HARK version 1.2.0 Document

奥乃 博
中臺 一博
高橋 徹
武田 龍
中村 圭佑
水本 武志
吉田 尚水
大塚 琢馬
柳楽 浩平
糸原 達彦

[ARIEL sings]
Come unto these yellow sands,
And then tale hands:
Curt'sied when you have, and kiss'd,
(The wild waves whist;)
Foot it featly hear and there;
And sweet sprites, the burden bear.
[Burden dispersedly.]
HARK, hark! bowgh-wowgh: the watch-dogs bark,
Bowgh-wowgh.
Ariel. HARK, hark! I hear
The strain of strutting chanticleer
Cry cock-a-doodle-doo.

Ariel's Song, The Tempest, Act I, Scene II, William Shakespeare

目 次

第 1 章　はじめに **1**
1.1　ロボット聴覚ソフトウエアは総合システム 1
1.2　HARK の設計思想 1
1.3　HARK のモジュール群 6
1.4　HARK の応用 8
1.4.1　3 話者同時発話認識 9
1.4.2　口ジャンケンの審判 10
1.4.3　CASA 3D Visualizer 10
1.4.4　テレプレゼンスロボットへの応用 12
1.5　まとめ 14

第 2 章　ロボット聴覚とその課題 **15**
2.1　ロボット聴覚は聞き分ける技術がベース 15
2.2　音環境理解をベースにしたロボット聴覚 15
2.3　人のように 2 本のマイクロフォンで聞き分ける 16
2.4　自己生成音抑制機能 17
2.5　視聴覚情報統合による曖昧性解消 19
2.6　ロボット聴覚が切り開くキラーアプリケーション 19
2.7　まとめ 20

第 3 章　はじめての HARK **23**
3.1　ソフトウェアの入手方法 23
3.2　ソフトウェアのインストール方法 23
3.2.1　Linux 版のインストール方法 23
3.2.2　Windows 版のインストール方法 24
3.3　FlowDesigner 25
3.3.1　Linux 版 25
3.3.2　Windows 版 33

第 4 章　データ型 **38**
4.1　基本型 41
4.2　FlowDesigner オブジェクト型 42
4.2.1　Vector 42
4.2.2　Matrix 42
4.3　FlowDesigner 固有型 43
4.3.1　any 43
4.3.2　ObjectRef 43

4.3.3 Object . . . 43
4.3.4 subnet_param . . . 43
4.4 HARK 固有型 . . . 45
4.4.1 Map . . . 45
4.4.2 Source . . . 45
4.5 HARK 標準座標系 . . . 46

第 5 章 ファイルフォーマット **47**
5.1 HGTF ファイル形式 . . . 48
5.1.1 LocalizeMUSIC 用音源定位伝達関数 . . . 49
5.1.2 GHDSS 用音源分離伝達関数 . . . 49
5.1.3 GHDSS 用 分離行列 . . . 50
5.2 HARK テキスト形式 . . . 51
5.2.1 マイクロホン位置テキスト形式 . . . 51
5.2.2 ノイズ位置テキスト形式 . . . 53
5.3 その他のファイル形式 . . . 55
5.3.1 音源位置リスト情報 (srcinf) 形式 . . . 55
5.3.2 PCM バイナリ形式 . . . 59
5.3.3 float バイナリ . . . 59
5.3.4 定位結果テキスト . . . 60
5.3.5 Map テキスト . . . 60
5.3.6 音源定位用相関行列ファイル . . . 60

第 6 章 ノードリファレンス **62**
6.1 AudioIO カテゴリ . . . 63
6.1.1 AudioStreamFromMic . . . 63
6.1.2 AudioStreamFromWave . . . 71
6.1.3 SaveRawPCM . . . 73
6.1.4 SaveWavePCM . . . 76
6.1.5 HarkDataStreamSender . . . 78
6.2 Localization カテゴリ . . . 85
6.2.1 CMLoad . . . 85
6.2.2 CMSave . . . 87
6.2.3 CMChannelSelector . . . 89
6.2.4 CMMakerFromFFT . . . 91
6.2.5 CMMakerFromFFTwithFlag . . . 93
6.2.6 CMDivideEachElement . . . 96
6.2.7 CMMultiplyEachElement . . . 98
6.2.8 CMInverseMatrix . . . 100
6.2.9 CMMultiplyMatrix . . . 102
6.2.10 CMIdentityMatrix . . . 104
6.2.11 ConstantLocalization . . . 106
6.2.12 DisplayLocalization . . . 108
6.2.13 LocalizeMUSIC . . . 110

6.2.14 LoadSourceLocation . . . 119
6.2.15 SaveSourceLocation . . . 121
6.2.16 SourceIntervalExtender . . . 123
6.2.17 SourceTracker . . . 125
6.3 Separation カテゴリ . . . 129
6.3.1 BGNEstimator . . . 129
6.3.2 CalcSpecSubGain . . . 133
6.3.3 CalcSpecAddPower . . . 135
6.3.4 EstimateLeak . . . 137
6.3.5 GHDSS . . . 139
6.3.6 HRLE . . . 149
6.3.7 PostFilter . . . 153
6.3.8 SpectralGainFilter . . . 167
6.4 FeatureExtraction カテゴリ . . . 169
6.4.1 Delta . . . 169
6.4.2 FeatureRemover . . . 172
6.4.3 MelFilterBank . . . 174
6.4.4 MFCCExtraction . . . 178
6.4.5 MSLSExtraction . . . 181
6.4.6 PreEmphasis . . . 185
6.4.7 SaveFeatures . . . 187
6.4.8 SaveHTKFeatures . . . 189
6.4.9 SpectralMeanNormalization . . . 191
6.5 MFM カテゴリ . . . 193
6.5.1 DeltaMask . . . 193
6.5.2 DeltaPowerMask . . . 196
6.5.3 MFMGeneration . . . 198
6.6 ASRIF カテゴリ . . . 201
6.6.1 SpeechRecognitionClient . . . 201
6.6.2 SpeechRecognitionSMNClient . . . 203
6.7 MISC カテゴリ . . . 205
6.7.1 ChannelSelector . . . 205
6.7.2 CombineSource . . . 207
6.7.3 DataLogger . . . 209
6.7.4 MatrixToMap . . . 211
6.7.5 MultiDownSampler . . . 213
6.7.6 MultiFFT . . . 218
6.7.7 MultiGain . . . 222
6.7.8 PowerCalcForMap . . . 224
6.7.9 PowerCalcForMatrix . . . 226
6.7.10 SegmentAudioStreamByID . . . 228
6.7.11 SourceSelectorByDirection . . . 230
6.7.12 SourceSelectorByID . . . 232
6.7.13 Synthesize . . . 234

6.7.14 WhiteNoiseAdder 236
6.8 Flow Designer に依存しないモジュール 238
6.8.1 JuliusMFT 238

第 7 章 サポートツール 245
7.1 HARKTOOL 245
7.1.1 概要 245
7.1.2 インストール方法 247
7.1.3 起動方法 247
7.1.4 作業画面説明 248
7.1.5 インパルス応答リストファイル作成方法 250
7.1.6 TSP 応答リストファイル作成方法 252
7.1.7 マイクロホン位置情報ファイル作成方法 255
7.1.8 ノイズ位置情報ファイル作成方法 257
7.1.9 定位用伝達関数ファイルの作成 258
7.1.10 分離用伝達関数ファイルの作成 264
7.1.11 コマンド実行形式 270
7.2 wios 273
7.2.1 概要 273
7.2.2 インストール方法 273
7.2.3 使用方法 273

第 8 章 HARK 対応マルチチャネル A/D 装置の紹介と設定 275
8.1 System In Frontier, Inc. RASP シリーズ 276
8.1.1 無線 RASP 276
8.1.2 RASP-24 277
8.2 RME Hammerfall DSP シリーズ Multiface AE 279
8.2.1 Multiface の PC への接続 279
8.2.2 Multiface を用いた HARK での録音テスト 281
8.3 東京エレクトロンデバイス TD-BD-16ADUSB 286
8.3.1 16ADUSB の PC への接続 286
8.3.2 16ADUSB 用ソフトウェアのインストールと設定 286
8.3.3 TD-BD-16ADUSB を用いた HARK での録音テスト 286

第1章　はじめに

本ドキュメントは，ロボット聴覚ソフトウエア HARK (HRI-JP Audition for Robots with Kyoto Univ., hark は listen を意味する中世英語) に関する情報の集大成である．第 1 章では，HARK の設計思想，設計方針，個々の技術の概要，HARK の応用について述べるとともに，HARK を始めとするロボット聴覚ソフトウエア，ロボット聴覚機能が切り開く新しい地平について概観する．

1.1　ロボット聴覚ソフトウエアは総合システム

人は，色々な音が聞こえる多様な環境で音を「聞き分けて」処理を行い，人とコミュニケーションを行ったり，TV，音楽，映画などを楽しんだりしている．このような聞き分ける処理を提供するロボット聴覚機能は，実環境で聞こえる多様な音を様々なレベルで処理するための機能を包含する必要があり，ロボットビジョンの機能と同様に一言で定義できない．実際，オープンソース画像処理ソフトウエア OpenCV が膨大な処理モジュールの集合体であるように，ロボット聴覚ソフトウエアも最低限必要な機能を含んだ集合体を成していることが不可欠である．

ロボット聴覚ソフトウエア HARK は『聴覚の OpenCV』を目指したシステムである．OpenCV のように「聞き分ける」ために必要なモジュールをデバイスレベルから信号処理アルゴリズム，測定ツール，GUI まで包含するだけでなく，さらに，オープンソースとして公開をしている．

音情報を基に音環境を理解する音環境理解（Computational Auditory Scene Analysis）研究での 3 つの主要課題は，音源定位 (sound source localization)，音源分離 (sound source separation)，及び，分離音声の音声認識 (automatic speech recognition) である．HARK 第 1 版は，これらの研究の成果として開発してきた．現在，研究用にはオープンソースとして無償公開[1]を行っている．

以下，第 2 節で HARK の設計思想について述べ，HARK が現在ミドルウエアとして利用している FlowDesigner について概説する．第 3 節で HARK のモジュール群について概説する．第 4 節で今後の開発予定を述べる．

1.2　HARK の設計思想

ロボット聴覚ソフトウエア HARK の設計思想を以下にまとめる．

1. **入力から音源定位・音源分離・音声認識までの総合機能の提供**：ロボットに装備するマイクロフォンからの入力，マルチチャネル信号処理による音源定位，音源分離，雑音抑制，分離音認識にわたる総合性能の保証，
2. **ロボットの形状への対応**：ユーザの要求するマイク配置への対応と信号処理への組込，
3. **マルチチャネル A/D 装置への対応**：価格帯・機能により多様なマルチチャネル A/D 装置をサポート，

[1] http://winnie.kuis.kyoto-u.ac.jp/HARK/

4. **最適な音響処理モジュールの提供と助言**：信号処理アルゴリズムはそれぞれアルゴリズムが有効な前提を置いており，同一機能に対して複数のアルゴリズムを開発し，その使用経験を通じて最適なモジュールを提供，

5. **実時間処理**：音を通じたインタラクションや挙動を行うためには不可欠である．

このような設計思想の下に，オープンソースとして HARK の公開を行ってきた．改良，機能向上，バグフィックスには，hark-support に寄せられたユーザからの声が多く反映されている．

時期	バージョン	主な機能
2008 年 4 月	HARK 0.1.7	オープンソースとして公開開始
2009 年 11 月	HARK 1.0.0 プレリリース	改良，バグフィックス，ドキュメント充実化
2010 年 10 月	HARK 1.0.0 確定版	バグフィックス，周辺ツール提供など
2012 年 2 月	HARK 1.1	機能向上，64bit サポート，ROS サポート，バグフィックス
2013 年 3 月	HARK 1.2	3D 音源定位，Windows サポート，英語音響モデル提供，バグフィックス

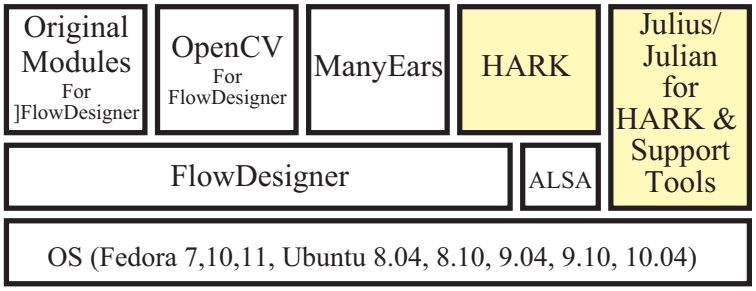


図 1.1: ロボット聴覚ソフトウエア HARK とミドルウエア FlowDesigner，OS との関係

HARK は，図 1.1 に示すように，音声認識部 (Julius) やサポートツールを除き，FlowDesigner [2] をミドルウエアとして用いている．

図 1.1 から分かるように，Linux 系の OS しかサポートされていない．この 1 つの理由は，複数のマルチチャネル A/D 装置をサポートするために ALSA (Advanced Linux Sound Architecture) という API を使用しているためである．最近 PortAudio が Windows 系で利用されるようになっているので，PortAudio を使用した HARK も開発中である．

ミドルウエア FlowDesigner

ロボット聴覚では，音源定位データを基に音源分離し，分離した音声に対して音声認識を行うことが多い．各処理は，アルゴリズムが部分的に置換できるよう複数モジュールで構成する方が柔軟である．このため，効率のよいモジュール間統合が可能なミドルウェアの導入が不可欠である．しかし，統合するモジュール数が多くなると，モジュール間接続の総オーバヘッドが増大し，実時間性が損なわれる．モジュール間接続時にデータのシリアライズを必要とする CORBA (Common Object Request Broker Architecture) のような一般的な枠組みではこうした問題への対応は難しい．実際，HARK の各モジュールでは，同じ時間フレームであれば，同じ音響データを用いて処理を行う．この音響データを各モジュールがいちいちメモリコピーを行って使っていたのでは，速度的にもメモリ効率的にも不利である．

このような問題に対応できるミドルウエアとして，我々は，データフロー指向の GUI 開発環境である FlowDesigner [2] を採用した．FlowDesigner は，CORBA 等汎用的にモジュール統合に用いることが可能な枠組みと比較して，処理が高速で軽い．

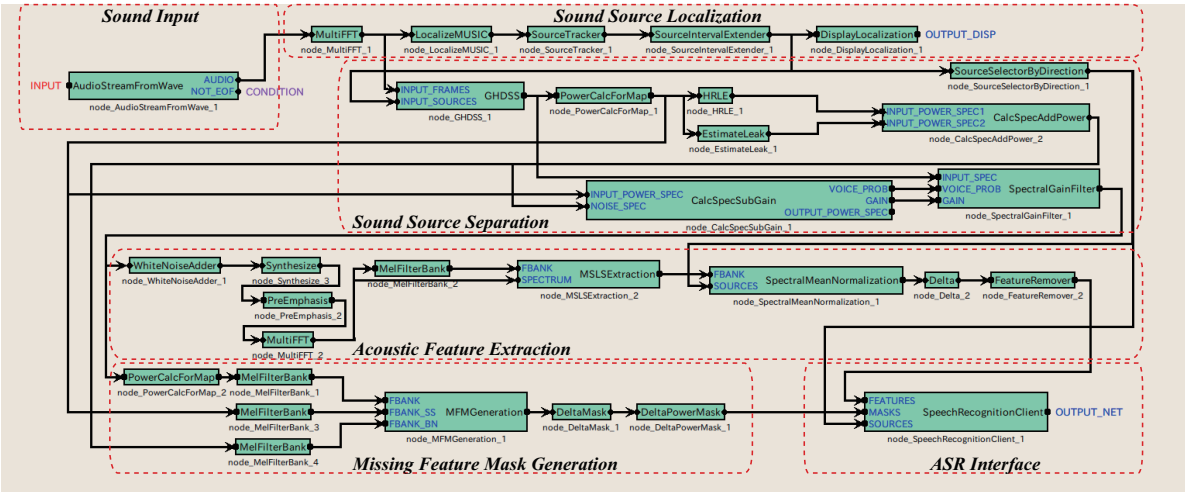


図 1.2: HARK を用いた典型的なロボット聴覚の FlowDesigner 上でのモジュール構成

FlowDesigner は，単一コンピュータ内の利用を前提とすることで[2]，高速・軽量なモジュール統合を実現したデータフロー指向の GUI 開発環境を備えたフリー（LGPL/GPL）のミドルウエアである．FlowDesigner では，各モジュールは C++ のクラスとして実現される．これらのクラスは，共通のスーパークラスを継承するため，モジュール間のインタフェースは自然と共通化される．モジュール間接続は，各クラスの特定メソッドの呼び出し（関数コール）で実現されるため，オーバヘッドが小さい．データは，参照渡しやポインタで受け渡されるため，前述の音響データのような場合でも，高速にかつ少ないリソースで処理できる．つまり，FlowDesigner の利用によって，モジュール間のデータ通信速度とモジュール再利用性の両立が可能である．

我々は，これまでの使用経験に基づき，メモリリーク等のバグに対処するとともに，操作性の向上（主に属性設定部）を図った FlowDesigner も同時に公開している[3]．

HARK を用いた典型的なロボット聴覚に対する FlowDesigner のネットワークを図 1.2 に示す．ファイル入力によりマルチチャネル音響信号を取得し，音源定位・音源分離を行う．得られた分離音から音響特徴量を抽出し，ミッシングフィーチャマスク (MFM) 生成を行い，これらを音声認識 (ASR) に送る．各モジュールの属性は，属性設定画面で設定することができる（図 1.3 は GHDSS の属性設定画面の例）．

HARK で現在提供している HARK モジュールと外部ツールを表 1.1 に示す．次節では，各モジュールの概要をその設計方針とともに説明をする．

入力装置

HARK では複数のマイク (マイクアレイ) をロボットの耳として搭載して処理を行う．ロボットの耳の設置例を図 4 に示す．この例では，いずれも 8 チャネルのマイクアレイを搭載しているが，HARK では，任意のチャネル数のマイクアレイが利用可能である．HARK がサポートするマルチチャネル A/D 変換装置は，下記のとおりである．

- システムインフロンティア社製，RASP シリーズ，
- ALSA ベースの A/D 変換装置，例えば，RME 社製 Hammerfall DSP シリーズ，Multiface AE.

[2]コンピュータをまたいだ接続は，HARK における音声認識との接続部のようにネットワーク接続用のモジュールを作成することで実現可能である．

[3]FlowDesigner のオリジナルは，http://flowdesigner.sourceforge.net/ から，FlowDesigner 0.9.0 の機能向上版は，http://winnie.kuis.kyoto-u.ac.jp/HARK/ からそれぞれダウンロードできる．

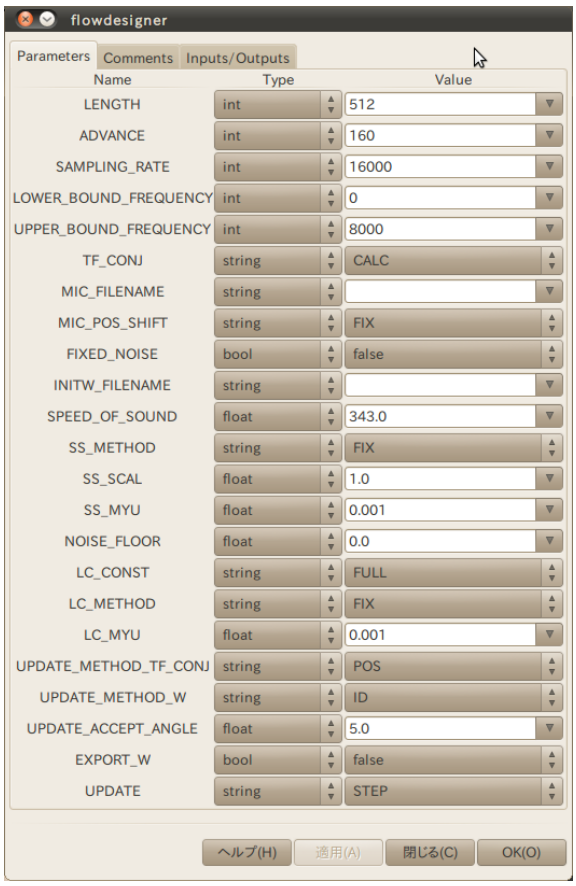


図 1.3: GHDSS の属性設定画面の例

SIG2 (whole body)

Robovie-R2

HRP-2 (8ch)

図 1.4: 3 種類のロボットの耳（マイクロフォン配置）

- Microsoft Kinect
- Sony PS-EYE
- Dev-Audio Microcone

これらの A/D 装置は入力チャネル数が異なるが，HARK での内部パラメータを変更することで対応できる．ただし，チャネル数が増加すれば，処理速度は低下する．また，量子化ビット数は 16 ビット, 24 ビットの両方に対応している．HARK の想定するサンプリングレートは，16kHz であるので，48KHz サンプリングデータに対しては，ダウンサンプリングモジュールが用意されている．なお，東京エレクトロンデバイス社製 TD-BD-16ADUSB（USB インタフェース）は，サポートするカーネルのバージョンが古いため，HARK 1.2 からサポート対象外となっている．

マイクは，安価なピンマイクで構わないが，ゲイン不足解消のため，プリアンプがあった方がよい．RME 社製からは OctaMic II が販売されている．ヤマハ製のマイクロフォンアンプの方が，収録音のノイズが少ないようである．TD-BD-16ADUSB や RASP は，プリアンプおよび，プラグインパワー対応の電源供給機能を有しているので，使い勝手がよい．

表 1.1: Nodes and Tools provided by HARK 1.2

機能	カテゴリ名	モジュール名	説明
音声入出力	AudioIO	AudioStreamFromMic	マイクから音を取得
		AudioStreamFromWave	ファイルから音を取得
		SaveRawPCM	音をファイルに格納
		SaveWavePCM	音を WAV 形式でファイルに格納
		HarkDataStreamSender	音をソケット通信で送信
音源定位・追跡	Localization	ConstantLocalization	固定定位値を出力
		DisplayLocalization	定位結果の表示
		LocalizeMUSIC	音源定位
		LoadSourceLocation	定位情報をファイルから取得
		SaveSourceLocation	定位情報をファイルに格納
		SourceIntervalExtender	追跡結果を前方に延長
		SourceTracker	音源追跡
		CMLoad	相関行列ファイルの読み込み
		CMSave	相関行列ファイルの保存
		CMChannelSelector	相関行列のチャネル選択
		CMMakerFromFFT	相関行列の生成
		CMMakerFromFFTwithFlag	相関行列の生成
		CMDivideEachElement	相関行列の成分ごとの除算
		CMMultiplyEachElement	相関行列の成分ごとの乗算
		CMInverseMatrix	相関行列逆行列演算
		CMMultiplyMatrix	相関行列の乗算
		CMIdentityMatrix	単位相関行列の出力
音源分離	Separation	BGNEstimator	背景雑音推定
		CalcSpecSubGain	ノイズスペクトラム減算＆最適ゲイン係数推定
		CalcSpecAddPower	パワースペクトラム付加
		EstimateLeak	チャネル間リークノイズ推定
		GHDSS	GHDSS による音源分離
		HRLE	ノイズズスペクトラム推定
		PostFilter	音源分離後ポストフィルター処理
		SpectralGainFilter	音声スペクトラム推定
特徴量抽出	FeatureExtraction	Delta	Δ 項計算
		FeatureRemover	項の削除
		MelFilterBank	メルフィルタバンク処理
		MFCCExtraction	MFCC 抽出
		MSLSExtraction	MSLS 抽出
		PreEmphasis	プリエンファシス
		SaveFeatures	特徴量を格納
		SaveHTKFeatures	特徴量を HTK 形式で格納
		SpectralMeanNormalization	スペクトル平均正規化
ミッシングフィーチャマスク	MFM	DeltaMask	Δ マスク項計算
		DeltaPowerMask	Δ パワーマスク項計算
		MFMGeneration	MFM 生成
ASR との通信	ASRIF	SpeechRecognitionClient	ASR に特徴量を送る
		SpeechRecognitionSMNClient	同上，特徴量 SMN 付
その他	MISC	ChannelSelector	チャネル選択
		DataLogger	データのログ出力
		MatrixToMap	`Matrix` → `Map` 変換
		MultiGain	マルチチャネルのゲイン計算
		MultiDownSampler	ダウンサンプリング
		MultiFFT	マルチチャネル FFT
		PowerCalcForMap	`Map` 入力のパワー計算
		PowerCalcForMatrix	行列入力のパワー計算
		SegmentAudioStreamByID	ID による音響ストリームセグメント選択
		SourceSelectorByDirection	方向による音源選択
		SourceSelectorByID	ID による音源選択
		Synthesize	波形変換
		WhiteNoiseAdder	白色雑音追加
機能	カテゴリ	ツール名	説明
データ生成	外部ツール	harktool4	データ可視化・各種設定ファイル作成
		wios	伝達関数作成用録音ツール

1.3 HARK のモジュール群

音源定位

音源定位には，これまでの経験から最も性能が良かった MUltiple SIgnal Classification (MUSIC) 法を提供している．MUSIC 法は，音源位置と各マイク間のインパルス応答 (伝達関数) を用いて，音源定位を行う手法である．インパルス応答は，実測値もしくは，マイクロフォンの幾何的位置を用いて計算により求めることができる．

HARK 0.1.7 では，音源定位として ManyEars [3] のビームフォーマが利用可能であった．このモジュールは，2D 極座標空間 (3D 極座標空間で方向情報が認識できるという意味で「2D」となっている) で，マイクアレイから 5 m 以内，かつ，音源間が 20° 以上離れていれば，定位誤差は約 1.4° であると報告されている．しかし，ManyEars のモジュール全体がもともと 48 kHz サンプリング用に作成されており，HARK で利用している 16 kHz サンプリングと合致しないこと，マイクロフォン配置からインパルス応答をシミュレーションする時にマイクロフォンが自由空間に配置されていることが前提となっており，ロボットの身体の影響を考慮できないこと，MUSIC のような適応ビームフォーマの方が一般的なビームフォーマよりも音源定位精度が高いことなどの理由から HARK 1.0.0 では，MUSIC 法のみをサポートしている．

HARK 1.1 では，MUSIC 法における部分空間に分解するアルゴリズムを拡張した GEVD-MUSIC と GSVD-MUSIC[7] のサポートを新たに行った．本拡張により，既知の雑音（ロボットのファン雑音等）を白色化をした上で音源定位を行うことができ，ロボットの自己雑音を初めとする，大きな雑音下においてもロバストに音源定位ができるようになった．

HARK 1.2 では，さらに 3 次元音源定位を行うことができるように拡張を行った．

音源分離

音源分離には，これまでの使用経験から種々の音響環境で最も総合性能の高い Geometric-Constrained High-order Source Separation (GHDSS) [8]，及び，ポストフィルタ PostFilter とノイズ推定法 Histogram-based Recursive Level Estimation HRLE を HARK 1.0.0 では提供している．現在，最も性能がよく，様々な音環境で安定しているのは，GHDSS と HRLE の組合せである．

これまでに，適応型ビームフォーマ (遅延和型，適応型)，独立成分分析 (ICA)，Geometric Source Separation (GSS) など様々な手法を開発し，評価実験を行ってきた．HARK で提供してきた音源分離手法を下記にまとめる：

1. HARK 0.1.7 で提供した遅延和型ビームフォーマ，
2. HARK 0.1.7 で外部モジュールとしてサポートした ManyEars Geometric Source Separation (GSS) と PostFilter の組合せ [4]，
3. HARK 1.0.0 プレリリースで提供した独自設計の GSS と PostFilter の組合せ [5]，
4. HARK 1.0.0 で提供する GHDSS と HRLE の組合せ [6, 8].

HARK 0.1.7 で利用していた ManyEars の GSS は，音源からマイクへの伝達関数を幾何制約として使用し，与えられた音源方向から到来する信号の分離を行う手法である．幾何学的制約は，音源から各マイクへの伝達関数として与えらると仮定し，マイク位置と音源位置との関係から伝達関数を求めている．本伝達関数の求め方ではマイク配置が同じでもロボットの形状が変わると伝達関数が変わるという状況においては，性能劣化の原因となっていた．

HARK 1.0.0 プレリリースでは，GSS を新たに設計し直し，実測の伝達関数を幾何学的制約として使用できるように拡張し，ステップサイズを適応的に変化させて分離行列の収束を早める等の改良を行った．さらに，GSS の属性設定変更により，遅延和型ビームフォーマが構成できようにもなった．このため，HARK 0.1.7 で提供されていた遅延和型ビームフォーマ DSBeamformer は廃止された．

音源分離一般に当てはまるのだが，音源分離手法の大部分は，ICA を除き，分離すべき音源の方向情報をパラメータとして必要とする．もし，定位情報が得られない場合には，分離そのものが実行されないことになる．一方，ロボット定常雑音は，方向性音源としての性質が比較的強いので，音源定位ができれば，定常雑音を除去することができる．しかし，実際にはそのような雑音に対する音源定位がうまく行かないことが少なからずあり，その結果，定常雑音の分離性能が劣化する場合があった．HARK 1.0.0 プレリリースの GSS および GHDSS には，特定方向に常に雑音源を指定する機能が追加され，定位されない音源でも常に分離し続けることが可能となっている．

一般に，GSS や GHDSS のような線形処理に基づいた音源分離では分離性能に限界があるので，分離音の音質向上のためにポストフィルタという非線形処理が不可欠である．ManyEars のポストフィルタを新たに設計し直し，パラメータ数を大幅に減らしたポストフィルタを HARK 1.0.0 プレリリース版および確定版で提供している．

ポストフィルタは，上手に使えばよく切れる包丁ではあるが，その使い方が難しく，下手な使い方をすれば逆効果になる．ポストフィルタの設定すべきパラメータ数は，PostFilter においても少なからずあるので，それらの値を適切に設定するのが難しい．さらに，ポストフィルタは確率モデルに基づいた非線形処理を行っているので，分離音には非線形スペクトラム歪が生じ，分離音に対する音声認識率の性能がなかなか向上しない．

HARK 1.0.0 では，HRLE (Histogram-based Recursive Level Estimation) という GHDSS に適した定常ノイズ推定法を提供している．GHDSS 分離アルゴリズムを精査して開発したチャンネル間リークエネルギーを推定する EstimateLeak と HRLE とを組み合わせて使用すると，従来よりも音質の向上した分離音が得られる．

MFT-ASR: MFT に基づく音声認識

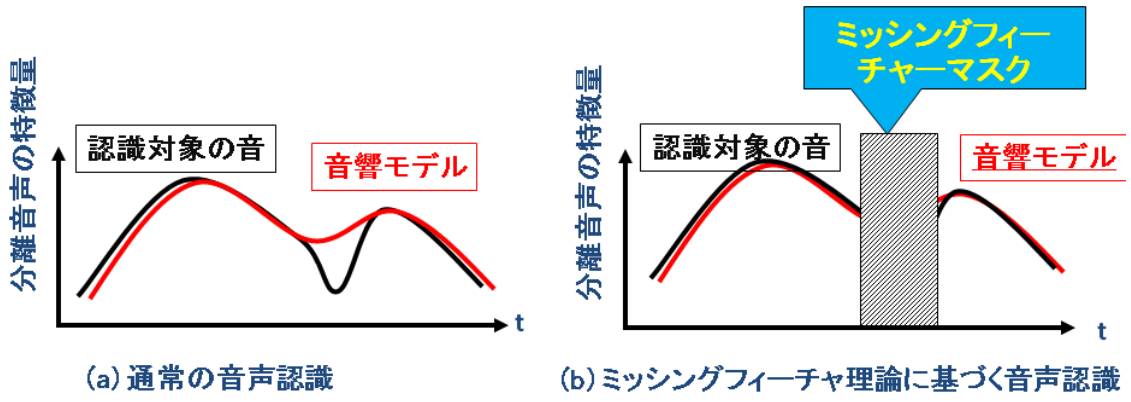


図 1.5: ミッシングフィーチャ理論による音声認識の概念図

混合音や分離など様々な要因によって引き起こされるスペクトル歪は，従来の音声認識コミュニティで想定されている以上のものであり，それに対処するためには，音源分離と音声認識とをより密に結合する必要がある．HARK では，ミッシングフィーチャ理論 (Missing Feature Theory，MFT) に基づいた音声認識 (MFT-ASR) により対処をしている [4].

MFT-ASR の概念を図 1.5 に示す．図中の黒い線は分離音の音響特徴量の時間変化を，赤い線は ASR システムが保持する対応する発話の音響モデルの時間変化を示す．分離音の音響特徴量は歪によりシステムのそれと大きく異なっている箇所がある (図 1.5(a))．MFT-ASR では，歪んでいる箇所をミッシングフィーチャマスク (MFM) でマスクすることにより，歪みの影響を無視する (図 1.5(b))．MFM とは，分離音の音響特徴量に対応する時間信頼度マップであり，通常は 2 値のバイナリーマスク (ハードマスクとも呼ばれる) が使用される．0〜1 の連続値をとるマスクはソフトマスクと呼ばれる．HARK では，MFM はポストフィルタから得られる定常雑音とチャネル間リークのエネルギーから求めている．

MFT-ASR は，一般的な音声認識と同様に隠れマルコフモデル (Hidden Markov Model，HMM) に基づいているが，MFM が利用できるよう HMM から計算する音響スコア (主に出力確率計算) に関する部分に変更を加えている．HARK では，東京工業大学古井研究室で開発されたマルチバンド Julius を MFT-ASR と解釈し直して使用している [13].

HARK 1.0.0 では，Julius 4 系のプラグイン機能を利用し，MFT-ASR の主要部分は Julius プラグインとして提供している．プラグインとして提供したことで，Julius のバージョンアップによる新しい機能を，そのまま利用できる．また，MFT-ASR は FlowDesigner から独立したサーバ/デーモンとして動き，HARK の音声認識クライアントからソケット通信で送信された音響特徴量とその MFM に対し，結果を出力する．

音響特徴量抽出と音響モデルの雑音適用

スペクトル歪を特定の音響特徴量だけに閉じ込めて，MFT の有効性を高めるために，音響特徴量には，メルスケール対数スペクトル特徴量 (Mel Scale Log Spectrum，MSLS) [4] を使用している．HARK では，音声認識で一般的に使用されるメル周波数ケプストラム係数 (Mel-Frequency Cepstrum Coefficient，MFCC) も提供しているが，MFCC では，歪がすべての特徴に拡散するので，MFT との相性が悪い．同時発話が少ない場合には，MFCC を用いて音声認識を行う方が認識性能がよい場合もある．

HARK 1.0.0 では，MSLS 特徴量で，新たに Δ パワー項を利用するためのモジュールを提供する [6]．Δ パワー項は，MFCC 特徴量でもその有効性が報告されている．各 13 次元の MSLS と Δ MSLS，及び，Δ パワーという 27 次元 MSLS 特徴量を使用した方が，HARK 0.1.7 で使用していた MSLS，Δ MSLS 各 24 次元の計 48 次元 MSLS 特徴量よりも性能がよいことを確認している．

HARK では，上述の非線形分離による歪の影響を，少量の白色雑音を付加することで緩和している．クリーン音声と白色雑音を付加した音声とを使ったマルチコンディション学習により音響モデルを構築するとともに，認識音声にも分離後に同量の白色雑音を付加してから 音声認識を行う．これにより，一話者発話では，S/N が -3 dB 程度でも，高精度な認識が可能である [6].

1.4 HARK の応用

我々は，これまでに 2 本のマイクロフォンを使用した両耳聴によるロボット聴覚機能を開発し，３話者同時発話認識を一種のベンチマークとして使用してきた．SIG や SIG2 という上半身ヒューマノイドロボット上でのロボット聴覚では，1m 離れた所から 30 度間隔に立つ３話者の同時発話認識がそれなりの精度で認識が可能となった [16]．しかし，このシステムは事前知識量や事前処理量が多く，どのような音環境でも手軽に使えるロボット聴覚として機能を備えるのは難しいと判断せざるを得なかった．この性能限界を突破するために，マイクロフォンの本数を増やしたロボット聴覚の研究開発を開始し，HARK が開発されたわけである．

したがって，HARK がベンチマークとして使用してきた 3 人が同時に料理の注文をするのを聞き分けるシステムに応用するのは必然であった．現在，Robovie-R2，HRP-2 等のロボット上で動いている．３話者同時発話認識の変形として，３人が口で行うじゃんけんの勝者判定を行う審判ロボットも Robovie-R2 上で開発を行った [17].

また，ロボットの応用ではないが，実時間で取得したデータ，あるいは，アーカイブされたデータに対して，HARK が定位・分離した音を可視化するシステムを開発してきた．音の提示において，多くの環境で正確な「音に気づかない」状況がしばしば見受けられる．この問題を，**聴覚的アウエアネス** (音の気づき) の欠如によるものと捉え，聴覚的アウエアネスを改善するために，音環境理解の支援を行う 3 次元音環境可視化システムを設計し，HARK を用いて実装を行った [18, 19].

1.4.1　3 話者同時発話認識

a) Robovie が注文をたずねる．　b) 3 人が同時に料理の注文を行う．　c) 1.9 秒後に Robovie が注文を反復し，合計金額を答える．

図 1.6: 3 人が料理を同時に注文するのを聞き分ける Robovie-R2

3 話者同時発話認識は，マイクロフォン入力，音源定位，音源分離，ミッシングフィーチャマスク生成，および，自動音声認識の一連の処理により，話者それぞれの発話認識結果を返す．この FlowDesigner でのモジュールネットワークは図 1.2 に示したものである．対話管理モジュールは，

1. ユーザの発話を聞き，注文依頼だと判定すると，次の処理を行う．
2. ロボット聴覚の一連の処理 – 音源定位・音源分離・ポストフィルタ処理・音響特徴量の抽出・ミッシングフィーチャマスク生成 – を行う．
3. 発話人数分の 音響特徴量とミッシングフィーチャマスクを音声認識エンジンに送り，音声認識結果を受け取る．
4. 音声認識結果を分析し，料理の注文である場合には，注文を復唱し，料理の金額の合計額を答える．
5. さらに注文を受け付ける．

音声認識での音響モデルは，不特定話者対象としている．言語モデルは文脈自由文法で記述しているので，文法を工夫すれば,「ラーメン　大盛り」や「ラーメン　ピリ辛　大盛り」,「ラーメン　ライス大盛り」なども可能である．

3 人の実話者全員が話し終えてから認識終了までに従来のファイル経由ベースの処理では，約 7.9 秒を要していたが，HARK の使用により，応答が約 1.9 秒に短縮された[4]．応答が速いため，全員の注文終了後，直ちにロボットがそれぞれの注文を復唱し，合計金額を答えるように感じられる．なお，モジュールの設定にも依存するが，ファイル入力の場合には，発話終了時が明確であるので，発話終了から認識を終え，ロボットが応答を始めるまでの遅延時間は 0.4 秒程度である．

また，復唱の時に，ロボットが発話者の方へ顔を振り向けることも可能である．HRP-2 では挙動付きの応答を行っている．ただし，身振り手振りを入れるとその準備のためにどうしても応答が遅れ，間の抜けた挙動となってしまうので，注意が必要である．

[4] デモは http://winnie.kuis.kyoto-u.ac.jp/SIG/

1.4.2　口ジャンケンの審判

３話者が同時に料理を注文するのは，デモとして不自然であるとのご意見があったので，同時発話が不可欠なゲームを対象とした．ジャンケンを言葉で行う「口ジャンケン」である.「口ジャンケン」の面白さは，相手に顔を見せずにジャンケンができたり，暗闇でもジャンケンができることにあるものの，問題を誰が勝ったのかがすぐに分からないことである．ロボット聴覚機能のついたロボットに，口ジャンケンの審判をさせようと言うわけである [17].

口ジャンケン審判のプログラムは，前述の 3 話者同時発話認識と対話戦略のところだけが異なっている．ジャンケンが正しく発話されたか，つまり，後出しをしたプレーヤはいないか，をチェックしてから，誰が勝ったのか，あるいは，勝負がアイコだったのか，の判定を行い，結果を知らせる．もし，勝負がつかない場合には，再度ジャンケンを行うようにプレーヤに指示をする．(ニュースサイエンティスト誌の記事を参照)

本システムの詳細は，ICRA-2008 の論文 [17] に書かれているので，興味のある方はそちらを参照していただきたい．

1.4.3　CASA 3D Visualizer

一般に，音声は，時間的・場所的空間を共有する人間同士のコミュニケーションメディアとして，根源的な役割を果たしており，我々は様々な環境で音声を通じて情報のやり取りを行っている．しかし，いろいろな音を聴き逃していることも多く，また，録音を高忠実に再生しても，そのような聞き逃しを回避することは難しい．これは，人生のすべてを記録しようというライフログで，音の再生上大きな問題となろう．このような問題の原因の１つは，録音からは音の気づき (アウエアネス) が得られない，すなわち聴覚的アウエアネスの欠如であると考えられる．

高忠実再生技術は，聴覚的アウエアネスを現実世界以上に改善するわけではない．現実世界で聞き分けられないものが，高忠実再生になったから解決できるとは考えられない．実際，心理物理学の観点から人は 2 つ以上の音を同時に認識することは難しい [20] とされており，複数話者など同時に複数の音が発生する時には，音を聞き分けて提示する等の施策が不可欠である．

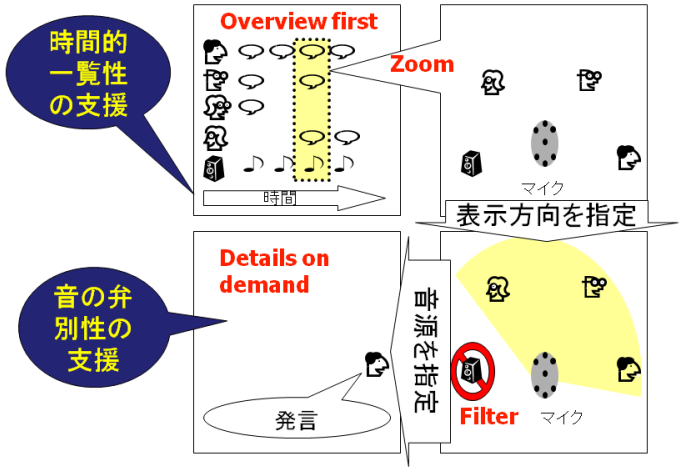


図 1.7: CASA 3D Visualizer: Visual Information-Seeking Matra “Overview first，zoom and filter, then details on demand” に従った HARK 出力の可視化

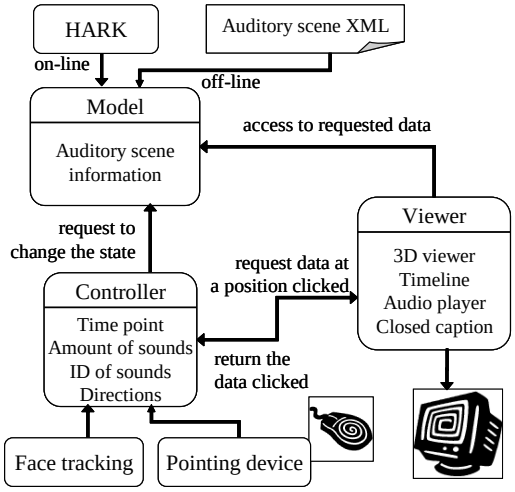


図 1.8: CASA 3D Visualizer の MVC (Model-View-Control) モデルを使用した実装法

我々は，**聴覚的アウエアネス** (音の気づき) の改善にするために，HARK を応用して，音環境理解の支援を行う 3 次元音環境可視化システムを設計し，実装を行った [18, 19]． GUI には Schneiderman が提唱した情報視覚化の指針 “overview first，zoom and filter，then details on demand” (図 1.7) を音情報提示に解釈し直し，以下のような機能を設計した．

1. Overview first: まず概観を見せる．
2. Zoom: ある特定の時間帯を詳しく見せる．
3. Filter: ある方向の音だけを抽出して，聞かせる．
4. Details on Demand: 特定の音だけ聞かせる．

このような GUI により，従来音情報を取り扱う上での課題であった時間的一覧性の支援と音の弁別性の支援の解決を図った．また，実装に関しては，Model-View-Control (MVC) モデルに基づいた設計 (図 1.8) をした．HARK から得られる情報は，まず AuditoryScene XML に変換される．次に，AuditoryScene XML 表現に対して，3D 可視化システムが表示を行う．

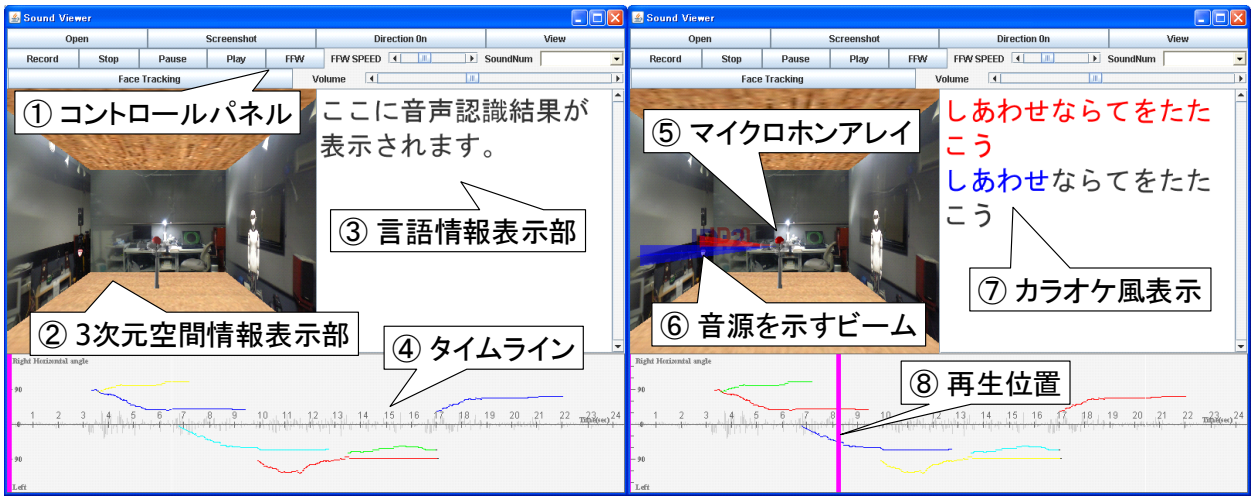


図 1.9: CASA 3D Visualizer の GUI

図 1.9 に表示画面を示す．3 次元空間情報表示では，拡大・縮小，回転が行える．音の再生時には，音源方向を示すビームが ID とともに表示される．また，矢印の大きさは音量の大きさに対応している．言語情報表示部には，音声認識結果が表示される．音声の再生時には対応する字幕がカラオケ風に表示される．タイムラインには，音源の定位の変化の overview 情報が表示され，音の再生時には，再生位置が表示される．表示と音響データとは対応付けが行われているので，ビームあるいはタイムラインの音源をマウスでクリックすると，対応する分離音が再生される．また，再生については早送りモードも提供されている．このように，音情報を見せることにより，聴覚的アウエアネスの改善を試みた．

HARK 出力の可視化のさらなる応用として次のようなシステムも試作されている．

1. ユーザの顔の動きに従って，GUI の表示や音の再生を変更 [18]，
2. Visualizer の結果をヘッドマウントディスプレィ (HMD) に表示 [21]．

上記で説明した GUI は，3D 音環境を鳥瞰する外部観察者のモードである．それに対して，1 番目の応用は，3D 音環境の満真中にいる没入モードの提供である．この 2 つの表示法は，Google Map のアナロジーをとると，

鳥瞰モードと street view モードに相当する．没入モードでは，顔を近づけると音量が大きくなり，顔を遠ざけるとすべての音が聞こえてくる．また，顔を上下左右に移動すると，そちらから聞こえる音が聞こえてくる，等の機能が提供されている．

2 番目の応用は，CASA 3D Visualizer を HMD に表示することで，音源方向を実時間で表示するとともに，その下部には，字幕を表示している．字幕の作成は音声認識ではなく，iptalk という字幕作成用ソフトウエアを使用している．聴覚障害者が字幕を頼りに講義を受ける場合，視線は字幕と黒板の板書をいったりきたりすることになる．これは，非常に負担が大きい上に，話が進んでいることに気がつかずに重要なことを見逃したりする場合が少なからず生ずる．本システムを利用すると，ディスプレイに音源の方向が表示されるので，話題の切り替えへの聴覚的アウエアネスが補強されると期待される．

1.4.4 テレプレゼンスロボットへの応用

2010 年の 3 月に，米国 Willow Garage 社のテレプレゼンスロボット Texai に，HARK と音環境を可視化するシステムを移植し，遠隔ユーザが音源方向をカメラ映像に表示し，特定方向の音源の音だけを聞く機能を実現した[5]．テレプレゼンスロボットでの音情報提示の設計は，前節で説明をした「聴覚的アウエアネスがキーテクノロジである」というこれまでの経験に基づいている．

図 1.10: Texai (中央) を通じて，remote operator が 2 人の話者と， 1 台の Texai とインタラクションを行う．なお，場所はカリフォルニア州であるが， 左側の Texai はインディアナ州から遠隔操作中.

