音響情報ハイディング演習
~大雑把なヒトと真正直なマシン~

薗田光太郎（長崎大）

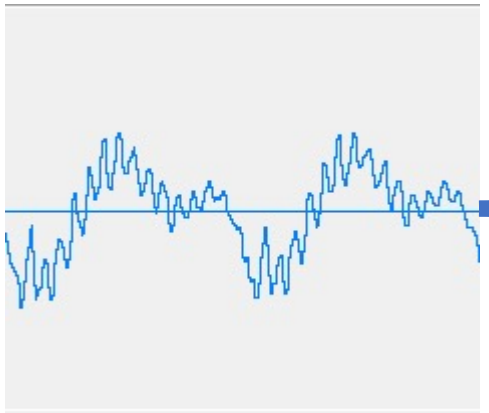
kotaro@nagasaki-u.ac.jp

1. ヒトは大雑把に音を理解している（マシンは真っ正直に音を記録している）

- ヒトは音波を完全には受け取っていない
  - 耳が物理的に無理。難聴になってさらに受け取れない。
  - マイクにも限界があるが、耳よりは受け取れる。

- ヒトは歪んだ音波を歪みがなかったかのように理解する
  - ヒトは知覚的に音波を補償する。
  - マシンは原則的には補償しない。歪んだまま。

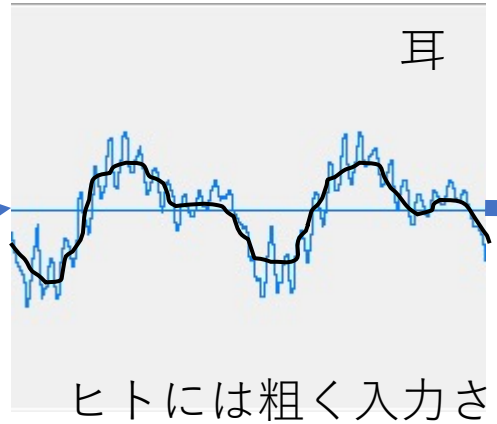
ヒトの耳と，キカイのマイク．



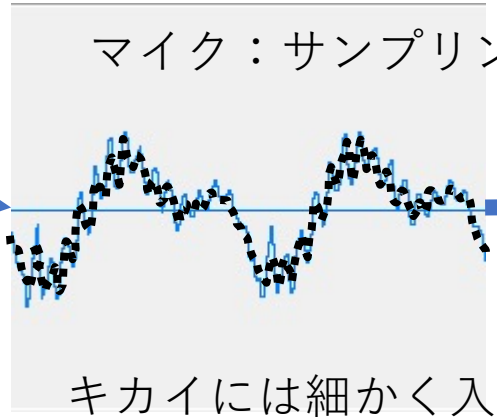
実際にはこんな音だったのに，，

ヒト
（アナログ）

キカイ
（デジタル）



ヒトには粗く入力される
（年をとるとさらに粗く
生体の限界）



キカイには細かく入力される
（ほとんど忠実）

ヒトによる音声認識
（Speech2Text）

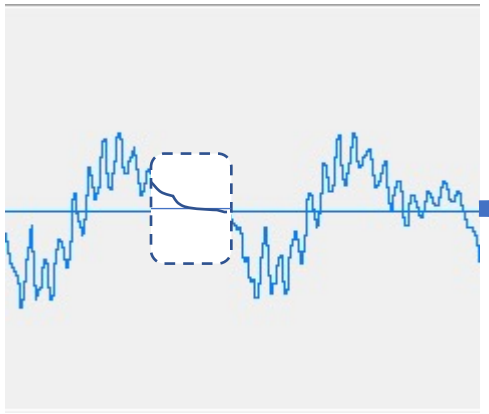
月がきれいですね

キカイによる認識
（ヒトの認識結果に合わせこむ）

もっと細かいこと
話者，感情，情景，
…までわかるのに
不要だって…

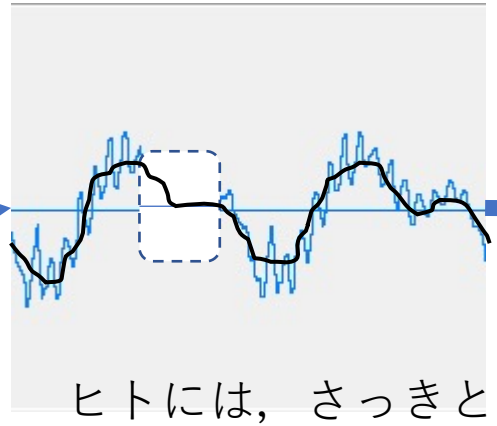
キカイ学習!!

