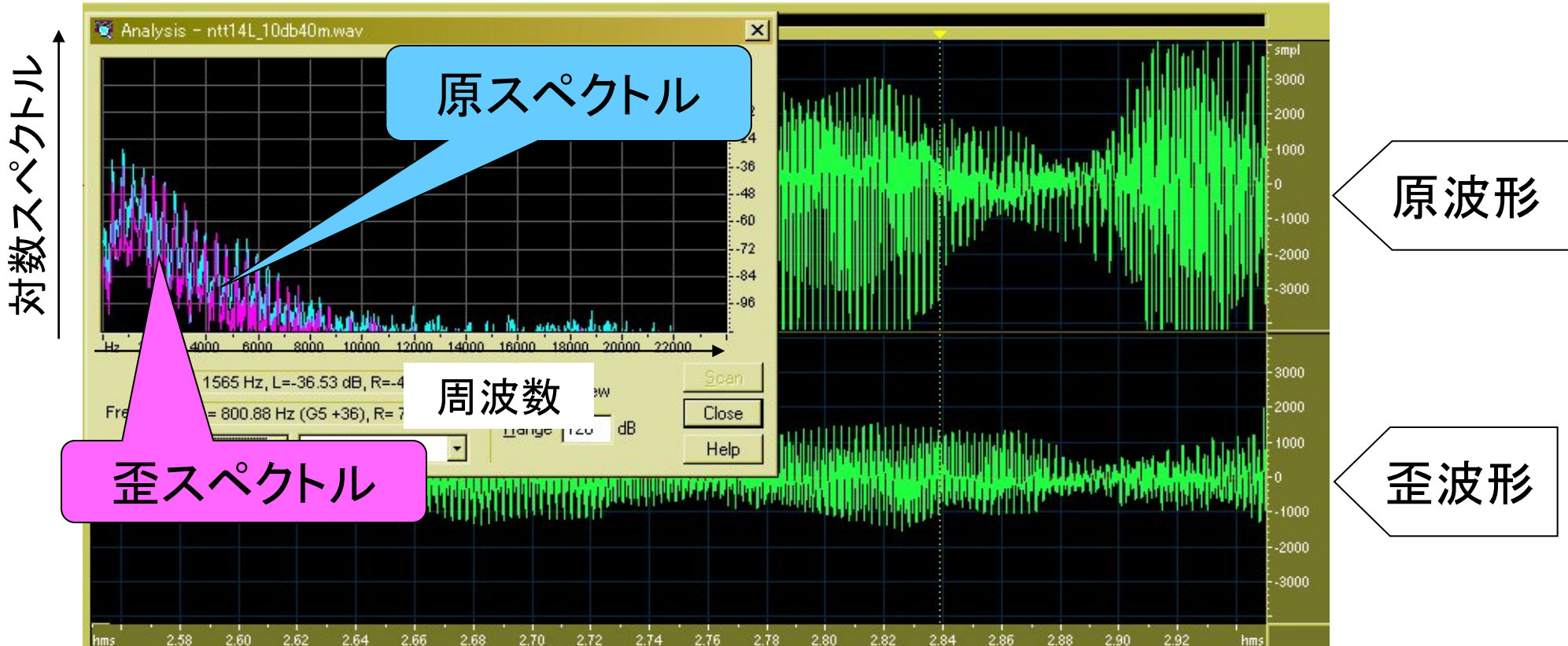
情報圧縮による歪



歪が原スペクトルでマスクされる

歪が目立たないような歪制御

エコー付加による歪



対数スペクトル

原スペクトル

原波形

歪スペクトル

周波数

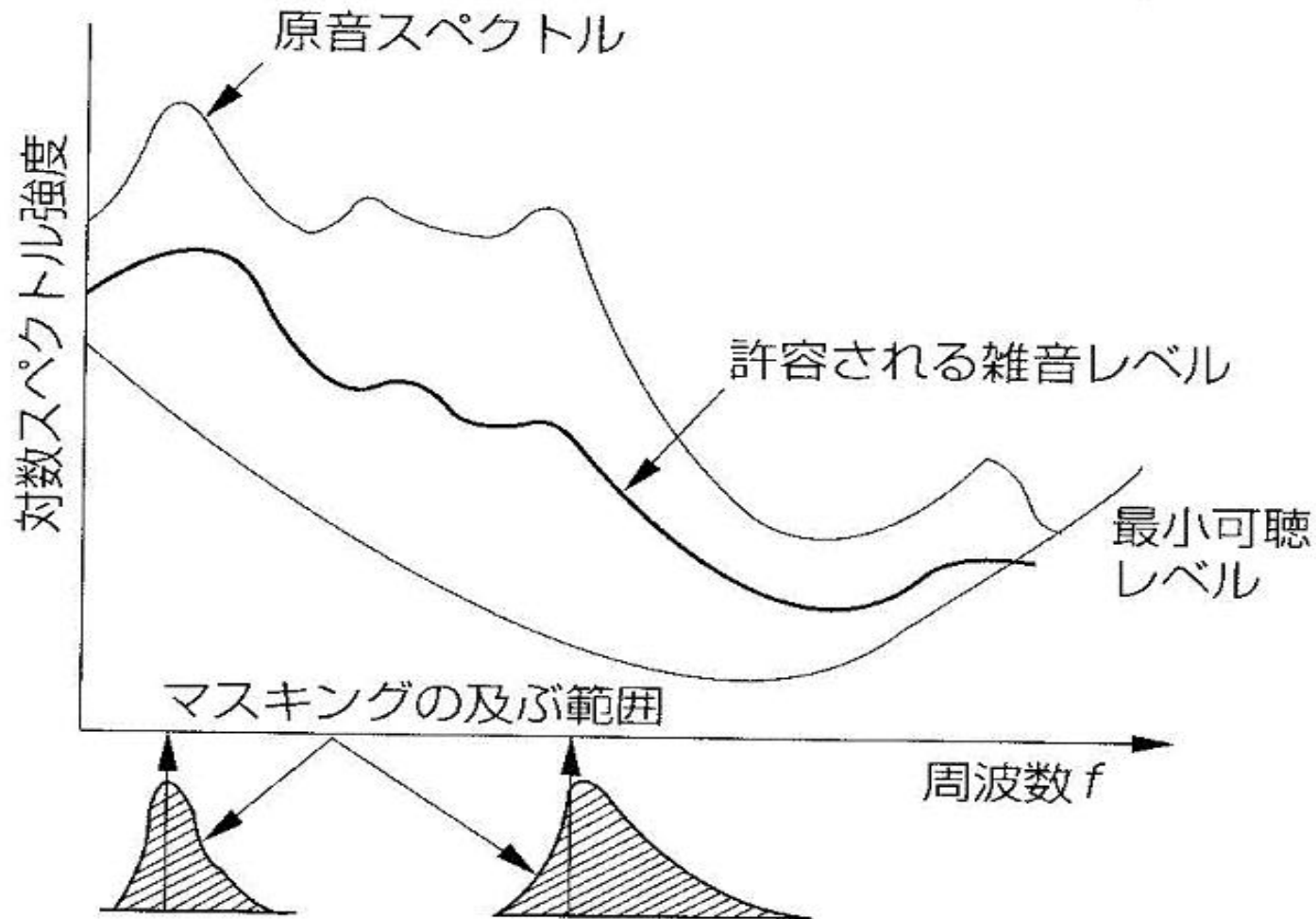
歪波形

時間

歪がめだたない

電子透かし