具体的な HARK の移植と Texai への HARK 関連モジュールの開発は次の 2 工程に分けられる．

1. Texai へのマイクロフォン搭載，インパルス応答の測定及び HARK の移植,
2. Texai 制御プログラムが走る ROS (Robot Operating System) への HARK インタフェースとモジュールの実装.

図 1.11 に最初に設置したマイクロフォンの設置状況を示す．このロボットを使用する講義室と大食堂に置き，それぞれ 5 度間隔でインパルス応答を測定し，音源定位の性能を測定した．次に，見栄え，さらには，マイクロフォン間のクロストークを減少させるために Texai に頭を付けることを検討した．具体的には，雑貨店で見つけた竹製のサラダボールである．最初に付けたものとほぼ同じ直径になる辺りに MEMS マイクロフォンを設置した (図 1.11)．同様にインパルス応答を測定し，音源定位性能について評価を行った．その結果，両者の性能はそれほど変わらないことが判明した．

[5] http://www.willowgarage.com/blog/2010/03/25/hark-texai

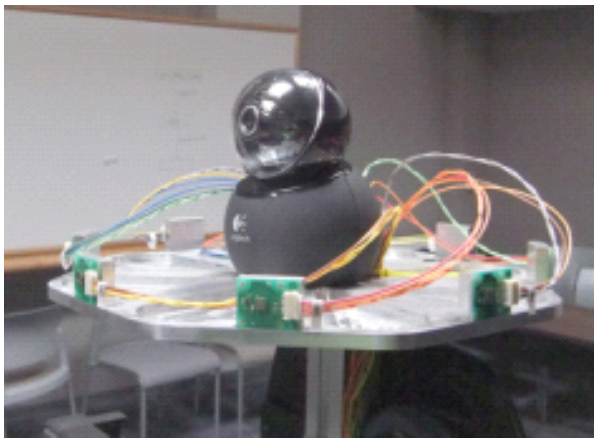


図 1.11: Texai の最初の頭部の拡大: 8 個の MEMS マイクロフォンを円盤上に設置

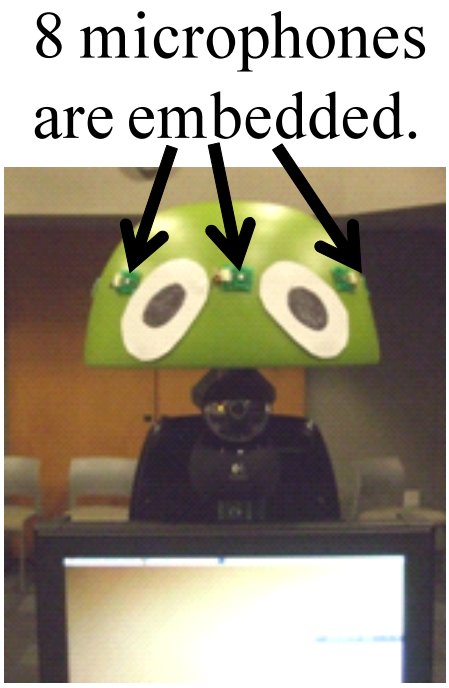


図 1.12: Texai の頭部の拡大: 8 個の MEMS マイクロフォンを円周状に設置

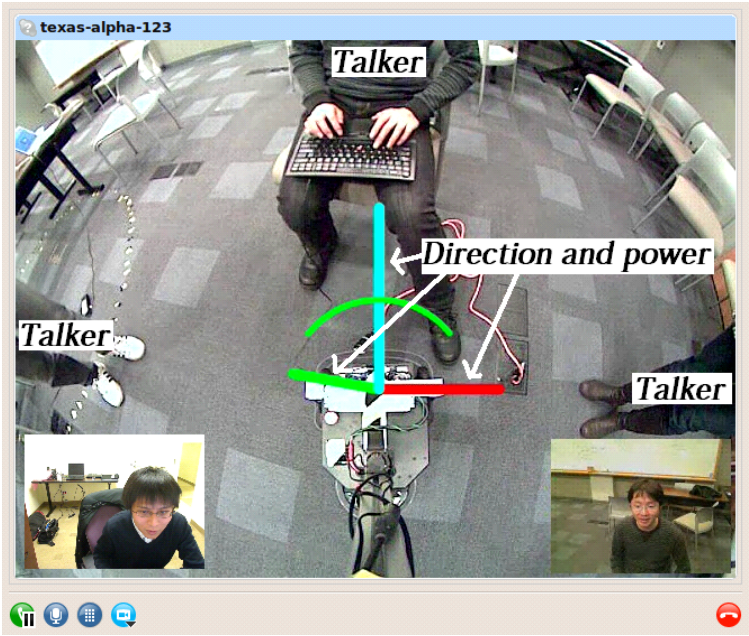


図 1.13: Texai を通じて，remote operator に見える画面

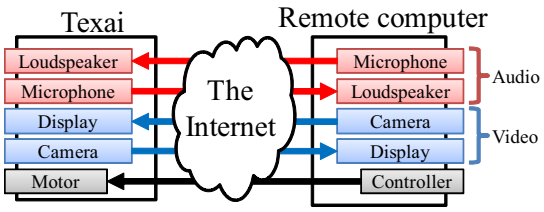


図 1.14: Texai の Teleoperation の方法

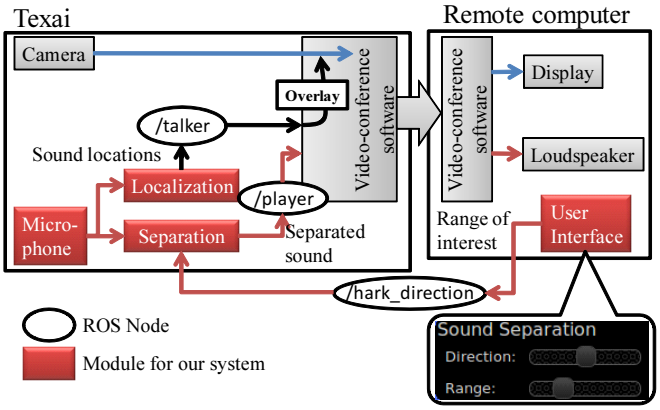


図 1.15: Texai への HARK の組込方法

GUI については，Visual Information-seeking matra の，overview と filter を実装した．図 1.13 に示した Texai 自身の斜め下の全方位の画像の中央から出ている矢印が，話者の音源方向である．矢印の長さは音量を表している．図中では３名の話者がしゃべっていることが分かる．Texai のもう 1 つのカメラの画像が右下に，リモートオペレータの画像が左下に示されている．図中の円弧は，filter で通過させる範囲を示す．この円弧内にある方位から届いた音は，リモートオペレータに送られる．データは図 1.14 に示したように The Internet を通じて行われる．

GUI と，リモートオペレータ用の操作コマンド群はすべて ROS モジュールとして実装されるので，図 1.15 に示した方法で HARK を組み込むようにした．図中の茶色が HARK システムである．ここで開発したモジュールは，ROS の Web サイトから入手可能である．

これら一連の作業は頭部の加工，インパルス応答の測定，予備実験，GUI と操作コマンド群の設計を含めて１週間で終了できた．HARK や ROS の高いモジュール性が，生産性向上に寄与したと考えられる．

1.5　まとめ

以上，HARK 1.0.0 の概要を報告した．ミドルウエア FlowDesigner を使って，音環境理解の基本機能である音源定位，音源分離，分離音認識をモジュールとして実現し，ロボットの耳への応用について概説した．

HARK 1.0.0 は，ロボット聴覚研究をさらに展開するための機能を提供している．例えば，移動音源処理に向けた機能，音源分離の各種パラメータの詳細設定機能，設定データ可視化・作成ツールなどである．また，Windows のサポート，OpenRTM へのインタフェースなども進行中である．

HARK は，ダウンロードし，インストールするだけでもある程度の認識は可能であるものの，個々のロボットの形状や使用環境に合わせたチューニングを行えば，さらに音源定位，音源分離，分離音認識の性能が向上する．このようなノウハウの顕在化には，HARK コミュニティの形成が重要である．本稿がロボット聴覚研究開発者のクリティカルマスを超えるきっかけとなれば幸いである．

第2章　ロボット聴覚とその課題

本章では，HARK の開発のきっかけとなったロボット聴覚研究，およびその課題について述べる．

2.1　ロボット聴覚は聞き分ける技術がベース

鉄腕アトム大事典（沖光正著，晶文社）によると鉄腕アトムには「スイッチひとつで聴力が千倍になり，遠くの人の声もよく聞こえ，さらに 2 千万ヘルツの超音波も聞きとる」サウンドロケータが装備されているという[1]．サウンドロケータは，1953 年に Cherry が発見した選択的に音声を聞き分ける「カクテルパーティ効果」を実現するスーパーデバイスなのであろう．

聴覚障害者や耳の聞こえが悪くなった高齢者からは「スーパーデバイスでなくても，常時同時発話が聞き分けられる機能じゃだめなの」という素朴な疑問がわく．日本書紀推古紀には,「一聞十人訴，以勿失能辨」とあり，同時に 10 人の訴えを聞き分けて裁いたという「聖徳太子」の逸話が紹介されている．動物や草木の言葉が聞こえるという「聞き耳頭巾」の昔話は子供たちの想像力をかき立てる．このような聞き分け機能をロボットに持たせることができれば，人との共生が大きく前進すると期待される．(日本書紀推古紀によれば,「一聞十人訴以勿失能辨兼知未然」豊聡耳厩戸皇子)

日常生活で最も重要なコミュニケーション手段が話声や歌声などを含めた音声であることは論を俟たない．音声コミュニケーションは，言葉獲得，非音声によるバックチャネルなどを包含し，その機能は極めて多彩である．実際，自動音声認識 (ASR，Automatic Speech Recognition) 研究の重要性は高く認識され，過去 20 年以上に渡り膨大な資金と労力が投入された．一方，ロボット自身に装着されたマイクロフォンで音を聞き分け，音声認識をするシステムの研究は麻生らの仕事を除き，ほとんど取り組まれてこなかった．

筆者らの研究スタンスは，事前知識最小の音の処理方式を開発することであった．そのために，音声だけでなく，音楽，環境音，さらにはそれらの混合音の処理を通じて音環境を分析理解する音環境理解の研究が重要であると考えた．この立場から，単一音声入力を仮定する現行の ASR がロボット学で重要な役割を果たせ切れていないことの説明が付く．

2.2　音環境理解をベースにしたロボット聴覚

音声に加えて音楽や環境音さらには混合音を含めた音一般を扱う必要があるという立場から，音環境理解 (Computational Auditory Scene Analysis) [9] 研究を進めてきた．音環境理解研究での重要な課題は，混合音の処理である．話者の口元に設置した接話型マイクロフォンを使用して混合音の問題を回避するのではなく，入力は混合音との立場から，混合音処理に直球で立ち向うのが音環境理解である．

音環境理解の主たる課題は，音源方向認識の**音源定位** (sound source localization)，**音源分離** (sound source separation)，**分離音の音声認識** (automatic speech recognition) の 3 つである．個々の課題に対してはこれまでに多種多様な技術が研究開発されている．しかし，いずれの技術もその能力を最大限発揮するためには何らかの条件を前提としている．ロボット聴覚でこれらの技術を組合せ，能力を最大限発揮させるためには，個別技術のインタフェース，すなわち，前提条件をうまく揃えて，システム化することが不可欠である．このためには，

[1] http://www31.ocn.ne.jp/ goodold60net/atm_gum3.htm

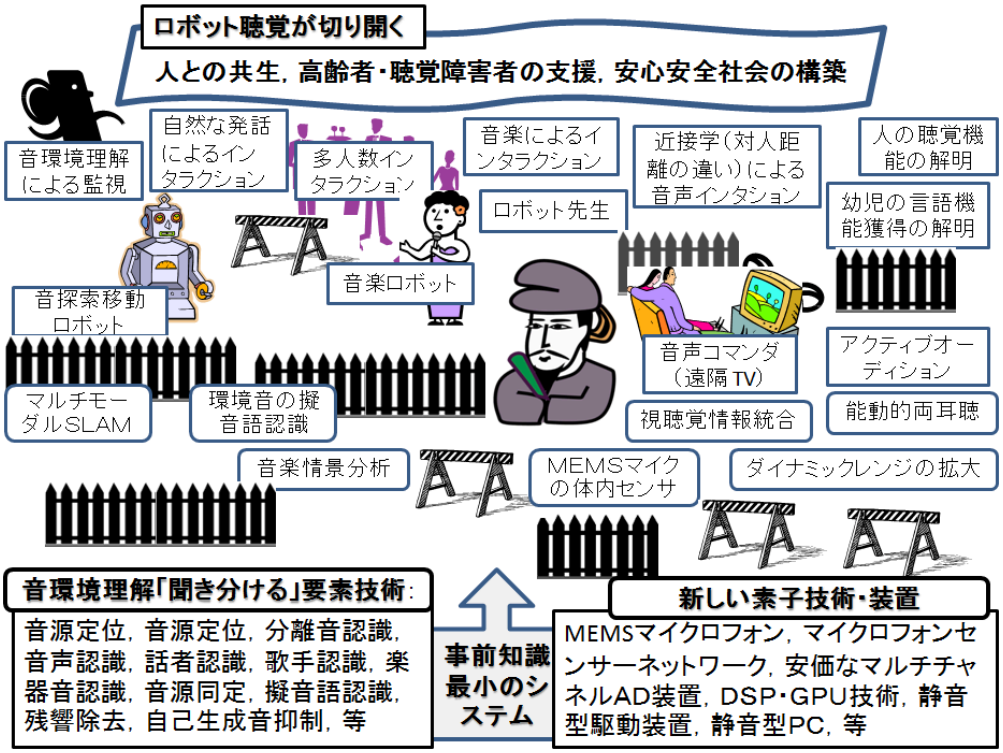


図 2.1: 音環境理解をベースとしたロボット聴覚の展開

ドベネックの桶 (リービッヒの最小律) ではないが，バランスの良い組合せを効率よく提供できるミドルウエアも重要となる．

ロボット聴覚ソフトウエア **HARK** は，FlowDesigner というミドルウエアの上に構築されており，8 本のマイクロフォンを前提として，音環境理解の機能を提供している．HARK は，事前知識を極力減らすという原則で設計されおり，“音響処理の OpenCV” を目指したシステムである．実際，3 人の料理の注文を聞き分けるロボットや口によるじゃんけんの審判ロボットなどが複数のロボットで実現されている．

一般には画像や映像が主たる環境センサとなっているものの，見え隠れや暗い場所には対応できず，必ずしも万能というわけではない．音情報を使って，画像や映像での曖昧性を解消し，逆に，音響情報での曖昧性を画像情報を使って解消する必要がある．例えば，2 本のマイクロフォンによる音源定位では，音源が前か後ろかの判断は極めて難しい．

2.3　人のように 2 本のマイクロフォンで聞き分ける

人や哺乳類は 2 つの耳で聞き分けを行っている．ただし，頭を固定した実験では高々2 音しか聞き分けれないことが報告されている．人の音源定位機能のモデルとしては，両耳入力に遅延フィルタをかけて和を取る Jeffress モデルと，両耳間相互相関関数によるモデルがよく知られている．中臺と筆者らは，ステレオビジョンにヒントを得て，調波構造を両耳で抽出し，同じ基本周波数の音に対して，両耳間位相差と両耳間強度差を求めて，音源定位を行っている [11, 12]．一対の参照点を求めるのに，ステレオビジョンではエピポーラ幾何を使用し，我々の方法は調波構造を使用する．

2 本のマイクロフォンによる混合音からの音源定位では，定位が安定せず大きくぶれることが少なからずあり，また，前後問題，とくに，真正面と真後ろにある音源を区別するのが難しい．中臺らは視聴覚情報統合に

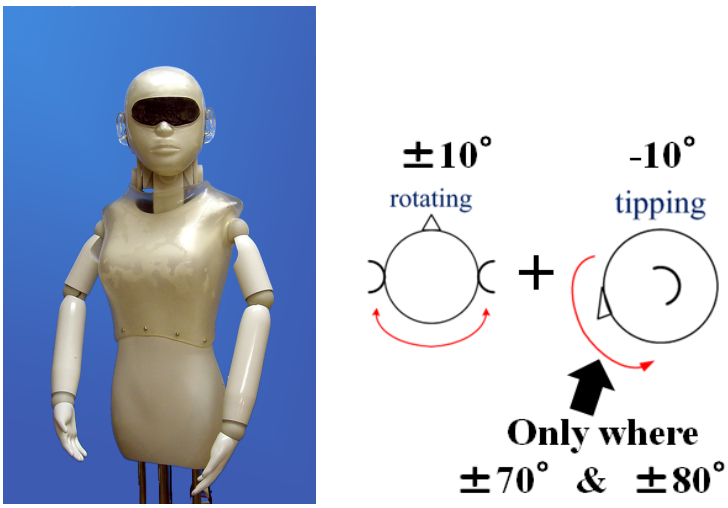


図 2.2: SIG2 のアクティブオーディション：周辺部の音に対しては首を左右と下に動かして前後問題の曖昧性を解消する．

より安定した音源定位を実現するとともに，SIG というロボットで呼びかけられたら振り向くロボットを実現している [14, 15, 27]．前後問題の曖昧性解消は百聞一見に如かず，というわけである．

金と奥乃らは，SIG2 というロボットに頭を動かすことにより音源定位の曖昧性の解消するシステムを実現している．単純に頭を左右に 10 度動かすだけでなく，音源が 70 度〜80 度にある時には，下向きに 10 度頷きを入れるとよい．実際，正面の音源同定では 97.6%と 1.1%の性能向上に過ぎないのに対して，後ろの音源同定では 75.6%と 10%大幅に性能が向上する (図 2.2)．これは， Blauert が “Spatial Hearing” で報告している人の前後問題の解消時の頭の動きとよく一致している．曖昧性の解消のために挙動を用いる方法は**アクティブオーディション**の 1 形態である．

公文のグループや中島のグループは，様々な耳介を用いて頭や耳介自身を動かすことで音源定位の性能向上に取り組んでいる [12]．ちょうど，ウサギの耳が通常は垂れ下がって広範囲な音を聞いており，異常音がすると耳が立ちあがり，特定方向の音を聞くために指向性を高める．このようなアクティブオーディションの実現法の基礎研究である．これが，ロボットだけでなく，様々な動物の聴覚機能の構成的解明に応用できると，新たなロボットの耳の設計開発につながっていくと期待される．とくに，両耳聴は，ステレオ入力装置がそのまま使えるので，高性能の両耳聴機能が実現できると，工学的な貢献が大きいと考えられる．

2.4 自己生成音抑制機能

アクティブオーディションでは，モータが動くことにより発生するモータ自身の音に加えてロボット自身の体の軋みから音が発生することがある．ロボットの動きに伴って発生する音は，小さい音であっても音源がマイクロフォンの近くにあるので，逆 2 乗則から外部の音源と比較して相対的に大きな音となる．

モデルベースによる自己生成音抑制

中臺らはロボット SIG の頭部内部にマイクロフォンを 2 本設置し，自己生成音の抑制を試みている．モータ音や機械音について簡単なテンプレートを持ち，モータの稼働中でテンプレートに合うような音が発生すると，ヒューリスティクスを用いて破壊されやすいサブバンドを破棄する．本手法を用いた理由は，FIR フィルタに基づくアクティブノイズキャンセラでは，左右の耳が別々に処理されるので両耳間位相差を正しく求めること

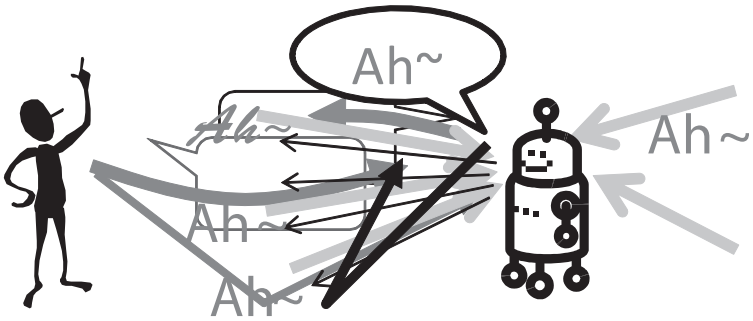


図 2.3: 自分の話声が残響を伴って自分の耳に入り，さらに，相手の割り込み発話 (バージイン) も聞こえる

ができないからであり，さらに，バースト性雑音の抑制に FIR フィルタがあまり効果がなかったからである．なお，SIG2 では，マイクロフォンが人の外耳道モデルに埋め込まれており，モータも静音型かなので，雑音抑制処理は行っていない．ソニーの QRIO でも体内に 1 本マイクロフォンを設置し，外部を向いた 6 本のマイクロフォンを使用して自分の出す雑音を抑制している．

Ince らは，自分の動きから生じる自己生成雑音を，関節角の情報から予測し，スペクトルサブトラクション法により削減する方法を開発している [12]．中臺らは，特定の方向からのモータ雑音を棄却する機能を HARK に組み込んでいる [12]．Even らは，体内に設置した 3 個の振動センサを使って，体表から放射される音の方向を推定し，その放射音方向と話者方向が一致しないように線形マイクロフォンアレイの角度を調節し，自己生成音の抑制を行っている [12]．

ロボットが人とインタラクションを取るときには，自己生成音の影響，環境による音への影響を勘案して，最もよく聞こえる位置に移動したり，体の向きを変えるといった「よりよく聞くための戦略」の開発が不可欠である．

セミブラインド分離による自己生成音抑制機能

ロボット聴覚では，自己発話信号がロボット自身に既知である点を活用した自己生成音抑制が可能である．武田らは，図 2.3 に示した状況において，自己発話を既知として，その残響成分を推定し，入力混合音から自己発話を抑制し，相手の発話を抽出する自己生成音抑制機能を独立成分分析 (ICA) に基づいたセミブラインド分離技術より開発している [12]．本技術の応用のプロトタイプとしてバージイン許容発話認識と音楽ロボット（後述）が開発されている．

バージイン許容発話とは，ロボットの発話中でも人が自由に発話ができる機能である．ロボットが項目を列挙して情報提供を行っているときに，ユーザが割り込んで「それ」「2 番目の」「アトム」と発話すると，本技術を応用して，発話内容や発話タイミングからどの項目が指定されたか従来よりは高性能で判定することができる．人とロボットが共生していくためには，交互に話すのではなく，いついかなる時でもお互いに自由に話すことができる混合主導型のインタラクションが不可欠であり，本自己生成音抑制機能によってそのような機能が容易に実現できる．

セミブラインド分離技術は，自己生成音が耳まで入るが，分離されると捨てられ，高次処理の対象となっていない．本庄の『言葉をきく脳しゃべる脳』によると，成人では自分の声が側頭葉の一次聴覚野までは入るが，大脳皮質の連合聴覚野には送られず，聞き流していることが観測されている．上述のセミブラインド分離による自己生成音抑制は一次聴覚野止まりの処理の工学的実現ととらえることもできよう．

2.5　視聴覚情報統合による曖昧性解消

ロボット聴覚は**要素技術ではなく，プロセスであり，複数のシステムから構成される**．構成部品となる要素技術は多数あり，しかも，構成部品の性能にはばらつきがあるので，プロセスではすべてがうまくかみ合って機能する必要がある．しかも，このかみ合わせがしっかりするほど，プロセスはうまく機能する．音響処理だけでは曖昧性が解消できないので，視聴覚情報統合がかみ合わせの重要な鍵となる．

情報統合のレベルには，時間的，空間的，メディア間，システム間があり，さらに，各レベル内でも，レベル間でも階層的な情報統合が必要である．中臺らは次のような視聴覚情報統合を提案している．最下位レベルでは音声信号と唇の動きから話者を検出する．その上のレベルでは，音素 (phoneme) 認識 と口形素 (viseme) 認識とを統合する．その上位レベルは，話者位置と顔の 3D 位置との統合である．最上位は，話者同定・検証と顔同定・検証との統合である．もちろん，同一レベルの情報統合だけでなく，ボトムアップ処理やトップダウン処理の相互作用が考えられる．

一般に混合音処理は不良設定問題であり，より完全な解を得るためには，何らかの前提，例えばスパースネスの仮定が必要となる．時間領域でのスパースネス，周波数領域でのスパースネス，3D 空間でのスパースネス，さらには特徴空間でのスパースネスなどが考えられる．情報統合の成否は，スパースネスの設計だけでなく，個々の要素技術の性能にも依存することに注意する必要がある．

2.6　ロボット聴覚が切り開くキラーアプリケーション

ロボット聴覚機能が充実しても，それは，個々の信号処理モジュールの統合であり，それからどのような応用が見えてくるのかは明らかでない．実際，音声認識は IT 事業の中でも非常に低い地位しか与えられていない．そのような現状から，本当に不可欠な応用を見つけるためには，まず，使えるシステムを構築し，経験を積んでいく必要があろう．

近接学によるインタラクション

インタラクションの基本原理として，対人距離に基づく近接学 (Proxemics) が知られている．すなわち，親密距離 (〜0.5 m)，個人距離 (0.5 m〜1.2 m)，社会距離 (1.2 m〜3.6 m)，公共距離 (3.6 m〜) に分け，各距離ごとにインタラクションの質が変っている．

近接学に対するロボット聴覚の課題は，マイクロフォンのダイナミックレンジが拡大することである．複数人インタラクションにおいて，個々の話者が同じ音量で話すとすると，遠方の話者の声は逆 2 乗則に従って小さくなる．従来の 16 ビット入力では不足し，24 ビット入力で対応することが不可欠である．システム全体を 24 ビット化するのは，計算資源や既存ソフトウェアとの整合性から難しい．荒井らは，情報欠損の少ない 16 ビットへのダウンサンプリング法を提案している [12]．また，マルチチャネル A/D 装置や携帯電話用 MEMS マイクロフォンなど，新しい装置の出現にも対応していく必要もある．

音楽ロボット

音楽を聴けば自然と体が動き，インタラクションが円滑になるので，音楽インタラクションへの期待は大きい．ロボットが音楽を扱えるようになるには,「聞き分ける」機能が不可欠である．テストベッドとして開発した音楽ロボット処理の流れを示す．

1. 自己生成音を入力音（混合音）から抑制あるいは分離,
2. 分離音のビート追跡からテンポ認識と次テンポ推定,

3. テンポに合わせて挙動（歌を歌う，動作）を実行.

ロボットは，スピーカから音楽が鳴るとすぐにテンポに合わせて足踏みを始め，音楽がなり終わると足踏みを終える.

自分の歌声を残響の影響を含めて入力混合音から分離するのために自己生成音抑制機能を使用している．ビート追跡やテンポ推定では誤りが避けられない．音楽ロボットでは，テンポ推定誤りから生ずる楽譜追跡時の迷子からいかに早く，かつ，スマートに合奏や合唱に復帰するかが重要であり，人とのインタラクションで不可欠な機能となっている.

視聴覚統合型 SLAM

佐々木・加賀美（産総研）らは，32 チャネルマイクロフォンアレイを装着した移動ロボットを開発し，室内の音環境理解の研究開発に取り組んでいる．事前に与えられたマップを使い，いくつかのランドマークをたどりながら定位とマップ作成を同時に行う SLAM (Simultaneous Localization And Mapping) の音響版である [1]. 従来の SLAM では，画像センサ，レーザレンジセンサ，超音波センサなどが使われるものの，マイクロフォン，つまり，可聴帯域の音響信号は使用されてこなかった．佐々木らの仕事は，従来の SLAM では扱えていなかった音響信号を SLAM に組み込む研究であり，重要な先駆的な研究である．これにより，見えないけれども音がする場合にも，SLAM あるいは音源探索が可能となり，真の情景理解 (Scene anaysis) や環境理解への道筋が開かれたことになると考えられる.

2.7 まとめ

ロボットが自分自身の耳で聞くという**ロボット聴覚研究** の筆者の考え方を述べるとともに，今後の展開への期待を述べた．ロボット聴覚研究は，ほとんど 0 からの立ち上げであったために，自分たちの研究だけでなく，当該研究の振興を図るべく浅野 (産総研，以下敬称略)，小林 (早大)，猿渡 (奈良先端大) らのアカデミア，NEC，日立，東芝，HRI-JP などのロボット聴覚を展開する企業，さらには，カナダ Sherbrooke 大学，韓国 KIST，フランス LAAS，ドイツ HRI-EU などの海外研究機関からの協力を得て，IEEE/RSJ IROS でこれまでに 6 年間ロボット聴覚 organized session を組み，ロボット学会学術講演会でも 5 年間特別セッションを組んでいる．さらに，2009 年には IEEE 信号処理部門の国際会議 ICASSP-2009 でロボット聴覚スペシャルセッションを開催した．このような研究コミュニティの育成により，世界的に徐々に研究者が増加し，その中でも日本のロボット聴覚研究のレベルの高さが輝いている．今後斯学の益々の発展を通じ，聖徳太子ロボットが聴覚障害者や高齢者の支援，安心できる社会の構築に寄与していくことを期待したい.

六十而耳順 (「論語・為政」)

60 にして耳に順う，というが，聴覚器官は加齢あるいは酷使されると高域周波数の感度が落ち，人の話が聞こえなくなり，耳に順いたくとも，順えなくなる.

関連図書

[1] 中臺，光永，奥乃 (編): ロボット聴覚特集，日本ロボット学会誌，Vol.28，No.1 (2010 年 1 月).

[2] C. Côté，et al.:Code Reusability Tools for Programming Mobile Robots, *IEEE/RSJ IROS 2004*，pp.1820–1825.

[3] J.-M. Valin，F. Michaud，B. Hadjou，J. Rouat: Localization of simultaneous moving sound sources for mobile robot using a frequency-domain steered beamformer approach. *IEEE ICRA 2004*，pp.1033–1038.

[4] S. Yamamoto，J.-M. Valin，K. Nakadai，T. Ogata，and H. G. Okuno. Enhanced robot speech recognition based on microphone array source separation and missing feature theory. *IEEE ICRA 2005*，pp.1427–1482.

[5] 奥乃，中臺: ロボット聴覚オープンソフトウエア HARK, 日本ロボット学会誌，Vol.28，No.1 (2010 年 1 月) 6–9，日本ロボット学会.

[6] K. Nakadai，T. Takahasi，H.G. Okuno，H. Nakajima，Y. Hasegawa，H. Tsujino: Design and Implementation of Robot Audition System ”HARK”，*Advanced Robotics*，Vol.24 (2010) 739-761，VSP and RSJ.

[7] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, “Intelligent Sound Source Localization for Dynamic Environments”, in *Proc. of IEEE/RSJ Int’l Conf. on Intelligent Robots and Systems (IROS 2009)*, pp. 664–669, 2009.

[8] H. Nakajima，K. Nakadai，Y. Hasegawa，H. Tsujino: Blind Source Spearation With Parameter-Free Adaptive Step-Size Method for Robot Audition, *IEEE Transactions on Audio*，*Speech*，*and Language Processing*, Vol.18，No.6 (Aug．2010) 1467–1485，IEEE.

[9] D. Rosenthal，and H.G. Okuno (Eds.): *Computational Auditory Scene Analysis*，Lawrence Erlbaum Associates，1998.

[10] Bregman，A.S.: *Auditory Scene Analysis – the Perceptual Organization of Sound*, MIT Press (1990).

[11] H.G. Okuno，T. Nakatani，T. Kawabata: Interfacing Sound Stream Segregation to Automatic Speech Recognition – Preliminary Results on Listening to Several Sounds Simultaneously，*Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-1996)*，1082–1089，AAAI，Portland，Aug．1996.

[12] 人工知能学会 AI チャレンジ研究会資料．Web より入手可能: http://winnie.kuis.kyoto-u.ac.jp/AI-Challenge/

[13] 西村 義隆，篠崎 隆宏，岩野 公司，古井 貞熙: 周波数帯域ごとの重みつき尤度を用いた音声認識の検討，日本音響学会 2004 年春季研究発表会講演論文集，日本音響学会，Vol.1，pp.117–118，2004.

[14] Nakadai，K.，Lourens，T.，Okuno，H.G.，and Kitano，H.: Active Audition for Humanoid. In *Proc．of AAAI-2000*，pp.832–839, AAAI，Jul．2000.

[15] Nakadai，K.，Hidai，T.，Mizoguchi，H.，Okuno，H.G.，and Kitano，H.: Real-Time Auditory and Visual Multiple-Object Tracking for Robots, In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-2001)*，pp.1425–1432，IJCAI，2001.

[16] Nakadai, K., Matasuura, D., Okuno, H.G., and Tsujino, H.: Improvement of recognition of simultaneous speech signals using AV integration and scattering theory for humanoid robots, *Speech Communication*, Vol.44, No.1–4 (2004) pp.97–112, Elsevier.

[17] Nakadai, K., Yamamoto, S., Okuno, H.G., Nakajima, H., Hasegawa, Y., Tsujino H.: A Robot Referee for Rock-Paper-Scissors Sound Games, *Proceedings of IEEE-RAS International Conference on Robotics and Automation (ICRA-2008)*, pp.3469–3474, IEEE, May 20, 2008. doi:10.1109/ROBOT.2008.4543741

[18] Kubota, Y., Yoshida, M., Komatani, K., Ogata, T., Okuno, H.G.: Design and Implementation of 3D Auditory Scene Visualizer towards Auditory Awareness with Face Tracking, *Proceedings of IEEE International Symposium on Multimedia (ISM2008)*, pp.468–476, Berkeley, Dec. 16. 2008. doi:10.1109/ISM.2008.107

[19] Kubota, Y., Shiramatsu, S., Yoshida, M., Komatani, K., Ogata, T., Okuno, H.G.: 3D Auditory Scene Visualizer With Face Tracking: Design and Implementation For Auditory Awareness Compensation, *Proceedings of 2nd International Symposium on Universal Communication (ISUC2008)*, pp.42–49, IEEE, Osaka, Dec. 15. 2008. doi:10.1109/ISUC.2008.59

[20] Kashino, M., and Hirahara, T.: One, two, many – Judging the number of concurrent talkers, *Journal of Acoustic Society of America*, Vol.99, No.4 (1996), Pt.2, 2596.

[21] 徳田 浩一，駒谷 和範，尾形 哲也，奥乃 博: 音源定位結果と音声認識結果を HMD に統合呈示する聴覚障害者向け音環境理解支援システム，情報処理学会第 70 回全国大会，5ZD-7，Mar. 2008.

[22] 奥乃 博，中臺 一博: ロボット聴覚の課題と現状，情報処理，Vol.44，No.11 (2003) pp.1138–1144，情報処理学会.

[23] 奥乃 博，溝口 博: ロボット聴覚のための情報統合の現状と課題，計測と制御，Vol.46，No.6 (2007) pp.415–419，計測自動制御学会.

[24] 奥乃 博，山本 俊一: 音環境理解コンピューティング，人工知能学会誌，Vol.22，No.6 (2007) pp.846–854，人工知能学会.

[25] Takeda, R., Nakadai, K., Komatani, K., Ogata, T., and Okuno, H.G.: Exploiting Known Sound Sources to Improve ICA-based Robot Audition in Speech Separation and Recognition, In *Proc. of IEEE/RSJ IROS-2007*, pp.1757–1762, 2007.

[26] Tasaki, T., Matsumoto, S., Ohba, H., Yamamoto, S., Toda, M., Komatani, K. and Ogata, T. and Okuno, H.G.: Dynamic Communication of Humanoid Robot with Multiple People Based on Interaction Distance, 人工知能学会論文誌，Vol.20，No.3 (Mar. 2005) pp.209–219，人工知能学会.

[27] H-D. Kim, K. Komatani, T. Ogata, H.G. Okuno: Binaural Active Audition for Humanoid Robots to Localize Speech over Entire Azimuth Range, *Applied Bionics and Biomechanics*, Special Issue on ”Humanoid Robots”, Vol.6, Issue 3 & 4(Sep. 2009) pp.355-368, Taylor & Francis 2009.

第3章　はじめての HARK

この章では，はじめて HARK を使う人を対象に，ソフトウェアの入手方法，インストール方法について述べ，基本的な操作について説明する．

3.1　ソフトウェアの入手方法

HARK の Web サイト (http://winnie.kuis.kyoto-u.ac.jp/HARK/) に入手方法が解説されている。Windows 版はこの URL からインストーラをダウンロードできる。Linux 版はこの URL にある手順に従ってリポジトリを登録し、インストールを行う。リポジトリを登録すればソースコードのダウンロードもできる。

はじめてインストールする人は，パッケージファイルをダウンロードして，インストールする方法を強く推奨する．　ソースコードをダウンロードして，インストールする方法は，上級者向けであり，本ドキュメントでは扱わない．

3.2　ソフトウェアのインストール方法

3.2.1　Linux 版のインストール方法

本項の説明で，インストール完了までの説明に，すべて作業例を示す．行頭の > はコマンドプロンプトを表す．作業例の太字の部分は，ユーザの入力を，イタリック部分は，システムからのメッセージを表す．例えば，

```
> echo Hello World!
 Hello World!
```

という作業例で，1 行目の先頭 > は，コマンドプロンプトを表している．作業環境によってプロンプトの表示が異なるので，各自の環境に合せて読み換える必要がある．1 行目のプロンプト以降の太字部分は，ユーザが実際に入力する部分である．ここでは， echo Hello World! の 17 文字（スペースを含む）がユーザの入力部分である．行末では，Enter キーを入力する．2 行目の斜字体部分は，システムの出力である． 1 行目の行末で Enter キーの入力後，表示される部分である．

ユーザの入力やシステムからのメッセージの一部には，バージョン番号やリリース番号が含まれている．そのため，実際にインストールするバージョンやリリースに応じて，読み換えて作業を進める必要がある．また，具体的な作業例で表示されるシステムからのメッセージは，オプションでインストール可能なライブラリの有無により，異なる．メッセージの内容が完全に一致しなくてもエラーメッセージが表示されない限り，作業を進めてよい．

For Ubuntu

Ubuntu 12.04 使用者は，パッケージからのインストールを利用できる．パッケージの配布サーバを設定ファイルに加えた後に，パッケージのインストールを行う．リポジトリへの追加方法は HARK installation instructions のページを参照。
次にパッケージのインストールを行う． 端末で以下のコマンドを実行する．

```
> sudo apt-get update
> sudo apt-get install harkfd harktool4 julius-4.2.2-hark-plugin
```

その他の環境を使っている場合は、ソースコードからコンパイルする必要がある。ここでは扱わないので、次のページ HARK installation instructions でソースコードを入手し、コンパイルを行う。

3.2.2 Windows 版のインストール方法

Windows ではインストーラーによりインストールを行う．このインストーラーでは，次のソフトウェアがインストールされる。

1. FlowDesigner for HARK
2. hark-fd
3. HARK-Julius
4. HARK-WIOS

`HARK_1.2.0_setup.exe` を実行する．Windows 8 でインストールする際には，インストーラーのアイコン上で右クリックし「管理者として実行」を選択する．インストーラーが起動すると，FlowDesigner for HARK のインストールを開始される．ライセンスが表示されるので、それを読み、「使用許諾契約の条項に同意します」を選択すると「次へ」ボタンが表示されるのでクリックして次へ進みインストールが始まる．

FlowDesigner for HARK のインストールが完了するとインストールの続行画面が表示されるので,「続行」をクリックすると hark-fdf のインストーラーが起動する．

同様に使用許諾契約の条項が表示され，続いて hark-fd がインストールされる．同様に，HARK Julius，HARK-WIOS のインストールが完了するとインストール完了となる．

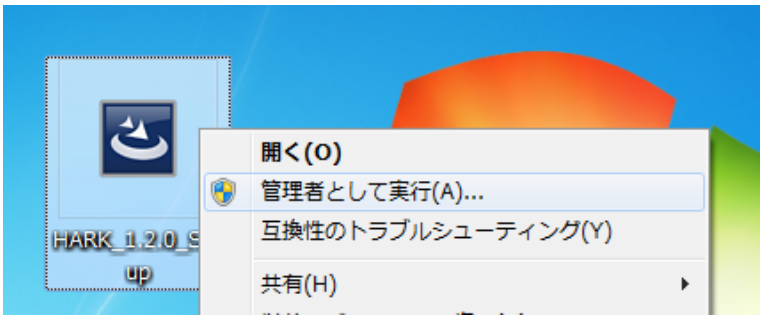


図 3.1: 管理者として実行

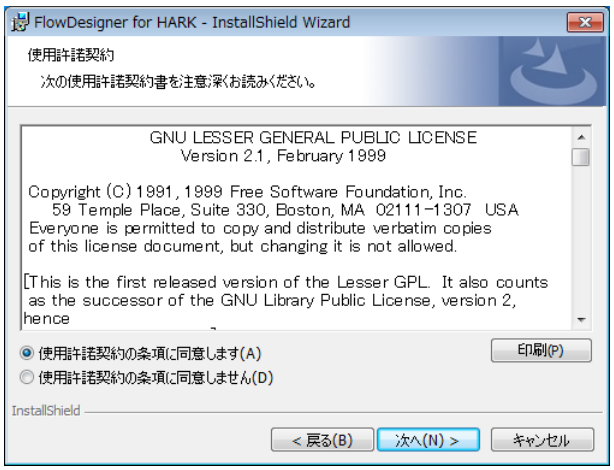


図 3.2: 使用許諾契約画面

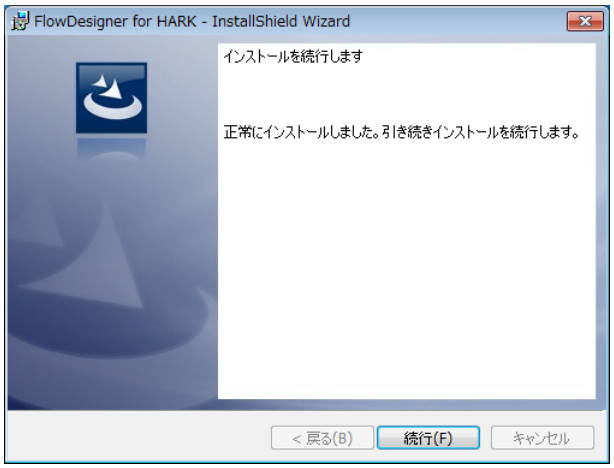


図 3.3: インストールの続行

3.3 FlowDesigner

FlowDesigner は，オープンソースのミドルウエアの一つである．現在 HARK がベースにしているミドルウエアは，FlowDesigner のみである．今後，他のミドルウエア上への実装も視野に入っている．

FlowDesigner の特徴は， Graphical User Interface (GUI) を通じてシステムを構築できる点である．simulink や LabView に見られるグラフィカルプログラミングスタイルが採用されており，プログラム作成経験の少いユーザも容易にプログラム作成可能である．FlowDesigner でのプログラム作成は，ノードの配置と結線，ノードのプロパティ値の設定によって行われる．

以下では、 3.3.1 節 で Linux 版 FlowDesigner, 3.3.2 節 で Windows 版 FlowDesigner の使い方を解説する。Windows 版の解説では Linux 版との差だけを解説するので、適宜 Linux 版を参照されたい。

3.3.1 Linux 版

図 3.4 に FlowDesigner の概観を示す．以下の作業で FlowDesigner を起動できる．

```
> flowdesigner
```

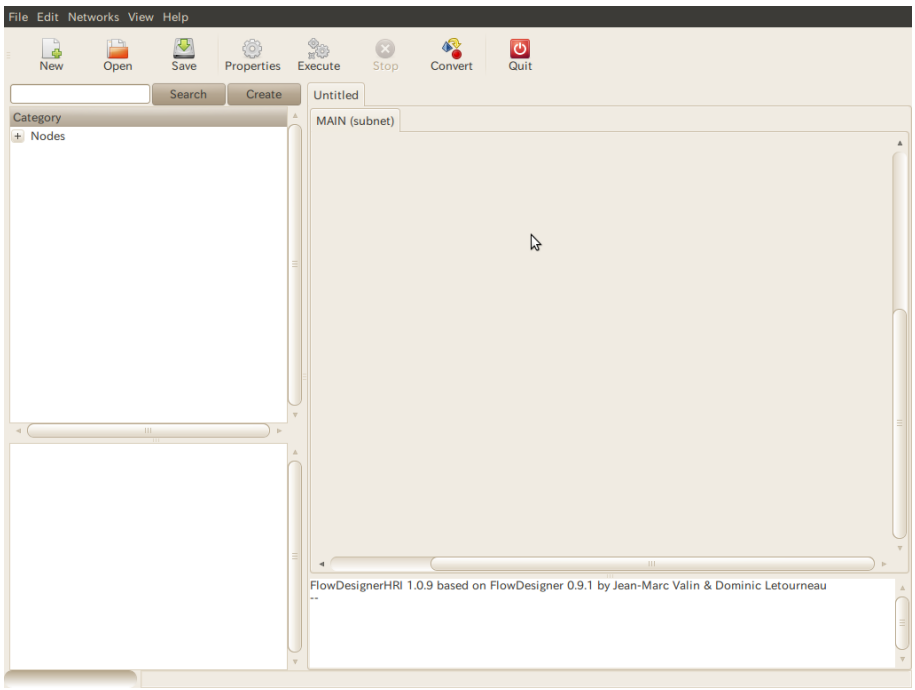


図 3.4: FlowDesigner の概観

FlowDesigner の基本操作

ノードは，データを処理する処理単位である．最初にノードの例を図 3.5 に示す．長方形で囲まれた緑色の部分が，ノードである．このノードの中央には，ノード名が表示される．この例では，PostFilter である．ノード下に表示される名前 node_Postfilter_1 は，ノードのインスタンス名である．同種のノードが複数ある場合には，ノードのインスタンスを生成した順番に node_Postfilter_2, node_Postfilter_3 と名付けられる．個々のノードを区別するのに役立つ．

ノードは，処理するデータの入力と出力の口になるターミナルをそれぞれ複数もつことが可能である．ノードの左右にある黒い点がターミナルである．左側のターミナルが入力ターミナルで，右側のターミナルが出力ターミナルである．ターミナルには名前が付けられている場合と名前が付けられていない場合がある．この例では，全てのターミナルに名前が付けられている．ノードは，少なくとも 1 つの出力ターミナルがある．入力ターミナルは 1 つも無いこともある．例えば FlowDesigner に標準で含まれる **constant** ノードは，入力がなく出力ターミナルを 1 つ持つ．

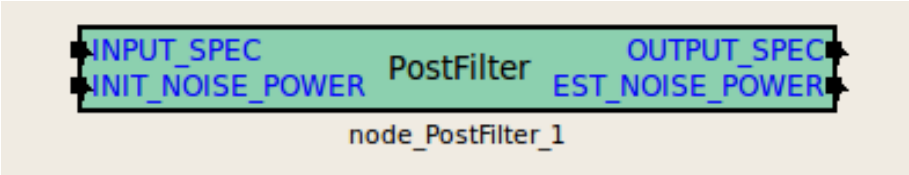


図 3.5: ノードの概観

複数のノードを配置し，ノードのターミナル同士を接続することによって，データの処理の流れを定義する．接続されてた一連のノードをネットワーク呼ぶ．ネットワークを作成すること自体が，プログラム作成になっている．ネットワークは，ファイルの保存メニューから保存できる．一度保存すれば，いつでも FlowDesigner

でロードできる．保存したファイルをテキストエディタで開くと XML ライクな記述を確認できる．慣れればテキストエディタで直接ネットワークを編集することもできるようになる．

ノードには，プロパティをもつものがあり，プロパティ値を設定することでノードの処理を詳細に指定可能である．プロパティ値の設定には，ノード上にマウスカーソルを移動させ，左ダブルクリックする．設定ダイアログウィンドウが開き，各種プロパティ値をキーボードで入力あるいは選択できる．

プロパティ値には値のデータ型あり，データ型と値をペアで設定する必要がある．データ型の詳細は，データ型の章を参考のこと．データ型は，プルダウンメニュー形式の中から選択する．値は，キーボードから値を入力する場合と，プルダウンメニュー形式で値を選択する場合がある．