音を粗くしても状況変わらない？



一部分を粗くしてみた，，

ヒト
（アナログ）

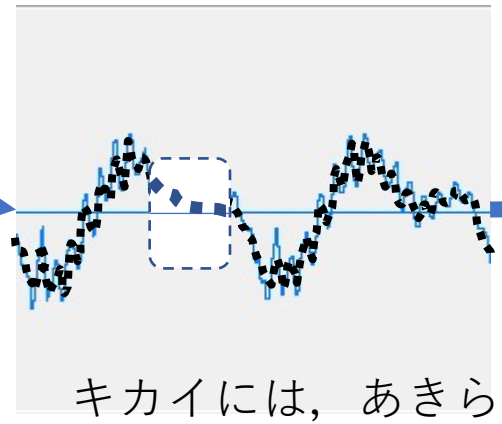


ヒトには，さっきと同じに聞こえる

ヒトによる認識

月がきれいですね

キカイ
（デジタル）

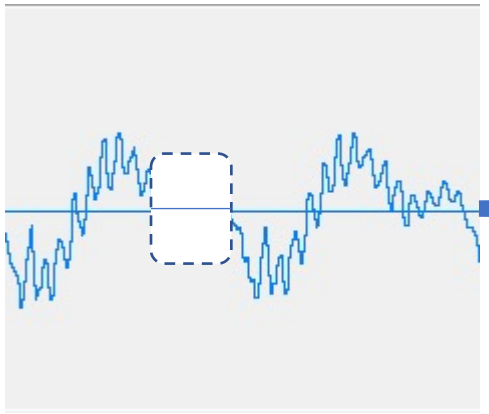


キカイには，あきらかに
さっきと違う音

キカイによる認識
（ヒトの認識結果に合わせこむ）

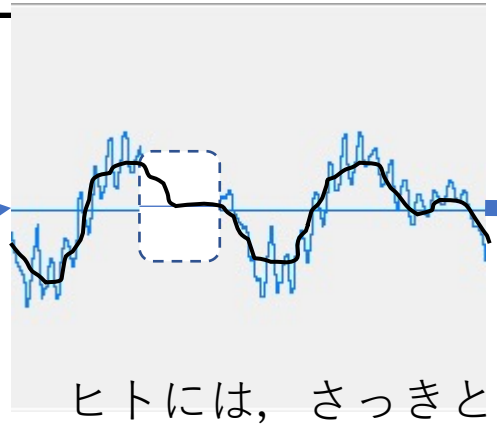
当然まったく問題ない
無駄が省けた
「データ圧縮」‼

粗くするどころか，いじってもヒトには気づかれない？



一部分を**無音**にしてみた，，

ヒト
（アナログ）

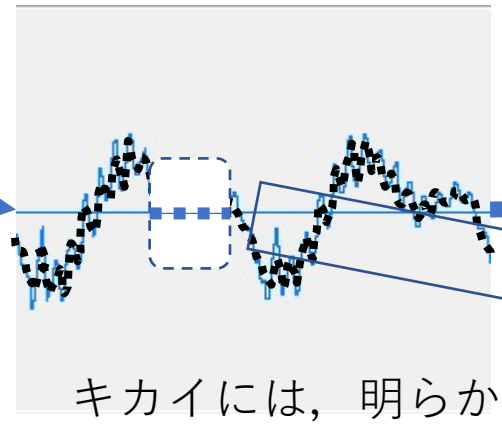


ヒトには，さっきと同じに聞こえる
（粗いどころか**音が無いのに**！！
聴覚補償、錯覚）

ヒトによる認識

月がきれいですね

キカイ
（デジタル）



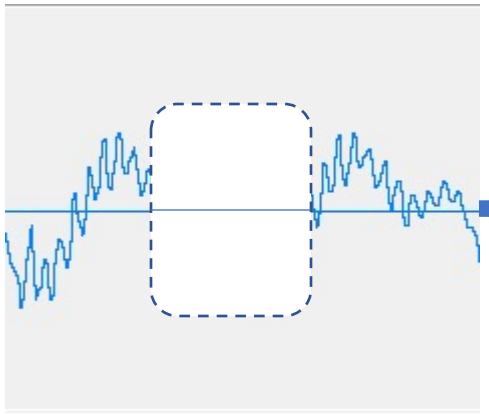
キカイには，明らかに無音

キカイによる認識
（ヒトの認識結果に合わせこむ）

ここは，どういじってもいいってこと？

