

知能情報メディア実験中間レポート

複数話者の同時発話音声からの個別音声の抽出

情報科学類, 3 年, 2 クラス, 学籍番号: 201311403 山岡 洸瑛, Yamaoka Kouei

2015/5/15(金)

現在までの進捗状況

課題 1: 瞬時混合 ICA の実装

課題 1 はすべて実装した。ここで学んだのは、混合の単純化されたモデルである瞬時混合であり、これに関しては大まかには理解できたと考えている。事実ここで得た知識が課題 2 において非常に役立った。ここで得た知識とは、音声の混合及び分離においてどのようなモデルで表すことができ、どのような式で表現できるのか、ということと、その理由である。しかし、分離行列の推定方法など理解しきれていない部分もあり、そういった部分の理解が今後の課題である。1,2 回授業資料を読んだ程度では理解出来なかったため、課題が進み時間に余裕がきたら手計算で確認してみようと考えている。

課題 2-1: たたみ込み混合の分析

課題 2 はまだ 1 つ目の課題すら終わっていない。しかし、追加で出された畳み込み信号をつくる課題は完了した。ここでは畳み込み積分やインパルス応答などまだ知らなかった多くの知識を必要とし、実装が難しかったが、TA の杉本さんなどに教えてもらいながら最終的には実装することができた。実装した今では知らなかった知識も得ることができ、理解することもできたと考えている。このレポートを書いている頃には、他の授業でもフーリエ変換や畳み込み積分を本格的に扱だし、より理解も深まっている。

現在の問題点

現在取り組んでいる課題は課題 2-1-1 である。この課題は畳み込み混合信号の stft 分析と逆 stft 分析を行う課題である。stft については理解し、そのために必要な手順も理解しているが、それを実装することに手間取っている。手間取っている点は、stft 分析し、推定された音源信号を時間領域に戻す点である。stft 分析の過程でハミング窓を使用し、フレームシフトをフレーム長の半分にした場合など、条件を固定してやれば実装することはできると考えているが、フレームシフトをフレーム長の $\frac{1}{4}$ に変更した場合などに動かない、汎用性の低いコードになってしまう。窓については窓関数だけを変えれば動くような、フレームシフトについてはフレーム長の半分、 $\frac{1}{4}$ どちらにも、あるいは可能ならばその他の一般的でない長さでも対応できるようなコードを作成したいと考えている。

具体的には音声信号と同じ長さの配列を用意し、窓関数で切り出すたびに同じインデックスに 1 加算する。これによって音声信号のどの位置を何回切り出したかを保存することができる。最後の段階で切り出された音声信号に逆 fft をし、足しあわせると、オーバーラップにより 2 回以上切り出された部分の振幅が、切り出された回数倍されてしまうが、この保存用配列で割ることで元の信号と同じ振幅に復元することができる、と考えている。

このように実装方針は立っているのでは実装するだけであるが、うまく行っていないのが現状である。従ってまずは汎用性の低いコードを作成し、それを拡張していく、という方針で今後は動いていく予定である。