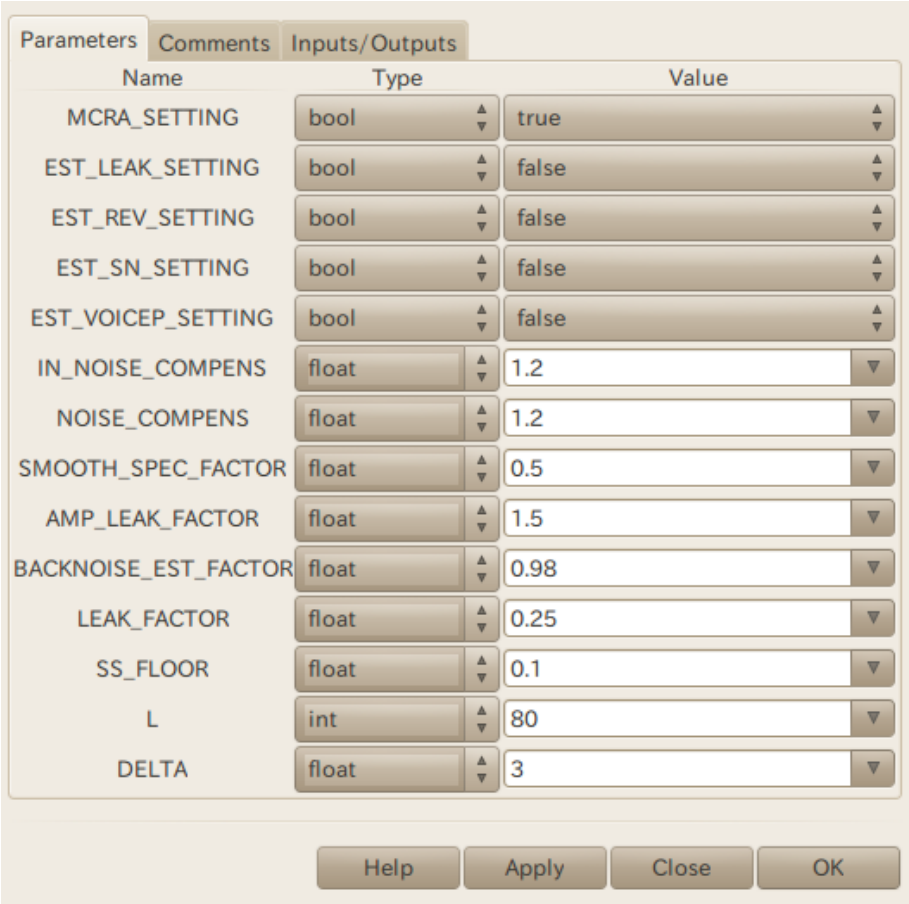


図 3.6: プロパティの概観

ノードの操作・ターミナルの接続・プロパティの設定

以下では，具体的なノードの操作方法，ターミナルの接続方法，プロパティの設定方法を述べる．

- **ノードの作成**
  ネットワーク構築タブでマウスの右ボタンをクリックすると，ノードのカテゴリ一覧が表示される．目的のカテゴリ名の上にマウスカーソルを移動させると，ノードが表示される．そのまま，目的のノード名の上にマウスカーソルを移動させ，左ボタンをクリックすると，ノードが作成される．カテゴリ一覧や，ノード一覧が表示された状態から，ノードの配置を中止する場合には，ネットワーク構築タブ内の地の部分を（左右どちらのボタンでも可）クリックする．

- **ノードの再配置**
  一度配置したノードを見やすい位置に再配置するには，ノードの上にマウスカーソルを移動し，左ボタンでドラッグする．

- **ノードの削除**
  ノードの上にマウスカーソルを移動し，右ボタンをクリックすると，ノードに対する操作メニューが表示される．Delete と表示された部分にマウスカーソルを移動し，左クリックするとノードを削除できる．

- **ノードの複製**
  Shift キーを押下したまま，ノード上で左クリックすると，そのノードを複製できる．属性値を含め，複製ノードが生成できるため，同様の処理を並列・並行して処理するネトワークを構築する場合に便利な機能である．

- **ノードの接続**
  ノード同士は，ターミナルを矢線で結線し，接続する．接続元のターミナル上で左ボタンを押下し，ドラッグして接続先のターミナルでボタンを放すとターミナル間が矢線で結線される．これで，ノードの接続が完了である．

  入力と出力ターミナルの接続が可能で，入力ターミナル同士や出力ターミナル同士を接続することはできない．１つの出力ターミナルから複数の接続を引出すことが可能である．ノードの処理結果が，複数のノードに送られる．一方，１つの入力ターミナルで複数の接続を受け入れることはできない．入力と出力ターミナルの接続でも，接続できないことがある．ターミナルが処理できるデータ型に対応していない場合は接続できず，矢線が赤色に表示される．ターミナルが処理できるデータ型の自動判定に失敗し，矢線が黒色であるにもかかわらず，プログラム実行時にエラーが起る場合もある．データ型の自動判定は，補助的なものと考え，正確な接続には，ノードリファレンスで，対応するデータ型を確認する方がよい．

- **ノードの切断**
  切断したい接続の矢線の始点または終点で，Shift キーを押下しながら，左クリックすることで接続を切断できる．

- **ノード接続線の再配置**
  ノードを移動させると，接続を維持したままノードが移動する．多数のノードを配置すると，ノードと接続線が重なり，見ずらいネットワークになる．そこで，線を曲，重なりの少い見やすいネットワークにするとよい．接続線の途中の部分の上にマウスカーソルを移動させ，そこでマウス左ボタンでドラッグすると，接続線がその点で折れ曲る．曲げた点上でマウス左ボタンでドラッグすると，点を移動できる．

- **プロパティ値の設定**
  ノードの上にマウスカーソルを移動し，右ボタンをクリックすると，ノードに対する操作メニューが表示される．Properties と表示された部分にマウスカーソルを移動し，左クリックするとプロパティ値の設定ダイアログが開く．ノードリファレンスを参考にして値とそのデータ型を入力する．Apply ボタンをクリックすれば設定が反映される． OK ボタンをクリックすれば設定が反映され，かつダイアログが閉じる．

以上で，ノードの基本操作の説明を終え，実際に音源定位ネットッワークを構築しながら，ネットワーク構築に必要な基本操作を述べる．

はじめてのネットワークの作成

はじめてネットワークを作成する人を対象に，音源定位ネットワークを構築する例を示す．メインネットワークタブの他に Iteration 用のサブネットワークタブが必要になる．まず，メインタブで音源波形読込部分を構築し，続いて，サブネットワークを作成し，サブネットワーク内に，音源定位部，結果表示部を構築する．

音源波形読込部分のノード配置，接続，プロパティの設定について解説する．FlowDesigner を起動すると，ウィンドウ内右側に MAIN (subnet) というタブが現れる．始めに，このタブの中で作業を進める．Constant ノードと， InputStream ノードを図 3.7 のように配置する．タブ内の地の部分を右クリックすると，プルダウンメニューが現れる．その中の New Node 上にマウスカーソルを移動すると， FlowDesigner に登録されているノード一覧が表示される．ここでは， General カテゴリを選ぶ．マウスカーソルをそのまま General という表示の上まで移動さる．ここで，プルダウンメニューが 1 段展開され，General カテゴリに登録されているノードの一覧が表示される．この中に Constant ノードがある．マウスカーソルをそのまま Constant という表示の上まで移動させると，MAIN (subnet) タブに Constant ノードを配置できる．以後，この様にプルダウンメニューからノードを選択する操作を，簡略化して,「New Node → General → Constant」と表記する．同様に，InputStream ノードを New Node → IO → InputStream と辿り，ノードを配置する．

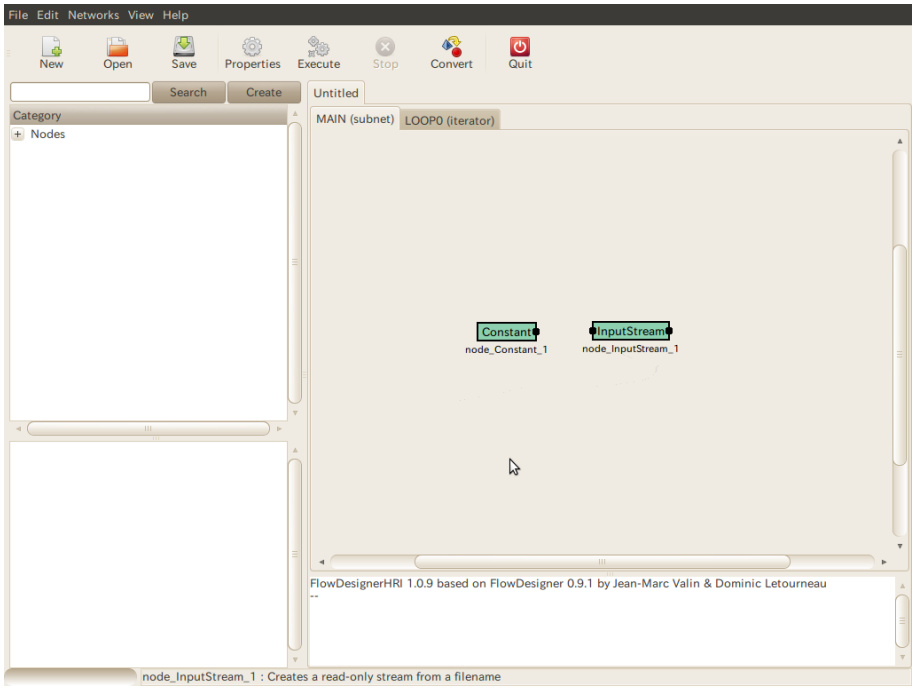


図 3.7: 2 つのノードを配置した様子

次に Constant ノードと， InputStream ノードを図 3.8 のように接続する．

次に ノードのプロパティの設定を行う．ここでは， Constant ノードのみ設定する．InputStream ノードには設定可能なプロパティ値がない．

Constant ノードを右クリックすると，プルダウンメニューが現れる．Properties という表示の上にマウスカーソルを移動させ Properties を左クリックすると，プロパティ設定ダイアログが開く．ダイアログには， Parameters タブ，Comments タブ， Inputs/Ouptuts タブがある．デフォルトでは Parameters タブが開かれている．Comments タブ， Inputs/Ouptuts タブを使用するこはない．

Parameters タブ内には，NAME，Type，Value という 3 つの項目がある．NAME は，プロパティ名であり，Type は，そこに設定する値のデータ型を表す，Value は，属性に指定する値である．Constant ノードには，設

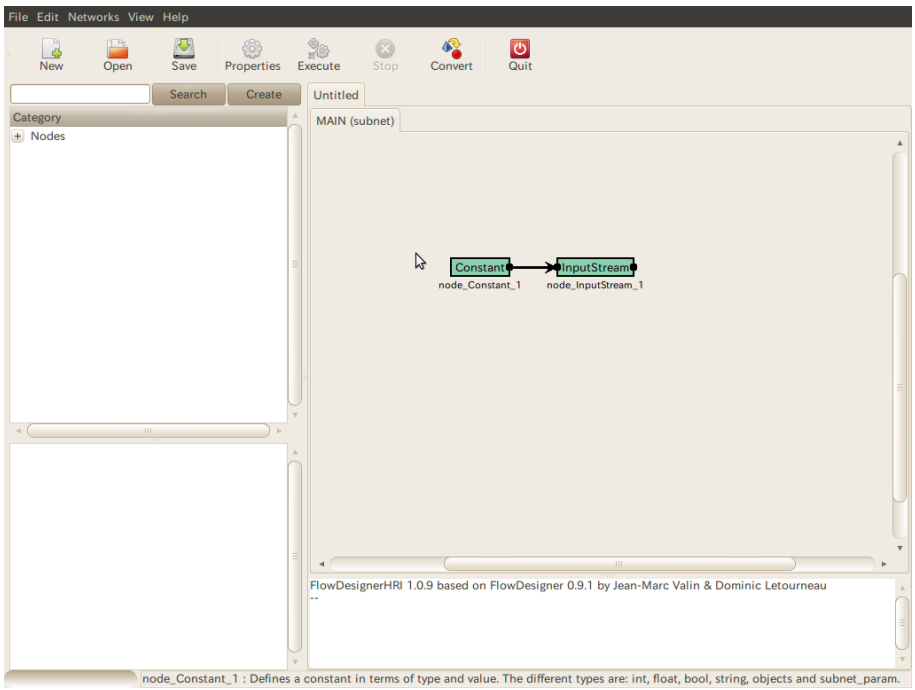


図 3.8: 2 つのノードを接続した様子

定できるプロパティは 1 つしかない．VALUE という属性名である．このノードで，ファイル名を表したいので，プロパティの値にファイル名を tutorial1.wav とする．データ型は `string` である．最後にプロパティウンドウの適用を押すことで設定が反映される．閉じるを押すと設定内容が破棄される．OK を押すと適用を押した後に，閉じると押した動作と等価な処理が行われる．プロパティウィンドウ表示中は，FlowDesigner の別のウィンドウの操作が無効状態になっているので，操作へ復帰するには，プロパティ設定を終える必要がある．以上でプロパティの設定は完了である．これで音源波形読込部分のネットワークが完成した．

続いて，サブネットワーク作成について述べる．サブネットワークは， FlowDesigner のウィンドウメニュー，Networks から add Iterator を選択して作成する．add Iterator を選択すると，ダイアログウィンドウが起動し，追加する Iterator タブの名前の入力が促される．デフォルトの LOOP0 でよければ OK ボタンを押す．キャンセルボタンを押すとサブネットワーク作成を中止する．

LOOP0 で OK すると，MAIN (subnet) タブの隣に，LOOP0 というタブが現れ， LOOP0 タブがアクティブになる．このタブの中に音源定位部と結果表示部を構築する．

音源定位部を，ファイルの読込，FFT，MUSIC 法による音源定位，音源追従，定位結果の表示のためのノードで構築する．以下の手順でノードを配置すると，図 3.9 のようになる．

```
New Node -> HARK -> AudioIO      -> AudioStreamFromWave
New Node -> HARK -> MISC         -> MultiFFT
New Node -> HARK -> Localization -> LocalizeMUSIC
New Node -> HARK -> Localization -> SourceTracker
New Node -> HARK -> Localization -> DisplayLocalization
```

これらのノードを図 3.10 の様に接続する．

接続が終ったら，ノードのプロパティを設定する．処理する音声ファイルが 16000 [Hz] であることを前提に設定する．AudioStreamFromWave ノードには，LENGTH，ADVANCE，USE_WAIT の 3 個所のプロパティ値の設定場所がある．LENGTH と ADVANCE は，音声の分析フレーム長とフレームシフト長を単位サンプルで表している．データ型は `int` である．デフォルト値 512，160 が設定されており，それぞれ 32 msec， 10 msec に

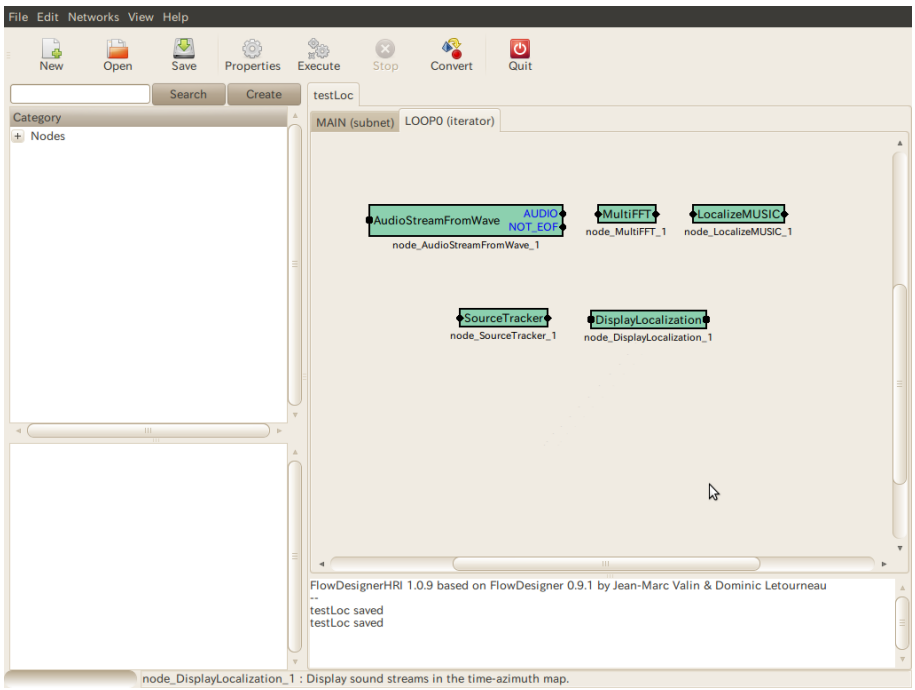


図 3.9: 音源定位部分のノードを配置した様子

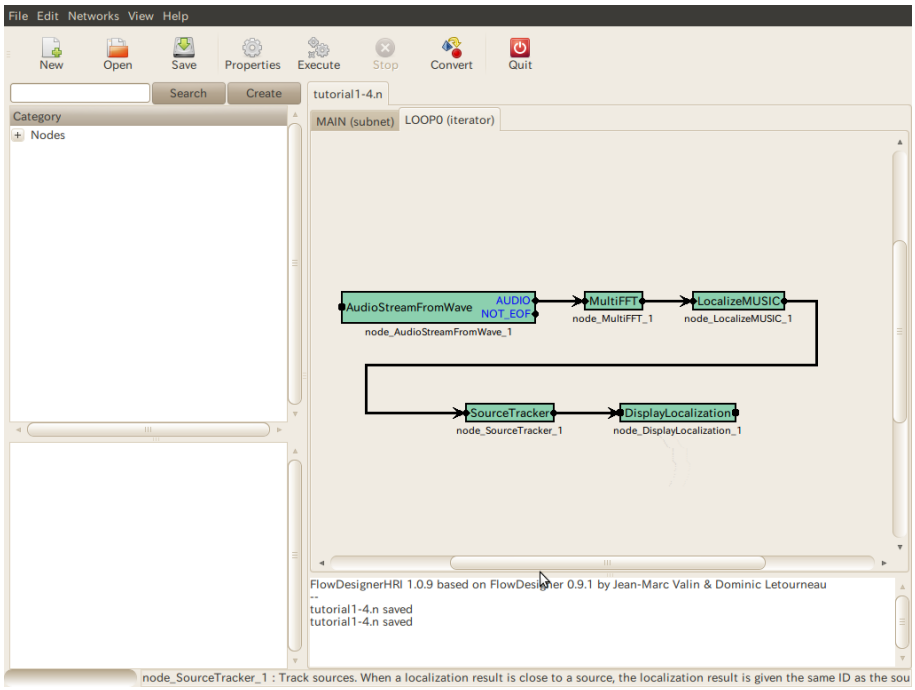


図 3.10: 音源定位部分のノードを接続した様子

対応している．変更する必要はない．データ型 bool の USE_WAIT にもデフォルト値が設定されており，false が設定されている．変更する必要はない．

MultiFFT ノードには，LENGTH，WINDOW，WINDOW_LENGTH の 3 個所のプロパティ値の設定場所がある．音声の分析フレーム長と窓関数の種類，窓の長さを単位サンプルで表している．デフォルト値 512，CONJ，

512 が設定されており，LENGTH と WINDOW_LENGTH は， 32 msec に対応し，データ型は， `int` である．CONJ は，分析窓に CONJ 窓を使うことを指定している．データ型は，`string` である．これらの値を変更する必要はない．

LocalizeMUSIC ノードは，MUSIC_ALGORITHM, TF_CHANNEL_SELECTION, LENGTH, SAMPLING_RATE, A_MATRIX, ELEVATION, WINDOW, PERIOD, NUM_SOURCE, MIN_DEG, MAX_DEG, LOWER_BOUND_FREQUENCY, UPPER_BOUND_FREQUENCY, SPECTRUM_WEIGHT_TYPE, A_CHAR_SCALING, MANUAL_WEIGHT_SPLINE, MANUAL_WEIGHT_SQUARE, ENABLE_EIGENVALUE_WEIGHT, DEBUG の 19 個所のプロパティ値の設定場所がある．各パラメータの詳細は LocalizeMUSIC のパラメータの項を参照されたい．以下は各パラメータの設定例を示す．

- **MUSIC_ALGORITHM** : SEVD
- **TF_CHANNEL_SELECTION** : `<Vector<int> 0 1 2 3 4 5 6 7>`
- **LENGTH** : 512[samples]
- **SAMPLING_RATE** : 16000[Hz]
- **A_MATRIX** : 用意した伝達関数のファイル名
- **ELEVATION** : 16.7[deg]
- **WINDOW** : 50[frames]
- **PERIOD** : 50[frames]
- **NUM_SOURCE** : 2[sources]
- **MIN_DEG** : -180[deg]
- **MAX_DEG** : 180[deg]
- **LOWER_BOUND_FREQUENCY** : 500[Hz]
- **UPPER_BOUND_FREQUENCY** : 2800[Hz]
- **SPECTRUM_WEIGHT_TYPE** : Uniform
- **A_CHAR_SCALING** : 非表示
- **MANUAL_WEIGHT_SPLINE** : 非表示
- **MANUAL_WEIGHT_SQUARE** : 非表示
- **ENABLE_EIGENVALUE_WEIGHT** : true
- **DEBUG** : false

SourceTracker ノードは，THRESH，PAUSE_LENGTH，MIN_SOURCE_INTERVAL，DEBUG の 4 個所のプロパティ値の設定場所がある．

- **THRESH**
  音源のパワーが THRESH より大きい間，音源を検出する．データ型は， `float` であり，デフォルト値はない．ここでは，とりあえず 25 を設定する．値のチューニング方法は，ノードリファレンスを参照のこと．

- **PAUSE_LENGTH**
  発話中のポーズ区間で，音源のパワーが一時的に THRESH を下回る場合がある．その都度，音源検出が途絶えては，検出後の処理が困難である．ポーズと見做される時間内であれば一時的に THRESH を下回っても音源追従を継続し，音源があるとみなす方が都合がよい．データ型は `float` であり，デフォルト値は， 800 である．これを変更する必要はない．

  PAUSE_LENGTH を長くとると，音源検出が継続しやすい傾向になるが，発話終了時刻が PAUSE_LENGTH 分延長される．次の発話開始とオーバーラップし， 2 つの発話が 1 つの発話に見えるという問題が発生する．短過ぎると， 1 つの発話と看做してほしい発話が，時間的に細切れに検出される傾向が強くなる．

- **MIN_SOURCE_INTERVAL**
  音源を追従する上で，同一音源が常に同一方向に検出され続けることは稀である．検出方向がゆらぐため，揺らぎ幅内であれば同一音源みなさなす必要がある．許容する揺らぎ幅を角度で指定する．データ型は，`float` で，デフォルト値が 20 である． これを変更する必要はない．

DisplayLocalization ノードは，LOG_IS_PROVIDED， 1 個所のプロパティ値設定場所がある．

音源位置のログが欲しいときに， `true` にする．データ型は， `bool` でデフォルト値は `false` である．これを変更する必要はない．以上で，音源定位部のプロパティ設定が完了である．

続いて，Iterator の設定を行う．Iterator は，MAIN (subnet) からサブルーチンのように使用される．以下の設定を行い MAIN (subnet) との関係を設定する必要がある．

Iterator に必須な項目は，ネットワークに INPUT と OUTPUT と CONDITION があることである．INPUT はデータの入力部分で，OUTPUT は，データの出力部分である．サブルーチンの入力と出力と対応付けて考えると理解しやすい．Iterator は，サブルーチンでありながら繰り返し処理を行う構造になっている．そのため，CONDITION つまり繰り返し処理の停止条件を記述する必要がある．

図 3.11 のように設定する．INPUT の設定は，入力ターミナル部分で Shift を押下しながら，左クリックすると，ダイアログが現れ，名前を変更することができる．OK を押すと，設定され，入力部分に赤字で INPUT と表示される．OUTPUT の設定は，出力ターミナル部分で Shift を押下しながら，左クリックすると，ダイアログが現れ，名前を変更することができる．OK を押すと，設定され，入力部分に青字で OUTPUT と表示される．CONDITION の設定は，出力ターミナル部分で Control を押下しながら，左クリックすると，設定される．紫色で CONDITION と表示される．

意図しない場所に INPUT，OUTPUT，CONDITION を設定した場合は，これらの文字の上で Shift を押下しながら，左クリックすると，解除できる．

最後に，subnet と iterator を統合する．MAIN (subnet) タブを押し，MAIN (subnet) のネットワーク表示に戻る．地を左クリックし， New Node 上にマウスカーソル移動させると，ノードカテゴリ名に新しく subnet という項目ができている．subnet カテゴリ上にマウスカーソル移動させると， LOOP0 という項目が現れる．ノードを配置するのと同様に LOOP0 上で左クリックすると， LOOP0 という 1 入力，1 出力のノードが配置される．つまり，これまでの作業で，LOOP0 (iterator) 内に記述したサブネットワークを仮想的なノードにしたことになる．

図 3.12 の様に，InputStream と LOOP0 を接続し， LOOP0 に OUTPUT を設定し，音源定位ネットワークの完成である．

3.3.2 Windows 版

本節では Linux 版の FlowDesigner との差異をメインに Windows 版 FlowDesigner の操作方法について説明する．

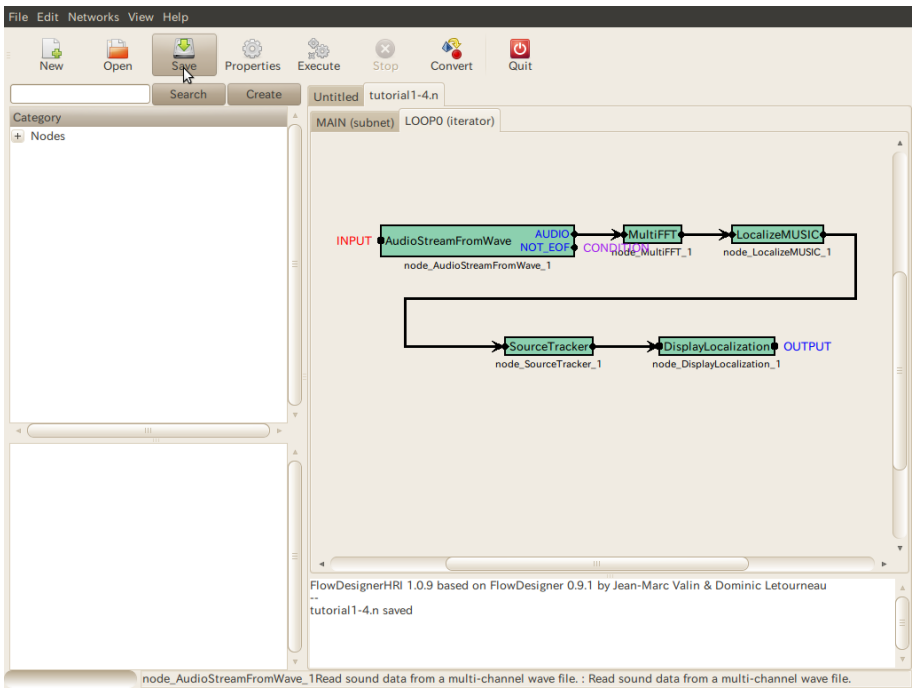


図 3.11: 音源定位部分のノードに INPUT，OUTPUT，CONDITION を設定した様子

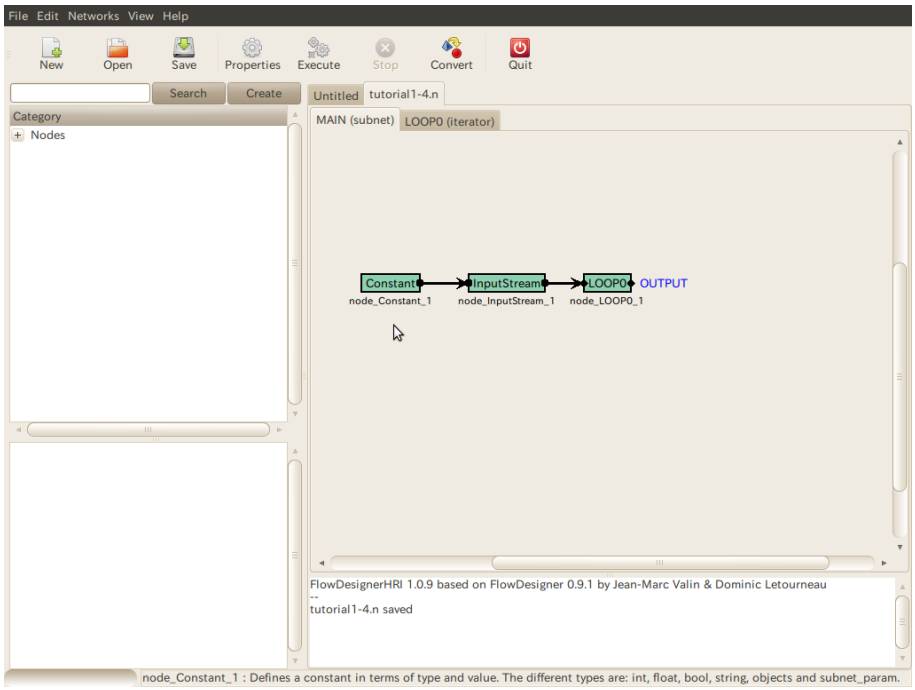


図 3.12: 完成した音源定位ネットワーク

FlowDesigner の起動

Windows 版の FlowDesigner ではデスクトップ上のアイコン，または [スタート] → [プログラム] → [HARK] から FlowDesigner for HARK を起動する．

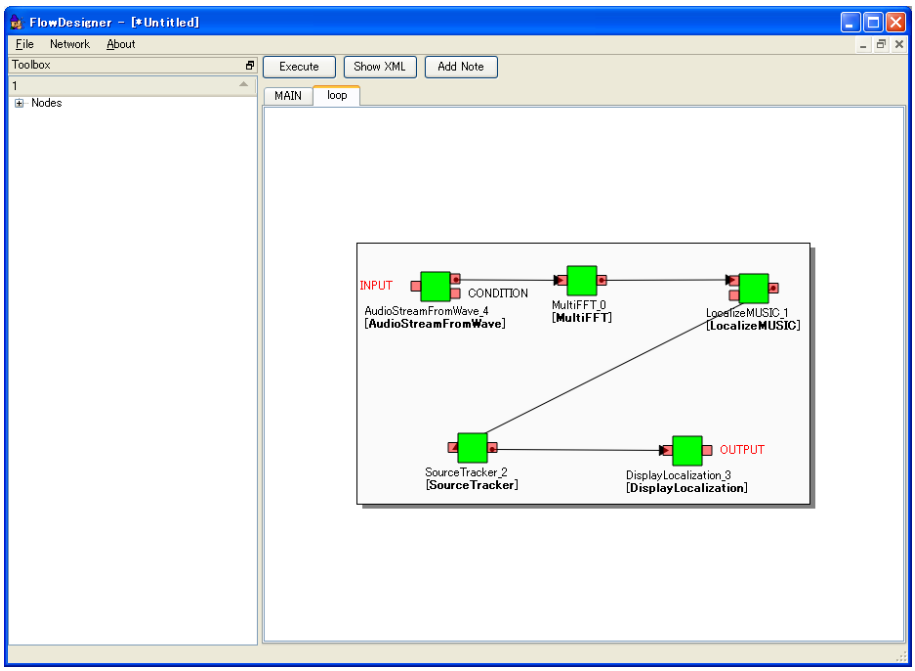


図 3.13: FlowDesigner for HARK 概観

図 3.14: FlowDesigner for HARK アイコン

コマンドプロンプトから起動したい場合は以下の操作で起動できる．

```
> flowdesigner
```

ノードの操作

- **ノードの追加**
  ノードの追加はウィンドウ左側にある，Toolbox というサブウィンドウにある Nodes というツリーを展開し，ノードをドラッグアンドドロップしてネットワークに追加する．

  例えば SpectralGainFilter を追加したい場合は，Nodes → HARK → Separation とツリーを展開し，SpectralGainFilter をドラッグアンドドロップしてネットワークに追加する．

- **ノードの削除**
  ノードを削除する場合はノードをクリックして選択状態にした後 Delete ボタンで削除する．

- **ノードの複製**
  ノードの複製には対応していない．

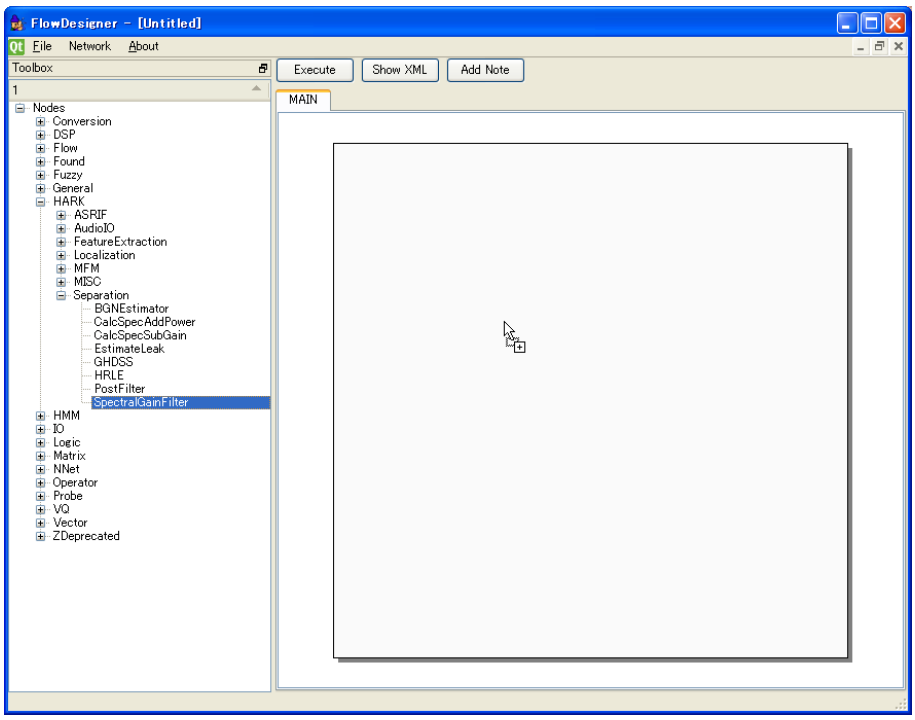


図 3.15: ノードの追加

ターミナルの接続

- **ノードの接続**
  ターミナルの接続方法は Linux 版と同様，ターミナルを矢線で結線し，接続する．接続元のターミナル上で左ボタンを押下し，ドラッグして接続先のターミナルでボタンを放すとターミナル間が矢線で結線される．

- **ノードの切断**
  ターミナル切断は，矢線を左クリックし，選択状態にした後，Delete ボタンで切断する．

- **ターミナル名の表示**
  Windows 版では，ターミナル名が常時表示されていないが，カーソルをターミナルに当てるとヒントとしてターミナル名が表示される．

  また，ターミナルの接続先選択時では，ターミナル上にカーソルを当てるとターミナル名と対応する型が表示される．

- **未対応の機能**
  Linux 版ではターミナルが処理できるデータ型に対応していない接続の場合，ノード接続線が赤色に表示されるが，Windows 版では対応していない．

  ノード接続線の再配置についても対応していない．

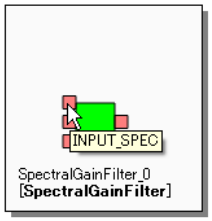


図 3.16: ターミナル名の表示

第4章　データ型

本章では，FlowDesigner と HARK のノード群で使用するデータ型について述べる．HARK でデータ型を意識する必要があるのは，以下の 2 つのケースである．

- ノードのプロパティ設定
- ノード同士の接続（ノード間通信）

ノードのプロパティ設定で用いるデータ型

ノードのプロパティとして現状で指定できるデータ型は，以下の 5 種類である．

型	意味	データ型レベル
int	整数型	基本型
float	単精度浮動小数点型	基本型
string	文字列型	基本型
bool	論理型	基本型
Object	オブジェクト型	FlowDesigner 固有型
subnet_param	サブネットパラメータ型	FlowDesigner 固有型

int , float , string , bool , については，C++ の基本データ型をそのまま利用しているので，仕様は C++ に準じる．Object ， subnet_param については，FlowDesigner 固有のデータ型である．Object は, FlowDesigner 内部の Object 型を継承しているクラスとして定義されるデータ型の総称となっている．HARK では，プロパティとして指定できる Object は，Vector もしくは Matrix であるが，後述のように基本型以外は Object 型を継承しているため，Vector や Matrix のようにテキスト形式での記述が実装されていればプロパティとして指定することができる．基本型であっても，Object 型を継承したオブジェクトを利用することで (例えば <Int 1 > など)， Object として指定することも可能である．subnet_param は，複数のノード間で一つのパラメータをラベルを用いて共有する際に用いられる特殊なデータ型である．

ノード同士の接続の際に用いるデータ型

ノードの接続（ノード間通信）は，異なるノードのターミナル（ノードの左右に黒点として表示される）を FlowDesigner の GUI 上で線で結ぶことによって，実現される．この際に用いられるデータ型は，以下の通りである．

型	意味	データ型レベル
any	Any 型	FlowDesigner 固有型
int	整数型	基本型
float	単精度浮動小数点実数型	基本型
double	倍精度浮動小数点実数型	基本型
complex<float>	単精度浮動小数点複素数型	基本型
complex<double>	倍精度浮動小数点複素数型	基本型
char	文字型	基本型
string	文字列型	基本型
bool	論理型	基本型
Vector	配列型	FlowDesigner オブジェクト型
Matrix	行列型	FlowDesigner オブジェクト型
Int	整数型	FlowDesigner オブジェクト型
Float	単精度浮動小数点実数型	FlowDesigner オブジェクト型
String	文字列型	FlowDesigner オブジェクト型
Complex	複素数型	FlowDesigner オブジェクト型
TrueObject	論理型（真）	FlowDesigner オブジェクト型
FalseObject	論理型（偽）	FlowDesigner オブジェクト型
nilObject	オブジェクト型 (nil)	FlowDesigner オブジェクト型
ObjectRef	オブジェクト参照型	FlowDesigner 固有型
Map	マップ型	HARK 固有型
Source	音源情報型	HARK 固有型

any はあらゆるデータ型を含む抽象的なデータ型であり，FlowDesigner 固有で定義されている．int , float , double , complex<float> , complex<double> , char , string , bool は，C++ の基本データ型を利用している．これらの仕様は対応する C++のデータ型の仕様に準ずる．基本型を，Object のコンテキストで使おうとすると自動的に GenericType<T> に変換され，Int , Float のように，Object を継承した先頭が大文字になったクラスとして扱うことができる．ただし，String , Complex は，GenericType ではなく，それぞれ std:string, std:complex に対するデファインとして定義されているが，同様に string , complex を Object 型として使う際に用いられる．このように基本型に対して，FlowDesigner の Object を継承する形で定義されているデータ型を FlowDesigner オブジェクト型と呼ぶものとする．TrueObject , FalseObject , nilObject もそれぞれ，true, false, nil に対応する Object として定義されている．FlowDesigner オブジェクト型で最もよく使われるものは，Vector , Matrix であろう．これらは，C++ の STL を継承した FlowDesigner オブジェクト型であり，基本的には C++ の STL の対応するデータ型の仕様に準ずる．

ObjectRef は，オブジェクト型へのスマートポインタとして実現されている FlowDesigner 固有のデータ型であり，Vector , Matrix , Map の要素として用いられることが多い．

Map も，C++ の STL を継承しているが FlowDesigner ではなく，HARK 固有のデータ型である．Source は，音源情報型として定義される HARK 固有のデータ型である．

ノードのターミナルの型 は，FlowDesigner 上で型を知りたいノードのターミナルにカーソルをフォーカスすると， FlowDesigner の最下部に表示される．図 4.1 に，AudioStreamFromMic ノードの AUDIO ターミナルにマウスをフォーカスした例を示す．FlowDesigner の最下部に AUDIO (Matrix<float> Windowed multi-channel sound data. A row index is a channel, and a column index is time. が表示され，AUDIO ポートが Matrix<float> をサポートしていること，窓掛けされた音声波形を出力すること，行列の行がチャネルを表し，列が時刻を表し

ていることがわかる．一般的に，ノードのターミナル同士は，データ型が同じである，もしくは受け側のター

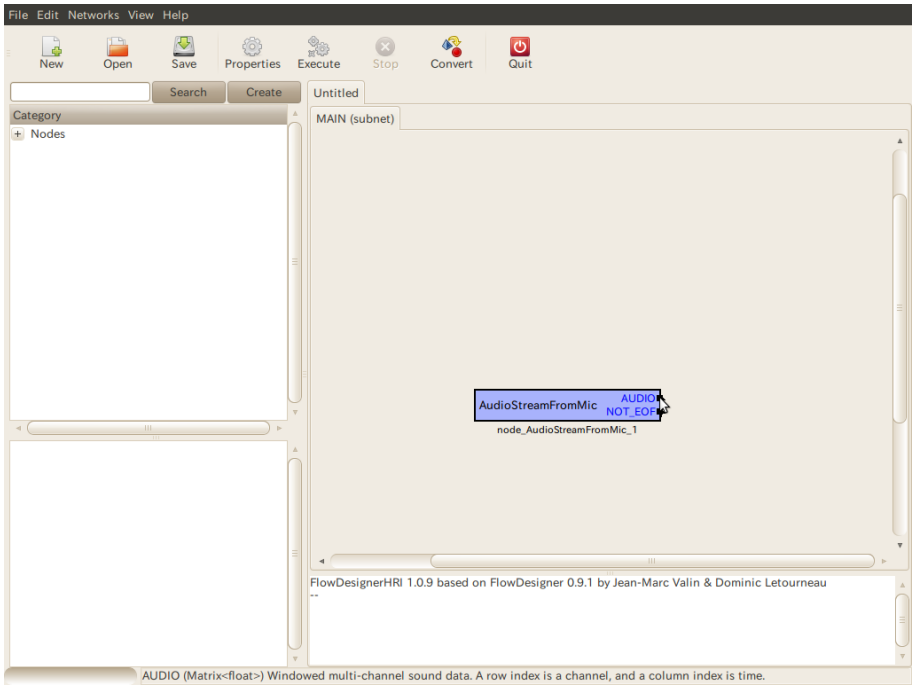


図 4.1: メッセージバーの表示例

ミナルが，送り側のターミナルのデータ型を包含していれば，正常に接続することができ，黒矢印で表示される．この条件を満たさないターミナル同士を接続した場合は，警告の意味で赤矢印で表示される．

以降の節では，上述について基本型，FlowDesigner オブジェクト型，FlowDesigner 固有型 HARK 固有型に分けて説明を行う．

4.1 基本型

int , float , double , bool , char , string , complex (complex<float> , complex<double>) は，前述のように C++ データ型を引き継いだ基本型である．HARK では，番号など，必ず整数であると分かっているもの (音源数や FFT の窓長など) は int が，それ以外の値（角度など）にはすべて float が用いられる．フラグなど，真偽の 2 値のみが必要な場合は bool が用いられる．ファイル名などの文字列が必要な場合は string が使われる．HARK はスペクトル単位の処理や特定長の時間ブロック（フレーム）ごとの処理を行うことが多いため，ノードのターミナルのデータ型としては，直接基本型を用いる場合は少なく，Matrix や Vector , Map の要素として用いることが普通である．complex<float> も，単独で用いることは少なく，スペクトルを表現するために，Vector , Matrix の要素として用いることが多い．倍精度の浮動小数点 (double 型) は，FlowDesigner としてはサポートしているが，HARK では，Source を除いて利用していない．

この型に変換するノード: Conversion カテゴリの To* ノードが各型に変換する．int は ToInt , float は ToFloat , bool は ToBool , string は ToString を用いる．

4.2 FlowDesigner オブジェクト型

Int , Float , String , Complex はそれぞれ，int , float , string , complex の Object 型である．TrueObject , FalseObject は bool 型の true, false に対応する Object であり，nilObject は，nil に対応する Object である．これらの説明は省略する．C++ の標準テンプレートライブラリ (STL) を継承する形で FlowDesigner 内で Object 型として再定義されているものとして，Vector , Matrix が挙げられる．これらに関して以下で説明する．

4.2.1 Vector

データの配列を格納する型を表す．Vector には何が入っていてもよく代表的には ObjectRef を要素にもつ Vector< Obj > ，値 (int ， float) を要素にもつ Vector< int > ， Vector< float > などが使われる．

複数の値を組にして用いるので，ConstantLocalization での角度の組の指定や，LocalizeMUSIC の出力である定位結果の組を表すのに用いられる．

以下に，Vector 型の定義を示す．BaseVector は，FlowDesigner 用のメソッドを実装した型である．下に示すように，Vector 型は STL の vector 型を継承している．

```
template<class T> class Vector : public BaseVector, public std::vector<T>
```

Conversion カテゴリにある ToVect ノードは，int ， float などの入力をとり，入力された値を 1 つだけ要素に持つ Vector を出力する．

また，ノードのパラメータとして Vector を使いたい時は，パラメータのタイプを Object にし，以下のようにテキストで入力することができる．

例えば，二つの要素 3，4 を持つ int 型の Vector をパラメータに入力したい時は，図 4.2 に示すように文字列を入力すればよい．ただし，文字列は初めの文字 < の次に，スペースを開けずに Vector を書く必要があるなどの注意が必要である．

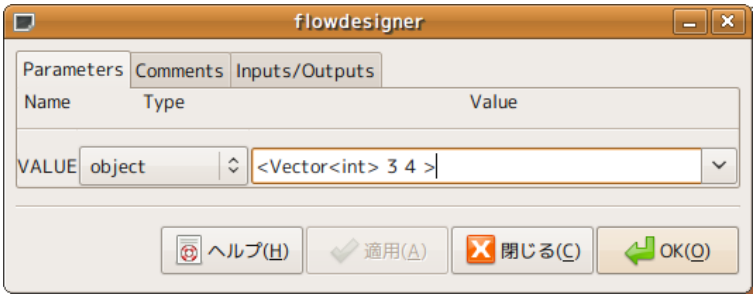


図 4.2: Vector の入力例

4.2.2 Matrix

行列を表す．代表的な型は Matrix<complex<float> > 型と Matrix<float> 型である．それぞれ，複素数を要素に持つ行列と，実数を要素に持つ行列である．

Matrix をノード間通信に用いるノードとしては，MultiFFT (周波数解析)，LocalizeMUSIC (音源定位) などが挙げられるなお，HARK を用いた典型的な音源定位・追従・分離と音声認識の機能を有するロボット聴覚システムでは，音源追従 (SourceTracker) で，音源に ID が付与されるので，それ以前の処理では，ノード間通信に Matrix が用いられ，それ以降の処理では，Map が用いられることが多い．

4.3 FlowDesigner 固有型

FlowDesigner 固有型には，any , ObjectRef , Object , subnet_param が挙げられる．

4.3.1 any

any はあらゆるデータ型の総称となっており，ノードのターミナルが any 型である場合は，他のノードのターミナルがどんな型であっても警告なしに（黒線で）接続することができる．ただし，実際に通信が可能かどうかはノード内部の実装に左右されてしまうため，自ら実装を行う際には極力利用しないことが望ましい．

HARK では，MultiFFT , DataLogger , SaveRawPCM , MatrixToMap といった汎用的に用いるノードに使用を限定している．

4.3.2 ObjectRef

FlowDesigner 内で定義されている Object 型を継承するデータ型への参照を表すデータ型である．

具体的には，Object 型へのスマートポインタとして定義されている．FlowDesigner オブジェクト型，FlowDesigner 固有型 HARK 固有型はいずれも Object を継承したデータ型であるので，これらのデータ型はどれでも参照することができる．基本データ型も，前述のように，ObjectRef に代入しようとした際に最終的に NetCType<T> で変換され，Object のサブクラスとなるため，利用可能である．

4.3.3 Object

主にノードのプロパティで用いられるデータ型である．HARK では，ChannelSelector などで用いられている．基本的には，事前に用意されている int , float , string , bool , subnet_param 以外のデータ型をプロパティとして設定する際に用いるデータ型である．4.3.2 節で述べたように，基本データ型を含めて Object 型として利用可能なため，原理的には，すべてのデータ型を Object として指定できることになるが，実際に入力できるのはテキストでの入出力が実装されているデータ型に限られる．Vector や, Matrix も指定できるように実装されているが，Map はテキスト入出力を実装していないため，Object として入力することは現時点ではできない．

参考までに，以下に，入力できる例，できない例を挙げる．

OK	<Vector<float> 0.0 1.0>	一般的な Vector の入力法
OK	Complex <float (0,0)>	complex も Complex とすれば入力できる
NG	<Vector<complex<float> > (0.0, 1.0) >	complex の入力はサポートされていないためＮＧ
OK	<Vector<ObjectRef> <Complex <float > (0.0, 1.0)> >	Complex として入力すれば問題ない
OK	<Int 1>	int も Int として入力できる

4.3.4 subnet_param

ノードのプロパティで用いられるデータ型である．subnet の複数のノードに同じパラメータをプロパティとして設定する際に，subnet_param を指定して，共通のラベルを記述すれば，MAIN(subnet) 上で，このラベルの値を書き換えることによって，該当の箇所すべての値を修正することができる．

例えば，Iterator ネットワークを作成して (名前を LOOP0 とする）その上に LocalizeMUSIC , GHDSS といったサンプリング周波数を指定する必要のあるノードを配置した際に，これらのプロパティである SAMPLING_RATE の型を subnet_param とし，“SAMPLINGRATE” というラベルを指定しておけば，MAIN(subnet) 上に仮想ネッ

トワーク LOOP0 を配置すると，そのプロパティに SAMPLINGRATE が現れる．このプロパティを int 型にして 16000 を記入しておけば，LocalizeMUSIC , GHDSS の SAMPLING_RATE は常に同一であることが保証できる．