！！情報ハイディング！！

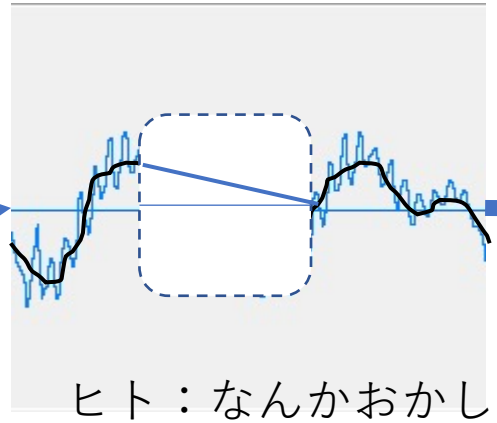
無音をもっと長くすると，さすがにバレる



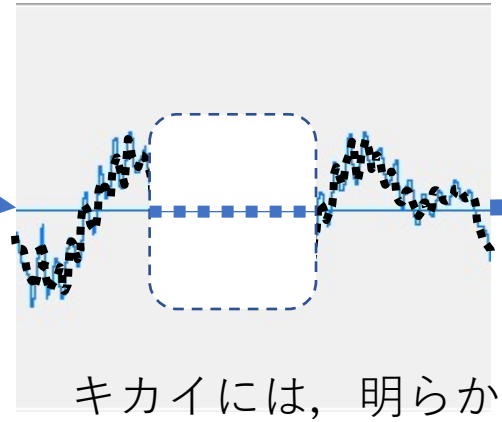
無音をもっと長くしてみた，，

ヒト
（アナログ）

キカイ
（デジタル）

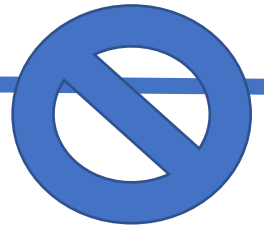


ヒト：なんかおかしい！
（やっと気づく）



キカイには，明らかに無音

ヒトによる認識



月がきれいですね

キカイによる認識
（ヒトの認識結果に合わせこむ）

長さの限界は？？
！！検知限！！

工学への応用

- 音信号ファイルの超絶サイズ圧縮（MP3など）
  - マイクで録音した音だとサイズが大きいが，ヒトの聴こえの粗さに合わせて無駄な細かさを省くとサイズが超絶小さくなる（10分の1！）

- 電子透かし：音に別の情報をひそかに隠蔽
  - ヒトに気づかれない程度の波形を別の波形に．別の波形に意味を持たせる
  - 究極の暗号：
    - ↑の「別の波形」が秘密文。
    - ヒト的には普通の音なので，誰も暗号だときづかない．
      (普通の暗号は，暗号化すると明らかに怪しい文)