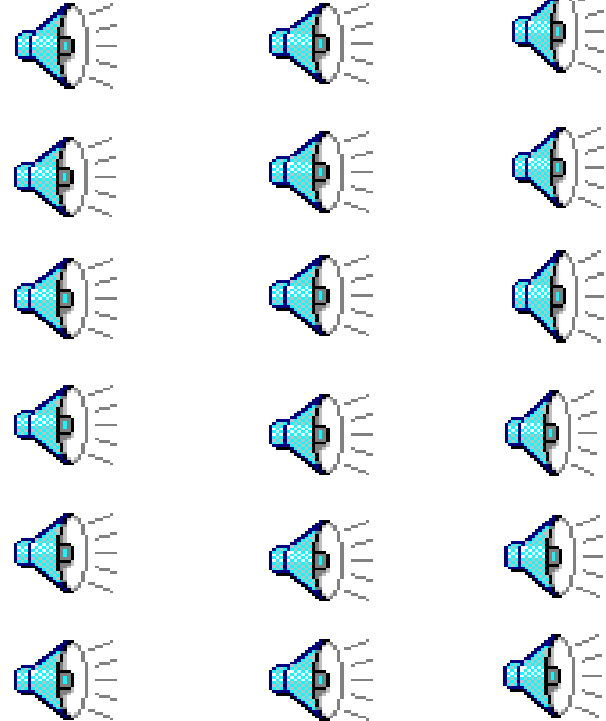
マスキング効果



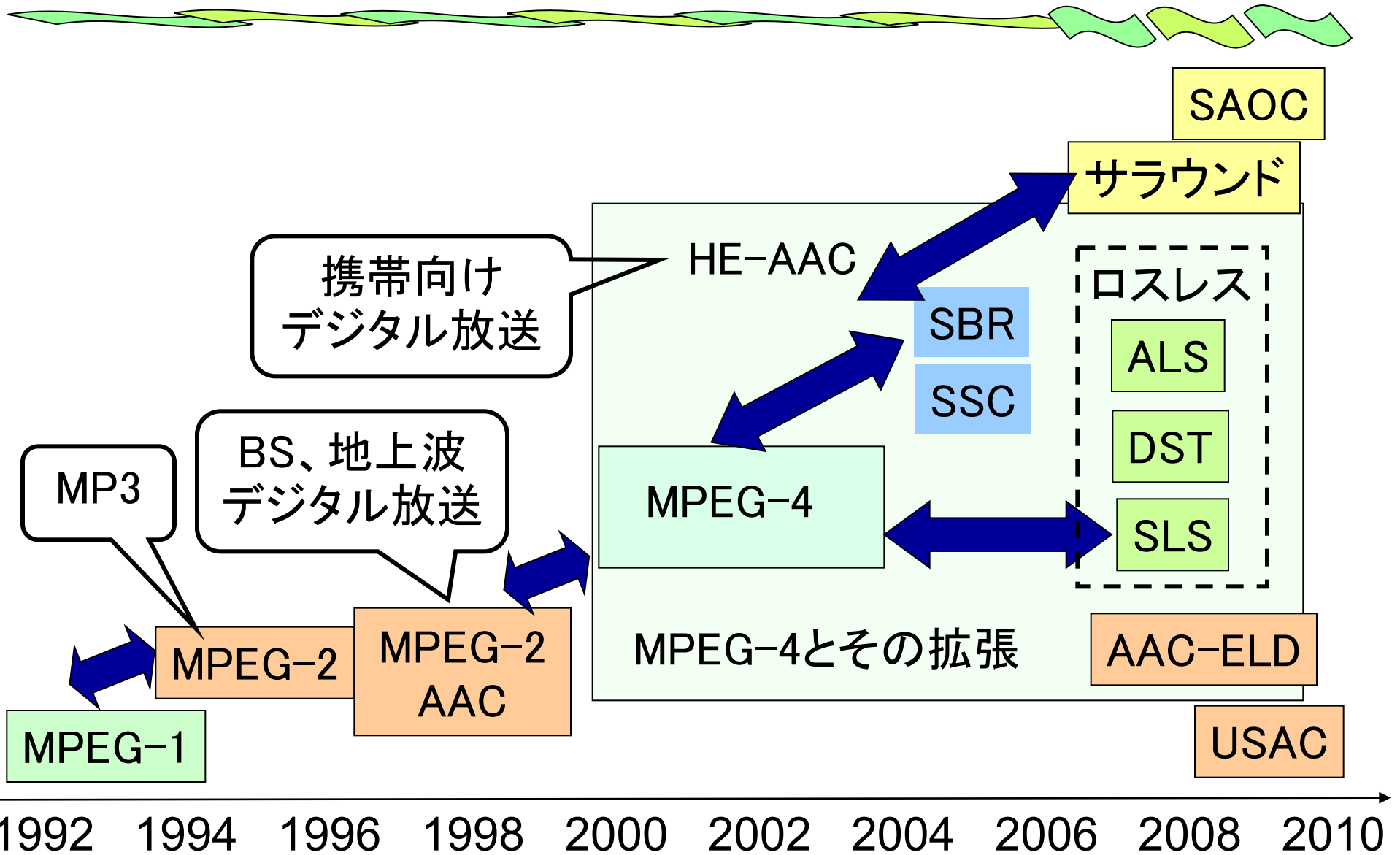
音声デモ



- ITU-T G.711 64 kbit/s
- ITU-T G.726 32 kbit/s
- ITU-T G.729 8 kbit/s
- PDC Half 3.45 kbit/s
- MPEG4 HVXC 2 kbit/s
- MPEG4 TwinVQ 8kbit/s



MPEGの歴史



聴覚圧縮符号化から歪のない符号化

- 歪のない圧縮
 - 高品質・高サンプルレート・マルチチャンネル
 - 過去の大量のアナログデータの永久保存
- ブロードバンドサービスには必要
- 100年後にも解凍できなくてはいけない
- 媒体の大容量化速度 < コンテンツ増大速度