また，別の使い方として，MAIN(subnet) 上のノードのプロパティを subnet_param 型し，名前を ARGx (x は引数の番号) にすると，そのパラメータをバッチ実行の際に引数として指定することができるようになる（例えば，名前を ARG1 として，subnet_param を指定すると，バッチ実行の際の第一引数として利用できる）．引数として指定できるかどうかは，FlowDesigner のプロパティをクリックすることによって確認できる．このプロパティのダイアログに値を記述しておくと，バッチ実行の際にデフォルト値として指定した値が用いられる．

4.4 HARK 固有型

HARK が独自に定義しているデータ型は，Map 型と Source 型である．

4.4.1 Map

Map 型は，キーと ObjectRef 型をセットにしたデータ型である．ObjectRef は，Matrix，Vector，Source といった Object を継承するデータ型へのポインタを指定する．HARK では，音声認識機能も提供しているため，発話単位で処理を行うことが多い．この際に，発話単位の処理を実現するため，発話 ID（音源 ID）をキーとした Map<int, ObjectRef> を用いている．例えば，GHDSS（音源分離）の出力は，Map<int, ObjectRef> となっており，発話 ID がキーであり，ObjectRef には，分離した発話のスペクトルを表す Vector< complex > へのポインタが格納されている．

こうしたノードと，通常，音源追従処理より前に使う Matrix ベースで通信を行うノードを接続するために，MatrixToMap が用意されている．

4.4.2 Source

音源定位情報を表す型であり，HARK では，LocalizeMUSIC（出力），SourceTracker（入出力），GHDSS（入力）という音源定位から音源分離に至る一連の流れの中で Map<int, ObjectRef> の ObjectRef が指し示す情報として用いられる．

Source 型は，次のような情報を持っている．

1. **ID :** int 型．音源の ID
2. **パワー:** float 型．定位された方向のパワー．
3. **座標:** float 型の長さ 3 の配列．音源定位方向に対応する，単位球上の直交座標．
4. **継続時間:** double 型．定位された音源が終了するまでのフレーム数，対応する音源が検出されなければ時間とともに減っていき，この値が 0 になった場合，その音源は消滅する．この変数は，SourceTracker でのみ使用される内部変数である．

Problem

MFCCExtraction や SpeechRecognitionClient などのノードの入出力に使われているデータ型「Map < ·, · >」について知りたいときに読む．

Solution

Map 型は，キーとそのキーに対応するデータの組からなる型である．例えば 3 話者同時認識を行う場合，音声認識に用いる特徴量は話者毎に区別する必要がある．そのため，特徴量がどの話者の何番目の発話に対応するのかを表した ID をキーとし，そのキーとデータをセットにして扱うことで話者・発話を区別する．

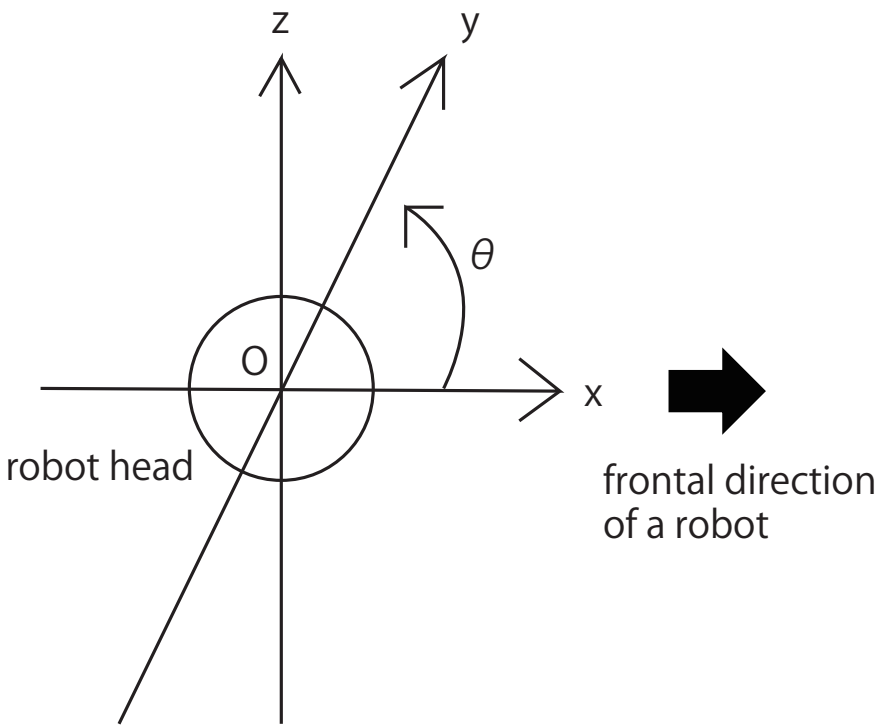


図 4.3: HARK 標準座標系

4.5 HARK 標準座標系

HARK で用いる座標は，指定した原点を中心とし（通常はマイクロホンアレイの中心），x 軸の正方向が正面，y 軸の正方向が左，z 軸の正方向が上になるようにしている．単位はメートルで記述する．また，角度は正面を 0[deg] として反時計回りとすることを前提にしている．(例えば，左方向が 90[deg]．) また，仰角は，その座標の方向ベクトルが XY 平面となす角として定義する．

第5章　ファイルフォーマット

本節では，HARK で利用するファイルの種類およびその形式について述べる．HARK では，ノードの入出力やプロパティ設定でファイルを指定することができる．表 5.1 に一覧を示す

表 5.1: ファイル入出力を指定する HARK ノード一覧

ノード名	使用箇所	ファイル種類	ファイル形式
SaveRawPCM	出力	PCM raw ファイル	PCM バイナリ
LocalizeMUSIC	プロパティ設定	音源定位伝達関数ファイル	HGTF バイナリ
SaveSourceLocation	出力	音源定位結果ファイル	定位結果テキスト
LoadSourceLocation	プロパティ	音源定位結果ファイル	定位結果テキスト
GHDSS	プロパティ設定	音源分離伝達関数ファイル	HGTF バイナリ
	プロパティ設定	マイクロホン位置ファイル	HARK テキスト
	プロパティ設定	定常ノイズ位置ファイル	HARK テキスト
	プロパティ設定	初期分離行列ファイル	HGTF バイナリ
	出力	分離行列ファイル	HGTF バイナリ
SaveFeatures	出力	特徴量ファイル	`float` バイナリ
SaveHTKFeatures	出力	特徴量ファイル	HTK 形式
DataLogger	出力	`Map` データファイル	`Map` テキスト
CMSave	プロパティ	相関行列ファイル	相関行列テキスト
CMLoad	出力	相関行列ファイル	相関行列テキスト
JuliusMFT	起動時引数	設定ファイル	jconf テキスト
	設定ファイル中	音響モデル・音素リスト	julius 形式
	設定ファイル中	言語モデル・辞書	julius 形式
harktool	harktool	音源位置リストファイル	srcinf テキスト
	harktool	インパルス応答ファイル	`float` バイナリ

これらのファイルのうち，Julius 形式については Julius のファイルフォーマットに基本的に準ずる．オリジナルの Julius との違いに関しては JuliusMFT の説明を参照してほしい．

以降では，Julius 形式以外のファイル形式について説明する．なお，実装面では，HGTF (HARK General Transfer Function) バイナリ形式，HARK テキスト形式は，HARK の入出力ライブラリ libharkio で統合的に入出力をサポートしており，総称して，HARK ファイル形式と呼ぶ．それ以外の srcinf テキスト形式については，PCM バイナリ，`float` バイナリ，定位結果テキスト，`Map` テキスト，相関行列テキスト形式についても，順次 harkio でサポートを行うよう変更していく予定である．

表 5.2: HGTF(汎用伝達関数) のデータ構造

<table>
<tr><td colspan="2">識別子 “HARK”</td><td>4 bytes</td></tr>
<tr><td colspan="2">サイズフォームタイプ以降のサイズ</td><td>4 bytes</td></tr>
<tr><td rowspan="4">データチャンク</td><td>フォームタイプ “HGTF”</td><td>4bytes</td></tr>
<tr><td colspan="2">fmt サブチャンク (必須)
(フォーマット ID, マイクロホン数
サンプリング周波数などの基本データ)</td></tr>
<tr><td colspan="2">M2PG/GTF/SM サブチャンク（排他的)</td></tr>
<tr><td colspan="2">SPEC サブチャンク（M2PG があるときのみ存在)
（付加情報）</td></tr>
</table>

5.1 HGTF ファイル形式

HGTF(HARK General Transfer Function) ファイル形式は RIFF (Resource Interchange File Format) ライクなバイナリファイル形式であり，LocalizeMUSIC の音源定位伝達関数ファイル，GHDSS の音源分離伝達関数ファイル，および，分離行列ファイルを表すために用いる表現である．データ構造を表 5.2 に示す．

フォームタイプ “HGTF” は，以下の 3 種類のデータをサポートしている．

- M2PG 1.0
- GTF 1.0
- SM 1.0

M2PG (MUSIC 2D Polar Grid) は，LocalizeMUSIC の音源定位伝達関数ファイルのために用いられる．GTF (GHDSS Transfer Function) は，GHDSS の音源分離伝達関数ファイルにまた，SM (Separation `Matrix`) は，GHDSS の分離行列ファイルを表すために用いる．

fmt は，基本的なデータを入力するために使われる必須のサブチャンクであり，詳細を．表 5.3 に示す．なお，チャンクのサイズは，チャンク名の 4 バイトとチャンクサイズの 4 バイトを除いた部分のサイズの合計である．

表 5.3: fmt サブチャンクの定義

型	値
char[4]	“ fmt “
`int`	チャンクのサイズ，24+12*Nmic [bytes]
short[2]	fmt バージョン（メジャー，マイナー）
`int`	フォーマット ID (0: 未定義，1:M2PG, 2:GTF, 3:SM)
`int`	マイクロホン数 (Nmic)
`int`	FFT 長 [サンプル] (Nfft)
`int`	インパルスレスポンス長 [サンプル]
`int`	サンプリング周波数 [Hz]
Vector3[nmic]	マイクロホン位置の座標 [m]

M2PG/GTF/SM サブチャンクは，個々のファイル（音源定位伝達関数ファイル，音源分離伝達関数ファイル，分離行列ファイル）のデータが格納されるチャンクである．次節以降で，別途説明を行う．

表 5.4: SPEC サブチャンクの定義

型	値
char[4]	“ SPEC ”
int	チャンクのサイズ，1551 [bytes]
short[2]	SPEC バージョン（メジャー，マイナー）
char[11]	作成日 (例：2009/05/19)
char[128]	場所（例：Smart room）
char[128]	ロボット（例：HRP-2, SIG2, ···）
char[128]	マイクロホンアレイのタイプ（例：8ch circular，8ch cubic, ···）
char[128]	作成者（例：Kazuhiro Nakadai）
char[1024]	備忘録（MUSIC の TF 作成のパラメータなど）

SPEC は，ファイル作成条件など付加的な情報を入力するために用いられる．表 5.4 に SPEC のフォーマットを示す．各項目は，0 終端文字列とし，文字コードは ASCII （基本的に英数字など 1 バイト文字しか利用しない），改行コードは LF とする．

5.1.1 LocalizeMUSIC 用音源定位伝達関数

音源定位伝達関数は，音源からマイクロホンアレイまでの伝達関数を表し，M2PG チャンクを用った HGTF 形式のファイルとして表現する，音源定位伝達関数ファイルは，LocalizeMUSIC のプロパティ A_MATRIX に指定する．具体的にどのように使われるかは，LocalizeMUSIC を参照されたい．

M2PG の定義を表 5.5 に示す．なお，多次元配列のメモリの並びは row major (C 言語の場合と同じ) である．

表 5.5: M2PG サブチャンクの定義

型	値
char[4]	”M2PG”
int	チャンクのサイズ， $12 + 12 * Nd * Nr + 8 * Nd * Nr * nmic * (Nfft/2 + 1)$ [bytes]
short[2]	M2PG バージョン（メジャー，マイナー）
int	水平方向の要素数 (Nd)
int	距離の要素数 (Nr)
Vector3[Nd][Nr]	音源位置の座標
complex<float> [Nd][Nr][nmic][Nfft/2+1]	伝達関数

本ファイルは，インパルス応答形式のファイルから，サポートツールである harktool を使用して作成する．具体的な作成方法は 7.1 節で述べる．

5.1.2 GHDSS 用音源分離伝達関数

音源分離用伝達関数も，音源定位用伝達関数と同様に音源からマイクロホンアレイまでの伝達関数を表し，GTF サブチャンクを用った HGTF 形式のファイルとして表現する，音源分離伝達関数ファイルは，GHDSS の

プロパティ TF_CONJ が true の時に，TF_CONJ_FILENAME に指定する．具体的にどのように使われるかは，GHDSS を参照されたい．

GTF サブチャンクの定義を表 5.6 に示す．多次元配列のメモリの並びは row major (C 言語の場合と同じ) である．

表 5.6: GTF サブチャンクの定義

型	値
char[4]	”GTF ”
int	チャンクのサイズ， 8+12*Ns+4+4*Nf+8*Ns*nmic*Nf [bytes]
short[2]	GTF バージョン（メジャー，マイナー）
int	音源位置数 (Ns)
Vector3[Ns]	音源位置の座標 [m]
int	周波数ライン数 (Nf)
int[Nf]	インデックスから周波数ビンへのマップ
complex<float> [Ns][nmic][Nf]	伝達関数

本ファイルは，インパルス応答形式のファイル，もしくは，マイクロホン位置と音源位置リストファイルからサポートツールである harktool を使用して作成する．具体的な作成方法は，7.1 節で述べる．

5.1.3 GHDSS 用 分離行列

分離行列は，SM サブチャンクを用った HGTF 形式のファイルとして表現する，音源分離伝達関数ファイルは，GHDSS のプロパティ INITW_FILENAME に指定することができる．また，EXPORT_W を true にした際に現れる EXPORT_W_FILENAME を指定することで分離行列ファイルを出力することができる．具体的にどのように使われるかは，GHDSS を参照されたい．

SM サブチャンクの定義を表 5.7 に示す．多次元配列のメモリの並びは row major (C 言語の場合と同じ) である．

表 5.7: SM サブチャンクの定義

型	値
char[4]	”SM ”
int	チャンクのサイズ， 8+12*Ns+4+4*Nf+8*Ns*nmic*Nf [bytes]
short[2]	SM バージョン（メジャー，マイナー）
int	音源位置数 (Ns)
Vector3[Ns]	音源位置の座標 [m]
int	周波数ライン数 (Nf)
int[Nf]	インデックスから周波数ビンへのマップ
complex<float> [Ns][nmic][Nf]	分離行列

5.2 HARK テキスト形式

HARK テキスト形式は，マイクロホン位置情報，ノイズ位置情報をサポートしているテキスト形式ファイルである．

記述形式は、hark1.1 までは、TEXT 形式、hark1.2 以降は、XML 形式である。

5.2.1 マイクロホン位置テキスト形式

マイクロホン位置テキスト形式は，GHDSS や harktool で使用するマイクロホンアレイの各マイクロホン位置を記述したファイルである．GHDSS モジュールの TF_CONJ プロパティで CALC を選択すると，MIC_FILENMAE プロパティが現れる．この欄にマイクロホン位置形式で保存されたマイクロホンアレイの各マイクロホン座標を与えると，インパルス応答をシミュレーションで生成，音源分離を行う．

XML 形式で使用する各タグの説明をする。

表 5.8: マイクロホン位置テキストのタグ定義

タグ名	説　明
hark_config	ファイル種別 id : ”MicArrayLocation” 固定 major: メジャー番号 (2) minor: マイナー番号 (0)
Comment	コメント
Mic size	マイク数
MicIndex	Mic 番号
Position	音源位置座標 マイクロホンアレイの中心（原点）から測定場所への方向ベクトル (x，y，z) を表す． ベクトルの大きさ $\sqrt{x^2+y^2+z^2}$ が 1.0 になるように記述する． unit: meter type: cartesian
X	音源位置座標: X 軸 (上記 Position タグ参照)
Y	音源位置座標: Y 軸 (上記 Position タグ参照)
Z	音源位置座標: Z 軸 (上記 Position タグ参照)

マイク位置ファイルの記述例

下記に 4 チャンネルのマイクロフォンアレイのマイク位置ファイルの例を示す。

```
<hark_xml>
  <hark_config id="MicArrayLocation" major="2" minor="0">
    <Comment>created by harktool4</Comment>
    <Mics size="4">
      <Mic>
        <MicIndex>0</MicIndex>
        <Position unit="meter" type="cartesian">
          <X>-0.957823</X>
          <Y>8.37355e-08</Y>
          <Z>0.287361</Z>
        </Position>
      </Mic>
      <Mic>
        <MicIndex>1</MicIndex>
        <Position unit="meter" type="cartesian">
          <X>1.14219e-08</X>
          <Y>-0.957823</Y>
          <Z>0.287361</Z>
        </Position>
      </Mic>
      <Mic>
        <MicIndex>2</MicIndex>
        <Position unit="meter" type="cartesian">
          <X>0.957823</X>
          <Y>0</Y>
          <Z>0.287361</Z>
        </Position>
      </Mic>
      <Mic>
        <MicIndex>3</MicIndex>
        <Position unit="meter" type="cartesian">
          <X>-4.18678e-08</X>
          <Y>0.957823</Y>
          <Z>0.287361</Z>
        </Position>
      </Mic>
    </Mics>
  </hark_config>
</hark_xml>
```

5.2.2 ノイズ位置テキスト形式

ノイズ位置テキスト形式は，GHDSS で固定方向からの定常ノイズ方向を指定するためのファイルである．GHDSS モジュールの FIXED_NOISE プロパティで `true` を選択すると， FIXED_NOISE_FILENAME プロパティが現れる．この欄にノイズ源形式の音源座標を与えると，その方向のノイズを既知ノイズ源として処理する．

原点からのノイズ源方向を，3 次元のベクトル（大きさ 1）で示す．なお，指定できるノイズ源の数は，1 つである．

下記に各タグの説明をする。

表 5.9: ノイズ位置テキストのタグ定義

タグ名	説　明
hark_config	ファイル種別 id : ”NoiseLocation” 固定 major: メジャー番号 (2) minor: マイナー番号 (0)
Comment	コメント
NoiseSources size	”1” 固定
NoiseIndex	”0” 固定
Position	音源位置座標 マイクロホンアレイの中心（原点）から測定場所への方向ベクトル (x，y，z) を表す． ベクトルの大きさ $\sqrt{x^2+y^2+z^2}$ が 1.0 になるように記述する． unit: meter type: cartesian
X	音源位置座標: X 軸 (上記 Position タグ参照)
Y	音源位置座標: Y 軸 (上記 Position タグ参照)
Z	音源位置座標: Z 軸 (上記 Position タグ参照)

ノイズ位置ファイルの記述例

```
<hark_xml>
<hark_config id="NoiseLocation" major="2" minor="0">
  <Comment>created by harktool4</Comment>
  <NoiseSources size="1">
    <NoiseSource>
      <NoiseIndex>0</NoiseIndex>
      <Position unit="meter" type="cartesian">
        <X>-4.37114e-08</X>
        <Y>-3.82137e-15</Y>
        <Z>1</Z>
      </Position>
      </NoiseSource>
    </NoiseSources>
  </hark_config>
</hark_xml>
```

5.3 その他のファイル形式

その他のファイル形式には，srcinf テキスト形式，PCM バイナリ，`float` バイナリ，定位結果テキスト，`Map` テキスト形式が挙げられる．これらについて説明する．

5.3.1 音源位置リスト情報 (srcinf) 形式

音源定位伝達関数ファイル（LocalizeMUSIC の A_MATRIX に設定）や音源分離伝達関数ファイル（GHDSS の TF_CONJ_FILE に設定）を作成するため harktool で用いるファイルである．

記述形式は、hark1.1 までは、TEXT 形式、hark1.2 以降は、XML 形式である。

XML 形式で使用する各タグの説明をする。

表 5.10: 音源位置リスト情報のタグ定義 1

タグ名	説　明
hark_config	ファイル種別 id : ”ImpulseResponseList” または ”TSPList” 固定 major: メジャー番号 (2) minor: マイナー番号 (0)
NumMic	マイク数
SourceFileType	float または wav
Sources size	要素数
Comment	コメント
Position	音源位置座標 マイクロホンアレイの中心（原点）から測定場所への方向ベクトル (x，y，z) を表す． ベクトルの大きさ $\sqrt{x^2+y^2+z^2}$ が 1.0 になるように記述する． unit: meter type: cartesian
X	音源位置座標: X 軸 (上記 Position タグ参照)
Y	音源位置座標: Y 軸 (上記 Position タグ参照)
Z	音源位置座標: Z 軸 (上記 Position タグ参照)
Files size	ファイル数

表 5.11: 音源位置リスト情報のタグ定義 2

タグ名	説　明
Name	ファイル名 測定したインパルス応答のファイル名または，TSP 応答のファイル名を記述する． **インパルス応答の場合**: example_D**ddd**_E**eee**_R**rrr**_ch**cc**.flt (“example” 部分は任意，**ddd** は方位角，**eee** は仰角,**rrr** は半径,**cc** はチャネル番号) の形式で記述する． なお,flt はリトルエンディアンの 32bit の実数 `float` フォーマットファイルである． 例: 方位角:180 度 仰角:16.7 度 半径:1m の場合 /path/example_D180_E16.7_R100_ch00.flt **TSP 応答の場合:** example_D**ddd**_E**eee**_R**rrr**_ch**cc**.wav (“example” 部分は任意，**ddd** は方位角，**eee** は仰角,**rrr** は半径,**cc** はチャネル番号) の形式で記述する． 各ファイルは，RIFF 形式の PCM wave フォーマットのみが利用可能である． 例: 方位角:180 度 仰角:16.7 度 半径:1m の場合 /path/example_D180_E16.7_R100_ch00.wav
MicIndices size	MicIndex 数 (総 Mic 数)
MicIndex	Mic 番号
CutStart	浮動小数点ベクトルのカット開始インデックス 何番目のデータ（サンプル）から読込みを開始するかを指定する． 例えば 1 を指定すれば，1 サンプル目から読み込みを開始する．通常は 1 で問題はない． オフセットは正の整数とする必要がある．
CutEnd	浮動小数点ベクトルのカット終了インデックス 開始オフセットから，何番目のデータ（サンプル）まで読込むかを指定する． 例えば, 開始オフセットを 100 に指定し，終了オフセットを 200 に指定すれば, 100 サンプル目から 200 サンプル目までを読み込む． オフセットは正の整数とする必要がある． この項目に”0”を記載すると、全サンプルを読込みます。
SynchronousAverge	同期平均の数
OriginalImpulseFile	オリジナルの TSP ファイル
TSPOffset	サンプル中の TSP 信号のオフセット
TSPLength	サンプルの１つの TSP の長さ
SignalMax	TSP 信号の最大振幅
NoiseSources size	要素数
NoiseIndex	Mic 番号

ImpulseResponseList ファイルの記述例

```
<hark_xml>
  <hark_config id="ImpulseResponseList" major="2" minor="0">
    <NumMic>8</NumMic>
    <SourceFileType>float</SourceFileType>
    <Comment>created by harktool4</Comment>
    <Sources size="72">
      <Source>
        <Position unit="meter" type="cartesian">
          <X>-0.957823</X>
          <Y>8.37355e-08</Y>
          <Z>0.287361</Z>
        </Position>
        <Files size="8">
          <File>
            <Name>/YOUR_PATH/D-180_E16.7_R100_ch00.flt</Name>
            <MicIndices size="1">
              <MicIndex>0</MicIndex>
            </MicIndices>
          </File>
          <File>
            <Name>/YOUR_PATH/D-180_E16.7_R100_ch01.flt</Name>
            <MicIndices size="1">
              <MicIndex>1</MicIndex>
            </MicIndices>
          </File>
            ・
            ・
        </Files>
        <CutStart>1</CutStart>
        <CutEnd>0</CutEnd>
      </Source>
            ・
            ・
    </Sources>
  </hark_config>
</hark_xml>
```

TSPList ファイルの記述例

```
  <hark_config id="TSPList" major="2" minor="0">
      <SynchronousAverage>16</SynchronousAverage>
      <OriginalImpulseFile type="float">
        /usr/bin/harktool4_utils/16384.little_endian.tsp
      </OriginalImpulseFile>
      <TSPOffset>16384</TSPOffset>
      <TSPLength>16384</TSPLength>
      <SignalMax>32768</SignalMax>
      <NumMic>8</NumMic>
      <SourceFileType>float</SourceFileType>
      <Comment>created by harktool4</Comment>
      <Sources size="72">
          <Source>
              <Position unit="meter" type="cartesian">
                <X>-0.957823</X>
                <Y>8.37355e-08</Y>
                <Z>0.287361</Z>
              </Position>
              <Files size="1">
                <File>
                  <Name>/YOUR_PATH/D180_E16.7_R100.wav</Name>
                    <MicIndices size="8">
                      <MicIndex>0</MicIndex>
                      <MicIndex>1</MicIndex>
                      <MicIndex>2</MicIndex>
                      <MicIndex>3</MicIndex>
                      <MicIndex>4</MicIndex>
                      <MicIndex>5</MicIndex>
                      <MicIndex>6</MicIndex>
                      <MicIndex>7</MicIndex>
                    </MicIndices>
                  </File>
                </Files>
              <CutStart>1</CutStart>
            <CutEnd>0</CutEnd>
          </Source>
            .
            .
      </Sources>
    </hark_config>
</hark_xml>
```

5.3.2 PCM バイナリ形式

音声波形を PCM 形式で表したファイル形式．分離音の時間波形をファイル保存する際などに用いる．各サンプル値は 16 ビットあるいは 24 ビット符号付整数をリトルエンディアンで表し，ヘッダなど付加的情報は用いない．HARK では，SaveRawPCM が出力し，標準の拡張子は .sw である．複数チャネル存在する音声波形の場合には，チャネルでマルチプレクスして表現する．従って，本形式のファイルを別プログラムで読み書きする際には，sox などを用いて wav ファイル等に変換して使用するか，適切なサンプリング周波数とトラック数，量子化ビット数を指定する必要がある．

5.3.3 float バイナリ

実数を IEEE 758 の 32 ビット浮動小数点数，リトルエンディアンで格納したファイルであり，HARK では，主に，特徴ベクトルやインパルス応答データに用いられるファイル形式である．

特徴ベクトル

特徴ベクトルは，実数ベクトルである．SaveFeatures で出力され，拡張子は .spec が用いられる．ベクトルの次元要素は低次元から高次元方向に書き出される．ベクトル値がもつ意味は，ユーザの使用に依存する．HARK で通常用いる特徴は，パワースペクトル，MFCC，MSLS, ミシングフィーチャマスクである．SaveFeatures ではヘッダ情報は記録されないため，特徴ベクトルの次元数や，ベクトルの数は，記録されないことに注意されたい．ヘッダ情報の記録に関し，HARK では HTK(The Hidden Markov Model Toolkit) 形式のヘッダを付与する SaveHTKFeatures をサポートしており，用途に応じて SaveFeatures と SaveHTKFeatures を使い分けることができる．

インパルス応答形式

音源位置とマイクロホン間のインパルス応答を表す．標準の拡張子は .flt であり，harktool で伝達関数の行列を作成するために用いられる．

本形式のファイルをテキストに変換するには，例えば次のようなプログラムを作成する．

```
#include <stdio.h>
#include <fcntl.h>

int main(int argc, char *argv[])
{
  int fd=0;
  float data=0;

  fd = open(argv[1], O_RDONLY);
  while(1){
    int ret = read(fd, &data, 4); /* 4 [byte] = 32 [bit] */
    if( ret != 4){
      break;
    }
```

```
    printf("%f\n", data);
  }
  close(fd);
  return (0);
}
```

5.3.4 定位結果テキスト

音源定位結果を記録するファイル形式である．LoadSourceLocation , SaveSourceLocation モジュールで用いられる．プロパティの FILENAME に本形式で保存されたファイルを指定することで読み書きできる．

音源定位結果フォーマットは，次の情報で構成されている．フレーム番号を先頭に，N 個分の音源情報を 1 行に表示する構成となっている．

例えば，N 個の音源が M フレーム分定位された場合，次の形式のテキストファイルが保存される．

```
0 id0 x0 y0 z0 id1 x1 y1 z1  ...  idN xN yN zN
1 id0 x0 y0 z0 id1 x1 y1 z1  ...  idN xN yN zN
    .
    .
    .
M id0 x0 y0 z0 id1 x1 y1 z1  ...  idN xN yN zN
```

1 列目は入力されたフレームの通し番号で，それ以降は，4 列を 1 セット（1 つ分の音源情報）として，同時に検出された音源数に応じて列が増える．音源情報 1 セット中の 1 列目は音源の ID，2-4 列目は，単位球上の直交座標系として表した音源の方向である．

5.3.5 Map テキスト

DataLogger で出力されるファイル形式である．`Map<int, float>`，`Map<int, int>` または `Map<int, ObjectRef>` をサポートしている（`Map<int, ObjectRef>` の `ObjectRef` は， `Vector< float >` または `Vector< complex >` のみ）．

出力される形式は以下の通りである．

ラベル フレームカウント キー 1 値 1 キー 2 値 2 . . .

1 フレーム 1 行形式で，上記のように最初にパラメータで指定したラベル，次にフレームカウント，その後に `Map` 型のキーと値を全てスペース区切りで出力する．値が `Vector` の時は，すべての要素がスペース区切りで出力される．

5.3.6 音源定位用相関行列ファイル

周波数領域における入力信号のマイク間の相関行列を格納するためのファイル形式である．LocalizeMUSIC における，既知の雑音情報の白色化などに用いる．現在，このファイルを保存・読み込むモジュールは，CMSave と CMLoad である．

マイク数を M，入力信号の周波数解析後の周波数ビン数を N とすると，相関行列は M 次の複素正方行列で，周波数ビン毎に N 個演算される．n 番目の周波数ビン $(1 \leq n \leq N)$ の相関行列 $\boldsymbol{R}(n)$ を以下で表すとする．

$$\boldsymbol{R}(n) = \begin{bmatrix} r_{11}(n) & \cdots & r_{1M}(n) \\ \vdots & \ddots & \vdots \\ r_{M1}(n) & \cdots & r_{MM}(n) \end{bmatrix} \tag{5.1}$$

ここで，$r_{11}(n), \cdots, r_{MM}(n)$ は複素数である．

相関行列がファイルに保存される時は実部と虚部に分けられ，以下のように保存される．

相関行列ファイル（実部）

$$\begin{array}{ccccccccc} \mathrm{Re}[r_{11}(1)] & \mathrm{Re}[r_{12}(1)] & \cdots & \mathrm{Re}[r_{1M}(1)] & \mathrm{Re}[r_{21}(1)] & \mathrm{Re}[r_{22}(1)] & \cdots & \mathrm{Re}[r_{MM}(1)] \\ \mathrm{Re}[r_{11}(2)] & \mathrm{Re}[r_{12}(2)] & \cdots & \mathrm{Re}[r_{1M}(2)] & \mathrm{Re}[r_{21}(2)] & \mathrm{Re}[r_{22}(2)] & \cdots & \mathrm{Re}[r_{MM}(2)] \\ \mathrm{Re}[r_{11}(N)] & \mathrm{Re}[r_{12}(N)] & \cdots & \mathrm{Re}[r_{1M}(N)] & \mathrm{Re}[r_{21}(N)] & \mathrm{Re}[r_{22}(N)] & \cdots & \mathrm{Re}[r_{MM}(N)] \end{array}$$

相関行列ファイル（虚部）

$$\begin{array}{ccccccccc} \mathrm{Im}[r_{11}(1)] & \mathrm{Im}[r_{12}(1)] & \cdots & \mathrm{Im}[r_{1M}(1)] & \mathrm{Im}[r_{21}(1)] & \mathrm{Im}[r_{22}(1)] & \cdots & \mathrm{Im}[r_{MM}(1)] \\ \mathrm{Im}[r_{11}(2)] & \mathrm{Im}[r_{12}(2)] & \cdots & \mathrm{Im}[r_{1M}(2)] & \mathrm{Im}[r_{21}(2)] & \mathrm{Im}[r_{22}(2)] & \cdots & \mathrm{Im}[r_{MM}(2)] \\ \mathrm{Im}[r_{11}(N)] & \mathrm{Im}[r_{12}(N)] & \cdots & \mathrm{Im}[r_{1M}(N)] & \mathrm{Im}[r_{21}(N)] & \mathrm{Im}[r_{22}(N)] & \cdots & \mathrm{Im}[r_{MM}(N)] \end{array}$$

各成分間はスペース区切りとなる．保存されるファイルは N 行 $M \times M$ 列の行列ファイルになる．読み込まれる際は，そのファイルを $\boldsymbol{R}(n)$ として解釈する．

第6章　ノードリファレンス

本章では，各ノードの詳細な情報を示す．はじめに，ノードリファレンスの読み方について述べる．

ノードリファレンスの読み方

1. **ノードの概要:** そのノードが何の機能を提供しているのかについて述べる．大まかに機能を知りたいときに読むとよい．
2. **必要なファイル:** そのノードを使用するのに要求されるファイルについて述べる．このファイルは 5 節 の記述とリンクしているので，ファイルの詳しい内容は 5 節で述べる．
3. **使用方法:** どんなときにそのノードを使えばよいのかと，具体的な接続例について述べる．とにかく当該ノードを使って見たいときは，その例をそのまま試してみるとよい．
4. **ノードの入出力とプロパティ:** ノードの入力ターミナルと出力ターミナルの型と意味について述べる．また，設定すべきパラメータを表に示している．詳しい説明が必要なパラメータについては表の後にパラメータごとに解説を加えている．
5. **ノードの詳細:** そのノードの理論的背景や実装の方法を含む詳細な解説を述べる．詳しく当該ノードを知りたいときはこの部分を読むとよい．

記号の定義

本ドキュメントで用いる記号を表 6.1 の通り定義する．また，暗黙的に，次のような表記を用いる．

- 小文字は時間領域，大文字は周波数領域を意味する．
- ベクトルと行列は太字で表記する．
- 行列の転置は ^{T} ，エルミート転置は ^{H} で表す．(X^T, X^H)
- 推定値にはハットをつける．(例： x の推定値は $\hat{x}$)
- 入力は x ，出力は y を用いる．
- 分離行列は $\boldsymbol{W}$ を，伝達関数行列は $\boldsymbol{H}$ を用いる．
- チャネル番号は下付き文字で表す． (例: 3 チャネル目の信号源は s_3)
- 時間，周波数は () 内に書く． (例: $X(t, f)$)

表 6.1: 記号のリスト

変数名	説明
m	マイクロホンのインデックス
M	マイクロホンの数
$m_1, \cdots, m_M$	各マイクロホンを示す記号
n	音源のインデックス
N	音源の数
$s_1, \cdots, s_N$	各音源を示す記号
i	周波数ビンのインデックス
K	周波数ビンの数
$k_0 \cdots k_{K-1}$	周波数ビンを示す記号
$NFFT$	FFT ポイント数
$SHIFT$	シフト長
$WINLEN$	窓長
π	円周率
j	虚数単位

6.1 AudioIO カテゴリ

6.1.1 AudioStreamFromMic

モジュールの概要

マイクロホンアレーからマルチチャネル音声波形データを取り込む．サポートするオーディオインタフェースデバイスは，システムインフロンティア製 RASP シリーズ，東京エレクトロンデバイス製 TD-BD-16ADUSB, ALSA ベースのデバイス (例，RME Hammerfall DSP Multiface シリーズ) である．また，デバイス以外にも，IEEE Float 形式のマルチチャネルの音響信号の raw データを TCP/IP ソケット接続で受信することも可能である．各種デバイスの導入は，第 8 章を参照のこと．

必要なファイル

無し．

使用方法

どんなときに使うのか

このモジュールは，HARK システムへの入力として，マイクロホンアレイから得られた音声波形データを用いる場合に使用する．

典型的な接続例

図 6.1 に AudioStreamFromMic モジュールの使用例を示す．

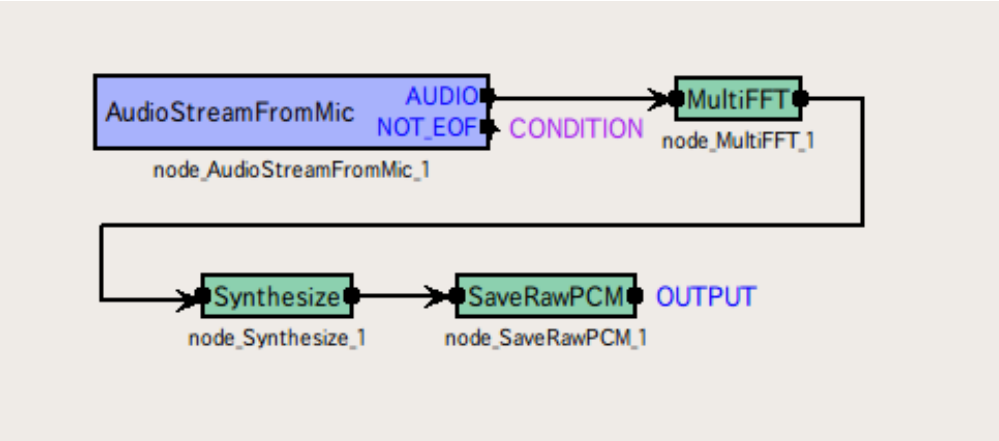


図 6.1: AudioStreamFromMic の接続例

デバイスの写真

AudioStreamFromMic モジュールがサポートするデバイスのうち，以下を写真で紹介する．

1. 無線 RASP,
2. RME Hammerfall DSP シリーズ Multiface (ALSA 対応デバイス).

１．無線 RASP　図 6.2 は無線 RASP の外観である．HARK システムとの接続には，無線 LAN による Ethernet を通じて行う．無線 RASP への電源供給は，付属の AC アダプタで行う．

無線 RASP は，プラグインパワーに対応しており，プラグインパワー供給のマイクロホンをそのまま端子に接続できる．マイクロホンプリアンプを使用せずに手軽に録音ができる利点がある．

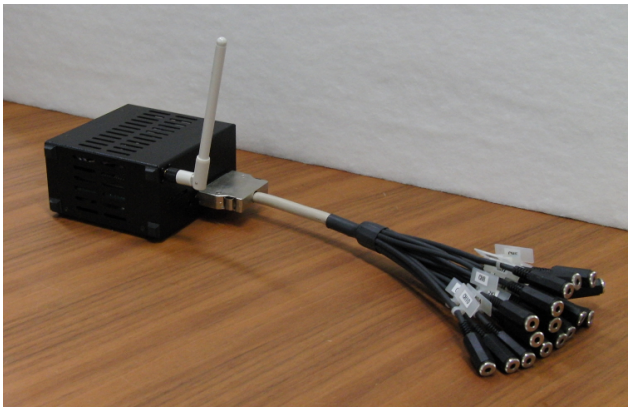


図 6.2: 無線 RASP

2．RME Hammerfall DSP Multiface シリーズ　図 6.3，6.4 は RME Hammerfall DSP シリーズ Multiface の外観である．32bit CardBus を通じてホスト PC と通信を行う．6.3 mm TRS 端子を通じてマイクロホンを接続できるが，入力レベルを確保するために，別途マイクロホンアンプを使用する (図 6.4)．例えば，マイクロホンを RME OctaMic II に接続し，OctaMic II と Multiface を接続する．OctaMic II は，ファンタム電源供給をサポートしており，ファンタム電源を必要とするコンデンサマイクロホン（例えば，DPA 社 4060-BM）を直接接続可

能である．しかし，プラグインパワー供給機能がないため，プラグインパワー供給型のマイクロホンを接続するためには，別途プラグインワパー用の電池ボックスが必要である．例えば，電池ボックスは Sony EMC-C115 や audio-technica AT9903 に附属している．

図 6.3: RME Hammerfall DSP Multiface の正面

図 6.4: RME Hammerfall DSP Multiface の背面

モジュールの入出力とプロパティ

表 6.2: AudioStreamFromMic のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	処理を行う基本単位となるフレームの長さ．
ADVANCE	int	160	[pt]	フレームのシフト長．
CHANNEL_COUNT	int	8	[ch]	使用するデバイスのマイクロホン入力チャネル数．
SAMPLING_RATE	int	16000	[Hz]	取り込む音声波形データのサンプリング周波数．
DEVICETYPE	string	WS		使用するデバイスの種類．
GAIN	string	0dB		RASP を使用する場合のゲイン．
DEVICE	string	127.0.0.1		デバイスへのアクセスに必要な文字列．“plughw:0,1” などのデバイス名や，RASP を使用する時は IP アドレスなど．

入力

無し．

出力

AUDIO : `Matrix<float>` 型．行がチャネル，列がサンプルのインデックスである，マルチチャネル音声波形データ．列の大きさはパラメータ LENGTH に等しい．

NOT_EOF : `bool` 型．まだ波形の入力があるかどうかを表す．録音波形に対する繰り返し処理の終了フラグとして用いる．`true` のとき，波形の取り込みを続行し，`false` のとき，読み込みを終える．常に `true` を出力する．

パラメータ

LENGTH : `int` 型．512 がデフォルト値．処理の基本単位であるフレームの長さをサンプル数で指定する．値を大きくすれば，周波数解像度が上がる一方，時間解像度は下がる．音声波形の分析には，20 ~ 40 [ms] に相当する長さが適切であると言われている．サンプリング周波数が 16000 [Hz] のとき，デフォルト値は 32 [ms] に相当する．

ADVANCE : `int` 型．160 がデフォルト値．フレームのシフト長をサンプル数で指定する．サンプリング周波数が 16000 [Hz] のとき，デフォルト値はフレーム周波数 10 [ms] に相当する．

CHANNEL_COUNT : `int` 型．使用するデバイスのチャネル数．

SAMPLING_RATE : `int` 型．16000 がデフォルト値．取り込む波形のサンプリング周波数を指定する．処理の中で ω [Hz] までの周波数が必要な場合，サンプリング周波数は 2ω [Hz] 以上の値を指定する．サンプリング周波数を大きくとると，一般にデータ処理量が増えるので，実時間処理が困難になる．

DEVICETYPE : `string` 型．ALSA, RASP, WS, TDBD16ADUSB, RASP24-16, RASP24-32, RASP-LC, NETWORK から選択する．ALSA ベースのドライバをサポートするデバイスを使用する場合には ALSA を選択する．RASP-2 を使用する場合には RASP を選択する．無線 RASP を使用する場合は WS を選択する．TD-BD-16ADUSB を使用する場合には TDBD16ADUSB を選択する．RASP-24 を 16bit 量子化ビット数の録音モードで使用する場合には RASP24-16 を選択する．RASP-24 を 24bit 量子化ビット数の録音モードで使用する場合には RASP24-32 を選択する．RASP-LC を PC と無線 LAN による接続で使用する場合には RASP-LC を選択する．（RASP-LC を PC に直接 USB 接続する場合は ALSA で構わない．）TCP/IP 接続を介して IEEE float 形式の raw データを受信したい場合は NETWORK を選択する．

GAIN : `string` 型．0dB がデフォルト値．マイクの録音ゲインを設定する．0dB, 12dB, 24dB, 36dB, 48dB の中から選択する．RASP-24 を録音デバイスとして利用する時のみ有効になる．

DEVICE : `string` 型．DEVICETYPE 毎に入力内容が異なるため，以下の説明を参考のこと．

モジュールの詳細

HARK がサポートするオーディオデバイスは，以下の通り．

1. システムインフロンティア製 RASP シリーズ
   - RASP-2
   - 無線 RASP
   - RASP-24
   - RASP-LC
2. 東京エレクトロンデバイス製 TD-BD-16ADUSB．

3. ALSA ベースのデバイス．以下は例．
   - Microsoft 製 Kinect Xbox
   - Sony 製 PlayStation Eye
   - Dev-Audio 製 Microcone
   - RME Hammerfall DSP シリーズ Multiface
4. TCP/IP ソケット接続で送られる音響信号 (IEEE float wav 形式)

以下ではそれぞれのデバイスを用いる際のパラメータ設定を記す．

RASP シリーズ:

- **RASP-2** のパラメータ設定

CHANNEL_COUNT	8
DEVICETYPE	WS
DEVICE	**RASP-2** の IP アドレス

- **無線 RASP** のパラメータ設定

CHANNEL_COUNT	16
DEVICETYPE	WS
DEVICE	**無線 RASP** の IP アドレス
備考	RASP シリーズはモデルによって 16 チャネル中マイクロホン入力とライン入力が混在しているものがある．混在する場合には，ChannelSelector モジュールを，AudioStreamFromMic モジュールの AUDIO 出力に接続しマイクロホン入力チャネルだけを選択する必要がある．

- **RASP-24** のパラメータ設定

CHANNEL_COUNT	9 の倍数
DEVICETYPE	RASP24-16 または RASP24-32
DEVICE	**RASP-24** の IP アドレス
備考	録音の量子化ビット数を 16bit にする場合は DEVICETYPE=RASP24-16 を，量子化ビット数を 24bit にする場合は DEVICETYPE=RASP24-32 を指定する．CHANNEL_COUNT は，9 の倍数を指定する．録音チャネルは 0 番目〜7 番目のチャネルはマイクロホン入力，8 番目のチャネルはライン入力となる．マイクアレイ処理には，ChannelSelector モジュールを，AudioStreamFromMic モジュールの AUDIO 出力の後段に接続しマイクロホン入力チャネルだけを選択する必要がある．

- **RASP-LC** のパラメータ設定

CHANNEL_COUNT	8
DEVICETYPE	ALSA または RASP-LC
DEVICE	DEVICETYPE=ALSA に指定した場合は plughw:a,b と指定する．設定の詳細は下記の ALSA 対応デバイスを参照．DEVICETYPE=RASP-LC に指定した場合は RASP-LC の IP アドレスを指定する．
備考	RASP-LC の USB インターフェイスを直接 PC に接続する場合は DEVICETYPE=ALSA に設定する．RASP-LC を 無線 LAN を通じて PC と接続する場合は DEVICETYPE=RASP-LC に設定する．録音チャネルは全てマイクロホン入力となる．

東京エレクトロンデバイス製デバイス:

- **TD-BD-16ADUSB** のパラメータ設定

CHANNEL_COUNT	16
DEVICETYPE	TDBD16ADUSB
DEVICE	TDBD16ADUSB

ALSA 対応デバイス:

ALSA 対応デバイスの場合，DEVICE パラメータは plughw:a,b と指定する．a と b には正の整数が入る．a には，arecord -l で表示されるカード番号を入れる．音声入力デバイスが複数接続されている場合には，カード番号が複数表示される．使用するカード番号を入れる．b には arecord -l で表示されるサブデバイス番号を入れる．サブデバイスが複数あるデバイスの場合，使用するサブデバイスの番号を入れる．サブデバイスが複数ある場合の例としては，アナログ入力とディジタル入力を持ったデバイスが該当する．

- **Kinect Xbox** のパラメータ設定

CHANNEL_COUNT	4
DEVICETYPE	ALSA
DEVICE	plughw:a,b

- **PlayStation Eye** のパラメータ設定

CHANNEL_COUNT	4
DEVICETYPE	ALSA
DEVICE	plughw:a,b

- **Microcone** のパラメータ設定

CHANNEL_COUNT	7
DEVICETYPE	ALSA
DEVICE	plughw:a,b

- **RME Hammerfall DSP シリーズ Multiface** のパラメータ設定

CHANNEL_COUNT	8
DEVICETYPE	ALSA
DEVICE	plughw:a,b

ソケット接続 (DEVICETYPE=NETWORK を選択した場合):

DEVICE パラメータは音響信号を送信する側のマシンの IP アドレスを指定する．その他のパラメータは送られてくる音響信号に合わせる必要がある．送信したい音響信号が M チャネルで，１フレーム毎に T サンプルのデータを取得できる場合，以下のように送信すれば良い．

```
WHILE(1){
    X = Get_Audio_Stream (Suppose X is a T-by-M matrix.)
    FOR t = 1 to T
        FOR m = 1 to M
            DATA[M * t + m] = X[t][m]
        ENDFOR
    ENDFOR
    send(soket_id, (char*)DATA, M * T * sizeof(float), 0)
}
```

ここで，X は IEEE float wav 形式であり，$-1 \leq X \leq 1$ である．

Windows 版 DirectSound，ASIO 対応デバイス:

Windows 版 HARK では，無線 RASP，RASP-24，ソケット接続に加えて DirectSound および ASIO に対応したオーディオインタフェースデバイスを取り扱うことができる．

DirectSound，ASIO 対応デバイスの指定は基本的に DEVICE パラメータにデバイス名を入力することになっている．

デバイス名はデバイスマネージャなどで確認できるが，Windows 版 HARK では [スタート] → [プログラム] → [HARK] にある Sound Device List を起動すると，図 6.5 に示すように現在接続中のデバイス名が表示される．

たとえば RME Hammerfall DSP シリーズ Multiface を利用したい場合は，DEVICETYPE に”ASIO”，DEVICE に”ASIO Hammerfall DSP”と設定すればよい．DEVICE パラメータは部分一致でも設定可能となっているので図 6.5 の場合であれば単に”Hammerfall”を設定するだけでもよい．