科学研究としての意味

- ヒトの音を聴くしくみの解明
  - 限界はどこか？聴覚補償？を解明
  - 「粗い」「細かい」だけではない錯覚の発見

音知覚

- 大きな音の直後の小さい音は物理的にあっても人には聞こえてない
  - 大きな音の直前の小さい音も聞こえない

- 大きな音に近い周波数の小さい音は聞こえない.
- 周波数の多少の違いはわからない
  - 聴覚フィルタ

- 音の強度の多少の違いはわからない

- 位相聾（絶対的な位相はわからない．相対的にしかわからない）
- 音の空隙：隙間が短いと連続して聞こえる（補償）
- ．．．

秘匿通信としての情報ハイディング

- 普通の暗号
  - X：「ABCQ」-- E(X,k)→ Y：「@*$#」--D(Y,k)→ X
  - Yは明らかに暗号文（攻撃者をそそる．危ない！）

- 情報ハイディング
  - X：「ABCQ」→E(X,k) → Y :  ---D(Y,k) → X
  - Yが暗号文にみえない（攻撃者をそそらない．安全）

2.LSB置換法

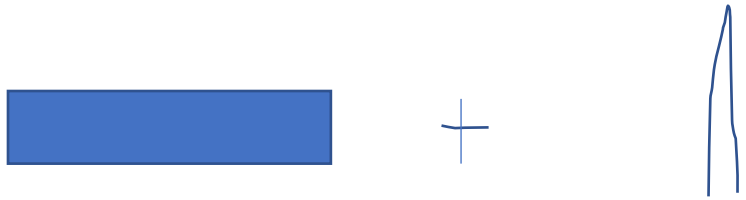
- 音の強度の多少の違いはわからない
- WAVファイル16bit量子化しているがLSBの1ビットを，秘密情報の0/1に置換

  Host　　Stego
  0101:1001:0000:0100 -(1)-> 0101:1001:0000:0101
  0100:1011:0010:1001 -(1)-> 0100:1011:0010:1001
  1110:1111:0010:0010 –(0)-> 1110:1111:0010:0010
- 検出するときはLSBを読めばいい.
- 1サンプルに1ビット埋め込める.
- 空気伝搬，雑音や圧縮などなどですぐ消える

3. スペクトラム拡散法

- 音の強度の多少の違いはわからない
- 音にとても小さい乱数系列（ホワイトノイズ）を順位相・逆位相で加える（0と1を表す）

- 検出するときは，stego音と乱数系列（埋め込み時に使用したもの）との相互相関を計算（ランダム射影）．もとの音波形が乱数系列に，埋め込んだ乱数系列が強いインパルスになる．



4. エコーハイディング

- ヒトは、エコーを不自然に感じない。
  - エコーまでの遅れが
    - 長すぎると、不自然
    - 適度に長いと、エコーなしとの違いはわかるが自然
    - 短いと、エコーなしとの違いにも気づかない
- ある遅延 τ のエコー生成フィルタをひそかにかける。
  - ヒトは気づかないが、マシンには検出できる。
    - 複数の τ のフィルタがあれば、テキストを表現できる。（2種類あれば0と1を表せる。）
    - 1曲に16フレームあれば16ビット隠せる。
- 検出は、エコー検出
  - ケプストラム上で τ に対応するケフレンシーにおける強度が十分に大きいことを判定
  - ※ケプストラムCとは…
    - $Y(\omega) = F[y(t)]$
    - $C(q) = F[\log X(\omega)]$
    - $y(t)=h(t) * x(t)$だとすると、
      - $Y(\omega)=H(\omega) \cdot X(\omega)$、
      - $\log Y(\omega)=\log H + \log X$
    - ケプストラムはスペクトラムY（実際は対数化）を時間波形かのように扱ってフーリエ変換したもの
      - スペクトラムの周期性をとらえている。
      - エコーフィルタが低い周波数（ケプストラム的にはケフレンシーと呼ぶ）にあらわれる

エコーハイディングで暗号化

- やってみよう
- ABCQ2021AIH
  - https://github.com/helmenov/ABCQ2021/release/tag/v08.30.1
    - ABCQ2021AIH.pdf　←このスライド
    - EchoHiding.ipynb　←演習用jupyter notebook
    - wmtest.wav　←テスト用のサンプル楽曲
  - JupyterやGoogleColabなどでEchoHiding.ipynbを開いてください
    - GoogleColabの場合は、左端のファイルタブを選び、wmtest.wavをアップロードしてください。

エコーハイディングの問題

- 鍵が推測されてしまう
  - ケプストラムを観察すると、エコー成分が見えてしまう。
  - →鍵であるτ_0 τ_1の予想がつく
- エコーを減らす処理をされると検出の妨害となる。

- 対策
  - 単発エコーではなく、複数の乱発エコー（多段エコー）にする。
    - →エコー拡散法
    - 遅延τだけでなく、乱数列(PN系列)も鍵に組み入れる。
    - ケプストラムとPN系列との相関の有無で検出