国際標準が必要

歪なし

歪あり

非圧縮

圧縮

ロスありとロスなし

- 信号にロス(歪)が生じる圧縮符号化
 - MPEGレイヤーIII (MP3)、AAC、ミニディスク等
 - 原音の情報量を1/5から1/10に圧縮
 - 聴感上の劣化は殆どないが、再生波形は異なる

圧縮優先

品質ベストエフォート

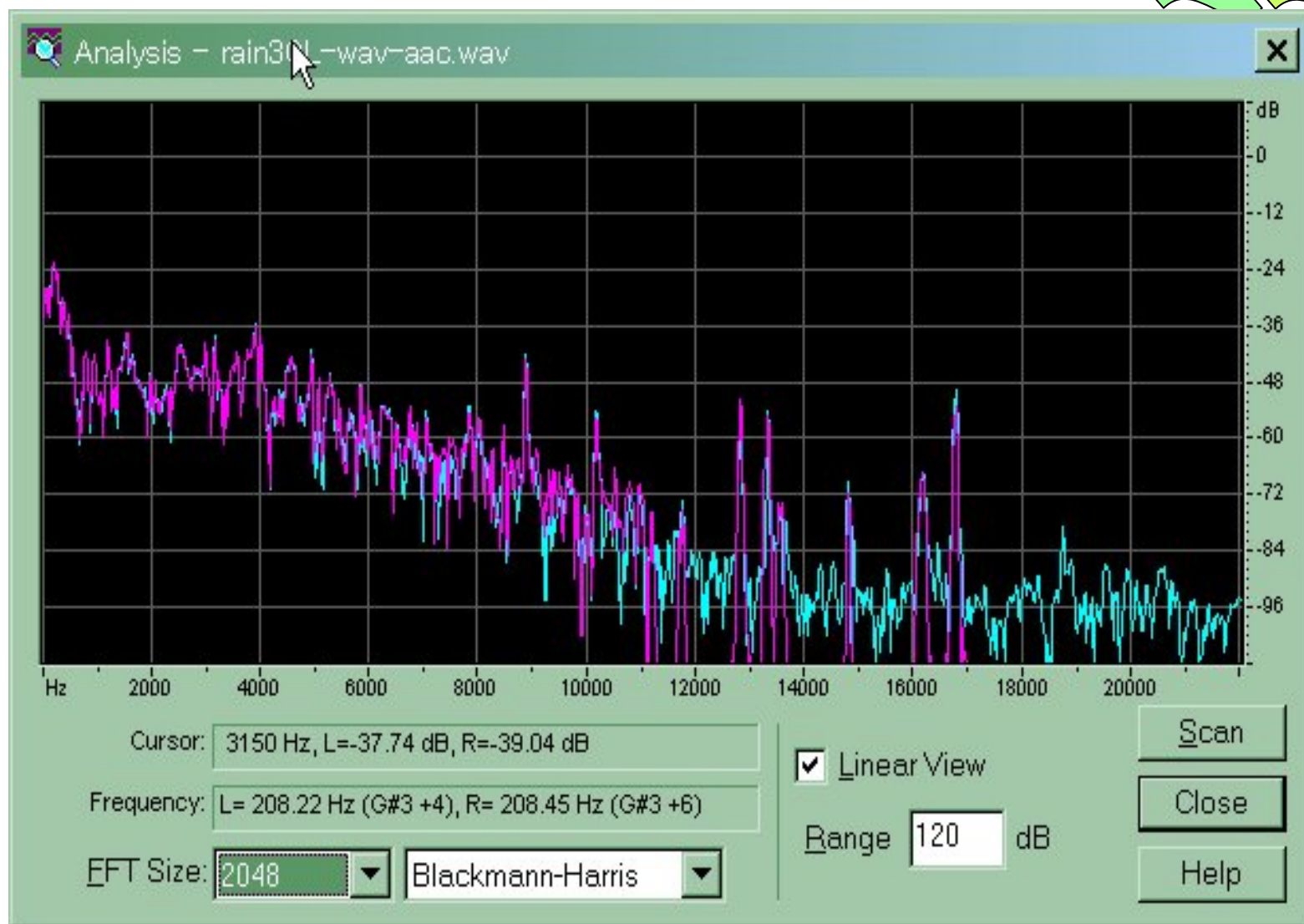
- 信号にロス(歪)が生じない圧縮符号化
 - 元のデータを復元可能(無歪)
 - マスターデータ・編集用素材の保存にも最適