ただし，複数のデバイス名が部分一致した場合は上位にリストアップされているデバイスが選択されるので注意が必要である．

また，マルチバイト文字の入力には対応していない．

Kinect Xbox， PlayStation Eye，Microcone に関しては正確なデバイス名の入力の必要はなく，下記に示すパラメータを設定することで利用可能となっている．

- **Kinect Xbox** のパラメータ設定

CHANNEL_COUNT	4
DEVICETYPE	DS
DEVICE	kinect

- **PlayStation Eye** のパラメータ設定

CHANNEL_COUNT	4
DEVICETYPE	DS
DEVICE	pseye

- **Microcone** のパラメータ設定

CHANNEL_COUNT	7
DEVICETYPE	DS
DEVICE	microcone

- **RME Hammerfall DSP シリーズ Multiface** のパラメータ設定

CHANNEL_COUNT	8
DEVICETYPE	ASIO
DEVICE	ASIO Hammerfall DSP

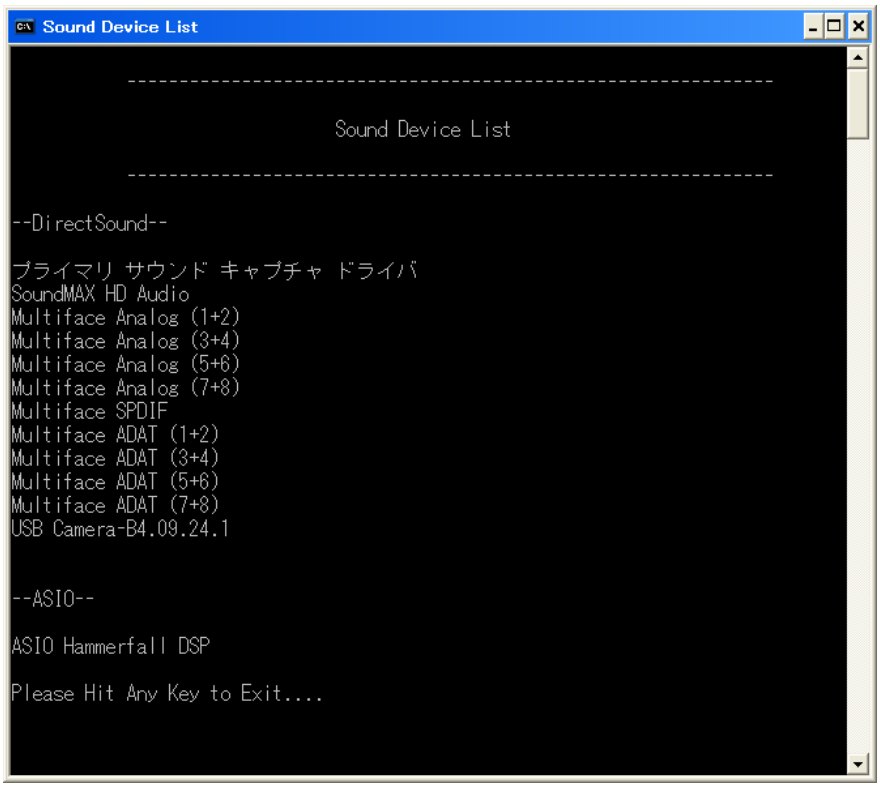


図 6.5: デバイス名の確認

6.1.2 AudioStreamFromWave

ノードの概要

音声波形データを WAVE ファイルから読み込む．読み込んだ波形データは，`Matrix<float>` 型で扱われる．行がチャネル，列が波形の各サンプルのインデックスとなる．

必要なファイル

RIFF WAVE フォーマットの音声ファイル．チャネル数，サンプリング周波数に制約はない．量子化ビット数は，16 bit または 24 bit の符号付き整数の，リニア PCM フォーマットを仮定する．

使用方法

どんなときに使うのか

このノードは，HARK システムへの入力として，WAVE ファイルを読み込ませたいときに使う．

典型的な接続例

図 6.6，6.7 に AudioStreamFromWave ノードの使用例を示す．

図 6.6 は，AudioStreamFromWave がファイルから読み込んだ `Matrix<float>` 型のマルチチャネル波形を MultiFFT ノードによって周波数領域に変換している例である．

AudioStreamFromWave でファイルを読み込むには，図 6.7 のように Constant ノード (FlowDesigner の標準ノード) でファイル名を指定し，InputStream ノードでファイルディスクリプタを生成する．そして，InputStream ノードの出力を，AudioStreamFromWave など HARK の各種ノードのネットワークがある iterator サブネットワーク（図 6.7 中の LOAD_WAVE）に接続する．

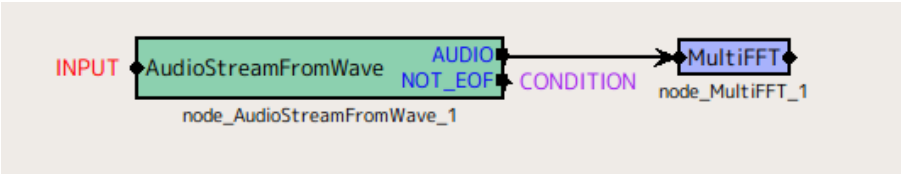


図 6.6: AudioStreamFromWave の接続例: LOAD_WAVE の内部

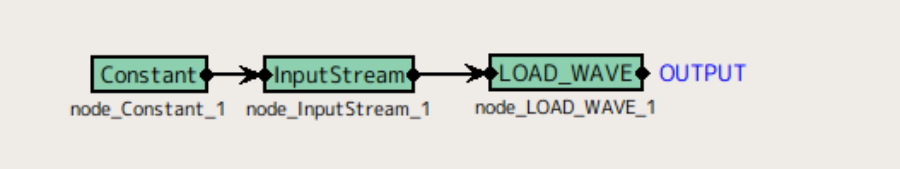


図 6.7: AudioStreamFromWave の接続例: MAIN

表 6.3: AudioStreamFromWave のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	処理を行う基本単位となるフレームの長さ．
ADVANCE	int	160	[pt]	イタレーション毎にフレームをシフトさせる長さ．
USE_WAIT	bool	false		処理を実時間で行うかどうか．

ノードの入出力とプロパティ

入力

INPUT ：Stream 型．FlowDesigner 標準ノードの，IO カテゴリにある InputStream ノードから入力を受け取る．

出力

AUDIO ：Matrix<float> 型．行がチャネル，列がサンプルのインデックスである，マルチチャネル音声波形データ．列の大きさはパラメータ LENGTH に等しい．

NOT_EOF ：bool 型．まだファイルを読めるかどうかを表す．ファイルに対する繰り返し処理の終了フラグとして用いる．ファイルの終端に達したとき false を出力し，それ以外のとき true を出力する．

パラメータ

LENGTH ：int 型．512 がデフォルト値．処理の基本単位であるフレームの長さをサンプル数で指定する．値を大きくすれば，周波数解像度が上がる一方，時間解像度は下がる．音声波形の分析には，20 ~ 40 [ms] に相当する長さが適切であると言われている．サンプリング周波数が 16000 [Hz] のとき，デフォルト値は 32 [ms] に相当する．

ADVANCE ：int 型．160 がデフォルト値．音声波形に対する処理のフレームを，波形の上でシフトする幅をサンプル数で指定する．サンプリング周波数が 16000 [Hz] のとき，10 [ms] に相当する．

USE_WAIT ：bool 型．false がデフォルト値．通常，HARK システムの音響処理は実時間よりも高速に動作する．処理に “待ち” を加えて，入力ファイルに対して実時間で処理を行いたい場合は true に設定する．ただし，実時間よりも遅い場合は，true にしても効果はない．

ノードの詳細

対応するファイルフォーマット: RIFF WAVE ファイルを読み込むことができる．チャネル数，量子化ビット数はファイルのヘッダから読み込むが，サンプリング周波数，量子化手法を表すフォーマット ID は無視する．チャネル数，サンプリング周波数は任意の形式に対応する．サンプリング周波数が処理を行う上で必要になる場合は，パラメータとして要求するノードがある (GHDSS ， MelFilterBank など)．量子化手法とビット数は，16 または 24 ビット符号付き整数によるリニア PCM を仮定する．

パラメータの目安: 処理の目的が音声の分析 (音声認識など) の場合，LENGTH には 20 ~ 40 [ms] 程度，ADVANCE には LENGTH の 1/3 ~ 1/2 程度が良いとされている．サンプリング周波数が 16000 [Hz] の時，LENGTH，ADVANCE のデフォルト値はそれぞれ，32，10 [ms] に対応する．

6.1.3 SaveRawPCM

ノードの概要

時間領域の音声データをファイルに保存する．音声データは，Raw PCM 形式に基づき 16 [bit] または 24 [bit] 整数でファイルに書き出される．その際，入力データの型に応じて多チャネル音声データ，または，モノラル音声データとして記録される．

SaveRawPCM ノードにより書き出されるファイルを WAVE ファイルに変換するためには，ヘッダをファイルの先頭に追加すればよい．例えば SoX などのソフトウェアを使用することでヘッダを追加することができる．しかし，SaveWavePCM ノードを使用するとこのヘッダの追加を自動的に行うため，WAVE ファイルが必要な場合は SaveWavePCM ノードを使うとよい．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源分離における性能を確認するために，分離音を聞いてみたい場合や，AudioStreamFromMic ノードと組みわせて，マイクロホンアレイを用いて多チャンネルの音声データ録音を行う場合に用いる．分離音を保存したい場合は，この SaveWavePCM ノードに接続する前に Synthesize ノードで時間領域の音声データにしておく必要がある．

典型的な接続例

図 6.8，6.9 に SaveRawPCM の使用例を示す．図 6.8 は，AudioStreamFromMic からの多チャネル音声データを SaveRawPCM ノードでファイルに保存する例である．この場合，各チャンネルの音声データは別々のファイルにそれぞれモノラル音声データとして保存される．図 6.8 では，MatrixToMap ノードを取り除き，AudioStreamFromMic ノードと SaveRawPCM ノードを直接つなぐことができる．この場合は全チャンネルの音声データが一つのファイルに保存される．

図 6.9 は，分離音を SaveRawPCM ノードによって保存する例である．GHDSS ノードや，分離後のノイズ抑圧を行う PostFilter ノードから出力される分離音は周波数領域にあるので，Synthesize ノードによって時間領域の波形に変換したのち，SaveRawPCM ノードに入力される．WhiteNoiseAdder ノードは，分離音の音声認識率向上のため通例用いられるもので，SaveRawPCM の使用に必須ではない．

ノードの入出力とプロパティ

入力

INPUT : `Map<int, ObjectRef>` または `Matrix<float>` 型．前者は分離音など，音源 ID と音声データの構造体，後者は多チャネルの音声データ行列．

出力

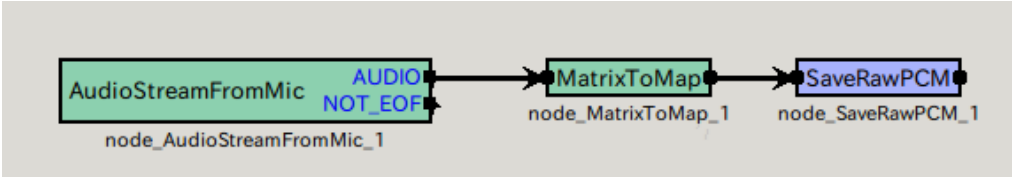


図 6.8: SaveRawPCM の接続例 1

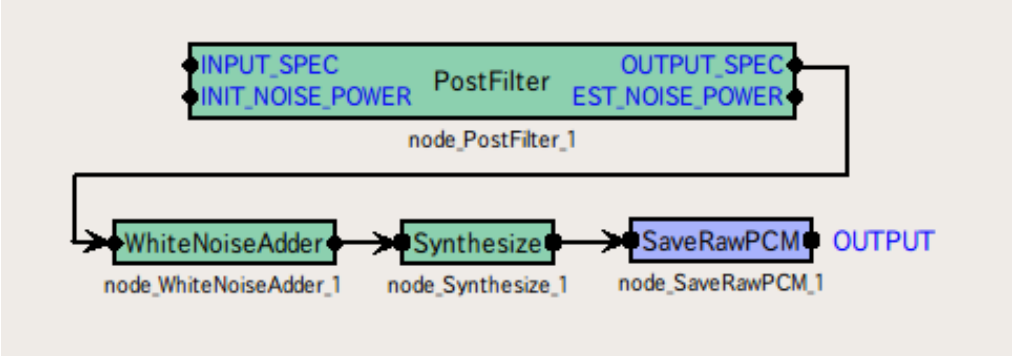


図 6.9: SaveRawPCM の接続例 2

表 6.4: SaveRawPCM のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
BASENAME	string	sep_		保存するファイル名のプレフィックス．
ADVANCE	int	160	[pt]	ファイルに保存する音声波形の分析フレームのシフト長．
BITS	int	16	[bit]	ファイルに保存する音声波形の量子化ビット数．
				16 または 24 を指定可．

OUTPUT ：`Map<int, ObjectRef>` または `Matrix<float>` 型．入力と同じものが出力される．

パラメータ

BASENAME ：`string` 型．デフォルトは sep_．ファイル名のプレフィックスを指定する．出力されるファイル名は，音源 ID が付いている場合は “BASENAME_ID.sw” となる．3 つの混合音を分離した結果の分離音のファイル名は，BASENAME が sep_ のとき，`sep_0.sw`，`sep_1.sw`，`sep_2.sw` などとなる．

ADVANCE ：`int` 型．他のノードの ADVANCE の値と揃える必要がある．

BITS ：`int` 型．ファイルに保存する音声データの量子化ビット数．16 または 24 を指定可．

ノードの詳細

保存されるファイルのフォーマット: 保存されるファイルは，ヘッダ情報を持たない Raw PCM 音声データとして記録される．したがって，ファイルを読む際には，適切なサンプリング周波数とトラック数，量子化ビット数を 16 [bit] または 24 [bit] に指定する必要がある．

また，入力の型によって書き出されるファイルは次のように異なる．

Matrix<float> 型 このとき書き出されるファイルは，入力の行の数だけチャネルを持った多チャネル音声データファイルとなる．

Map<int, ObjectRef> 型 このとき書き出されるファイルは，BASENAME の後に ID 番号が付与されたファイル名で，各 ID ごとにモノラル音声データファイルが書き出される．

6.1.4 SaveWavePCM

ノードの概要

時間領域の音声データをファイルに保存する．SaveRawPCM ノードとの違いは，出力されるのがヘッダーを持つ WAVE 形式のファイルである点である．そのため，例えば audacity や wavesurfer などで読み込む際に簡単である．また，ファイルを AudioStreamFromWave ノードで開きたい場合は，この SaveWavePCM で保存する．

必要なファイル

無し．

使用方法

どんなときに使うのか

SaveRawPCM ノードと同様に，分離音を聞いてみたい場合や，多チャンネルの音声データ録音を行う場合に用いる．

典型的な接続例

使い方は，サンプリング周波数をパラメータとして指定する必要がある以外は SaveRawPCM ノードと同じである．図 6.8，6.9 の例において SaveRawPCM ノードを SaveWavePCM ノードに入れ替えて使うことができる．

ノードの入出力とプロパティ

表 6.5: SaveWavePCM のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
BASENAME	`string`	sep_		保存するファイル名のプレフィックス．
ADVANCE	`int`	160	[pt]	ファイルに保存する音声波形の分析フレームのシフト長．
SAMPLING_RATE	`int`	16000	[Hz]	サンプリング周波数．ヘッダを作成するために使用する．
BITS	`string`	int16	[bit]	ファイルに保存する音声波形の量子化ビット数． int16 または int24 を指定可．

入力

INPUT ：`Map<int, ObjectRef>` または `Matrix<float>` 型．前者は分離音など，音源 ID と音声データの構造体，後者は多チャネルの音声データ行列．

出力

OUTPUT ：`Map<int, ObjectRef>` または `Matrix<float>` 型．入力と同じものが出力される．

パラメータ

BASENAME : `string` 型．デフォルトは sep_．ファイル名のプレフィックスを指定する．出力されるファイル名は，音源 ID が付いている場合は “BASENAME_ID．wav” となる．3 つの混合音を分離した結果の分離音のファイル名は，BASENAME が sep_ のとき，`sep_0.wav`，`sep_1.wav`，`sep_2.wav` などとなる．

ADVANCE : `int` 型．他のノードの ADVANCE の値と揃える必要がある．

SAMPLING_RATE : `int` 型．他のノードの SAMPLING_RATE の値と揃える必要がある．この値はヘッダに書き込むために用いられるだけであり，このパラメータを変更しても A/D 変換におけるサンプリングレートを変更することはできない．

BITS : `string` 型．ファイルに保存する音声波形の量子化ビット数．int16 または int24 を指定可．

ノードの詳細

保存されるファイルのフォーマット: 保存されるファイルは，ヘッダ情報を持つ WAVE ファイルとして記録される．したがって，ファイルを読む際には，特にサンプリング周波数とトラック数，量子化ビット数を指定しなくてもよい．

また，入力の型によって書き出されるファイルは次のように異なる．

`Matrix<float>` 型 このとき書き出されるファイルは，入力の行の数だけチャネルを持った多チャネル音声データを含む WAVE ファイルとなる．

`Map<int, ObjectRef>` 型 このとき書き出されるファイルは，BASENAME の後に ID 番号が付与されたファイル名で，各 ID ごとにモノラル音声データが書き出される (1 つのファイルには 1 つの ID に対応する音声データのみ)．

6.1.5 HarkDataStreamSender

ノードの概要

以下の音響信号処理結果をソケット通信で送信するノードである．

- 音響信号
- STFT 後の周波数スペクトル
- 音源定位結果のソース情報
- 音響特徴量
- ミッシングフィーチャーマスク

必要なファイル

無し．

使用方法

どんなときに使うのか

上記のデータの中で必要な情報を TCP/IP 通信を用いて，HARK 外のシステムに送信するために用いる．

典型的な接続例

図 6.10 の例では全ての入力端子に接続している．送信したいデータに合わせて入力端子を開放することも可能である．入力端子の接続と送信されるデータの関係については「ノードの詳細」を参照．

ノードの入出力とプロパティ

表 6.6: HarkDataStreamSender のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
HOST	string	localhost		データの送信先サーバのホスト名/IP アドレス
PORT	int	5530		ネットワーク送出用ポート番号
ADVANCE	int	160	[pt]	フレームのシフト長
BUFFER_SIZE	int	512		ソケット通信のために確保する float 配列のサイズ
FRAMES_PER_SEND	int	1	[frm]	ソケット通信を何フレームに一回行うか
DEBUG_PRINT	bool	false		デバッグ情報出力の ON/OFF
SOCKET_ENABLE	bool	true		ソケット出力をするかどうかを決めるフラグ

入力

MIC_WAVE ：Matrix<float> 型．音響信号（チャンネル数 × 各チャンネルの STFT の窓長サイズの音響信号）

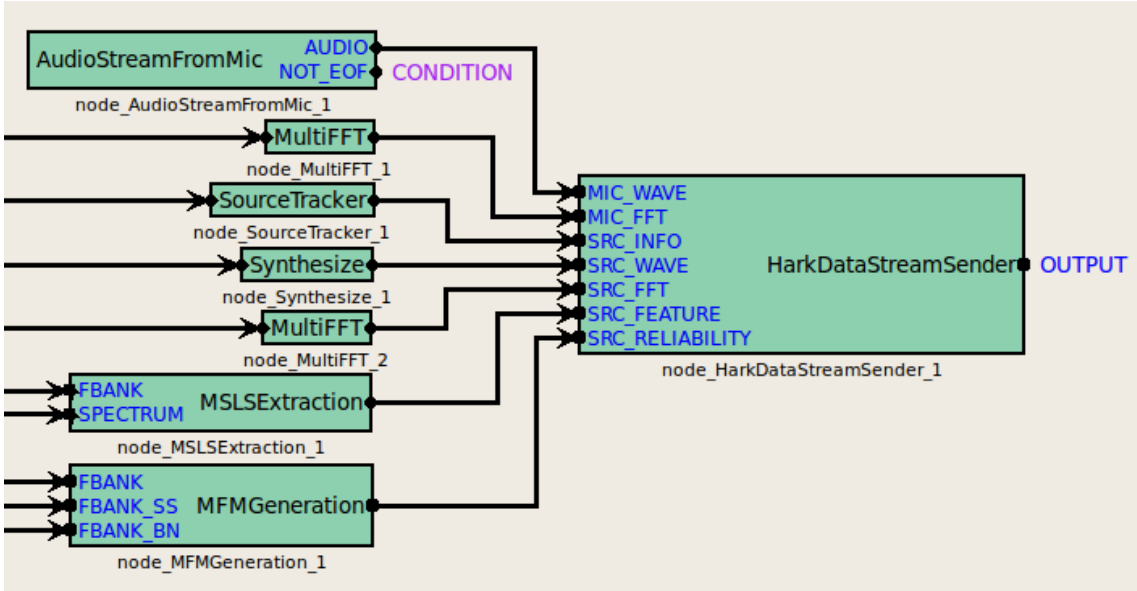


図 6.10: HarkDataStreamSender の接続例

MIC_FFT : `Matrix<complex<float> >` 型．周波数スペクトル（チャンネル数 × 各チャンネルのスペクトル）

SRC_INFO : `Vector<ObjectRef>` 型．音源数個の音源定位結果のソース情報

SRC_WAVE : `Map<int, ObjectRef>` 型．音源 ID と音響信号の `Vector<float>` 型のデータのペア．

SRC_FFT : `Map<int, ObjectRef>` 型．音源 ID と周波数スペクトルの `Vector<complex<float> >` 型のデータのペア．

SRC_FEATURE : `Map<int, ObjectRef>` 型．音源 ID と音響特徴量の `Vector<float>` 型のデータのペア．

SRC_RELIABILITY : `Map<int, ObjectRef>` 型．音源 ID とマスクベクトルの `Vector<float>` 型のデータのペア．

出力

OUTPUT : `ObjectRef` 型．入力と同じものが出力される．

パラメータ

HOST : `string` 型．データ送信先ホストの IP アドレス．SOCKET_ENABLED が `false` の場合は無効．

PORT : `int` 型．ソケット番号． SOCKET_ENABLED が `false` の場合は無効．

ADVANCE : `int` 型．フレームのシフト長．前段処理と同じ値にする．

BUFFER_SIZE : `int` 型．ソケット通信のために確保するバッファサイズ．

FRAMES_PER_SEND : `int` 型．ソケット通信を何フレームに一回行うかを指定する．デフォルトは 1．

DEBUG_PRINT : `bool` 型．デバッグ標準出力の ON/OFF

SOCKET_ENABLE : `bool` 型．`true` でデータをソケットに転送し，`false` で転送しない．

ノードの詳細

(A) パラメータの説明

HOST は，データを送信する外部プログラムが動作するホストのホスト名，または IP アドレスを指定する．
PORT は，データを送信するネットワークポート番号を指定する．
ADVANCE はフレームのシフト長であり，前段処理の設定値と同じにする．
BUFFER_SIZE はソケット通信のために確保するバッファサイズ．BUFFER_SIZE * 1024 の float 型の配列が初期化時に確保される．送信するデータより大きく確保する．
FRAMES_PER_SEND はソケット通信を何フレームに一回行うかを指定する．通常は 1 で問題ないが，通信量を削減したい時に使用できる．
DEBUG_PRINT はデバッグ標準出力の表示の可否である．送信データの一部と同じ情報が出力される．表示される内容については表 6.13 の「 Debug 」を参照.
SOCKET_ENABLED が `false` のときは，データを外部システムに送信しない．これは，外部プログラムを動かさずに HARK のネットワーク動作チェックを行うために使用する．

(B) データ送信の詳細

(B-1) データ送信用構造体

データの送信は，各フレーム毎に幾つかに分けられて行われる．データ送信のために定義されている構造体を下記にリストアップする．

- **HD_Header**
  説明 : 送信データの先頭で送信される基本情報が入ったヘッダ
  データサイズ : `3 * sizeof(int)`

表 6.7: HD_Header のメンバ

変数名	型	説明
type	int	送信データの構造を示すビットフラグ．各ビットと送信データとの関係については表 6.8 参照.
advance	int	フレームのシフト長
count	int	HARK のフレーム番号

表 6.8: HD_Header の type の各ビットと送信データ

桁数	関係入力端子	送信データ
1 桁目	MIC_WAVE	音響信号
2 桁目	MIC_FFT	周波数スペクトル
3 桁目	SRC_INFO	音源定位結果ソース情報
4 桁目	SRC_INFO, SRC_WAVE	音源定位結果ソース情報 + 音源 ID 毎の音響信号
5 桁目	SRC_INFO, SRC_FFT	音源定位結果ソース情報 + 音源 ID 毎の周波数スペクトル
6 桁目	SRC_INFO, SRC_FEATURE	音源定位結果ソース情報 + 音源 ID 毎の音響特徴量
7 桁目	SRC_INFO, SRC_RELIABILITY	音源定位結果ソース情報 + 音源 ID 毎のミッシングフィーチャーマスク

HarkDataStreamSender は入力端子の開放の可否によって送信されるデータが異なり，データ受信側では type によって，受信データを解釈できる．以下に例を挙げる．送信されるデータの更なる詳細については (B-2) に示す．

例 1) MIC_FFT 入力端子のみが接続されている場合，type は 2 進数で表すと 0000010 となる．また，送信されるデータはマイク毎の周波数スペクトルのみとなる．

例 2) MIC_WAVE, SRC_INFO, SRC_FEATURE の 3 つの入力端子が接続されている場合，type は 2 進数で表すと 0100101 となる．送信されるデータはマイク毎の音響信号，音源定位結果のソース情報，音源 ID 毎の音響特徴量となる．

注) SRC_WAVE, SRC_FFT, SRC_FEATURE, SRC_RELIABILITY の 4 つの入力端子については，音源 ID ごとの情報になるため，SRC_INFO の情報が必須である．もし，SRC_INFO を接続せずに，上記 4 つの入力端子を接続したとしても，何も送信されない．その場合，type は 2 進数で 0000000 となる．

- **HDH_MicData**
  説明：2 次元配列を送信するための，サイズに関する配列の構造情報
  データサイズ：`3 * sizeof(int)`

表 6.9: HDH_MicData のメンバ

変数名	型	説明
nch	int	マイクチャネル数（送信する 2 次元配列の行数）
length	int	データ長（送信する 2 次元配列の列数）
data_bytes	int	送信データバイト数． float 型の行列の場合は `nch * length * sizeof(float)` となる．

- **HDH_SrcInfo**
  説明：音源定位結果のソース情報
  データサイズ：`1 * sizeof(int) + 4 * sizeof(float)`

表 6.10: HDH_SrcInfo のメンバ

変数名	型	説明
src_id	int	音源 ID
x[3]	float	音源 3 次元位置
power	float	LocalizeMUSIC で計算される MUSIC スペクトルのパワー

- **HDH_SrcData**
  説明：1 次元配列を送信するための，サイズに関する配列の構造情報
  データサイズ：`2 * sizeof(int)`

(B-2) 送信データ

送信データは各フレーム毎に，表 6.12，表 6.13 の (a)-(q) のように，分割されて行われる．表 6.12 に，送信データ (a)-(q) と，接続された入力端子の関係を，表 6.13 に，送信データの説明を示す．

表 6.11: HDH_SrcData のメンバ

変数名	型	説明
length	int	データ長（送信する 1 次元配列の要素数）
data_bytes	int	送信データバイト数．float 型のベクトルの場合は `length * sizeof(float)` となる．

表 6.12: 送信順のデータリストと接続入力端子（○ 記号の箇所が送信されるデータ，○* は，SRC_INFO 端子が接続されていない場合は送信されないデータ）

送信データ詳細			入力端子と送信データ						
	型	サイズ	MIC_WAVE	MIC_FFT	SRC_INFO	SRC_WAVE	SRC_FFT	SRC_FEATURE	SRC_RELIABILITY
(a)	HD_Header	`sizeof(HD_Header)`	○	○	○	○	○	○	○
(b)	HDH_MicData	`sizeof(HDH_MicData)`	○						
(c)	float[]	`HDH_MicData.data_bytes`	○						
(d)	HDH_MicData	`sizeof(HDH_MicData)`		○					
(e)	float[]	`HDH_MicData.data_bytes`		○					
(f)	float[]	`HDH_MicData.data_bytes`		○					
(g)	int	`1 * sizeof(int)`			○	○*	○*	○*	○*
(h)	HDH_SrcInfo	`sizeof(HDH_SrcInfo)`			○	○*	○*	○*	○*
(i)	HDH_SrcData	`sizeof(HDH_SrcData)`				○*			
(j)	short int[]	`HDH_SrcData.data_bytes`				○*			
(k)	HDH_SrcData	`sizeof(HD_SrcData)`					○*		
(l)	float[]	`HDH_SrcData.data_bytes`					○*		
(m)	float[]	`HDH_SrcData.data_bytes`					○*		
(n)	HDH_SrcData	`sizeof(HD_SrcData)`						○*	
(o)	float[]	`HDH_SrcData.data_bytes`						○*	
(p)	HDH_SrcData	`sizeof(HD_SrcData)`							○*
(q)	float[]	`HDH_SrcData.data_bytes`							○*

(B-3) 送信アルゴリズム

HARK のネットワークファイルを実行する際に繰り返し演算される部分のアルゴリズムを以下に示す．

表 6.13: 送信データ詳細

	説明	Debug
(a)	送信データヘッダ．表 6.7 参照．	○
(b)	音響信号の構造（マイク数，フレーム長，送信バイト数）を表す構造体．表 6.9 参照．	○
(c)	音響信号（マイク数 × フレーム長 の float 型の行列）	
(d)	周波数スペクトルの構造（マイク数，周波数ビン数，送信バイト数）を表す構造体．表 6.9 参照．	○
(e)	周波数スペクトルの実部（マイク数 × 周波数ビン数 の float 型の行列）	
(f)	周波数スペクトルの虚部（マイク数 × 周波数ビン数 の float 型の行列）	
(g)	検出された音源個数	○
(h)	音源定位結果のソース．表 6.10 参照．	○
(i)	音源 ID 毎の音響信号の構造（フレーム長，送信バイト数）を表す構造体．表 6.11 参照．	○
(j)	音源 ID 毎の音響信号（フレーム長 の float 型の一次元配列）	
(k)	音源 ID 毎の周波数スペクトルの構造（周波数ビン数，送信バイト数）を表す構造体．表 6.11 参照．	○
(l)	音源 ID 毎の周波数スペクトルの実部（周波数ビン数 の float 型の一次元配列）	
(m)	音源 ID 毎の周波数スペクトルの虚部（周波数ビン数 の float 型の一次元配列）	
(n)	音源 ID 毎の音響特徴量の構造（特徴量次元数，送信バイト数）を表す構造体．表 6.11 参照．	○
(o)	音源 ID 毎の音響特徴量（特徴量次元数 の float 型の一次元配列）	
(p)	音源 ID 毎の MFM の構造（特徴量次元数，送信バイト数）を表す構造体．表 6.11 参照．	○
(q)	音源 ID 毎の MFM （特徴量次元数 の float 型の一次元配列）	

```
calculate{

    Send (a)

    IF MIC_WAVE is connected
        Send (b)
        Send (c)
    ENDIF

    IF MIC_FFT is connected
        Send (d)
        Send (e)
        Send (f)
    ENDIF

    IF SRC_INFO is connected

        Send (g) (Let the number of sounds 'src_num'.)

        FOR i = 1 to src_num (This is a sound ID based routine.)

            Send (h)

            IF SRC_WAVE is connected
                Send (i)
                Send (j)
            ENDIF

            IF SRC_FFT is connected
                Send (k)
                Send (l)
                Send (m)
            ENDIF

            IF SRC_FEATURE is connected
                Send (n)
                Send (o)
            ENDIF

            IF SRC_RELIABILITY is connected
                Send (p)
                Send (q)
            ENDIF

        ENDFOR

    ENDIF

}
```

ここで，コード内の (a)-(q) が，表 6.12 と表 6.13 の (a)-(q) に対応している．

6.2 Localization カテゴリ

6.2.1 CMLoad

ノードの概要

音源定位のための相関行列をファイルから読み込む.

必要なファイル

CMSave で保存する形式のファイル.

使用方法

どんなときに使うのか

CMSave で保存した音源定位用の相関行列を読み込む時に使用する.

典型的な接続例

図 6.11 に CMLoad ノードの使用例を示す.

FILENAMER と FILENAMEI は `string` 型の入力で，それぞれ相関行列の実数部と虚数部の入った読み込みファイル名を表す．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列を読み込むタイミングを指定する．使用例では FILENAMER, FILENAMEI, OPERATION_FLAG の全ての入力に対して，Constant ノードを接続しており，実行時のパラメータは不変となっているが，前段のノードの出力を動的にすることで，使用する相関行列を変更することが可能である.

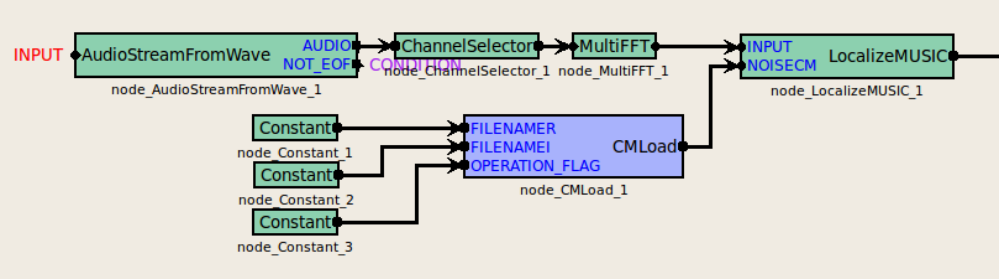


図 6.11: CMLoad の接続例

ノードの入出力とプロパティ

表 6.14: CMLoad のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
ENABLE_DEBUG	`bool`	false		デバッグ情報出力の ON/OFF

入力

FILENAMER ：string 型．読み込む相関行列の実部のファイル名．

FILENAMEI ：string 型．読み込む相関行列の虚部のファイル名．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時，かつファイル名が変更した時にのみ相関行列が読み込まれる．

出力

OUTPUTCM ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個出力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

パラメータ

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が読み込まれる時に標準出力にその旨が出力される．

ノードの詳細

周波数ビン毎の M 次の複素正方行列である相関行列の実部と虚部をそれぞれ二つのファイルから Matrix<float> 形式で読み込む．

周波数ビン数を k とする ($k = NFFT/2+1$) と，読み込みファイルはそれぞれ k 行 M^2 列の行列で構成されている．読み込みは OPERATION_FLAG が 1 もしくは真の時に，ネットワーク動作直後，または読み込みファイル名が変化した時に限り行われる．

6.2.2 CMSave

ノードの概要

音源定位のための相関行列をファイルに保存する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT や CMMakerFromFFTwithFlag 等から作成した音源定位用の相関行列を保存する時に使用する．

典型的な接続例

図 6.12 に CMSave ノードの使用例を示す．

INPUTCM 入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．FILENAMER と FILENAMEI は `string` 型の入力で，それぞれ相関行列の実数部と虚数部の保存ファイル名を表す．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列を保存するタイミングを指定する（図 6.12 では，一例として Equal ノードを接続しているが，`int` 型や `bool` 型を出力できるノードであれば何でも構わない）．

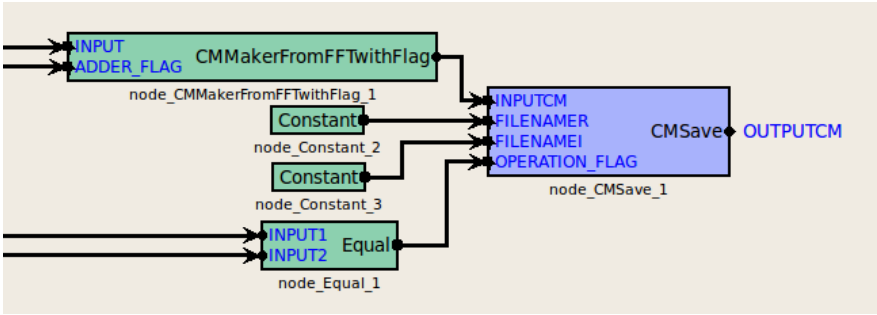


図 6.12: CMSave の接続例

ノードの入出力とプロパティ

表 6.15: CMSave のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
ENABLE_DEBUG	bool	false		デバッグ情報出力の ON/OFF

入力

INPUTCM ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

FILENAMER ：string 型．保存する相関行列の実部のファイル名．

FILENAMEI ：string 型．保存する相関行列の虚部のファイル名．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時にのみ相関行列が保存される．

出力

OUTPUTCM ：Matrix<complex<float> > 型．INPUTCM に同じ．

パラメータ

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が保存される時に，標準出力に保存した時のフレーム番号が出力される．

ノードの詳細

周波数ビン毎の M 次の複素正方行列である相関行列を Matrix<float> 形式に直し，指定したファイル名で保存する．

保存ファイルは相関行列の実部と虚部に分割される．周波数ビン数を k とする ($k = NFFT/2+1$) と，保存ファイルはそれぞれ k 行 M^2 列の行列が格納される．保存は OPERATION_FLAG が 1 もしくは真の時に限り行われる．

6.2.3 CMChannelSelector

ノードの概要

マルチチャネルの相関行列から，指定したチャネルのデータだけを指定した順番に取り出す．

必要なファイル

無し．

使用方法

どんなときに使うのか

入力されたマルチチャネルの相関行列の中から，必要のないチャネルを削除したいとき，あるいは，チャネルの並びを入れ替えたいとき．あるいは，チャネルを複製したいとき．

典型的な接続例

図 6.13 に CMChannelSelector ノードの使用例を示す．

入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．

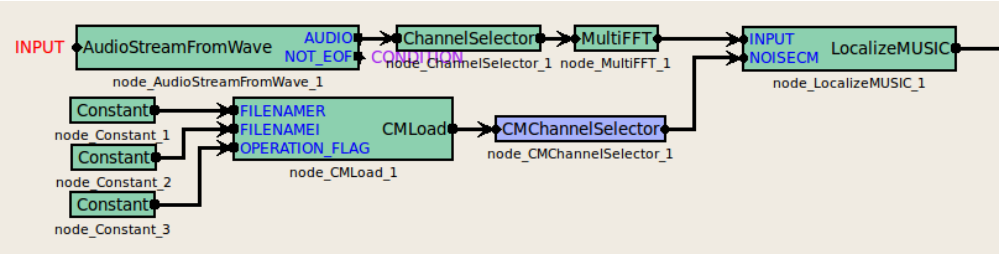


図 6.13: CMChannelSelector の接続例

ノードの入出力とプロパティ

表 6.16: CMChannelSelector のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
SELECTOR	`Vector<int>`	`<Vector<int> >`		出力するチャネルの番号を指定

入力

INPUTCM ：`Matrix<complex<float> >` 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．`Matrix<complex<float> >` の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

出力

OUTPUTCM : Matrix<complex<float> > 型．INPUTCM に同じ．

パラメータ

SELECTOR : Vector<int> 型，デフォルト値は無し（<Vector<int> >）．使用するチャネルの，チャネル番号を指定する．チャネル番号は 0 からはじまる．

例: 5 チャネル (0-4) のうち 2，3，4 チャネルだけを使うときは <Vector<int> 2 3 4> のように，3 チャネルと 4 チャネルを入れ替えたい時は <Vector<int> 0 1 2 4 3 5> のように指定する．

ノードの詳細

相関行列が格納された入力データの $k \times M \times M$ 型の複素三次元配列から指定したチャネルの相関行列だけを抽出し，新たな $k \times M' \times M'$ 型の複素三次元配列のデータを出力する．ただし，k は周波数ビン数 ($k = NFFT/2+1$)，M は入力チャネル数，M' は出力チャネル数．

6.2.4 CMMakerFromFFT

ノードの概要

MultiFFT ノードから出力されるマルチチャネル複素スペクトルから，音源定位のための相関行列を一定周期で生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

LocalizeMUSIC ノードの音源定位において，雑音等の特定の音源を抑圧したい場合は，あらかじめ雑音情報を含む相関行列を用意する必要がある．本ノードは，MultiFFT ノードから出力されるマルチチャネル複素スペクトルから，相関行列を一定周期で生成する．本ノードの出力を LocalizeMUSIC ノードの NOISECM 入力端子に接続することで，一定周期前の情報を常に雑音とみなして抑圧した音源定位が実現できる．

典型的な接続例

図 6.14 に CMMakerFromFFT ノードの使用例を示す．

INPUT 入力端子へは，MultiFFT ノードから計算される入力信号の複素スペクトルを接続する．型は `Matrix<complex<float> >` 型である．本ノードは入力信号の複素スペクトルから周波数ビン毎にチャネル間の相関行列を計算し出力する．出力の型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している．

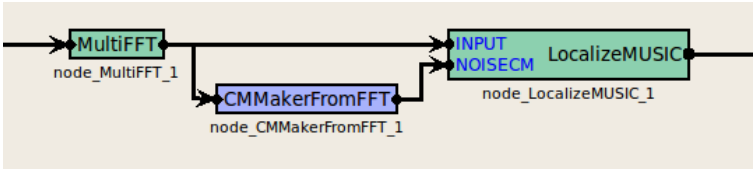


図 6.14: CMMakerFromFFT の接続例

ノードの入出力とプロパティ

表 6.17: CMMakerFromFFT のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
WINDOW	`int`	50		相関行列の平滑化フレーム数
PERIOD	`int`	50		相関行列の更新フレーム周期
ENABLE_DEBUG	`bool`	false		デバッグ情報出力の ON/OFF

入力

INPUT : `Matrix<complex<float> >`, 入力信号の複素スペクトル表現 $M \times (NFFT/2+1)$.

出力

OUTPUT : `Matrix<complex<float> >` 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個出力される．`Matrix<complex<float> >` の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M * M$ 列) を表す．

パラメータ

WINDOW : `int` 型．50 がデフォルト値．相関行列計算時の平滑化フレーム数を指定する．ノード内では，入力信号の複素スペクトルから相関行列を毎フレーム生成し，WINDOW で指定されたフレームで加算平均を取ったものが新たな相関行列として出力される．PERIOD フレーム間は最後に計算された相関行列が出力される．この値を大きくすると，相関行列が安定するが計算負荷が高い．

PERIOD : `int` 型．50 がデフォルト値．相関行列の更新フレーム周期を指定する．ノード内では，入力信号の複素スペクトルから相関行列を毎フレーム生成し，WINDOW で指定されたフレームで加算平均を取ったものが新たな相関行列として出力される．PERIOD フレーム間は最後に計算された相関行列が出力される．この値を大きくすると，相関行列の時間解像度が改善される計算負荷が高い．

ENABLE_DEBUG : `bool` 型．false がデフォルト値．true の場合は相関行列が生成される時に，標準出力に生成した時のフレーム番号が出力される．

ノードの詳細

MultiFFT ノードから出力される入力信号の複素スペクトルを以下のように表す．

$$\boldsymbol{X}(\omega, f) = [X_1(\omega, f), X_2(\omega, f), X_3(\omega, f), \cdots, X_M(\omega, f)]^T \tag{6.1}$$

ここで，ω は周波数ビン番号，f は HARK で扱うフレーム番号，M は入力チャネル数を表す．

入力信号 $\boldsymbol{X}(\omega, f)$ の相関行列は，各周波数，各フレームごとに以下のように定義できる．

$$\boldsymbol{R}(\omega, f) = \boldsymbol{X}(\omega, f)\boldsymbol{X}^*(\omega, f) \tag{6.2}$$

ここで，$()^*$ は複素共役転置演算子を表す．理論上は，この $\boldsymbol{R}(\omega, f)$ をそのまま以降の処理で利用すれば問題はないが，実用上，安定した相関行列を得るため，HARK では，次のように時間方向に平均したものを使用している．

$$\boldsymbol{R}'(\omega, f) = \frac{1}{\mathrm{WINDOW}} \sum_{i=0}^{\mathrm{WINDOW}-1} \boldsymbol{R}(\omega, f+i) \tag{6.3}$$

$\boldsymbol{R}'(\omega, f)$ が CMMakerFromFFT ノードの OUTPUT 端子から PERIOD で指定したフレーム周期ごとに出力される．

6.2.5 CMMakerFromFFTwithFlag

ノードの概要

MultiFFT ノードから出力されるマルチチャネル複素スペクトルから，音源定位のための相関行列を入力フラグで指定した区間で生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT ノードと使用する場面は同じであり，詳細は CMMakerFromFFT ノードの項を参照されたい．相違点は，相関行列の計算区間である．CMMakerFromFFT ノードでは，一定周期 (PERIOD) 毎に相関行列の更新が行われたが，本ノードでは，入力端子から得られるフラグの値に合わせて指定した区間の相関行列を生成することができる．

典型的な接続例

図 6.15 に CMMakerFromFFTwithFlag ノードの使用例を示す．

INPUT 入力端子へは，MultiFFT ノードから計算される入力信号の複素スペクトルを接続する．型は `Matrix<complex<float> >` 型である．ADDER_FLAG は `int` 型，または `bool` 型の入力で，相関行列計算に関するイベントを制御する．イベント制御の詳細はノードの詳細の項に譲る．本ノードは入力信号の複素スペクトルから周波数ビン毎にチャネル間の相関行列を計算し出力する．出力の型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している．

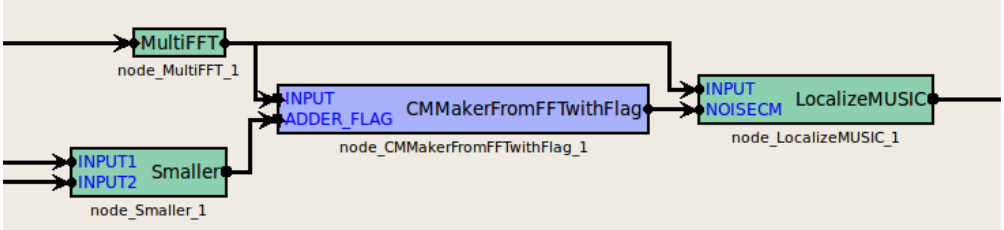


図 6.15: CMMakerFromFFTwithFlag の接続例

ノードの入出力とプロパティ

入力

INPUT : `Matrix<complex<float> >`, 入力信号の複素スペクトル表現 $M \times (NFFT/2+1)$.

ADDER_FLAG : `int` 型，または `bool` 型．相関行列計算に関するイベントを制御する．イベント制御の詳細についてはノードの詳細の項を参照されたい．

表 6.18: CMMakerFromFFTwithFlag のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
DURATION_TYPE	string	FLAG_PERIOD		フラグ値に従うかフレーム周期に従うかの選択
WINDOW	int	50		相関行列の平滑化フレーム数
PERIOD	int	50		相関行列の更新フレーム周期
MAX_SUM_COUNT	int	100		相関行列の平滑化フレーム数の最大値
ENABLE_ACCUM	bool	false		過去の相関行列との加算平均を取るかの選択
ENABLE_DEBUG	bool	false		デバッグ情報出力の ON/OFF

出力

OUTPUT : Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個出力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

パラメータ

DURATION_TYPE : string 型．FLAG_PERIOD がデフォルト値．相関行列を更新する周期と加算平均する区間をフラグ値に従うか，フレーム周期に従うかを選択する．詳細はノードの詳細の項を参照されたい．

WINDOW : int 型．50 がデフォルト値．DURATION_TYPE=FRAME_PERIOD の時のみ指定する．相関行列計算時の平滑化フレーム数を指定する．ノード内では，入力信号の複素スペクトルから相関行列を毎フレーム生成し，WINDOW で指定されたフレームで加算平均を取ったものが新たな相関行列として出力される．PERIOD フレーム間は最後に計算された相関行列が出力される．この値を大きくすると，相関行列が安定するが計算負荷が高い．

PERIOD : int 型．50 がデフォルト値．DURATION_TYPE=FRAME_PERIOD の時のみ指定する．相関行列の更新フレーム周期を指定する．ノード内では，入力信号の複素スペクトルから相関行列を毎フレーム生成し，WINDOW で指定されたフレームで加算平均を取ったものが新たな相関行列として出力される．PERIOD フレーム間は最後に計算された相関行列が出力される．この値を大きくすると，相関行列の時間解像度が改善される計算負荷が高い．

MAX_SUM_COUNT : int 型．100 がデフォルト値．DURATION_TYPE=FLAG_PERIOD の時のみ指定する．相関行列計算時の平滑化フレーム数の最大値を指定する．本ノードは，ADDER_FLAG によって相関行列の平滑化フレーム数を制御できる．このため，ADDER_FLAG が 常に 1 の場合は，相関行列の加算のみが行われ，いつまでも出力されないことになってしまう．そこで，MAX_SUM_COUNT を正しく設定することで，平滑化フレーム数の上限に来た時に強制的に相関行列を出力することができる．この機能を OFF にするには MAX_SUM_COUNT = 0 を指定すれば良い．

ENABLE_ACCUM : bool 型．false がデフォルト値．DURATION_TYPE=FLAG_PERIOD の時のみ指定する．過去に生成した相関行列も含めて加算平均を取るかどうかを指定できる．

ENABLE_DEBUG : bool 型．false がデフォルト値．true の場合は相関行列が生成される時に，標準出力に生成した時のフレーム番号が出力される．

ノードの詳細

CMMakerFromFFT ノードと相関行列算出のアルゴリズムは同じであり，詳細は CMMakerFromFFT のノードの詳細を参照されたい．CMMakerFromFFT ノードとの相違点は相関行列の平滑化フレームを ADDER_FLAG 入力端子のフラグによって制御できることである．

CMMakerFromFFT　ノードでは，PERIOD で指定したフレーム数によって以下の式で相関行列を算出していた．

$$\boldsymbol{R}'(\omega, f) = \frac{1}{\mathrm{PERIOD}} \sum_{i=0}^{\mathrm{PERIOD}-1} \boldsymbol{R}(\omega, f+i) \tag{6.4}$$

本ノードでは DURATION_TYPE=FLAG_PERIOD の場合，ADDER_FLAG の値によって次のように相関行列を生成する．

A) ADDER_FLAG が 0（または偽）から 1（または真）に変化する時

- 相関行列を零行列に戻し，PERIOD を 0 に戻す．
  $\boldsymbol{R}'(\omega) = \boldsymbol{O}$
  PERIOD = 0
  ただし，$\boldsymbol{O} \in \mathbb{C}^{(NFFT/2+1)\times M\times M}$ は零行列を表す．

B) ADDER_FLAG が 1（または真）を保持する区間

- 相関行列を加算する．
  $\boldsymbol{R}'(\omega) = \boldsymbol{R}'(\omega) + \boldsymbol{R}(\omega, f+i)$
  PERIOD = PERIOD + 1

C) ADDER_FLAG が 1（または真）から 0（または偽）に変化する時

- 加算した相関行列の平均を取って OUTPUT から出力する．
  $\boldsymbol{R}_{out}(\omega, f) = \frac{1}{\mathrm{PERIOD}}\boldsymbol{R}'(\omega)$

D) ADDER_FLAG が 0（または偽）を保持する区間

- 最後に生成した相関行列を保持する．
  $\boldsymbol{R}_{out}(\omega, f)$

ここで，$\boldsymbol{R}_{out}(\omega, f)$ が OUTPUT 端子から出力される相関行列となる．つまり，新たな相関行列が $\boldsymbol{R}_{out}(\omega, f)$ に格納されるのは C) のフェイズとなる．

また，DURATION_TYPE=FRAME_PERIOD の場合，ADDER_FLAG が 1（または真）の時のみ，式 (6.4) が実行され，相関行列の更新が起こる．

6.2.6 CMDivideEachElement

ノードの概要

音源定位のための二つの相関行列を成分ごとに除算する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT，CMMakerFromFFTwithFlag から作成した音源定位用の相関行列の演算ノードの一つで，成分毎に除算する機能を持つ．

典型的な接続例

図 6.16 に CMDivideEachElement ノードの使用例を示す．

CMA 入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．CMB 入力端子も CMA と同じく相関行列を接続する．除算の際は，CMA ./ CMB が演算される．ただし ./ は成分ごとの割り算を表す．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列の除算を計算するタイミングを指定する．

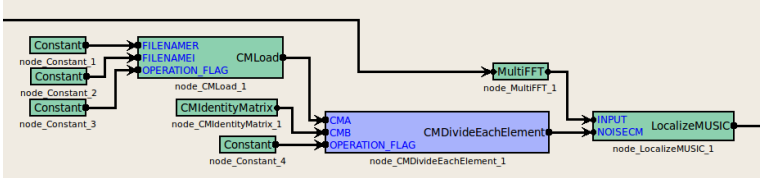


図 6.16: CMDivideEachElement の接続例

ノードの入出力とプロパティ

表 6.19: CMDivideEachElement のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FIRST_FRAME_EXECUTION	`bool`	false		1 フレーム目だけ演算を実行するかの選択
ENABLE_DEBUG	`bool`	false		デバッグ情報出力の ON/OFF

入力

CMA ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

CMB ：Matrix<complex<float> > 型．CMA に同じ．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時にのみ相関行列の演算が実行される．

出力

OUTPUTCM ：Matrix<complex<float> > 型．CMA ./ CMB に相当する除算後の相関行列が出力される．

パラメータ

FIRST_FRAME_EXECUTION ：bool 型．false がデフォルト値．true の場合は OPERATION_FLAG が常に 0 または偽であった場合にも 1 フレーム目のみ演算が実行される．

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が除算される時に，標準出力に計算した時のフレーム番号が出力される．

ノードの詳細

二つの相関行列の成分毎の除算を行う．相関行列の行列としての除算でないことに注意されたい．相関行列は $k \times M \times M$ の複素三次元配列であり，$k \times M \times M$ 回の除算が以下のように行われる．ただし，k は周波数ビン数 ($k = NFFT/2+1$)，M は入力信号のチャネル数である．

```
OUTPUTCM = zero_matrix(k,M,M)

calculate{

   IF OPERATION_FLAG
       FOR i = 1 to k
           FOR i = 1 to M
               FOR i = 1 to M

                   OUTPUTCM[i][j][k] = CMA[i][j][k] / CMB[i][j][k]

               ENDFOR
           ENDFOR
       ENDFOR
   ENDIF

}
```

OUTPUTCM 端子から出力される行列は，零行列として初期化され，以降は最後の演算結果を保持する．

6.2.7 CMMultiplyEachElement

ノードの概要

音源定位のための二つの相関行列を成分ごとに乗算する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT，CMMakerFromFFTwithFlag から作成した音源定位用の相関行列の演算ノードの一つで，成分毎に乗算する機能を持つ．

典型的な接続例

図 6.17 に CMMultiplyEachElement ノードの使用例を示す．

CMA 入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．CMB 入力端子も CMA と同じく相関行列を接続する．乗算の際は，CMA .∗ CMB が演算される．ただし .∗ は成分ごとの乗算を表す．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列の演算を実行するタイミングを指定する．

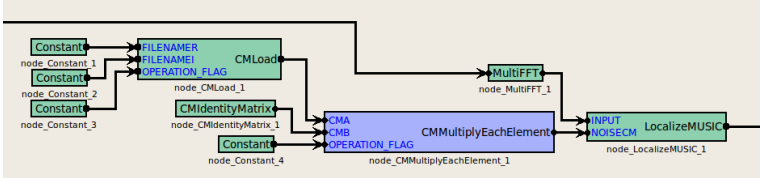


図 6.17: CMMultiplyEachElement の接続例

ノードの入出力とプロパティ

表 6.20: CMMultiplyEachElement のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FIRST_FRAME_EXECUTION	`bool`	false		1 フレーム目だけ演算を実行するかの選択
ENABLE_DEBUG	`bool`	false		デバッグ情報出力の ON/OFF

入力

CMA ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

CMB ：Matrix<complex<float> > 型．CMA に同じ．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時にのみ相関行列の演算が実行される．

出力

OUTPUTCM ：Matrix<complex<float> > 型．CMA .* CMB に相当する乗算後の相関行列が出力される．

パラメータ

FIRST_FRAME_EXECUTION ：bool 型．false がデフォルト値．true の場合は OPERATION_FLAG が常に 0 または偽であった場合にも 1 フレーム目のみ演算が実行される．

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が乗算される時に，標準出力に乗算した時のフレーム番号が出力される．

ノードの詳細

二つの相関行列の成分毎の乗算を行う．相関行列の行列としての乗算でないことに注意されたい（行列としての乗算は CMMultiplyMatrix を参照）．相関行列は $k \times M \times M$ の複素三次元配列であり，$k \times M \times M$ 回の乗算が以下のように行われる．ただし，k は周波数ビン数 ($k = NFFT/2+1$)，M は入力信号のチャネル数である．

```
OUTPUTCM = zero_matrix(k,M,M)

calculate{

   IF OPERATION_FLAG
       FOR i = 1 to k
           FOR i = 1 to M
               FOR i = 1 to M

                   OUTPUTCM[i][j][k] = CMA[i][j][k] * CMB[i][j][k]

               ENDFOR
           ENDFOR
       ENDFOR
   ENDIF

}
```

OUTPUTCM 端子から出力される行列は，零行列として初期化され，以降は最後の演算結果を保持する．

6.2.8 CMInverseMatrix

ノードの概要

音源定位のための相関行列の逆行列を演算する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT，CMMakerFromFFTwithFlag から作成した音源定位用の相関行列の演算ノードの一つで，相関行列の逆行列を演算する機能を持つ．

典型的な接続例

図 6.18 に CMInverseMatrix ノードの使用例を示す．

INPUTCM 入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列の逆行列を計算するタイミングを指定する．

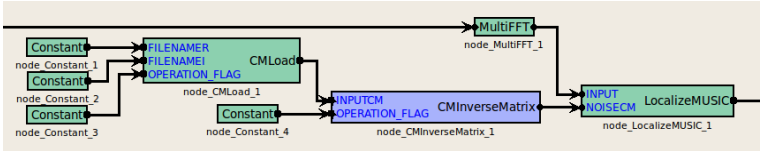


図 6.18: CMInverseMatrix の接続例

ノードの入出力とプロパティ

表 6.21: CMInverseMatrix のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FIRST_FRAME_EXECUTION	`bool`	false		1 フレーム目だけ演算を実行するかの選択
ENABLE_DEBUG	`bool`	false		デバッグ情報出力の ON/OFF

入力

INPUTCM ：`Matrix<complex<float> >` 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．`Matrix<complex<float> >` の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時にのみ相関行列の演算が実行される．

出力

OUTPUTCM ：Matrix<complex<float> > 型．逆行列演算後の相関行列が出力される．

パラメータ

FIRST_FRAME_EXECUTION ：bool 型．false がデフォルト値．true の場合は OPERATION_FLAG が常に 0 または偽であった場合にも 1 フレーム目のみ演算が実行される．

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が計算される時に，標準出力に計算した時のフレーム番号が出力される．

ノードの詳細

相関行列の逆行列の演算を行う．相関行列は $k \times M \times M$ の複素三次元配列であり，k 回の逆行列演算が以下のように行われる．ただし，k は周波数ビン数 ($k = NFFT/2 + 1$)，M は入力信号のチャネル数である．

```
OUTPUTCM = zero_matrix(k,M,M)

calculate{

   IF OPERATION_FLAG
       FOR i = 1 to k
           OUTPUTCM[i] = inverse( INPUTCM[i] )
       ENDFOR
   ENDIF

}
```

OUTPUTCM 端子から出力される行列は，零行列として初期化され，以降は最後の演算結果を保持する．

6.2.9 CMMultiplyMatrix

ノードの概要

音源定位のための二つの相関行列を周波数ビン毎に乗算する．

必要なファイル

無し．

使用方法

どんなときに使うのか

CMMakerFromFFT，CMMakerFromFFTwithFlag から作成した音源定位用の相関行列の演算ノードの一つで，周波数ビン毎の相関行列を乗算する機能を持つ．

典型的な接続例

図 6.19 に CMMultiplyMatrix ノードの使用例を示す．

CMA 入力端子へは，CMMakerFromFFT や CMMakerFromFFTwithFlag 等から計算される相関行列を接続する（型は `Matrix<complex<float> >` 型だが，相関行列を扱うため，三次元複素配列を二次元複素行列に変換して出力している）．CMB 入力端子も CMA と同じく相関行列を接続する．乗算の際は，CMA ∗ CMB が周波数ビン毎に演算される．OPERATION_FLAG は `int` 型，または `bool` 型の入力で，相関行列の演算を実行するタイミングを指定する．

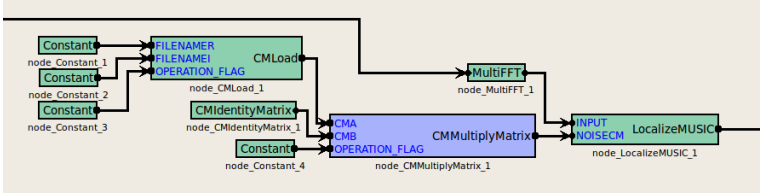


図 6.19: CMMultiplyMatrix の接続例

ノードの入出力とプロパティ

表 6.22: CMMultiplyMatrix のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FIRST_FRAME_EXECUTION	bool	false		1 フレーム目だけ演算を実行するかの選択
ENABLE_DEBUG	bool	false		デバッグ情報出力の ON/OFF

入力

CMA ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

CMB ：Matrix<complex<float> > 型．CMA に同じ．

OPERATION_FLAG ：int 型，または bool 型．本入力端子が 1 もしくは真の時にのみ相関行列の演算が実行される．

出力

OUTPUTCM ：Matrix<complex<float> > 型．CMA $*$ CMB に相当する乗算後の相関行列が出力される．

パラメータ

FIRST_FRAME_EXECUTION ：bool 型．false がデフォルト値．true の場合は OPERATION_FLAG が常に 0 または偽であった場合にも 1 フレーム目のみ演算が実行される．

ENABLE_DEBUG ：bool 型．false がデフォルト値．true の場合は相関行列が乗算される時に，標準出力に乗算した時のフレーム番号が出力される．

ノードの詳細

周波数ビン毎の二つの相関行列の乗算を行う．相関行列は $k \times M \times M$ の複素三次元配列であり，k 回の行列の乗算が以下のように行われる．ただし，k は周波数ビン数 ($k = NFFT/2+1$)，M は入力信号のチャネル数である．

```
OUTPUTCM = zero_matrix(k,M,M)

calculate{

   IF OPERATION_FLAG
       FOR i = 1 to k

           OUTPUTCM[i] = CMA[i] * CMB[i]

       ENDFOR
   ENDIF

}
```

OUTPUTCM 端子から出力される行列は，零行列として初期化され，以降は最後の演算結果を保持する．

6.2.10 CMIdentityMatrix

ノードの概要

単位行列が格納された相関行列を出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

LocalizeMUSIC ノードの NOISECM 入力端子に接続することで，LocalizeMUSIC ノードが持つ雑音抑圧機能を OFF にすることができる．

典型的な接続例

図 6.20 に CMIdentityMatrix ノードの使用例を示す．

本ノードは全ての周波数ビンに対して単位行列を持つ相関行列データをノード内で生成するため，入力端子は存在しない．出力端子から生成された相関行列が出力される．

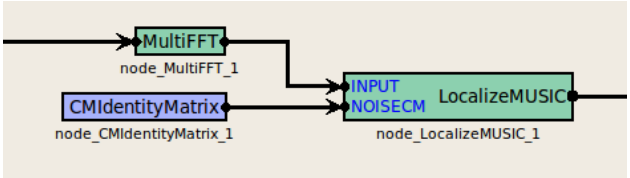


図 6.20: CMIdentityMatrix の接続例

ノードの入出力とプロパティ

表 6.23: CMIdentityMatrix のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
NB_CHANNELS	int	8		入力信号のチャネル数 M
LENGTH	int	512		処理を行う基本単位となるフレームの長さ $NFFT$

入力

無し．

出力

OUTPUTCM ：Matrix<complex<float> > 型．各周波数ビン毎の相関行列．M 次の単位行列である相関行列が $NFFT/2+1$ 個出力される．Matrix<complex<float> > の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．

パラメータ

NB_CHANNELS : int 型．入力信号のチャネル数．相関行列の次数と等価．前段で使用していた相関行列の次元を合わせる必要がある．8 がデフォルト値．

LENGTH : int 型．512 がデフォルト値．フーリエ変換の際の FFT 点数．前段までの FFT 点数と合わせる必要がある．

ノードの詳細

周波数ビン毎の M 次の複素正方行列である相関行列に対して，単位行列を格納し Matrix<complex<float>> 形式に直して出力する．

6.2.11 ConstantLocalization

ノードの概要

一定の音源定位結果を出力し続けるノード．パラメータは，ANGLES, ELEVATIONS, POWER, MIN_ID であり，それぞれに音源が到来する方位角 (ANGLES)，仰角 (ELEVATIONS)，パワー (POWER)，音源番号 (MIN_ID) を設定する．各パラメータは `Vector` なので，複数の定位結果を出力させることも可能である．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源定位結果が既知の場合の評価をするときに用いる．例えば，音源分離の処理結果を評価する場合に，分離処理に問題があるのか，音源定位誤差に問題があるのかを判断したいときや，音源定位を同じ条件にした状態での音源分離の性能評価をしたい場合など．

典型的な接続例

図 6.21 に接続例を示す．このネットワークでは，一定の定位結果を Iterate のノードパラメータに設定した回数だけ表示する．

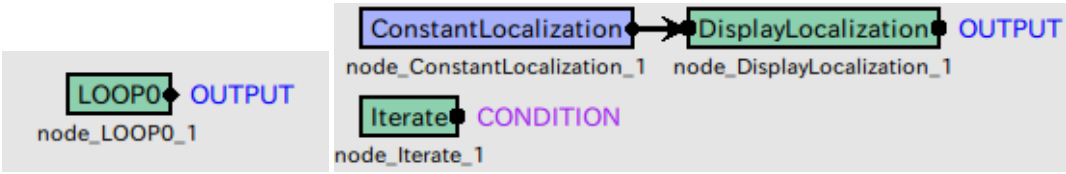


図 6.21: ConstantLocalization の接続例: 左が MAIN サブネットワーク, 右が Iterator サブネットワーク．

ノードの入出力とプロパティ

入力

無し．

出力

SOURCES : `Vector< ObjectRef >` 型．固定の音源定位結果を出力する．`ObjectRef` が参照するのは，`Source` 型のデータである．

パラメータ

ANGLES : `Vector< float >` 型．音源が到来している方向の，方位角 (左右) を表す．角度の単位は degree.

ELEVATIONS : `Vector< float >` 型．音源が到来している方向の，仰角 (上下) を表す．角度の単位は degree.

表 6.24: ConstantLocalization のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
ANGLES	Object	<Vector<float> >	[deg]	音源の方位角 (左右の向き)
ELEVATIONS	Object	<Vector<float> >	[deg]	音源の仰角 (上下の向き)
POWER	Object	<Vector<float> >	[dB]	音源のパワー
MIN_ID	int	0		音源番号

POWER : Vector< float > 型．到来音源のパワーを指定する．LocalizeMUSIC が算出する空間スペクトルと同じ次元であり，単位は [dB]．指定しない場合は自動的に 1.0[dB] が代入される．

MIN_ID : int 型．各音源に割り振られる最小音源番号を表す．各音源は後段処理で異なる音源と識別されるよう，音源番号はユニークである必要がある．音源番号は ANGLES と ELEVATIONS で指定した第一成分から MIN_ID を最小番号として順に割り振られる．例えば MIN_ID = 0 とし，ANGLES = <Vector<float> 0 30> とした場合，0 度方向の音源には 0 番が，30 度方向の音源には 1 番が振られる．

ノードの詳細

音源数を N，i 番目の音源の方位角 (ANGLE) を a_i，仰角 (ELEVATION) を e_i とする．このとき，パラメータは以下のように記述する．

ANGLES: <Vector<float> a_1 ... a_N >

ELEVATIONS: <Vector<float> e_1 ... e_N >

このように，入力は極座標系で行うが，実際に ConstantLocalization が出力するのは，単位球上の点に対応する，直交座標系の値 (x_i, y_i, z_i) である．極座標系から直交座標系への変換は，以下の式に基づいて行う．

$$x_i = \cos(a_i\pi/180)\cos(e_i\pi/180) \tag{6.5}$$

$$y_i = \sin(a_i\pi/180)\cos(e_i\pi/180) \tag{6.6}$$

$$z_i = \sin(e_i\pi/180) \tag{6.7}$$

なお, 音源の座標の他に， ConstantLocalization は音源のパワー (POWER で指定．未指定で 1.0 を自動的に代入) と音源番号 (MIN_ID + i) も出力する．

6.2.12 DisplayLocalization

ノードの概要

音源定位結果を GTK ライブラリを使って表示するノードである．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源定位結果を視覚的に確認したいときに用いる．

典型的な接続例

ConstantLocalization や LocalizeMUSIC などの，定位ノードの後に接続する．図 6.22 では，ConstantLocalization からの固定の定位結果を表示し続ける．

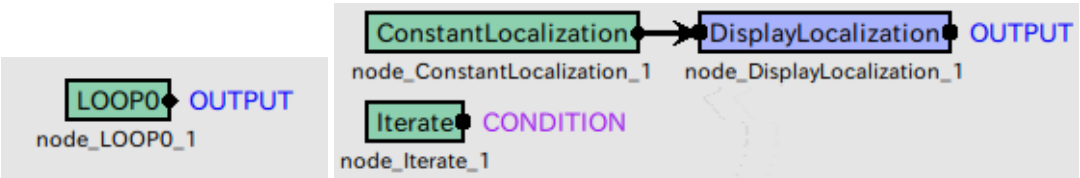


図 6.22: DisplayLocalization の接続例．（ConstantLocalization と同様）

ノードの入出力とプロパティ

入力

SOURCES ：`Vector< ObjectRef >` 型．音源位置を表すデータ (`Source` 型) を入力する．

出力

OUTPUT ：`Vector< ObjectRef >` 型．入力された値 (`Source` 型) をそのまま出力する．

パラメータ

WINDOW_NAME ：`string` 型．ウィンドウ名．

WINDOW_LENGTH ：`int` 型．デフォルトは 1000 なので, 表示されるウィンドウの幅は 1000 フレーム．このパラメータを調整することで, 表示する音源定位の時間幅を変えられる．

VERTICAL_RANGE ：`Vector< int >` 型．縦軸に表示されるデータの値域の範囲．第一成分に最小値，第二成分に最大値を指定する．デフォルトは `<Vector<int> -180 180>` なので，方位角推定結果を-180 度から 180 度の範囲で表示する．

表 6.25: DisplayLocalization のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
WINDOW_NAME	string	Source Location		音源定位結果を表示するウィンドウ名
WINDOW_LENGTH	int	1000	Frame	音源定位結果を表示するフレーム幅
VERTICAL_RANGE	Vector< int >	下記参照		縦軸に表示する値域の範囲
PLOT_TYPE	string	AZIMUTH		表示するデータの種類

PLOT_TYPE : string 型．表示されるデータの種類．AZIMUTH に指定すると方位角推定結果を，ELEVATION に指定すると仰角推定結果を表示する．

ノードの詳細

表示される色は、赤、青、緑の 3 色．ID の値が 1 増えるごとに順番に色が変わっていく．例えば、 ID が 0 なら赤、1 なら青, 2 なら緑、3 なら赤.

6.2.13 LocalizeMUSIC

ノードの概要

マルチチャネルの音声波形データから，MUltiple SIgnal Classification (MUSIC) 法を用いて，マイクロホンアレイ座標系で水平面方向での音源方向を推定する．HARK における音源定位のメインノードである．

必要なファイル

ステアリングベクトルからなる定位用伝達関数ファイルが必要．マイクロホンと音源の位置関係，もしくは，測定した伝達関数に基づき生成する．

使用方法

本ノードは MUSIC 法 によって，どの方向にどのくらいのパワーの音があるかを推定する．大きなパワーを持つ方向を各フレームで検出することで，音源の方向や，音源数，発話区間などをある程度知ることが可能である．本ノードから出力される定位結果が，後段の音源追跡や音源分離に利用される．

典型的な接続例

典型的な接続例を図 6.23 に示す．

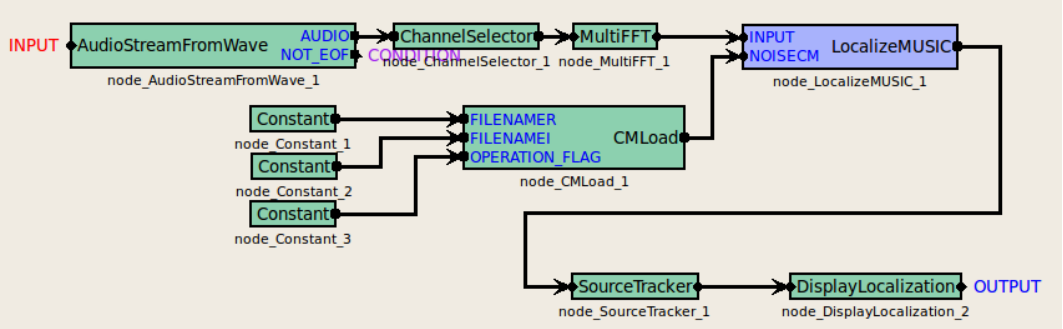


図 6.23: LocalizeMUSIC の接続例

ノードの入出力とプロパティ

入力

INPUT : `Matrix<complex<float> >`, 入力信号の複素周波数表現 $M \times (NFFT/2+1)$.

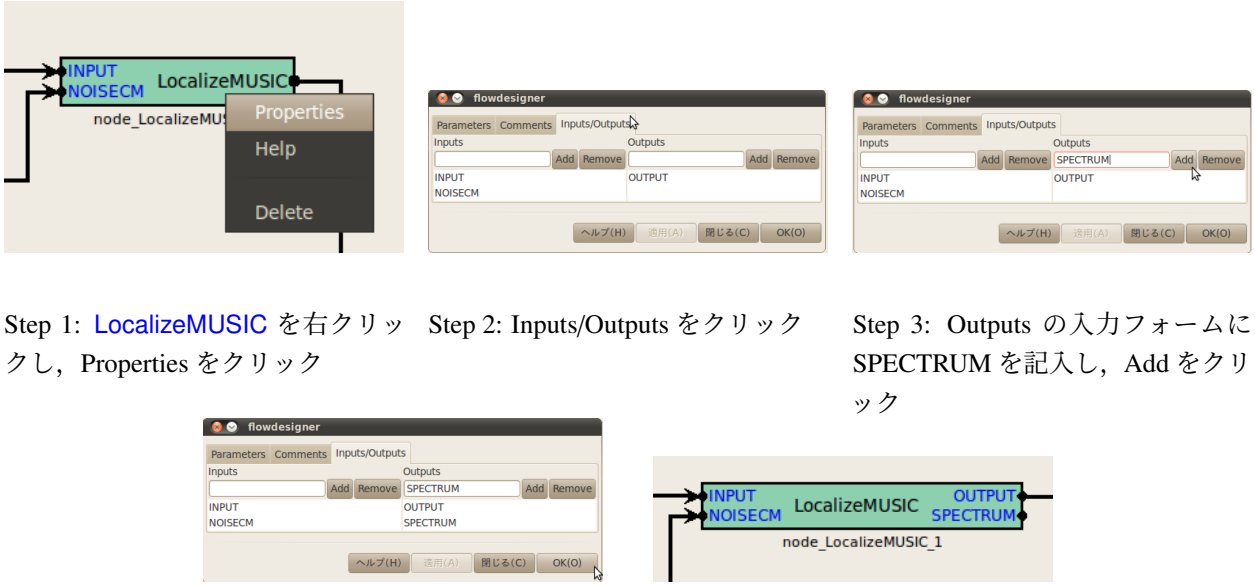
NOISECM : `Matrix<complex<float> >` 型．各周波数ビン毎の相関行列．M 次の複素正方行列である相関行列が $NFFT/2+1$ 個入力される．`Matrix<complex<float> >` の行は周波数 ($NFFT/2+1$ 行) を，列は複素相関行列 ($M*M$ 列) を表す．本入力端子は開放することも可能であり，開放した場合は相関行列に単位行列が用いられる．

出力

OUTPUT ：`Vector<ObjectRef>` 型で音源位置（方向）を表す．`ObjectRef` は，`Source` であり，音源位置とその方向の MUSIC スペクトルのパワーからなる構造体である．`Vector` の要素数は音源数 (N)．MUSIC スペクトルの詳細については，ノードの詳細を参照されたい．

SPECTRUM ：`Vector<float>` 型．各方向毎の MUSIC スペクトルのパワー．式 (6.16) の $\bar{P}(\theta)$ に相当する．三次元音源定位の場合は θ が三次元となる．出力形式についてはノードの詳細を参照．本出力端子は，デフォルトでは非表示である．

非表示出力の追加方法は図 6.24 を参照されたい．



Step 1: LocalizeMUSIC を右クリックし，Properties をクリック

Step 2: Inputs/Outputs をクリック

Step 3: Outputs の入力フォームに SPECTRUM を記入し，Add をクリック

Step 4: Outputs の出力一覧に SPECTRUM が追加されたことを確認し，OK をクリック

Step 5: ノードに SPECTRUM 出力端子が追加される

図 6.24: 非表示出力の使用例 : SPECTRUM 端子の表示

パラメータ

MUSIC_ALGORITHM ：`string` 型．MUSIC 法において，信号の部分空間を計算するために使うアルゴリズムの選択．SEVD は標準固有値分解を，GEVD は一般化固有値分解を，GSVD は一般化特異値展開を表す．LocalizeMUSIC は，NOISECM 端子から雑音情報を持つ相関行列を入力することで，その雑音を白色化（抑圧）した音源定位ができる機能を持つ．SEVD はその機能がついていない音源定位を実現する．SEVD を選択した場合は NOISECM 端子からの入力は無視される．GEVD と GSVD は共に NOISECM 端子から入力された雑音を白色化する機能を持つが，GEVD は GSVD に比べて雑音抑圧性能が良好だが計算時間がおよそ 4 倍かかる問題を持つ．使用したい場面や計算機環境に合わせて，三つのアルゴリズムを適宜使い分けられる．アルゴリズムの詳細については，ノードの詳細を参照されたい．

TF_CHANNEL_SELECTION ：`Vector<int>` 型．定位用伝達関数ファイルに格納されているマルチチャネルのステアリングベクトルの中で，指定したチャネルのステアリングベクトルを選択するパラメータである．ChannelSelector と同様に，チャネル番号は 0 から始まる．デフォルトでは 8 チャネルの信号処理を想定し，

表 6.26: LocalizeMUSIC のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
MUSIC_ALGORITHM	string	SEVD		MUSIC のアルゴリズム
TF_CHANNEL_SELECTION	Vector<int>	下記参照		使用チャネル番号
LENGTH	int	512	[pt]	FFT 点数 ($NFFT$)
SAMPLING_RATE	int	16000	[Hz]	サンプリングレート
A_MATRIX	string			定位用伝達関数ファイル名
WINDOW	int	50	[frame]	相関行列の平滑化フレーム数
WINDOW_TYPE	string	FUTURE		相関行列の平滑化区間
PERIOD	int	50	[frame]	定位結果を算出する周期
NUM_SOURCE	int	2		MUSIC で仮定する音源数
MIN_DEG	int	-180	[deg]	ピーク探索方位角の最小値
MAX_DEG	int	180	[deg]	ピーク探索方位角の最大値
LOWER_BOUND_FREQUENCY	int	500	[Hz]	使用周波数帯域の最小値
UPPER_BOUND_FREQUENCY	int	2800	[Hz]	使用周波数帯域の最大値
SPECTRUM_WEIGHT_TYPE	string	Uniform		定位周波数重みの種類
A_CHAR_SCALING	float	1.0		A 特性重みの伸展係数
MANUAL_WEIGHT_SPLINE	Matrix<float>	下記参照		スプライン重みの係数
MANUAL_WEIGHT_SQUARE	Matrix<float>	下記参照		矩形重みの周波数転換点
ENABLE_EIGENVALUE_WEIGHT	bool	true		固有値重みの有無
DEBUG	bool	false		デバッグ出力の ON/OFF

<Vector<int> 0 1 2 3 4 5 6 7> と設定されている．本パラメータの成分数 (M) を入力信号のチャネル数と合わせる必要がある．また，INPUT 端子に入力されるチャネルの順序と TF_CHANNEL_SELECTION のチャネル順序を合わせる必要がある．

LENGTH ：int 型．512 がデフォルト値．フーリエ変換の際の FFT 点数．前段までの FFT 点数と合わせる必要がある．

SAMPLING_RATE ：int 型．16000 がデフォルト値．入力音響信号のサンプリング周波数．LENGTH と同様，他のノードとそろえる必要がある．

A_MATRIX ：string 型．デフォルト値はなし．定位用伝達関数ファイルのファイル名を指定する．絶対パスと相対パスの両方に対応している．定位用伝達関数ファイルの作成方法については，harktool4 を参照．

WINDOW ：int 型．50 がデフォルト値．相関行列計算時の平滑化フレーム数を指定する．ノード内では，入力信号の複素スペクトルから相関行列を毎フレーム生成し，WINDOW で指定されたフレームで加算平均を取る．この値を大きくすると，相関行列が安定するが，区間が長い分，時間遅れが長くなる．

WINDOW_TYPE ：string 型．FUTURE がデフォルト値．相関行列計算時の平滑化フレームの使用区間を指定する．FUTURE に指定した場合，現在のフレーム f から $f + WINDOW - 1$ までが平滑化に使用される．MIDDLE に指定した場合，$f - (WINDOW/2)$ から $f + (WINDOW/2) + (WINDOW\%2) - 1$ までが平滑化に使用される．PAST に指定した場合，$f - WINDOW + 1$ から f までが平滑化に使用される．

PERIOD ：int 型．50 がデフォルト値．音源定位結果算出の周期をフレーム数で指定する．この値が大きいと，定位結果を得るための時間間隔が大きくなり，発話区間が正しく取りにくくなったり，移動音源の追

従性が悪くなる．ただし，小さくすると計算負荷がかかるため，計算機環境に合わせたチューニングが必要となる．

NUM_SOURCE ：`int` 型．2 がデフォルト値．MUSIC 法における信号の部分空間の次元数であり，実用上は，音源定位のピーク検出で強調すべき目的音源数と解釈できる．下記のノード詳細では N_s と表わされている．$1 \leq N_s \leq M-1$ である必要がある．目的音の音源数に合わせておくことが望ましいが，例えば目的音源数が 3 の場合にも，一つ一つの音源が発音している区間が異なるため，実用上はそれより少ない値を選択すれば十分である．

MIN_DEG ：`int` 型．-180 がデフォルト値．音源探索する際の最小角度であり，ノード詳細で θ_{min} として表わされている．0 度がロボット正面方向であり，負値がロボット右手方向，正値がロボット左手方向である．指定範囲は，便宜上 ±180 度としてるが，360 度以上の回り込みにも対応しているので，特に制限はない．

MAX_DEG ：`int` 型．180 がデフォルト値．音源探索する際の最大角度であり，ノード詳細で θ_{max} として表わされている．その他は，MIN_DEG と同様である．

LOWER_BOUND_FREQUENCY ：`int` 型．500 がデフォルト値．音源定位のピーク検出時に考慮する周波数帯域の下限であり，ノードの詳細では，ω_{min} で表わされている．$0 \leq \omega_{min} \leq$ SAMPLING_RATE/2 である必要がある．

UPPER_BOUND_FREQUENCY ：`int` 型．2800 がデフォルト値．音源定位のピーク検出時に考慮する周波数帯域の上限であり，下記では，ω_{max} で表わされている．$\omega_{min} < \omega_{max} \leq$ SAMPLING_RATE/2 である必要がある．

SPECTRUM_WEIGHT_TYPE ：`string` 型．Uniform がデフォルト値．音源定位のピーク検出時に使用する MUSIC スペクトルの周波数軸方向に対する重みの様式を指定する．Uniform は重みづけを OFF に設定する．A_Characteristic は人間の聴覚の音圧感度を模した重み付けを MUSIC スペクトルに与える．Manual_Spline は，MANUAL_WEIGHT_SPLINE で指定した点を補間点とした Cubic スプライン曲線に合わせた重み付けを MUSIC スペクトルに与える．Manual_Square は，MANUAL_WEIGHT_SQUARE で指定した周波数に合わせた矩形重みを生成し，MUSIC スペクトルに付与する．

A_CHAR_SCALING ：`float` 型．1.0 がデフォルト値．A 特性重みを周波数軸方向に伸展するスケーリング項を指定する．A 特性重みは人間の聴覚の音圧感度を模しているため，音声帯域外を抑圧するフィルタリングが可能である．A 特性重みは規格値があるが，雑音環境によっては，雑音が音声帯域内に入ってしまい，うまく定位できないことがある．そこで，A 特性重みを周波数軸方向に伸展し，より広い低周波帯域を抑圧するフィルタを構成する．

MANUAL_WEIGHT_SPLINE ：`Matrix<float>` 型．
`<Matrix<float> <rows 2> <cols 5> <data 0.0 2000.0 4000.0 6000.0 8000.0 1.0 1.0 1.0 1.0 1.0> >`
がデフォルト値．2 行 K 列の `float` 値で指定する．K はスプライン補間で使用するための補間点数に相当する．1 行目は周波数を，2 行目はそれに対応した重みを指定する．重み付けは補間点を通るスプライン曲線に合わせて行われる．デフォルト値では 0 [Hz] から 8000[Hz] までの周波数帯域に対して全て 1 となる重みが付与される．

MANUAL_WEIGHT_SQUARE ：`Vector<float>` 型．`<Vector<float> 0.0 2000.0 4000.0 6000.0 8000.0>`
がデフォルト値．MANUAL_WEIGHT_SQUARE で指定した周波数によって矩形重みを生成し，MUSIC スペクトルに付与する．MANUAL_WEIGHT_SQUARE の奇数成分から偶数成分までの周波数帯域は 1 の重みを，偶数成分から奇数成分までの周波数帯域は 0 の重みを付与する．デフォルト値では 2000 [Hz] から 4000[Hz]，6000 [Hz] から 8000[Hz] までの MUSIC スペクトルを抑圧することができる．

ENABLE_EIGENVALUE_WEIGHT : bool 型．true がデフォルト値．true の場合，MUSIC スペクトルの計算の際に，相関行列の固有値分解（または特異値分解）から得られる最大固有値（または最大特異値）の平方根を重みとして，付与する．この重みは，MUSIC_ALGORITHM に GEVD や GSVD を選ぶ場合は NOISECM 端子から入力される相関行列の固有値に依存して大きく変化するため，false にするのが良い．

DEBUG : bool 型．デバッグ出力の ON/OFF，デバッグ出力のフォーマットは，以下の通りである．まず，フレームで検出された音源数分だけ，音源のインデックス，方向，パワーのセットがタブ区切りで出力される．ID はフレーム毎に 0 から順番に便宜上付与される番号で，番号自身には意味はない．方向 [deg] は小数を丸めた整数が表示される．パワーは MUSIC スペクトルのパワー値（式 (6.16) の $\bar{P}(\theta)$）がそのまま出力される．次に，改行後，“MUSIC spectrum: ” と出力され，式 (6.16) の $\bar{P}(\theta)$ の値が，すべての θ について表示される．

ノードの詳細

MUSIC 法は，入力信号のチャネル間の相関行列の固有値分解を利用して，音源方向の推定を行う手法である．以下にアルゴリズムをまとめる．

伝達関数の生成:

MUSIC 法では，音源から各マイクロホンまでの伝達関数を計測または数値的に求め，それを事前情報として用いる．マイクロホンアレイからみて，θ 方向にある音源 $S(\theta)$ から i 番目のマイク M_i までの周波数領域での伝達関数を $h_i(\theta, \omega)$ とすると，マルチチャネルの伝達関数ベクトルは以下のように表せる．

$$\boldsymbol{H}(\theta, \omega) = [h_1(\theta, \omega), \cdots, h_M(\theta, \omega)] \tag{6.8}$$

この伝達関数ベクトルを，適当な間隔 $\Delta\theta$ 毎（非等間隔でも可）に，事前に計算もしくは計測によって用意しておく．HARK では，計測によっても数値計算によっても伝達関数ファイルを生成できるツールとして，harktool4 を提供している．具体的な伝達関数ファイルの作り方に関しては harktool4 の項を参照されたい（harktool4 より三次元の伝達関数に対応した）．LocalizeMUSIC ノードでは，この事前情報ファイル（定位用伝達関数ファイル）を A_MATRIX で指定したファイル名で読み込んで用いている．このように伝達関数は，音源の方向ごとに用意することから，方向ベクトル，もしくは，この伝達関数を用いて定位の際に方向に対して走査を行うことから，ステアリングベクトルと呼ぶことがある．

入力信号のチャネル間相関行列の算出:

HARK による音源定位の処理はここから始まる．まず，M チャネルの入力音響信号を短時間フーリエ変換して得られる周波数領域の信号ベクトルを以下のように求める．

$$\boldsymbol{X}(\omega, f) = [X_1(\omega, f), X_2(\omega, f), X_3(\omega, f), \cdots, X_M(\omega, f)]^T \tag{6.9}$$

ここで，ω は周波数，f はフレームを表す．HARK では，ここまでの処理を前段の MultiFFT ノードで行う．

入力信号 $\boldsymbol{X}(\omega, f)$ のチャネル間の相関行列は，各フレーム，各周波数ごとに以下のように定義できる．

$$\boldsymbol{R}(\omega, f) = \boldsymbol{X}(\omega, f)\boldsymbol{X}^*(\omega, f) \tag{6.10}$$

ここで $()^*$ は複素共役転置演算子を表す．理論上は，この $\boldsymbol{R}(\omega, f)$ をそのまま以降の処理で利用すれば問題はないが，実用上，安定した相関行列を得るため，HARK では，時間方向に平均したものを使用している．

$$\boldsymbol{R}'(\omega, f) = \frac{1}{\mathrm{WINDOW}} \sum_{i=W_i}^{W_f} \boldsymbol{R}(\omega, f + i) \tag{6.11}$$

平滑化に使用する区間は WINDOW_TYPE パラメータによって変更できる．WINDOW_TYPE=FUTURE の場合，$W_i = 0$, $W_f = \mathrm{WINDOW} - 1$ となる．WINDOW_TYPE=MIDDLE の場合，$W_i = \mathrm{WINDOW}/2$, $W_f = \mathrm{WINDOW}/2 + \mathrm{WINDOW}\%2 - 1$ となる．WINDOW_TYPE=PAST の場合，$W_i = -\mathrm{WINDOW} + 1$, $W_f = 0$ となる．

信号と雑音の部分空間への分解：

MUSIC 法では，式 (6.11) で求まった相関行列 $\boldsymbol{R}'(\omega, f)$ の固有値分解，もしくは特異値分解を行い，M 次の空間を，信号の部分空間と，それ以外の部分空間に分解する．

本節以降の処理は計算負荷が高いため，HARK では計算負荷を考慮し，数フレームに一回演算されるように設計されている．LocalizeMUSIC では，この演算周期を PERIOD で指定できる．

LocalizeMUSIC では，部分空間に分解する方法が MUSIC_ALGORITHM によって指定できる．

MUSIC_ALGORITHM を SEVD に指定した場合，以下の標準固有値分解を行う．

$$\boldsymbol{R}'(\omega, f) = \boldsymbol{E}(\omega, f)\boldsymbol{\Lambda}(\omega, f)\boldsymbol{E}^{-1}(\omega, f) \tag{6.12}$$

ここで，$\boldsymbol{E}(\omega, f)$ は互いに直交する固有ベクトルからなる行列 $\boldsymbol{E}(\omega, f) = [\boldsymbol{e}_1(\omega, f), \boldsymbol{e}_2(\omega, f), \cdots, \boldsymbol{e}_M(\omega, f)]$ を，$\boldsymbol{\Lambda}(\omega)$ は各固有ベクトルに対応する固有値を対角成分とした対角行列を表す．なお，$\boldsymbol{\Lambda}(\omega)$ の対角成分 $[\lambda_1(\omega), \lambda_2(\omega), \ldots, \lambda_M(\omega)]$ は降順にソートされているとする．

MUSIC_ALGORITHM を GEVD に指定した場合，以下の一般化固有値分解を行う．

$$\boldsymbol{K}^{-\frac{1}{2}}(\omega, f)\boldsymbol{R}'(\omega, f)\boldsymbol{K}^{-\frac{1}{2}}(\omega, f) = \boldsymbol{E}(\omega, f)\boldsymbol{\Lambda}(\omega, f)\boldsymbol{E}^{-1}(\omega, f) \tag{6.13}$$

ここで，$\boldsymbol{K}(\omega, f)$ は f フレーム目で NOISECM 端子から入力される相関行列を表す．$\boldsymbol{K}(\omega, f)$ との一般化固有値分解により $\boldsymbol{K}(\omega, f)$ に含まれる雑音由来の大きな固有値を白色化することができるため，雑音を抑圧した音源定位が実現できる．

MUSIC_ALGORITHM を GSVD に指定した場合，以下の一般化特異値分解を行う．

$$\boldsymbol{K}^{-1}(\omega, f)\boldsymbol{R}'(\omega, f) = \boldsymbol{E}(\omega, f)\boldsymbol{\Lambda}(\omega, f)\boldsymbol{E}_r^{-1}(\omega, f) \tag{6.14}$$

ここで，$\boldsymbol{E}(\omega, f), \boldsymbol{E}_r(\omega, f)$ は，それぞれ左特異ベクトル，右特異ベクトルからなる行列を表し，$\boldsymbol{\Lambda}(\omega)$ は各特異値を対角成分とした対角行列を表す．

分解によって得た固有ベクトル空間 $\boldsymbol{E}(\omega, f)$ に対応する固有値（または特異値）は音源のパワーと相関があることから，値が大きな固有値に対応した固有ベクトルを取ることで，パワーの大きな目的音の部分空間のみを選択することができる．すなわち，考慮する音源数を N_s とすれば，$[e_1(\omega), \cdots, e_{N_s}(\omega)]$ が音源に対応する固有ベクトル，$[e_{N_s+1}(\omega), \cdots, e_M(\omega)]$ が雑音に対応する固有ベクトルとなる．LocalizeMUSIC では N_s を NUM_SOURCE として指定できる．

MUSIC スペクトルの算出:

音源定位のための MUSIC スペクトルは，雑音に対応した固有ベクトルのみを用いて次のように計算される．

$$P(\theta, \omega, f) = \frac{|\boldsymbol{H}^*(\theta, \omega)\boldsymbol{H}(\theta, \omega)|}{\sum_{i=N_s+1}^{M} |\boldsymbol{H}^*(\theta, \omega)e_i(\omega, f)|} \tag{6.15}$$

右辺の分母は，入力のうち雑音に起因する固有ベクトルと伝達関数の内積を計算している．固有ベクトルによって張られる空間上では，小さい固有値に対応する雑音の部分空間と大きい固有値に対応する目的信号の部分空間は互いに直交するため，もし，伝達関数が目的音源に対応するベクトルであれば，この内積値は理論上 0 になる．よって，$P(\theta, \omega, f)$ は無限大に発散する．実際には，ノイズ等の影響により，無限大には発散しないが，遅延和などのビームフォーミングと比較すると鋭いピークが観測されるため，音源の抽出が容易になる．右辺の分子は正規化を行うための正規化項である．

$P(\theta, \omega, f)$ は，各周波数ごとに得られる MUSIC スペクトルであるため，以下のようにして周波数方向の統合を行う．

$$\bar{P}(\theta, f) = \sum_{\omega=\omega_{min}}^{\omega_{max}} W_\Lambda(\omega, f) W_\omega(\omega, f) P(\theta, \omega, f) \tag{6.16}$$

ここで，$\omega_{min}, \omega_{max}$ は，それぞれ MUSIC スペクトルの周波数方向統合で扱う周波数帯域の下限と上限を示し，LocalizeMUSIC では，それぞれ LOWER_BOUND_FREQUENCY, UPPER_BOUND_FREQUENCY として指定できる．

また，$W_\Lambda(\omega, f)$ は，周波数方向統合の際の固有値重みであり，最大固有値（または最大特異値）の平方根である．LocalizeMUSIC では，固有値重みの有無を ENABLE_EIGENVALUE_WEIGHT によって選択することができ，false の場合は $W_\Lambda(\omega, f) = 1$，true の場合は $W_\Lambda(\omega, f) = \sqrt{\lambda_1(\omega, f)}$ となる．

また，$W_\omega(\omega, f)$ は，周波数方向統合の際の周波数重みであり，LocalizeMUSIC では SPECTRUM_WEIGHT_TYPE でその種類を以下のように指定できる．

- SPECTRUM_WEIGHT_TYPE が Uniform の場合
  一様重みとなり，全ての周波数ビンに対して，$W_\omega(\omega, f) = 1$ となる．

- SPECTRUM_WEIGHT_TYPE が A_Characteristic の場合
  国際電気標準会議が規格している A 特性重み $\mathcal{W}(\omega)$ となる．図 6.25 に A 特性重みの周波数特性を示す．横軸は ω，縦軸は $\mathcal{W}(\omega)$ を表す．LocalizeMUSIC では，規格の周波数特性に対して，周波数方向の伸展スケーリング項 A_CHAR_SCALING を導入している．A_CHAR_SCALING を α とすれば，実際に使用する周波数重みは $\mathcal{W}(\alpha\omega)$ と表すことができる．図 6.25 では，例として，$\alpha = 1$ の場合と $\alpha = 4$ の場合をプロットしている．最終的に MUSIC スペクトルにかかる重みは $W_\omega(\omega, f) = 10^{\frac{\mathcal{W}(\alpha\omega)}{20}}$ となる．例として，図 6.26 に A_CHAR_SCALING = 1 の時の $W_\omega(\omega, f)$ を示す．