品質優先

圧縮ベストエフォート

AACと原音のスペクトルの差

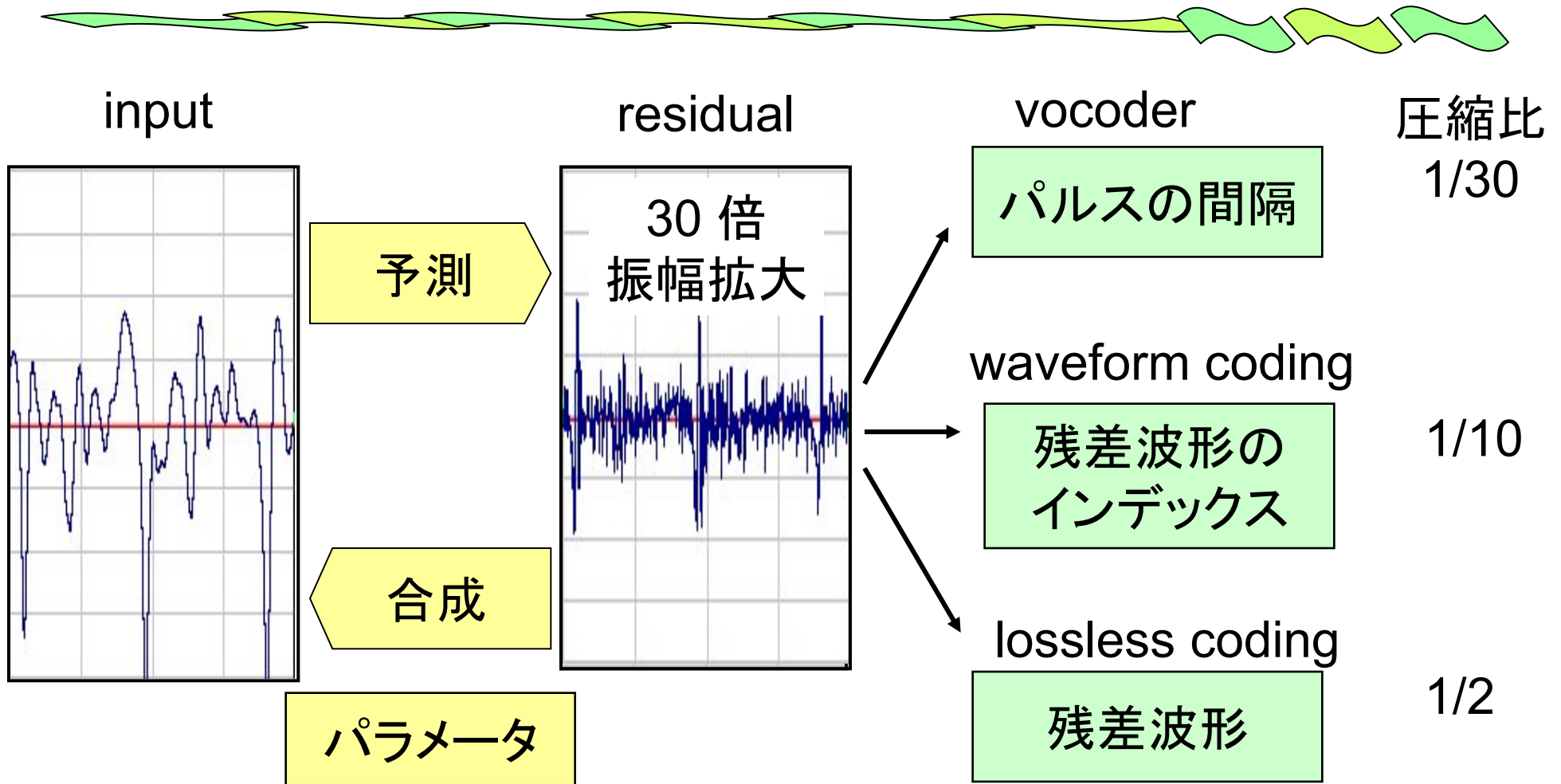
対数スペクトル



原音

AAC

予測符号化



今後の期待



- 符号化の課題
 - 音声と音楽に万能な低ビット符号化 (USAC)
 - 超多チャンネルの符号化・波面合成
- システム化
 - アーカイブフォーマット
 - ビデオとの連携
 - 著作権保護
- 世界戦略
 - 国際標準に日本の機関が大きな貢献
 - 家電製品などは日本のメーカーが大きなシェア
 - 欧米の特許権・ソフトウェア
 - アジア特に中国の生産力