- SPECTRUM_WEIGHT_TYPE が Manual_Spline の場合
  MANUAL_WEIGHT_SPLINE で指定した補間点に対してスプライン補間をした曲線にそった周波数重みとなる．MANUAL_WEIGHT_SPLINE は 2 行 k 列の `Matrix<float>` 型で指定し，一行目は周波数を，二行目はその周波数での重みを表す．補間点数 k は何点でも良い．例として，MANUAL_WEIGHT_SPLINE を，
  `<Matrix<float> <rows 2> <cols 3> <data 0.0 4000.0 8000.0 1.0 0.5 1.0> >`
  とした場合，補間点数は 3 で，周波数軸上の 0, 4000, 8000[Hz] の 3 つの周波数において，それぞれ，1, 0.5, 1 の重みをかけるスプライン曲線ができる．その時の $W_\omega(\omega, f)$ を図 6.27 に示す．

- SPECTRUM_WEIGHT_TYPE が Manual_Square の場合
  MANUAL_WEIGHT_SQUARE で指定した周波数で矩形が切り替わる矩形重みにそった周波数重みとなる．MANUAL_WEIGHT_SQUARE は k 次の `Vector<float>` 型で指定し，矩形を切り替えたい周波数を表す．切替点の個数 k は任意である．例として，MANUAL_WEIGHT_SQUARE を，
  `<Vector<float> 0.0 2000.0 4000.0 6000.0 8000.0>`
  とした場合の矩形重み $W_\omega(\omega, f)$ を図 6.28 に示す．この重みを使うことで，UPPER_BOUND_FREQUENCY と LOWER_BOUND_FREQUENCY だけでは指定できない複数の周波数領域を選択できる．

出力端子 SPECTRUM からは，式 (6.16) の $\bar{P}(\theta, f)$ が一次元配列として出力される．θ は三次元定位の場合は三次元となり，SPECTRUM からは三次元の $\bar{P}(\theta, f)$ が一次元配列として変換されて出力される．`Ne`, `Nd`, `Nr` をそれぞれ，仰角数，方位角数，半径数とすると，変換は以下のようになる．

```
FOR ie = 1 to Ne
  FOR id = 1 to Nd
```

```
    FOR ir = 1 to Nr
      SPECTRUM[ir + id * Nr + ie * Nr * Nd] = P[ir][id][ie]
    ENDFOR
  ENDFOR
ENDFOR
```

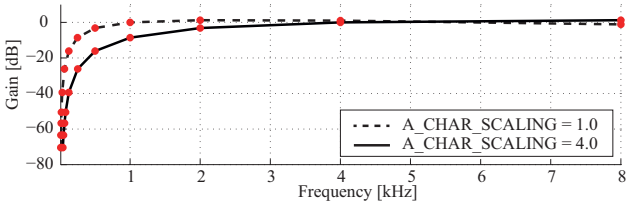


図 6.25: SPECTRUM_WEIGHT_TYPE = A_Charasteristic とした時の A 特性重みの周波数特性

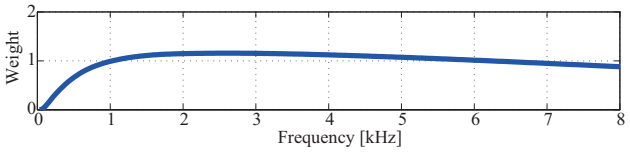


図 6.26: SPECTRUM_WEIGHT_TYPE = A_Charasteristic, A_CHAR_SCALING = 1 とした時の $W_\omega(\omega, f)$

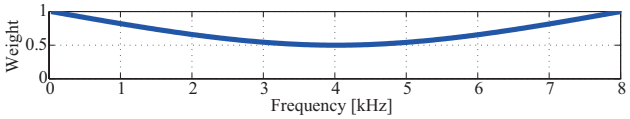


図 6.27: SPECTRUM_WEIGHT_TYPE = Manual_Spline とした時の $W_\omega(\omega, f)$

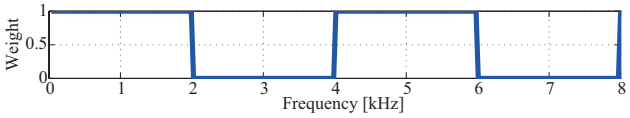


図 6.28: SPECTRUM_WEIGHT_TYPE = Manual_Square とした時の $W_\omega(\omega, f)$

音源の探索:

次に，式 (6.16) の $\bar{P}(\theta)$ について θ_{min} から θ_{max} までの範囲からローカルピークを検出し，値の大きい順に，上位 N_s 個について音源方向に対応する方向ベクトル，および，MUSIC スペクトルのパワーを出力する．また，ピークが N_s 個に満たない場合は，出力が N_s 以下になることもある．LocalizeMUSIC では，方位角の θ_{min} と θ_{max} をそれぞれ，MIN_DEG と MAX_DEG で指定できる．仰角と半径に関しては全てを用いる．

考察:

終わりに，MUSIC_ALGORITHM に GEVD や GSVD を選んだ場合の白色化が，式 (6.15) の MUSIC スペクトルに与える効果について簡単に述べる．

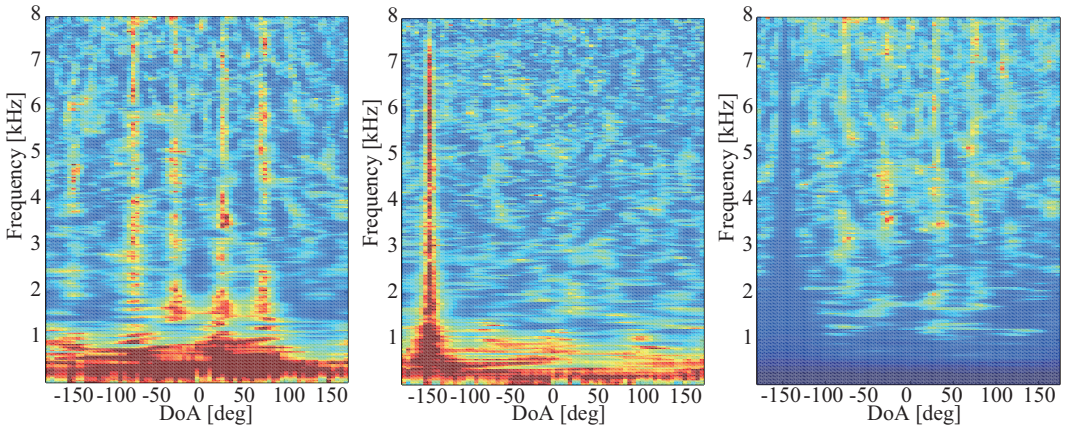
ここでは，例として，4 人（75 度，25 度，-25 度，-75 度方向）が同時に発話している状況を考える．

図 6.29(a) は，MUSIC_ALGORITHM に SEVD を選択し，雑音を白色化していない音源定位結果である．横軸が方位角，縦軸が周波数，値は式 (6.15) の $P(\theta, \omega, f)$ である．図のように低周波数領域に拡散性雑音と，-150 度方向に方向性雑音があり，正しく 4 話者の方向のみにピークを検出できていないことがわかる．

図 6.29(b) は，MUSIC_ALGORITHM に SEVD を選択し，4 話者が発話していない区間の MUSIC スペクトルである．図 6.29(a) で観察された拡散性雑音と方向性雑音が確認できる．

図 6.29(c) は，雑音情報として，図 6.29(b) の情報から $\boldsymbol{K}(\omega, f)$ を生成し，MUSIC_ALGORITHM に GSVD を選択して雑音を白色化した時の MUSIC スペクトルである．図のように，$\boldsymbol{K}(\omega, f)$ に含まれる拡散性雑音と方向性雑音が正しく抑圧され，4 話者の方向のみに強いピークが検出できていることが確認できる．

このように，既知の雑音に対して，GEVD や GSVD を用いることは有用である．



(a) MUSIC_ALGORITHM = SEVD の MUSIC スペクトル（話者数 4）

(b) $\boldsymbol{K}(\omega, f)$ を生成した雑音の MUSIC スペクトル（話者数 0）

(c) MUSIC_ALGORITHM = GSVD の MUSIC スペクトル（話者数 4）

図 6.29: MUSIC スペクトルの比較

参考文献

(1) Futoshi Asano *et. al*, “Real-Time Sound Source Localization and Separation System and Its Application to Automatic Speech Recognition.” in *Proc. of International Conference on Speech Processing (Eurospeech 2001)*, pp.1013–1016, 2001.

(2) 大賀 寿郎, 金田 豊, 山崎 芳男, “音響システムとディジタル処理,” 電子情報通信学会.

(3) K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, “Intelligent Sound Source Localization for Dynamic Environments”, in *Proc. of IEEE/RSJ Int’l Conf. on Intelligent Robots and Systems (IROS 2009)*, pp. 664–669, 2009.

6.2.14 LoadSourceLocation

ノードの概要

SaveSourceLocation ノードで保存された音源定位結果を読み込むノード.

必要なファイル

定位結果を保存したテキストファイル．形式は 5.3.4 参照．このファイルを生成するには，例えば SaveSourceLocation を使うとよい.

使用方法

どんなときに使うのか

音源定位した結果を再度使いたいときや, 完全に同じ音源定位結果を用いて複数の音源分離手法を評価するときなどに用いる. ファイル形式が揃えばよいので, 音源定位は別のプログラムで行い, その結果を LoadSourceLocation で渡すことも可能.

典型的な接続例

図 6.30 は，LoadSourceLocation のパラメータ FILENAME で指定した定位結果を読み込んで表示するネットワークである．一行ずつファイルを読み込み，ファイルの最後に到達すると終了する．このように，定位した結果を後で使う場合に用いる.

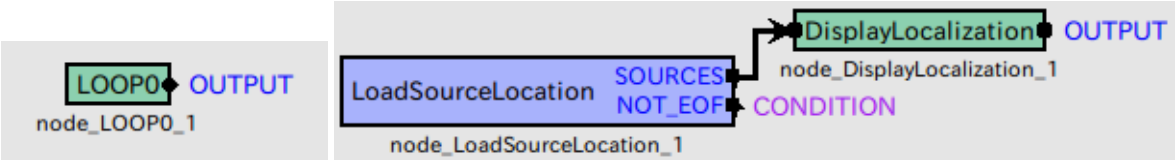


図 6.30: LoadSourceLocation の接続例: 左が MAIN サブネットワーク，右が Iterator サブネットワーク.

ノードの入出力とプロパティ

入力

無し.

出力

SOURCES : `Vector<ObjectRef>` 型．読み込まれた定位結果を，音源定位ノード (LocalizeMUSIC，ConstantLocalization など) と同様の形式で出力する．`ObjectRef` 型が参照するのは， `Source` 型のデータである.

NOT_EOF : `bool` 型．ファイルの終端まで読むと `false` になる出力端子なので，Iterator サブネットワークの終了条件端子に設定するとファイルを最後まで読ませられる.

パラメータ

表 6.27: LoadSourceLocation のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FILENAME	string			読み込むファイルのファイル名

FILENAME ：string 型．読み込むファイルのファイル名を設定する．

ノードの詳細

ノードが出力する音源定位結果は次の 5 つのメンバ変数を持ったオブジェクトの Vector である．

1. **パワー:** 100.0 で固定
2. **ID:** ファイルに保存されている，音源の ID
3. **音源位置の** x **座標:** 単位球上の，音源方向に対応する直交座標.
4. **音源位置の** y **座標:** 単位球上の，音源方向に対応する直交座標.
5. **音源位置の** z **座標:** 単位球上の，音源方向に対応する直交座標.

エラーメッセージとその原因は以下のとおり

`FILENAME is empty` ノードパラメータ FILENAME にファイル名が指定されていない．

`Can't open file name` ファイルのオープンに失敗した．原因は読み込み権限が無い，そのファイルが存在しないなど．

6.2.15 SaveSourceLocation

ノードの概要

音源定位結果をファイルに保存するノード．形式は 5.3.4 に定義されている．

必要なファイル

無し．

使用方法

どんなときに使うのか

定位結果の解析や視覚化などに利用するために，テキストファイルに保存したいときに使う．保存したファイルの読み込みには， LoadSourceLocation ノードを用いる．

典型的な接続例

図 6.31 に典型的な接続例を示す．例で示すネットワークは，ConstantLocalization のノードパラメータに定めた固定の定位結果をファイルに保存する．その他にも，本もジュールは音源定位結果を出力するノードなら何にでも接続できる．例えば LocalizeMUSIC ， ConstantLocalization ， LoadSourceLocation など．

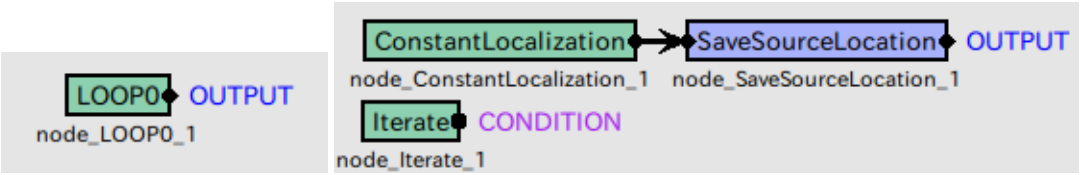


図 6.31: SaveSourceLocation の接続例: 左が MAIN サブネットワーク, 右が Iterator サブネットワーク

ノードの入出力とプロパティ

入力

SOURCES ：`Vector<ObjectRef>` 型．音源定位結果が入力される．`ObjectRef` 型が参照するのは `Source` 型．

出力

OUTPUT ：`Vector<ObjectRef>` 型．入力 (`Source` 型) がそのまま出力される．

パラメータ

FILENAME ：`string` 型．デフォルト値はなし．

表 6.28: SaveSourceLocation のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FILENAME	string			保存したいファイルの名前

ノードの詳細

エラーメッセージとその原因は以下のとおり

`FILENAME is empty` ノードパラメータ FILENAME にファイル名が指定されていない.

`Can't open file name` ファイルのオープンに失敗した．原因は書き込み権限が無いなど.

6.2.16 SourceIntervalExtender

ノードの概要

音源定位開始時刻を，実際よりも早めたいときに使う．ノードパラメータ PREROLL_LENGTH に与えた分だけ，実際よりも早く定位結果が出力される．

例えば， PREROLL_LENGTH が 5 なら，実際の定位結果が出力される 5 フレーム分前から定位結果が出力される．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源定位の後に音源分離を行うときに，音源定位と音源分離の間に挿入する．なぜなら，音源定位結果は音が発生してから出力するので，実際の音の開始時刻よりもやや遅れてしまう．従って，分離音の先頭が切れてしまう．そこで， SourceIntervalExtender を挿入して強制的に音源定位結果を早く出力させることで, この問題を解決する．

典型的な接続例

図 6.32 に典型的な接続例を示す．図のように，定位結果を元に分離したい場合には，SourceTracker の後に SourceIntervalExtender を挿入する．

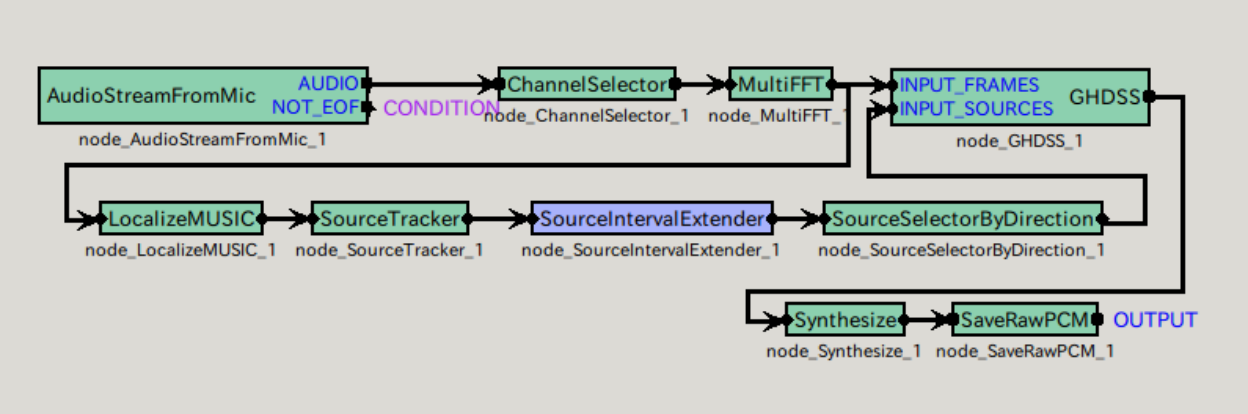


図 6.32: SourceIntervalExtender の接続例: Iterator サブネットワーク

ノードの入出力とプロパティ

入力

SOURCES ：`Vector<ObjectRef>` 型．`Source` 型で表現される音源定位結果の `Vector` が入力される．`ObjectRef` が参照するのは，`Source` 型のデータである．

出力

OUTPUT ：`Vector<ObjectRef>` 型．早められた出力された音源定位結果が出力される．`ObjectRef` が参照するのは，`Source` 型のデータである．

パラメータ

表 6.29: SourceIntervalExtender のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
PREROLL_LENGTH	`int`	50	[frame]	何フレームだけ早く定位結果を出力し始めるか．

PREROLL_LENGTH ：`int` 型．定位結果を実際よりどれだけ早く出力するかを決定する正の整数．値が小さすぎると分離音の先頭が出力されないので, 前段で用いる音源定位手法の遅延に合わせて設定する必要がある．

ノードの詳細

SourceIntervalExtender 無しで定位結果を元に音源分離を行ったとき，図 6.33 に示すように音源定位の処理時間の分だけ分離音の先頭部分が切れてしまう．特に音声認識の際，音声の先頭部分が切れていると認識性能に悪影響を及ぼすので，本ノードを使って定位結果を早める必要がある．

当然, 音源が定位される前に定位結果を出力するのは不可能である. そこで本ノード以降のネットワークの処理を PREROLL_LENGTH 分だけ遅らせることで “音源定位結果の出力を早める” 機能を実現する. こうして全体を遅らせたあと, 各繰り返しで PREROLL_LENGTH 分だけ入力を先読みし，そこ定位結果があれば，その時点から定位結果の出力を開始する (図 6.34 参照．)

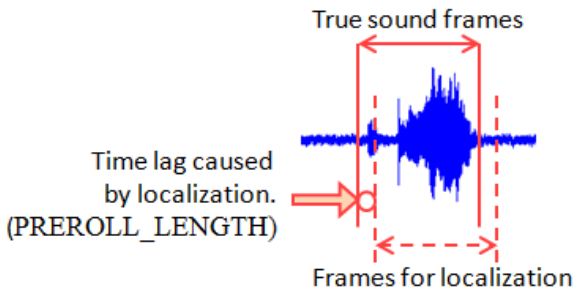


図 6.33: 音源定位結果を実際より早く出力する必要性

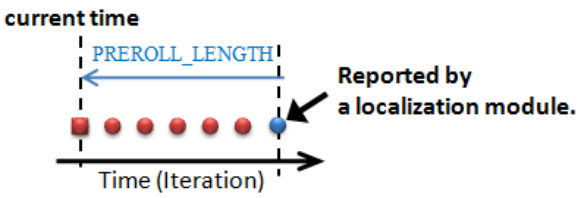


図 6.34: SourceIntervalExtender の動作: 定位結果を見つけると， PREROLL_LENGTH 分だけ事前に出力する．

6.2.17 SourceTracker

ノードの概要

時系列で入力された ID 無しの音源定位結果に対して，到来方向の近さに応じてクラスタリングを行い， ID を与えるノード．本ノードを通過した後の音源定位結果は，同じ音源か否かを ID のみから判定することが可能になる．ただし, 音長による信号の削除は行わない.

必要なファイル

無し.

使用方法

どんなときに使うのか

定位された音源の到来方向は，音源を固定していても（直立した人，固定したスピーカなど）通常同一方向であり続けることはない．従って，異なる時刻での到来方向が，同一音源となるように音源 ID を統一するためには，音源定位結果を追跡する必要がある．SourceTracker では，しきい値より近い到来方向の音源同士に対して同じ ID を与えるというアルゴリズムを用いている．本ノードを用いることで，音源に ID が付与され，ID 毎の処理が行える.

典型的な接続例

通常は，ConstantLocalization ，LocalizeMUSIC などの音源定位ノードの出力を本ノードの入力端子に接続する．すると適切な ID が定位結果に付加されるので，音源定位に基づく音源分離ノード GHDSS などや音源定位結果の表示ノード (DisplayLocalization) に接続する.

図 6.35 に接続例を示す．ここでは，固定の音源定位結果を， SourceTracker を通して表示している．このとき， ConstantLocalization の出力する定位結果が近ければ，それらは一つの音源にまとめて出力される．図中の ConstantLocalization に次のプロパティを与えた場合は，2 つの音源の成す角は MIN_SRC_INTERVAL のデフォルト値 20 [deg] より小さいので，1 つの音源だけが表示される.

ANGLES: `<Vector<float> 10 15>`

ELEVATIONS: `<Vector<float> 0 0>`

設定値の意味は ConstantLocalization を参照.

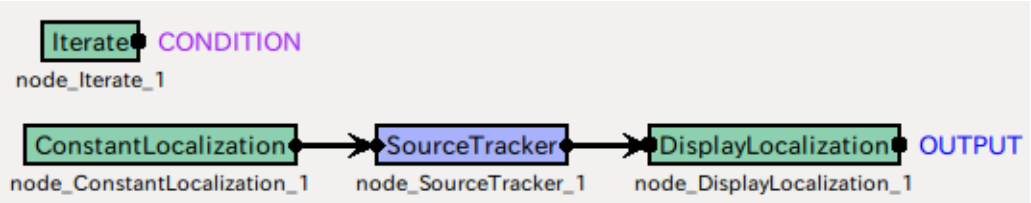


図 6.35: SourceTracker の接続例

ノードの入出力とプロパティ

入力

INPUT : `Vector<ObjectRef>` 型．ID が振られていない音源定位結果.

出力

OUTPUT : `Vector<ObjectRef>` 型．位置が近い音源に同じ ID を与えた音源定位結果

パラメータ

表 6.30: SourceTracker のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
THRESH	`float`			音源のパワーがこれより小さければ無視.
PAUSE_LENGTH	`float`	800	10 [frame]	音源の生存時間.
MIN_SRC_INTERVAL	`float`	20	[deg]	同一の音源とみなす角度差の閾値
MIN_ID	`int`	0		割り当てられる 音源 ID の最小値
DEBUG	`bool`	`false`		デバッグ出力の有無。

THRESH : `float` 型．音源定位結果を無視すべきノイズか否かを，そのパワーで判定する．パワーが THRESH より小さければノイズであると判断し，その定位結果は出力には反映されなくなる．小さくしすぎるとノイズを拾い，大きくしすぎると目的音の定位が困難になるので，このトレードオフを満たす値を見つける必要がある．

PAUSE_LENGTH : `float` 型．一度定位結果として出力した音源が，どれだけ長く続くと仮定するかを決めるパラメータ．一度定位した方向は，それ以降に音源定位結果が無くても，PAUSE_LENGTH / 10 [frame] の繰り返しの間だけ同一方向の定位結果を出力し続ける．デフォルト値は 800 なので，1 度定位した方向は，それ以降の 80 [frame] の繰り返しの間は定位結果を出力し続ける．

MIN_SRC_INTERVAL : `float` 型．音源の到来方向の差が MIN_SRC_INTERVAL より小さければ同一の音源とみなして片方の音源定位結果を削除する．こうして音源定位が揺らいでも追跡できる．

MIN_ID : `int` 型．定位結果ごとに割り振られる ID の開始番号を定める．

DEBUG : `bool` 型． `true` が与えられると, 音源定位結果が標準エラー出力にも出力される。

ノードの詳細

まず，本節で用いる記号を定義する．

1. ID : 音源の ID

2. パワー p: 定位された方向のパワー.

3. 座標 x, y, z: 音源定位方向に対応する，単位球上の直交座標.

4. 継続時間 r: 定位された音源がそれ以降どれだけ続くと仮定するかの指標.

定位された音源のパワーを p，音源方向に対応する単位球上の直交座標を x，y，z とする．

現在ノードが保持している音源数を N，新たに入力された音源数を M とする．また，直前の値には ^{last} を，現在の値には ^{cur} の添字をつける．例えば，i 番号めの新たに入力された音源のパワーは p_i^{cur} と表示する．

音源の近さを判定する指標の，音源同士の成す角を $\theta \in [0, 180]$ とする．

音源方向の近さの判定方法:

成す角 θ は，二つの音源方向を，$\boldsymbol{q}_1 = (x_1, y_1, z_1)$ と $\boldsymbol{q}_2 = (x_2, y_2, z_2)$ で表現すると次のように求まる．

$$\boldsymbol{q}_1 \cdot \boldsymbol{q}_2 = |\boldsymbol{q}_1||\boldsymbol{q}_2| \cos\theta \tag{6.17}$$

ここで逆余弦関数を用いると，θ が求まる．

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{q}_1 \cdot \boldsymbol{q}_2}{|\boldsymbol{q}_1||\boldsymbol{q}_2|}\right) = \cos^{-1}\left(\frac{x_1 \cdot x_2 + y_1 \cdot y_2 + z_1 \cdot z_2}{\sqrt{x_1^2 + y_1^2 + z_1^2}\sqrt{x_2^2 + y_2^2 + z_2^2}}\right) \tag{6.18}$$

以下では，表記を簡単にするために，第 i 音源と第 j 音源との成す角を θ_{ij} と表記する．

音源追跡方法:

SourceTracker の音源追跡アルゴリズムを図 6.36 にまとめる．また，青い丸が既にノードが持っている音源位置 (^{last})，緑の丸が新たに入力された音源位置 (^{cur}) を表す．

まず，すべての音源に対して，パワー p_i^{cur}，p_j^{last} が THRESH より小さければそれを削除する．次に，既にノードが持つ定位情報と新たに入力された音源位置を比較し，十分近い (=θ_{ij} が MIN_SRC_INTERVAL[deg] 以下) なら統合する．統合された音源には同じ ID が付与され，継続時間 r^{last} が PAUSE_LENGTH でリセットされる．統合は，1 つの音源を残して他のすべての音源位置を削除することで実現される．

θ_{ij} が MIN_SRC_INTERVAL [deg] より大きい音源は，異なる音源とみなされる．ノードが保持しているが，新たに入力されなかった音源位置は，r^{last} を 10 だけ減らす．もし，r^{last} が 0 を下回ったら，音源が消えたとみなして，その音源位置を削除する．新たに入力された音源位置が，既にノードが持つ音源位置のいずれとも異なる場合は，新たな ID を付与され，r^{cur} が PAUSE_LENGTH で初期化される．

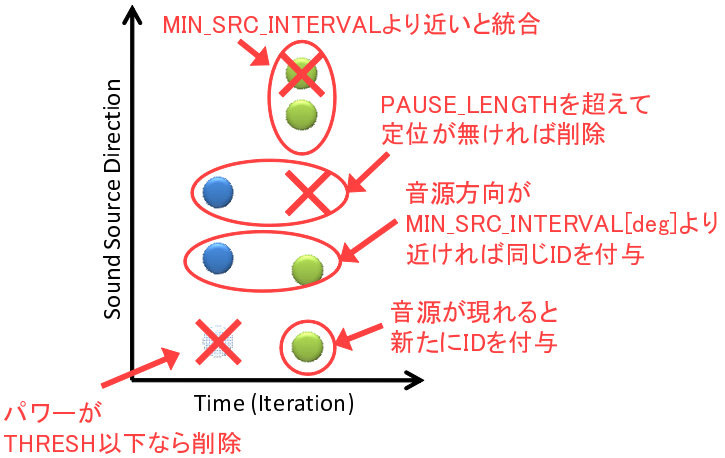


図 6.36: SourceTracker の音源追跡方法．横軸が時間 (=繰り返し回数) で，縦軸が音源方向を表す．

6.3 Separation カテゴリ

6.3.1 BGNEstimator

ノードの概要

マルチチャネル信号のパワースペクトルから，信号に含まれる定常ノイズ (例えば，ファンノイズや背景ノイズ，BackGround Noise) を推定する．推定した定常ノイズは，PostFilter ノードで使用される．

必要なファイル

無し．

使用方法

どんなときに使うのか

信号に含まれるフ定常ノイズ (ファンノイズや背景ノイズ，BackGround Noise) を推定する．この推定値が必要になるノードは，PostFilter である．PostFilter ノードでは，この背景ノイズと，PostFilter 内で推定されるチャネル間リークをもとに，分離処理でとりきれないノイズを抑制する．

典型的な接続例

BGNEstimator ノードの接続例を図 6.37 に示す．入力には，音声波形を周波数領域に変換し求めたパワースペクトルを入力する．出力は，PostFilter ノードで利用される．

ノードの入出力とプロパティ

表 6.31: BGNEstimator のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
DELTA	float	3.0		パワー比閾値
L	int	150	[frame]	検出時間幅
ALPHA_S	float	0.7		入力信号の平滑化係数
NOISE_COMPENS	float	1.0		定常ノイズの混入率
ALPHA_D_MIN	float	0.05		平滑化係数の最小値
NUM_INIT_FRAME	int	100	[frame]	初期化フレーム数

入力

INPUT_POWER : Matrix<float> 型． マルチチャネルパワースペクトル

出力

NOISE_POWER : Matrix<float> 型． 推定された定常ノイズのパワースペクトル．

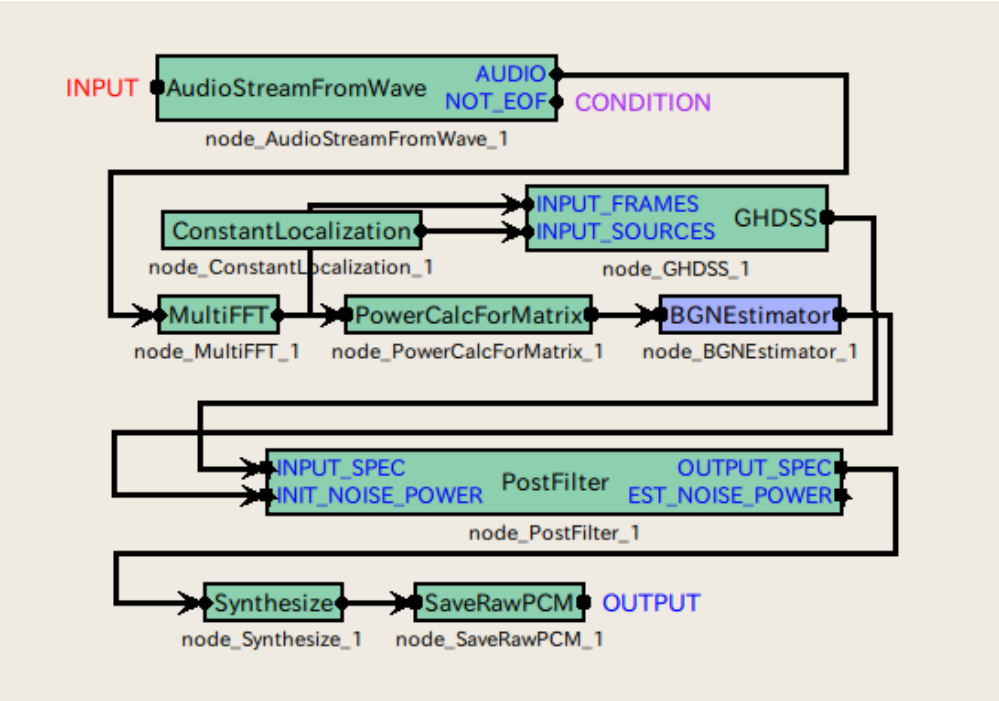


図 6.37: BGNEstimator の接続例

パラメータ

DELTA : `float` 型． 3.0 がデフォルト値． パワースペクトルの周波数ビンに音声などの目的音が含まれているかどうかの閾値．大きい値にすると，より多くのパワーを定常ノイズとみなす．

L : `int` 型． 150 がデフォルト値． 目的音が含まれるかの判断基準となる，過去最もパワーの小さいスペクトル (定常ノイズ成分) を保持する時間．AudioStreamFromWave ノードなどで指定される ADVANCE パラメータ分のシフトが行われる回数として指定する．

ALPHA_S : `float` 型． 0.7 がデフォルト値． 入力信号を時間方向に平滑化する際の係数． 値が大きいほど，過去のフレームの値の重みを大きく平滑化する．

NOISE_COMPENS : `float` 型． 1.0 がデフォルト値． 目的音が含まれないと判断されたフレームを，定常ノイズとして重みづけして加算する (定常ノイズの平滑化) ときの重み．

ALPHA_D_MIN : `float` 型． 0.05 がデフォルト値． 定常ノイズの平滑化処理で，目的音が含まれたと判断されたフレームのパワースペクトルを加える際の最小重み．

NUM_INIT_FRAME : `int` 型． 100 がデフォルト値． 処理を開始した際，このフレーム数だけ目的音の有無判定を行わず，全て定常ノイズとみなす．

ノードの詳細

以下では，定常ノイズを導出する過程を示す．時間，周波数，チャネルインデックスは 表 6.1 に準拠する．導出のフローは，図 6.38 の通り．

表 6.32: 変数表

変数名	対応パラメータ，または，説明
$S(f,k_i) = [S_1(f,k_i),...,S_M(f,k_i)]^T$	時間フレーム f，周波数ビン k_i の入力パワースペクトル
$\lambda(f,k_i) = [\lambda_1(f,k_i),...,\lambda_M(f,k_i)]^T$	推定されたノイズスペクトル
δ	DELTA，デフォルト 0.3
L	L，デフォルト 150
α_s	ALPHA_S，デフォルト 0.7
θ	NOISE_COMPENS，デフォルト 1.0
α_d^{min}	ALPHA_D_MIN，デフォルト 0.05
N	NUM_INIT_FRAME，デフォルト 100

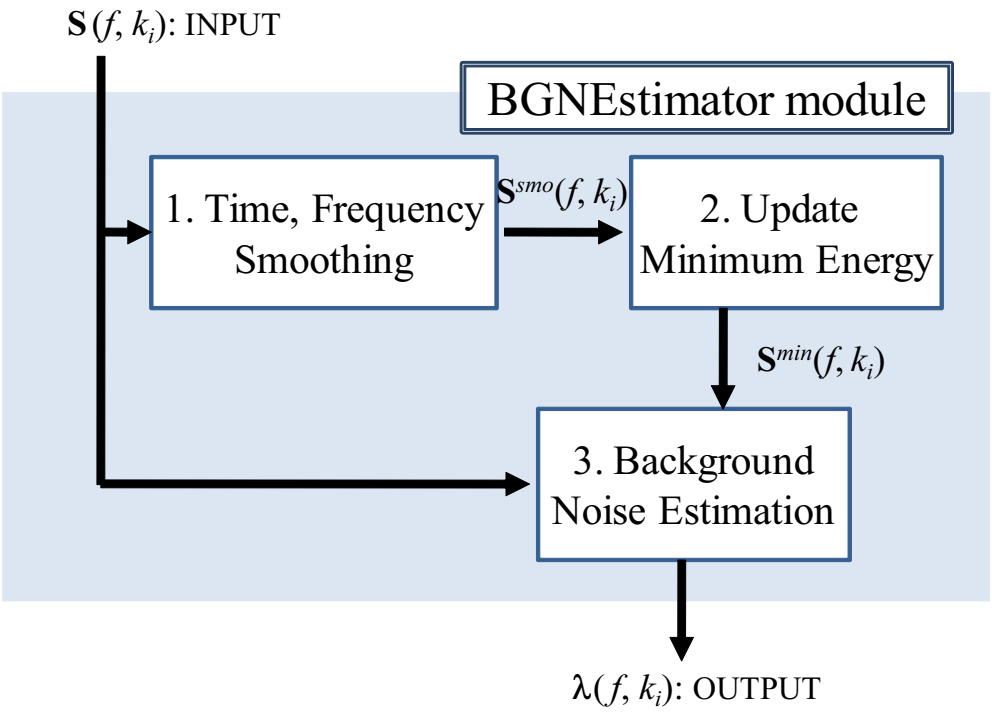


図 6.38: 定常ノイズ推定の流れ

1．時間方向，周波数方向平滑化: 時間方向の平滑化は，入力パワースペクトル $\boldsymbol{S}(f,k_i)$ と，前フレームの定常ノイズパワースペクトル $\boldsymbol{\lambda}(f-1,k_i)$ の内分により行う．

$$S_m^{smo,t}(f,k_i) = \alpha_s \lambda_m(f-1,k_i) + (1-\alpha_s) S_m(f,k_i) \tag{6.19}$$

周波数方向の平滑化は，時間平滑化された $S_m^{smo,t}(f,k_i)$ に対して行う．

$$S_m^{smo}(f,k_i) = 0.25 S_m^{smo}(f,k_{i-1}) + 0.5 S_m^{smo}(f,k_i) + 0.25 S_m^{smo}(f,k_{i+1}) \tag{6.20}$$

2．最小エネルギーの更新: 目的音の有無を判定するため，処理を開始してからの各チャネル，周波数ビンについての最小のエネルギー $\boldsymbol{S}^{min}$ を計算する．$\boldsymbol{S}^{min}$ は，各チャネル，周波数ビンごとの，処理を開始してからの最小エネルギーで，$\boldsymbol{S}^{tmp}$ は，L フレームごとに更新される暫定最小エネルギーである．

$$S_m^{tmp}(f,k_i) = \begin{cases} S_m^{smo}(f,k_i), & \text{if } f = nL \\ \min\{S_m^{tmp}(f-1,k_i), S_m^{smo}(f,k_i)\}, & \text{if } f \neq nL \end{cases} \tag{6.21}$$

$$S_m^{min}(f,k_i) = \begin{cases} \min\{S_m^{tmp}(f-1,k_i), S_m^{smo}(f,k_i)\}, & \text{if } f = nL \\ \min\{S_m^{min}(f-1,k_i), S_m^{smo}(f,k_i)\}, & \text{if } f \neq nL \end{cases} \tag{6.22}$$

ただし，n は任意の整数である．

3．定常ノイズ推定:

1. 目的音有無の判定
   以下にいずれかが成り立つ場合，該当する時間，周波数に目的音のパワーは存在せず，ノイズのみがあるとみなされる．

$$S_m^{smo}(f,k_i) < \delta S_m^{min}(f,k_i) \text{ または} \tag{6.23}$$

$$f < N \text{ または} \tag{6.24}$$

$$S_m^{smo}(f,k_i) < \lambda_m(f-1,k_i) \tag{6.25}$$

2. 平滑化係数の算出
   定常ノイズのパワーを計算する際に用いられる平滑化係数 α_d は,

$$\alpha_d = \begin{cases} \frac{1}{f+1}, & \text{if } (\frac{1}{f+1} \geq \alpha_d^{min}) \\ \alpha_d^{min}, & \text{if } (\frac{1}{t+1} < \alpha_d^{min}) \\ 0 & (\text{目的音が含まれるとき}) \end{cases} \tag{6.26}$$

   として計算する．定常ノイズは以下の式によって求める．

$$\lambda_m(f,k_i) = (1-\alpha_d)\lambda_m(f-1,k_i) + \alpha_d \theta S_m(f,k_i) \tag{6.27}$$

6.3.2 CalcSpecSubGain

ノードの概要

信号＋ノイズのパワースペクトルからノイズパワースペクトルを除去する時に，推定されたノイズのパワースペクトルをどの程度除去するべきかの最適ゲインを決定するノードである．その他，音声存在確率（6.3.7 節参照）を出力する．ただし，このノードは音声存在確率を常に 1 として出力する．分離音のパワースペクトルと推定ノイズのパワースペクトルの差分を出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

HRLE ノードを用いたノイズ推定時に用いる．

典型的な接続例

CalcSpecSubGain の接続例は図 6.39 の通り．入力は GHDSS で分離後のパワースペクトルおよび HRLE で推定されたノイズのパワースペクトル．出力は VOICE_PROB，GAIN を SpectralGainFilter に接続する．

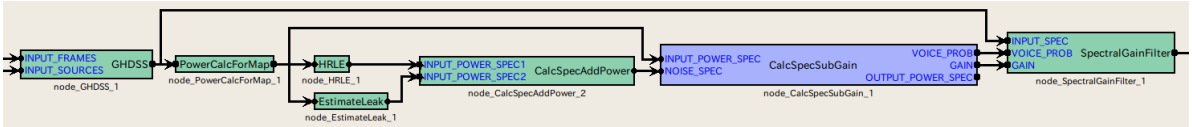


図 6.39: CalcSpecSubGain の接続例

ノードの入出力とプロパティ

表 6.33: CalcSpecSubGain のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
ALPHA	float	1.0		スペクトル減算のゲイン
BETA	float	0.0		最適ゲインの最小値
SS_METHOD	int	2		パワー・振幅スペクトル減算の選択

入力

INPUT_POWER_SPEC ：Map<int, ObjectRef> 型．音源 ID と分離音のパワースペクトルの Vector<float> 型データのペア．

NOISE_SPEC ：`Map<int, ObjectRef>` 型．音源 ID と推定ノイズのパワースペクトルの `Vector<float>` 型データのペア．

出力

VOICE_PROB ：`Map<int, ObjectRef>` 型．音源 ID と音声存在確率の `Vector<float>` 型データのペア．

GAIN ：`Map<int, ObjectRef>` 型．音源 ID と最適ゲインの `Vector<float>` 型データのペア．

OUTPUT_POWER_SPEC ：`Map<int, ObjectRef>` 型．音源 ID と分離音から推定ノイズを差し引いたパワースペクトル `Vector<float>` 型データのペア．

パラメータ

ALPHA ：スペクトル減算のゲイン

BETA ：最適ゲインの最小値

SS_METHOD ：パワースペクトル減算か振幅スペクトル減算かの選択

ノードの詳細

信号＋ノイズのパワースペクトルからノイズパワースペクトルを除去する時に，推定されたノイズのパワースペクトルをどの程度除去するべきかの最適ゲインを決定するノードである．音声存在確率（6.3.7 節参照）も出力する．ただし，このノードは音声存在確率を常に 1 として出力する．

分離音からノイズを差し引いたパワースペクトルを $Y_n(k_i)$，分離音のパワースペクトルを $X_n(k_i)$，推定されたノイズのパワースペクトルを $N_n(k_i)$ とすると， OUTPUT_POWER_SPEC からの出力は次のように表される．

$$Y_n(k_i) = X_n(k_i) - N_n(k_i) \tag{6.28}$$

ただし，n は，分析フレーム番号．k_i は，周波数インデクスを表す．最適ゲイン $G_n(k_i)$ は，次のように表される．

$$G_n(k_i) = \begin{cases} \mathrm{ALPHA}\frac{Y_n(k_i)}{X_n(k_i)}, & \text{if } Y_n(k_i) > \mathrm{BETA}, \\ \mathrm{BETA}, & \text{if otherwise.} \end{cases} \tag{6.29}$$

単純に $Y_n(k_i)$ を用いて処理すると，パワーが負になりえる．以後の処理で，パワースペクトルの取り扱いが困難になるので，予め，パワーが負にならないようにノイズのパワースペクトルを除去するためのゲインを計算するのが本ノードの狙いである．

6.3.3 CalcSpecAddPower

ノードの概要

2 つの入力パワースペクトルを加算したスペクトルを出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

HRLE ノードを用いたノイズ推定時に用いる．HRLE ノードで推定されたノイズのパワースペクトルと EstimateLeak で推定されたノイズのパワースペクトルを加算し，トータルのノイズパワースペクトルを求める．

典型的な接続例

CalcSpecAddPower の接続例は図 6.40 の通り．入力は HRLE で推定されたノイズのパワースペクトル及び，EstimateLeak で推定されたノイズのパワースペクトル．出力は CalcSpecSubGain に接続する．

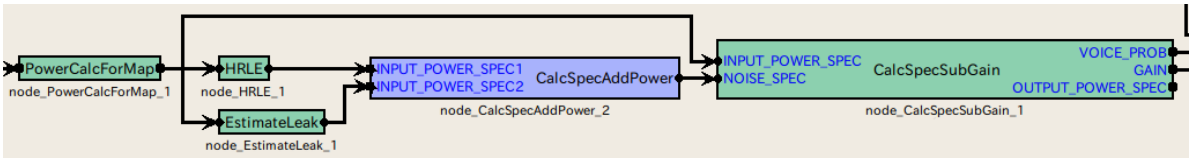


図 6.40: CalcSpecAddPower の接続例

ノードの入出力とプロパティ

入力

INPUT_POWER_SPEC1 ：`Map<int, ObjectRef>` 型．音源 ID とパワースペクトルの `Vector<float>` 型データのペア．

INPUT_POWER_SPEC2 ：`Map<int, ObjectRef>` 型．音源 ID とパワースペクトルの `Vector<float>` 型データのペア．

出力

OUTPUT_POWER_SPEC ：`Map<int, ObjectRef>` 型．音源 ID と 2 つの入力を加算したパワースペクトル `Vector<float>` 型データのペア．

パラメータ

無し．

ノードの詳細

本ノードは，2 つの入力パワースペクトルを加算したスペクトルを出力する．

6.3.4 EstimateLeak

ノードの概要

他チャネルからの漏れ成分の推定を行う．

必要なファイル

無し．

使用方法

どんなときに使うのか

GHDSS を使った音源分離後のノイズ除去に用いる．

典型的な接続例

EstimateLeak の接続例は図 6.41 の通り．入力は で音声のパワースペクトルで，GHDSS の出力である．出力は CalcSpecAddPower に接続する．

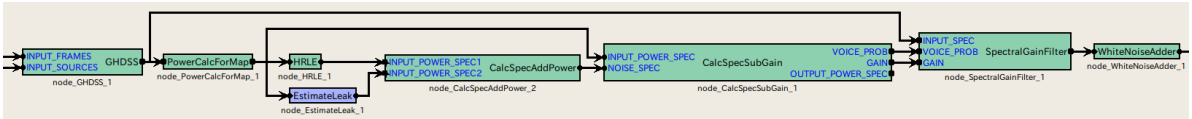


図 6.41: EstimateLeak の接続例

ノードの入出力とプロパティ

入力

INPUT_POWER_SPEC ：`Map<int, ObjectRef>` 型．音源 ID とパワースペクトルの `Vector<float>` 型データのペア．

出力

LEAK_POWER_SPEC ：`Map<int, ObjectRef>` 型．音源 ID と漏れノイズのパワースペクトル `Vector<float>` 型データのペア．

パラメータ

表 6.34: EstimateLeak のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LEAK_FACTOR	`float`	0.25		漏れ率．
OVER_CANCEL_FACTOR	`float`	1		漏れ率重み係数．

ノードの詳細

本ノードは，他チャネルからの漏れ成分の推定を行う．詳細は 6.3.7 節の PostFilter ノード 1-b) 漏れノイズ推定を参照のこと．

6.3.5 GHDSS

ノードの概要

GHDSS (Geometric High-order Dicorrelation-based Source Separation) アルゴリズムに基づいて，音源分離処理を行う．GHDSS アルゴリズムは，マイクロホンアレイを利用した処理で，

1. 音源信号間の高次無相関化
2. 音源方向へ指向性の形成

という 2 つの処理を行う．2. の指向性は，事前に与えられたマイクロホンの位置関係を幾何的制約として処理を行う．また，HARK に実装されている GHDSS アルゴリズムは，マイクロホンの位置関係に相当する情報として，マイクロホンアレイの伝達関数を利用することができる．

ノードの入力は，混合音のマルチチャネル複素スペクトルと，音源方向のデータである．また，出力は分離音ごとの複素スペクトルである．

必要なファイル

表 6.35: GHDSS に必要なファイル

対応するパラメータ名	説明
TF_CONJ_FILENAME	マイクロホンアレーの伝達関数
INITW_FILENAME	分離行列初期値

使用方法

どんなときに使うのか

所与の音源方向に対して，マイクロホンアレイを用いて当該方向の音源分離を行う．なお，音源方向として，音源定位部での推定結果，あるいは，定数値を使用することができる．

典型的な接続例

GHDSS ノードの接続例を図 6.42 に示す．入力は以下の 2 つが必要である．

1. INPUT_FRAMES には，混合音の多チャネル複素スペクトル，
2. INPUT_SOURCES には，音源方向．

出力である分離音声に対して音声認識を行うために，MelFilterBank などを利用して，音声特徴量に変換する以外に，以下のような音声認識の性能向上方法もある．

1. PostFilter ノードを利用して，音源分離処理によるチャネル間リークや拡散性雑音を抑圧する（図 6.42 右上参照）．
2. PowerCalcForMap ，HRLE ，SpectralGainFilter を接続して，音源分離処理によるチャネルリークや拡散性雑音を抑圧する（PostFilter と比較して，チューニングが容易）．

3. PowerCalcForMap，MelFilterBank，MFMGeneration を接続して，ミッシングフィーチャ理論を用いた音声認識を行うために，ミッシングフィーチャーマスクを生成する（図 6.42 右下参照）．

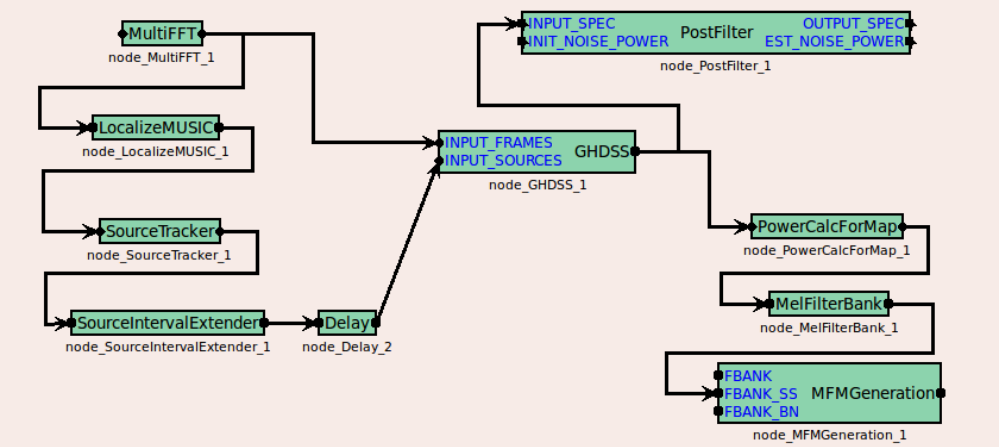


図 6.42: GHDSS の接続例

ノードの入出力とプロパティ

入力

INPUT_FRAMES：`Matrix<complex<float> >` 型．マルチチャネル複素スペクトル．行がチャネル，つまり，各マイクロホンから入力された波形の複素スペクトルに対応し，列が周波数ビンに対応する．

INPUT_SOURCES：`Vector<ObjectRef>` 型．音源定位結果等が格納された `Source` 型オブジェクトの `Vector` 配列である．典型的には，SourceTracker ノード，SourceIntervalExtender ノードと繋げ，その出力を用いる．

出力

OUTPUT：`Map<int, ObjectRef>` 型．分離音の音源 ID と，分離音の 1 チャネル複素スペクトル (`Vector<complex<float> >` 型) のペア．

パラメータ

LENGTH：`int` 型．分析フレーム長．前段階における値（AudioStreamFromMic，MultiFFT ノードなど）と一致している必要がある．

ADVANCE：`int` 型．フレームのシフト長．前段階における値（AudioStreamFromMic，MultiFFT ノードなど）と一致している必要がある．

SAMPLING_RATE：`int` 型．入力波形のサンプリング周波数．

LOWER_BOUND_FREQUENCY GHDSS 処理行う際に利用する最小周波数値であり，これより下の周波数に対しては処理を行わず，出力スペクトルの値は 0 となる．0 以上サンプリング周波数値の半分までの範囲で指定する．

表 6.36: GHDSS のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	分析フレーム長.
ADVANCE	int	160	[pt]	フレームのシフト長.
SAMPLING_RATE	int	16000	[Hz]	サンプリング周波数.
LOWER_BOUND_FREQUENCY	int	0	[Hz]	分離処理で用いる周波数の最小値
UPPER_BOUND_FREQUENCY	int	8000	[Hz]	分離処理で用いる周波数の最大値
TF_CONJ_FILENAME	string			マイクロホンアレーの伝達関数を記したファイル名.
INITW_FILENAME	string			分離行列の初期値を記述したファイル名.
SS_METHOD	string	ADAPTIVE		高次無相関化に基づくステップサイズの算出方法．FIX, LC_MYU, ADAPTIVE から選択．FIX は固定値，LC_MYU は幾何制約に基づくステップサイズに連動した値，ADAPTIVE は自動調節.
SS_METHOD==FIX				以下 SS_METHOD が FIX の時に有効.
SS_MYU	float	0.001		高次無相関化に基づく分離行列更新時のステップサイズ
SS_SCAL	float	1.0		高次相関行列計算におけるスケールファクタ
NOISE_FLOOR	float	0.0		入力信号をノイズとみなす振幅の閾値 (上限)
LC_CONST	string	DIAG		幾何制約の手法を決める．DIAG, FULL から選択．DIAG は目的音源方向に焦点を合わせるのみ．FULL では目的方向の焦点に加えて，非目的音源方向に死角を形成する.
LC_METHOD	string	ADAPTIVE		幾何制約に基づくステップサイズの算出方法．FIX，ADAPTIVE から選択．FIX は固定値，ADAPTIVE は自動調節.
LC_METHOD==FIX				以下 LC_METHOD が FIX の時に有効.
LC_MYU	float	0.001		高次無相関化に基づく分離行列更新時のステップサイズ
UPDATE_METHOD_TF_CONJ	string	POS		伝達関数を更新する手法を指定．POS，ID から選択.
UPDATE_METHOD_W	string	ID		分離行列を更新する手法を指定． ID，POS，ID_POS から選択.
UPDATE_SEARCH_AZIMUTH	float		[deg]	分離処理において，同一音源を検索する際の方位角方向の検索範囲.
UPDATE_SEARCH_ELEVATION	float		[deg]	分離処理において，同一音源を検索する際の仰角方向の検索範囲.
UPDATE_ACCEPT_ANGLE	float	5.0	[deg]	分離処理において，同一音源とみなす角度差の閾値.
EXPORT_W	bool	false		分離行列をファイルに書き出すかを指定.
EXPORT_W==true				以下 EXPORT_W が true の時に有効.
EXPORT_W_FILENAME	string			分離行列を書きだすファイル名.
UPDATE	string	STEP		分離行列の更新方法を決める．STEP，TOTAL から選択．STEP は高次無相関化に基づく更新を行った後に，幾何制約に基づく更新を行う．TOTAL では，高次無相関化に基づく更新と幾何制約に基づく更新を同時に行う.

UPPER_BOUND_FREQUENCY GHDSS 処理を行う際に利用する最大周波数値であり，これより上の周波数に対しては処理を行わず，出力スペクトルの値は 0 となる．LOWER_BOUND_FREQUENCY < UPPER_BOUND_FREQUENCY である必要がある.

TF_CONJ_FILENAME: string 型．伝達関数の記述されたバイナリファイル名を記す．ファイルフォーマットは 5.1.2 節を参照.

INITW_FILENAME ：string 型．分離行列の初期値を記したファイル名．事前の計算により，値の収束した

分離行列を初期値として与えることで，音が鳴り始めた最初の部分から精度よく分離することが可能となる．ここで与えるファイルは，EXPORT_W を `true` にすることで，予め用意しておく必要がある．ファイルフォーマットは 5.1.3 節を参照.

SS_METHOD : `string` 型．高次無相関化に基づくステップサイズの算出方法を選ぶ．ユーザが指定した値に固定する場合は FIX，幾何制約に基づくステップサイズに連動した値にする場合は LC_MYU，自動調節する場合は ADAPTIVE を指定.

1. FIX のとき: SS_MYU を設定する．
   SS_MYU: `float` 型．0.01 がデフォルト．高次無相関化に基づく分離行列更新時のステップサイズを指定する．この値と LC_MYU を 0 にし，Delay and Sum 型のビームフォーマの分離行列を INITW_FILENAME として渡すことで，GHDSS は，Delay and Sum 型のビームフォーマと等価な処理が可能となる．

SS_SCAL : `float` 型．1.0 がデフォルト．高次相関行列計算における双曲線正接関数 (tanh) のスケールファクタを指定する．0 より大きい正の実数を指定する．値が小さいほど非線形性が少なくなり通常の相関行列計算に近づく．

NOISE_FLOOR : `float` 型．0 がデフォルト．入力信号をノイズとみなす振幅の閾値 (上限) を指定する．入力信号の振幅がこの値以下の場合，ノイズ区間とみなされ，分離行列の更新がされない．ノイズが大きく，分離行列が安定して収束しない場合に，正の実数を指定する．

LC_CONST : `string` 型．幾何制約の手法を選択する．目的音源方向に焦点を合わせる幾何制約のみの場合は DIAG 目的方向の焦点に加えて，非目的音源方向に死角を形成する場合は FULL を指定する．死角は高次無相関化によって自動的に形成されるため， DIAG でも高精度な分離が可能．デフォルトは DIAG.

LC_METHOD : `string` 型．幾何制約に基づくステップサイズの算出方法を選ぶ．ユーザが指定した値に固定する場合は FIX，自動調節する場合は ADAPTIVE を指定.

1. FIX のとき: LC_MYU を設定する．
   LC_MYU: `float` 型．0.001 がデフォルト．幾何制約に基づく分離行列更新時のステップサイズを指定する．この値と LC_MYU を 0 にし，Delay and Sum 型のビームフォーマの分離行列を INITW_FILENAME として渡すことで，GHDSS は，Delay and Sum 型のビームフォーマと等価な処理が可能となる．

UPDATE_METHOD_TF_CONJ : `string` 型．ID または POS を指定する．POS がデフォルト．伝達関数の複素共役 TF_CONJ の更新をするかの判断を，各音源に付与された ID に基づいて行う (ID の場合) か，音源位置によって行う (POS の場合) かを指定する．

UPDATE_METHOD_W : `string` 型．ID，POS または ID_POS を指定．ID がデフォルト．音源位置情報が変わった際に，分離行列の再計算が必要となる．この時の音源位置情報が変わったとみなす方法を指定する．分離行列は，内部で対応する音源 ID や音源方向の角度とともに一定時間保存され，一度音が止んでも，同一の方向からの音源と判断される音が検出されると，再び保存された分離行列の値を用いて分離処理が行われる．このとき，分離行列の更新を行うかどうかの基準を設定する．ID を設定した場合，音源 ID によって同方向音源かどうか判断する．POS を設定した場合，音源方向を比較して判断する．ID_POS を設定した場合，音源 ID を比較し，同一と判断されなかった場合は，さらに音源方向の角度を比較して判断を行う．

UPDATE_SEARCH_AZIMUTH : `float` 型．空欄 がデフォルト．単位は [deg]．UPDATE_METHOD_TF_CONJ や，UPDATE_METHOD_W で POS，ID_POS を設定した場合の，同一音源判断時における方位角方向検

索範囲を設定する．値 x を設定した際は当該音源の方位角の前後 x °の範囲で検索が行われる．入力欄を空欄とした場合は検索範囲を限定しない．適度な値を入力することで計算処理時間を短縮できる．

UPDATE_SEARCH_ELEVATION ：`float` 型．空欄 がデフォルト．単位は [deg]．UPDATE_METHOD_TF_CONJ や，UPDATE_METHOD_W で POS，ID_POS を設定した場合の，同一音源判断時における仰角方向検索範囲を設定する．値 x を設定した際は当該音源の仰角の前後 x °の範囲で検索が行われる．入力欄を空欄とした場合は検索範囲を限定しない．適度な値を入力することで計算処理時間を短縮できる．

UPDATE_ACCEPT_ANGLE ： `float` 型．5 がデフォルト．単位は [deg]．UPDATE_METHOD_TF_CONJ や，UPDATE_METHOD_W で POS，ID_POS を設定した場合の，同一音源と判断する角度の許容誤差を設定する．

EXPORT_W ：`bool` 型． `false` がデフォルト．GHDSS により更新された分離行列の結果を出力するかどうかを設定．`true` のとき，EXPORT_W_FILENAME を指定．

EXPORT_W_FILENAME ：`string` 型．EXPORT_W が `true` のとき有効．分離行列を書きだすファイル名を指定．フォーマットは 5.1.3 節を参照．

ノードの詳細

音源分離の定式化: 音源分離問題の定式化で用いる記号を表 6.37 にまとめる．インデックスの意味は，表 6.1 に準拠する．演算は周波数領域において行われるため，各記号は周波数領域での，一般には複素数の値を表す．伝達関数以外は一般に時間変化するが，同じ時間フレームにおける演算の場合は，時間インデックス f を省略して表記する．また，以下の演算は周波数ビン k_i について述べる．実際には，K 個それぞれの周波数ビン $k_0, \ldots, k_{K-1}$ に対して演算が行われている．

表 6.37: 変数の定義

変数	説明
$S(k_i) = [S_1(k_i), \ldots, S_N(k_i)]^T$	周波数ビン k_i に対応する音源複素スペクトルのベクトル．
$X(k_i) = [X_1(k_i), \ldots, X_M(k_i)]^T$	マイクロホン観測複素スペクトルのベクトル，INPUT_FRAMES に対応．
$N(k_i) = [N_1(k_i), \ldots, N_M(k_i)]^T$	各マイクロホンに作用する加法性ノイズ．
$H(k_i) = [H_{m,n}(k_i)]$	反射，回折などを含む伝達関数行列 $(M \times N)$．
$H_D(k_i) = [H_{Dm,n}(k_i)]$	直接音の伝達関数行列 $(M \times N)$．
$W(k_i) = [W_{n,m}(k_i)]$	分離行列 $(N \times M)$．
$Y(k_i) = [Y_1(k_i), \ldots, Y_N(k_i)]^T$	分離音複素スペクトル．
μ_{SS}	高次無相関化に基づく分離行列更新時のステップサイズ，SS_MYU に対応．
μ_{LC}	幾何制約に基づく分離行列更新時のステップサイズ．LC_MYU に対応．

混合モデル N 個の音源から発せられた音は，その空間の伝達関数 $H(k_i)$ の影響を受け，M 個のマイクロホンを通じて式 (6.30) のように観測される．

$$X(k_i) = H(k_i)S(k_i) + N(k_i). \tag{6.30}$$

一般に，伝達関数 $H(k_i)$ は，部屋の形や，マイクロホンと音源の位置関係により変化するため，推定は困難である．

しかし，音の反射や回折を無視して，直接音のみに限定した伝達関数 $H_D(k_i)$ は，音源とマイクロホンの相対位置が分かっている場合は，次の式 (6.31) のように計算可能である．

$$H_{Dm,n}(k_i) = \exp\left(-j2\pi l_i r_{m,n}\right), \tag{6.31}$$

$$l_i = \frac{2\pi\omega_i}{c}, \tag{6.32}$$

ただし，c は音速で，l_i は周波数ビン k_i での周波数 ω_i に対応する波数とする．また，$r_{m,\,n}$ は，マイクロホン m から音源 n までの距離と，座標系の基準点 (たとえば原点) から音源 n までの距離の差である．つまり，音源から各マイクロホンまでの到達時間の差から生じる位相差として，$H_D(k_i)$ は定義される．

分離モデル　分離音の複素スペクトルの行列 $Y(k_i)$ は，

$$Y(k_i) = W(k_i)X(k_i) \tag{6.33}$$

として求める．GHDSS アルゴリズムは，$Y(k_i)$ が $S(k_i)$ に近づくように，分離行列 $W(k_i)$ を推定する．

モデルにおける仮定　このアルゴリズムで既知と仮定する情報は次の通り．

1. 音源数 N
2. 音源位置 (HARK では LocalizeMUSIC ノードが音源位置を推定する)
3. マイクロホン位置
4. 直接音成分の伝達関数 $H_D(k_i)$ (測定する or 式 (6.31) による近似)

未知の情報としては，

1. 観測時の実際の伝達関数 $H(k_i)$
2. 観測ノイズ $N(k_i)$

分離行列の更新式　GHDSS は，下記を満たすように分離行列 $W(k_i)$ の推定を行う．

1. 分離信号を高次無相関化
   すなわち，分離音 $Y(k_i)$ の高次相関行列 $R^{\phi(y)y}(k_i) = E[\phi(Y(k_i))Y^H(k_i)]$ の対角成分以外が 0 になるようにする．ここで ^{H} 作用素はエルミート転置を，$E[]$ は時間平均作用素を，$\phi()$ は非線形関数であり，本ノードでは下記で定義される双曲線正接関数を用いている．

$$\phi(Y) = [\phi(Y_1), \phi(Y_2), \ldots, \phi(Y_N)]^T \tag{6.34}$$

$$\phi(Y_k) = \tanh(\sigma|Y_k|)\exp(j\angle(Y_k)) \tag{6.35}$$

   ここで，σ はスケーリングファクタ（SS_SCAL に対応）である．

2. 直接音成分は歪みなく分離される (幾何的制約)
   分離行列 $W(k_i)$ と 直接音の伝達関数 $H_D(k_i)$ の積が単位行列になるようにする ($W(k_i)H_D(k_i) = I$).

上の 2 つの要素をあわせた評価関数は以下のようになる．簡単のため，周波数ビン k_i は略す．

$$J(W) = \alpha J_1(W) + \beta J_2(W), \tag{6.36}$$

$$J_1(W) = \sum_{i \neq j} |R^{\phi(y)y}_{i,j}|^2, \tag{6.37}$$

$$J_2(W) = \|WH_D - I\|^2, \tag{6.38}$$

ただし，α および β は重み係数である．また行列のノルムは $\|M\|^2 = tr(MM^H) = \sum_{i,j}|m_{i,j}|^2$ として定義される．

式 (6.36) を最小化するための分離行列の更新式は，複素勾配演算 $\frac{\partial}{\partial W^*}$ を利用した勾配法から，

$$W(k_i, f+1) = W(k_i, f) - \mu \frac{\partial J}{\partial W^*}(W(k_i, f)) \tag{6.39}$$

となる．ここで，μ は分離行列の更新量を調節するステップサイズである．通常，式 (6.39) の右辺にある複素勾配を求めると，$R^{xx} = E[XX^H]$ や $R^{yy} = E[YY^H]$ などの期待値計算に複数のフレームの値を要する．GHDSS ノードでの計算は，自己相関行列を求めず，1 つのフレームだけを用いた以下の更新式 (6.40) を用いる．

$$W(k_i, f+1) = W(k_i, f) - \left[\mu_{SS}\frac{\partial J_1}{\partial W^*}(W(k_i, f)) + \mu_{LC}\frac{\partial J_2}{\partial W^*}(W(k_i, f))\right], \tag{6.40}$$

$$\frac{\partial J_1}{\partial W^*}(W) = \left(\phi(Y)Y^H - \mathrm{diag}[\phi(Y)Y^H]\right)\tilde{\phi}(\mathrm{boldmath}WX)X^H, \tag{6.41}$$

$$\frac{\partial J_2}{\partial W^*}(W) = 2\,(WH_D - I)\,H_D^H, \tag{6.42}$$

ここで，$\tilde{\phi}$ は ϕ の偏微分であり，下記で定義される．

$$\tilde{\phi}(Y) = [\phi(\tilde{Y}_1), \phi(\tilde{Y}_2), \ldots, \phi(\tilde{Y}_N)]^T \tag{6.43}$$

$$\tilde{\phi}(Y_k) = \phi(Y_k) + Y_k\frac{\partial \phi(Y_k)}{\partial Y_k} \tag{6.44}$$

また，$\mu_{SS} = \mu\alpha$，$\mu_{LC} = \mu\beta$ であり，それぞれ高次無相関化および幾何制約に基づくステップサイズをあらわす．ステップサイズの調節を自動にした場合，ステップサイズは次式で計算される．

$$\mu_{SS} = \frac{J_1(W)}{2\|\frac{\partial J_1}{\partial W}(W)\|^2} \tag{6.45}$$

$$\mu_{LC} = \frac{J_2(W)}{2\|\frac{\partial J_2}{\partial W}(W)\|^2} \tag{6.46}$$

式 (6.41，6.42) では，各変数のインデックスを省略したが，いずれも (k_i, f) である．

分離行列の初期値は次のようにして求める．

$$W(k_i) = H_D^H(k_i)/M, \tag{6.47}$$

ただし，M はマイクロホン数である．

処理の流れ:

GHDSS ノードにおける，時間フレーム f における主な処理を図 6.43 に示す．より詳細には，以下のように固定ノイズに関する処理などが含まれる．

1. (直接音) 伝達関数の取得
2. 分離行列 W 推定
3. 式 (6.33) に従って音源分離処理を実行
4. 分離行列の書き出し (EXPORT_W が `true` のとき)

伝達関数の取得: パラメータ TF_CONJ_FILENAME で指定された伝達関数から，入力された音源定位結果の方向に最も近い位置にあるデータを検索する．

2 フレーム以降については，以下の通りである．

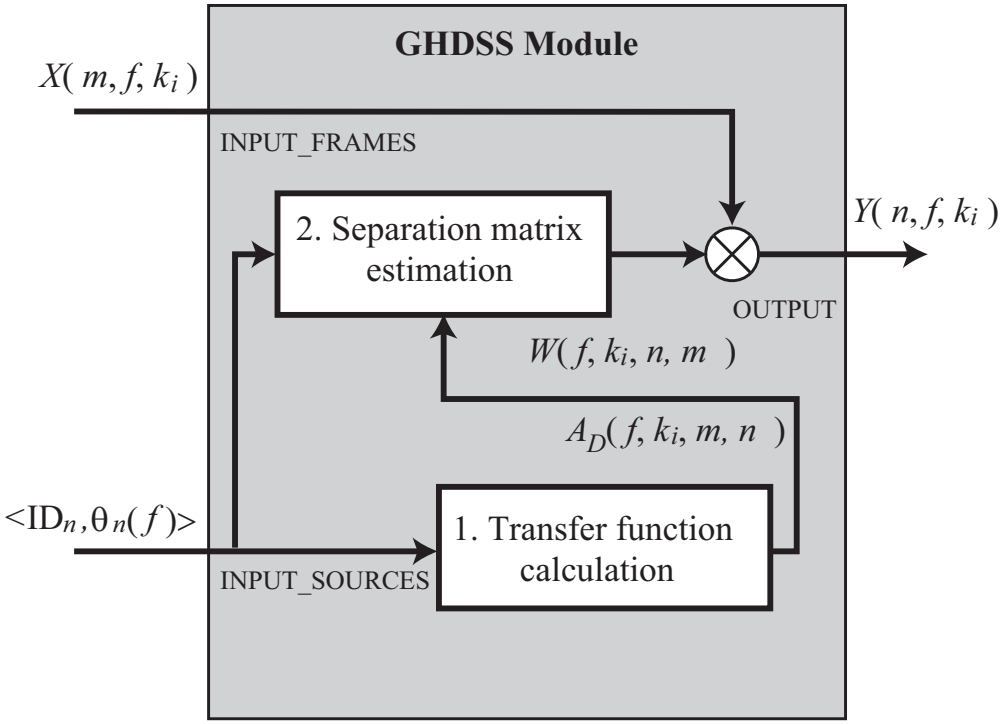


図 6.43: GHDSS の流れ図

- パラメータ UPDATE_METHOD_TF_CONJ の値によって以下のように，前フレームの伝達関数を引き継ぐか，ファイルから読み込むかを決定する．

UPDATE_METHOD_TF_CONJ が ID

1. 1 フレーム前の ID と取得した ID を比較
   - 同じ: 引き継ぐ
   - 異なる: 読み込む

UPDATE_METHOD_TF_CONJ が POS

1. 1 フレーム前の音源方向と取得した方向を比較
   - 誤差が UPDATE_ACCEPT_ANGLE 未満: 引き継ぐ
   - 誤差が UPDATE_ACCEPT_ANGLE 以上: 読み込む

分離行列の推定: 分離行列の初期値は，パラメータ INITW_FILENAME に値を指定するかによって異なる．パラメータ INITW_FILENAME が指定されていないときは，伝達関数 H_D から分離行列 W を計算する．パラメータ INITW_FILENAME が指定されているときは，指定された分離行列から，入力された音源定位結果の方向に最も近い位置にあるデータを検索する．

2 フレーム以降については，以下の通りである．

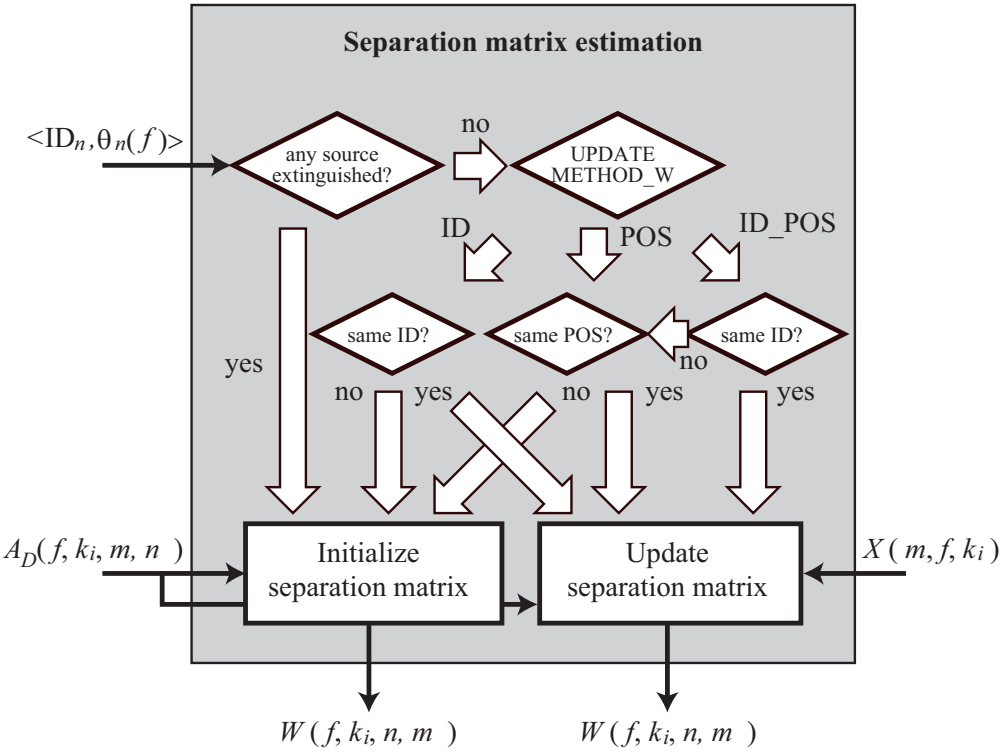


図 6.44: 分離行列推定の流れ図

分離行列を推定するまでの流れを図 6.44 に示す．ここでは，式 (6.40) に従って，前フレームの分離行列を更新するか，式 (6.47) を用いて，伝達関数を利用して分離行列の初期値の導出が行われる．

- 前フレームの音源定位情報を参照して，消滅した音源がある場合，分離行列は初期化される．
- 音源数に変化がない場合，UPDATE_METHOD_W の値によって分岐する．前フレームにおける，音源 ID，定位方向を，現在のフレームと比較して，分離行列を継続して使うか，初期化するかを決定する．

UPDATE_METHOD_W が ID

1. 前フレーム ID と比較
   - 同じ: W を更新
   - 異なる: W を初期化

UPDATE_METHOD_W が POS

1. 前フレーム定位方向と比較
   - 誤差が UPDATE_ACCEPT_ANGLE 未満: W を更新
   - 誤差が UPDATE_ACCEPT_ANGLE 以上: W を初期化

UPDATE_METHOD_W が ID_POS

1. 前フレーム ID と比較
   - 同じ: W を更新
2. ID が異なった場合，定位方向を比較
   - 誤差が UPDATE_ACCEPT_ANGLE 未満: W を更新
   - 誤差が UPDATE_ACCEPT_ANGLE 以上: W を初期化

分離行列の書き出し (EXPORT_W が `true` のとき): EXPORT_W が `true` のとき，収束した分離行列を EXPORT_W_FILENAME で指定したファイルに出力する．
複数の音源が検出された場合，それらの分離行列は全て１つのファイルに出力される．音源が消滅した時点で，その分離行列をファイルに書き出す．
ファイルに書き出す際は，既に保存されている音源の定位方向と比較して，既存音源を上書きするか，新たな音源として追加するかを決定する．

音源が消滅

1. 既に保存されている音源の定位方向と比較
   - 誤差が UPDATE_ACCEPT_ANGLE 未満: W を上書き保存
   - 誤差が UPDATE_ACCEPT_ANGLE 以上: W を追加保存

6.3.6 HRLE

ノードの概要

本ノードは，Histogram-based Recursive Level Estimation (HRLE) 法によって定常ノイズレベルを推定する．HRLE は，入力スペクトルのヒストグラム（頻度分布）を計算し，その累積分布とパラメータ Lx により指定した正規化累積頻度からノイズレベルを推定する．ヒストグラムは，指数窓により重み付けされた過去の入力スペクトルから計算され，１フレームごとに指数窓の位置は更新される．

必要なファイル

無し

使用方法

どんなときに使うのか

スペクトル減算によるノイズ抑圧を行うときに用いる．

典型的な接続例

図 6.45 に示すように，入力は GHDSS などの分離ノードの後に接続し，出力は CalcSpecSubGain などの最適ゲインを求めるノードに接続する．図 6.46 は，EstimateLeak を併用した場合の接続例である．

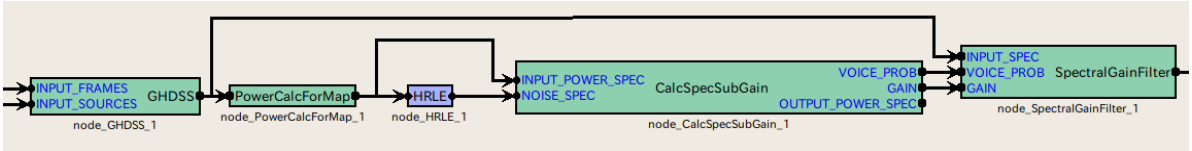


図 6.45: HRLE の接続例 1

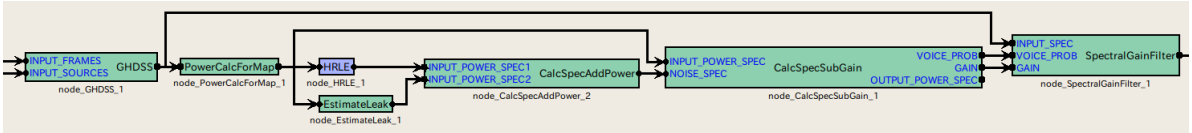


図 6.46: HRLE の接続例 2

ノードの入出力とプロパティ

入力

INPUT_SPEC ：`Map<int, float>` 型． 入力信号のパワースペクトル

出力

NOISE_SPEC ：`Map<int, float>` 型． 推定ノイズのパワースペクトル

パラメータ

表 6.38: HRLE のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LX	float	0.85		正規化累積頻度（Lx 値）.
TIME_CONSTANT	float	16000	[pt]	時定数.
NUM_BIN	float	1000		ヒストグラムのビン数.
MIN_LEVEL	float	-100	[dB]	ヒストグラムの最小レベル.
STEP_LEVEL	float	0.2	[dB]	ヒストグラムのビンの幅.
DEBUG	bool	false		デバッグモード.

LX : float 型．デフォルトは 0.85．累積頻度分布上の正規化累積頻度を 0–1 の範囲で指定する．0 を指定すると最小レベル，1 を指定すると最大レベル，0.5 を指定するとメジアン（中央値）を推定する．

TIME_CONSTANT : float 型．デフォルトは 16000．時定数（0 以上）を時間サンプル単位で指定する．

NUM_BIN : float 型．デフォルトは 1000．ヒストグラムのビン数を指定する．

MIN_LEVEL : float 型．デフォルトは -100．ヒストグラムの最小レベルを dB 単位で指定する．

STEP_LEVEL : float 型．デフォルトは 0.2．ヒストグラムのビンの幅を dB 単位で指定する．

DEBUG : bool デフォルトは false．デバッグモードを指定する．　デバッグモード (true) の場合，累積ヒストグラムの値が標準出力にコンマ区切りテキストファイル形式で 100 フレーム毎に 1 回出力される．　出力値は，複数の行と列を含む複素行列数値形式であり，行は周波数ビンの位置，列はヒストグラムの位置，各要素は丸括弧で区切られた複素数値（左側が実数，右側が虚数部）を示す.（累積ヒストグラムは，実数値であるため，通常では虚数部は 0 である．しかし今後のバージョンでも 0 であることは保障されない.）　1 つのサンプルに対する累積ヒストグラムの加算値は，1 ではなく指数的に増大している（高速化のため）．　そのため累積ヒストグラム値は，累積頻度そのものを表してはいない事に注意されたい．　各行の累積ヒストグラム値のほとんどが 0 で，最後の列に近い位置のみに値を含む場合，入力値が設定したヒストグラムのレベル範囲を超えて大きい状態（オーバーフロー状態）にあるので，NUM_BIN, MIN_LEVEL,STEP_LEVEL の一部またはすべてを高い値に設定しなおすべきである．　また逆に各行の累積ヒストグラム値がほとんど一定値で，最初の列に近い位置のみに異なる低い値が含まれる場合，入力値が設定したヒストグラムのレベル範囲より小さい状態（アンダーフロー状態）にあるので，MIN_LEVEL を低い値に設定しなおすべきである．出力の例：

```
---------- Compmat.disp() ----------
[(1.00005e-18,0), (1.00005e-18,0), (1.00005e-18,0), ..., (1.00005e-18,0);
(0,0), (0,0), (0,0), ..., (4.00084e-18,0);
...
(4.00084e-18,0), (4.00084e-18,0), (4.00084e-18,0), .., , (4.00084e-18,0)]^T
Matrix size = 1000 x 257
```

ノードの詳細

図 6.47 に HRLE の処理フローを示す．HRLE は，入力パワーからレベルのヒストグラムを求め，その累積分布から Lx レベルを推定する処理となっている．Lx レベルとは，図 6.48 に示すように，累積頻度分布上の正規化累積頻度が x になるレベルである．x は，パラメータであり，例えば，$x = 0$ であれば最小値，$x = 1$ であれば最大値，$x = 0.5$ であれば中央値を推定する処理となる．

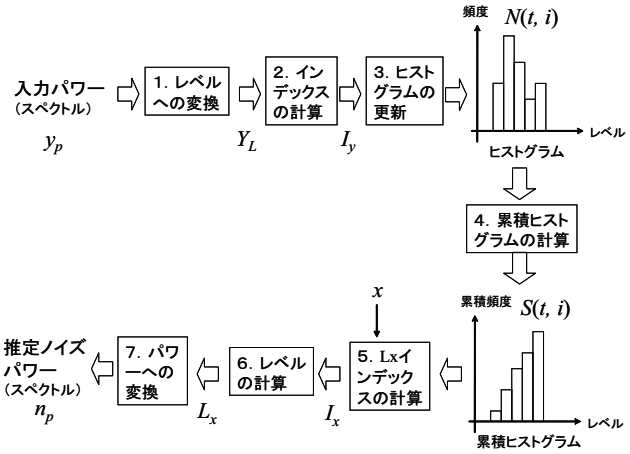


図 6.47: HRLE の処理フロー

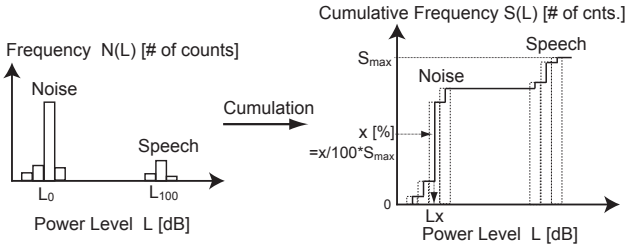


図 6.48: Lx 値の推定

HRLE の具体的な処理手順は，下記の 7 つの数式（図 6.47 の各処理に対応）で示すとおりである．式中で，t は時刻（フレーム単位），y_p は入力パワー (INPUT_SPEC)，n_p は推定ノイズパワー (NOISE_SPEC])，x，α，L_{min}，L_{step} はヒストグラムに関わるパラメータでそれぞれ正規化累積頻度 (LX)，時定数 (TIME_CONSTANT)，ビンの最小レベル (MIN_LEVEL)，ビンのレベル幅 (STEP_LEVEL)，$\lfloor a \rfloor$ は a 以下の a に最も近い整数を示している．また，パラメータを除く全ての変数は，周波数の関数であり，各周波数毎に独立して同じ処理が施される．式中では，簡略化のため周波数を省略した．

$$Y_L(t) = 10 \log_{10} y_p(t), \tag{6.48}$$

$$I_y(t) = \lfloor (Y_L(t) - L_{min})/L_{step} \rfloor, \tag{6.49}$$

$$N(t, l) = \alpha N(t-1, l) + (1-\alpha)\delta(l - I_y(t)), \tag{6.50}$$

$$S(t, l) = \sum_{k=0}^{l} N(t, k), \tag{6.51}$$

$$I_x(t) = \underset{I}{\operatorname{argmin}} \left[S(t, I_{max}) \frac{x}{100} - S(t, I) \right], \tag{6.52}$$

$$L_x(t) = L_{min} + L_{step} \cdot I_x(t), \tag{6.53}$$

$$n_p(t) = 10^{L_x(t)/10} \tag{6.54}$$

参考文献

(1) H. Nakajima, G. Ince, K. Nakadai and Y. Hasegawa: “An Easily-configurable Robot Audition System using Histogram-based Recursive Level Estimation”, Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), 2010 (to be appeared).

6.3.7 PostFilter

ノードの概要

このノードは，音源分離ノード GHDSS によって分離された複素スペクトルに対し，音声認識精度を向上するための後処理を行う．同時に，ミッシングフィーチャーマスクを生成するための，ノイズパワースペクトルの生成も行う．

必要なファイル

無し．

使用方法

どんなときに使うのか

このノードは，GHDSS ノードによって分離されたスペクトルの整形と，ミッシングフィーチャーマスクを生成するために必要なノイズスペクトルを生成する時に用いる．

典型的な接続例

PostFilter ノードの接続例は図 6.49 の通り．入力の接続として，INPUT_SPEC は GHDSS ノードの出力，INIT_NOISE_POWER は BGNEstimator ノードの出力と接続する．

出力について，図 6.49 では

1. 分離音 (OUTPUT_SPEC) の音声特徴抽出 (MSLSExtraction ノード),
2. 分離音と分離音に含まれるノイズのパワー (EST_NOISE_POWER) から音声認識時のミッシングフィーチャーマスク生成 (MFMGeneration ノード)

の接続例を示している．

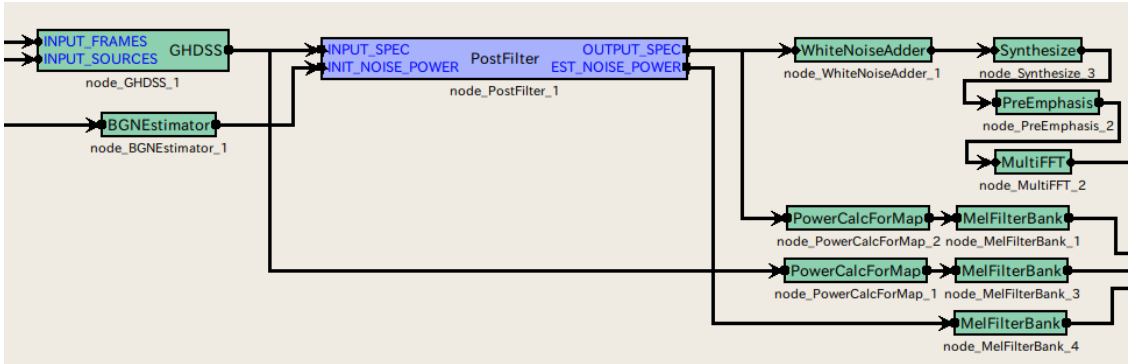


図 6.49: PostFilter の接続例

ノードの入出力とプロパティ

入力

INPUT_SPEC : Map<int, ObjectRef> 型．GHDSS ノードからの出力と同じ型．音源 ID と，分離音の複素スペクトルである Vector<complex<float> > 型データのペア．

INPUT_NOISE_POWER : Matrix<float> 型．BGNEstimator ノードによって推定された定常ノイズのパワースペクトル．

出力

OUTPUT_SPEC : Map<int, ObjectRef> 型．入力 INPUT_SPEC から，ノイズ除去がされた分離音の複素スペクトル．Object 部分は Vector<complex<float> > 型．

EST_NOISE_POWER : Map<int, ObjectRef> 型．OUTPUT_SPEC の各分離音に対して，含まれていると推定されたノイズのパワーが，Vector<float> 型データとして ID とペアになっている．

パラメータ

ノードの詳細

式で用いられる添字は，表 6.1 で定義されているものに準拠する．また，以降の式では，特に必要のない場合は，時間フレームインデックス f を省略して表記する．

図 6.50 は，PostFilter ノードの流れ図である．入力としては，GHDSS ノードからの分離音スペクトルと，BGNEstimator ノードの定常ノイズパワースペクトルが得られる．出力には，音声が強調された分離音スペクトルと，分離音に混入しているノイズのパワースペクトルである．

処理の流れは

1. ノイズ推定
2. SNR 推定
3. 音声存在確率推定
4. ノイズ除去

となっている．

1) ノイズ推定:

ノイズ推定処理の流れを図 6.51 に示す．PostFilter ノードが対処するノイズは，

a) マイクロホンの接点などが要因となる定常ノイズ ,

b) 除去しきれなかった別の音源の音 (漏れノイズ),

c) 前フレームの残響,

の 3 つである．

最終的な分離音に含まれるノイズ $\lambda(f, k_i)$ は,

$$\lambda(f,k_i) \quad = \quad \lambda^{sta}(f,k_i) + \lambda^{leak}(f,k_i) + \lambda^{rev}(f-1,k_i) \tag{6.55}$$

として求められる．ただし，$\lambda^{sta}(f,k_i), \lambda^{leak}(f,k_i), \lambda^{rev}(f-1,k_i)$ はそれぞれ，定常ノイズ，漏れノイズ，前フレームの残響を表す．

表 6.39: PostFilter のパラメータ表 (前半)

パラメータ名	型	デフォルト値	単位	説明
MCRA_SETTING	bool	false		ノイズ除去手法である，MCRA 推定に関するパラメータ設定項目を表示する時，true にする．
MCRA_SETTING				以下，MCRA_SETTING が true の時に表示される
STATIONARY_NOISE_FACTOR	float	1.2		定常ノイズ推定時の係数．
SPEC_SMOOTH_FACTOR	float	0.5		入力パワースペクトルの平滑化係数．
AMP_LEAK_FACTOR	float	1.5		漏れ係数．
STATIONARY_NOISE_MIXTURE_FACTOR	float	0.98		定常ノイズの混合比．
LEAK_FLOOR	float	0.1		漏れノイズの最小値．
BLOCK_LENGTH	int	80		検出時間幅．
VOICEP_THRESHOLD	int	3		音声存在判定の閾値．
EST_LEAK_SETTING	bool	false		漏れ率推定に関するパラメータ設定項目を表示する時，true にする．
EST_LEAK_SETTING				以下，EST_LEAK_SETTING が true の時に表示される．
LEAK_FACTOR	float	0.25		漏れ率．
OVER_CANCEL_FACTOR	float	1		漏れ率重み係数．
EST_REV_SETTING	bool	false		残響成分推定に関するパラメータ設定項目を表示する時，true にする．
EST_REV_SETTING				以下，EST_REV_SETTING が true の時に表示される．
REVERB_DECAY_FACTOR	float	0.5		残響パワーの減衰係数．
DIRECT_DECAY_FACTOR	float	0.2		分離スペクトルの減衰係数．
EST_SN_SETTING	bool	false		SN 比推定に関するパラメータ設定項目を表示する時，true にする．
EST_SN_SETTING				以下，EST_SN_SETTING が true の時に表示される．
PRIOR_SNR_FACTOR	float	0.8		事前 SNR と事後 SNR の比率．
VOICEP_PROB_FACTOR	float	0.9		音声存在確率の振幅係数．
MIN_VOICEP_PROB	float	0.05		最小音声存在確率．
MAX_PRIOR_SNR	float	100		事前 SNR の最大値．
MAX_OPT_GAIN	float	20		最適ゲイン中間変数 v の最大値．
MIN_OPT_GAIN	float	6		最適ゲイン中間変数 v の最小値．

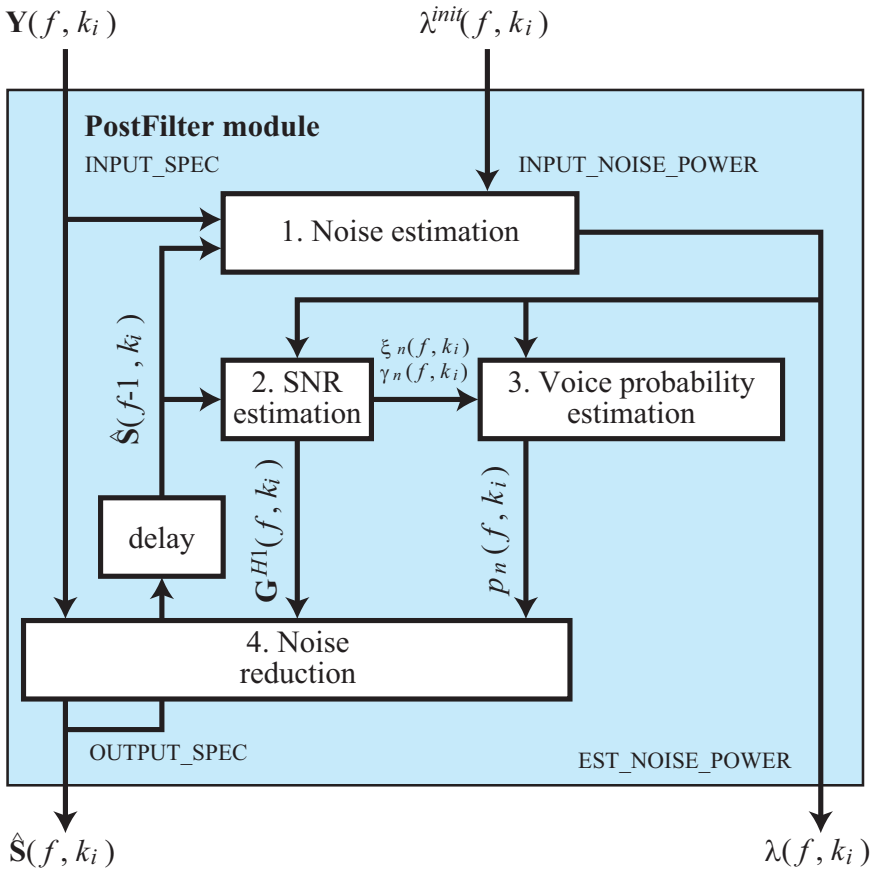


図 6.50: PostFilter の流れ図

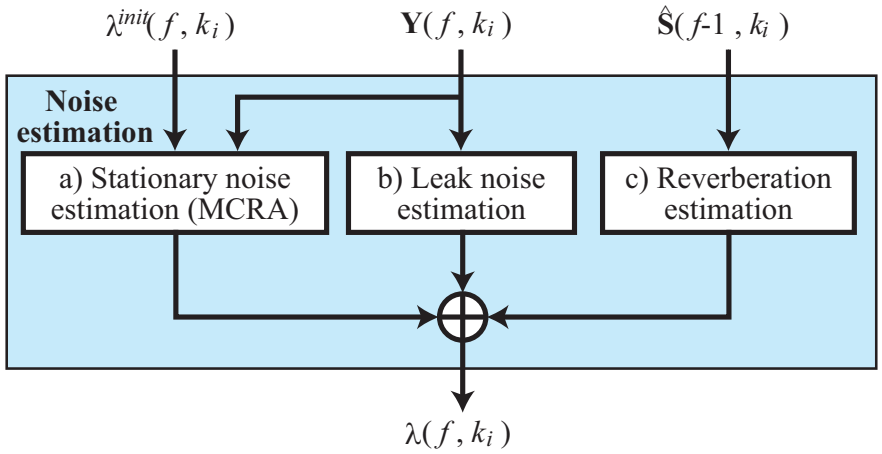


図 6.51: ノイズ推定の手順

1-a) MCRA 法による定常ノイズ推定　1-a) で用いる変数は 表 6.41 に基づく．

まず，入力スペクトルを 1 フレーム前のパワーと平滑化したパワースペクトル $S(f,k_i) = [S_1(f,k_i),\ldots,S_N(f,k_i)]$ を求める．

$$S_n(f,k_i) = \alpha_s S_n(f-1,k_i) + (1-\alpha_s)|Y_n(k_i)|^2 \tag{6.56}$$

次に，S^{tmp}，S^{min} を更新する．

$$S_n^{min}(f,k_i) = \begin{cases} \min\{S_n^{min}(f-1,k_i), S_n(f,k_i)\} & \text{if } f \neq nL \\ \min\{S_n^{tmp}(f-1,k_i), S_n(f,k_i)\} & \text{if } f = nL \end{cases}, \tag{6.57}$$

$$S_n^{min}(f,k_i) = \begin{cases} \min\{S_n^{tmp}(f-1,k_i), S_n(f,k_i)\} & \text{if } f \neq nL \\ S_n(f,k_i) & \text{if } f = nL \end{cases}, \tag{6.58}$$

ただし，n は任意の整数である．S^{min} はノイズ推定を始めてからの最小パワーを保持し，S^{tmp} は最近の L フレームの極小パワーを保持している．L フレームごとに S^{tmp} は更新される．

続いて，最小パワーと入力分離音のパワーの比から，音声が含まれるかどうかを判定する．

$$S_n^r(k_i) = \frac{S_n(k_i)}{S^{min}(k_i)}, \tag{6.59}$$

$$I_n(k_i) = \begin{cases} 1 & \text{if } S_n^r(k_i) > \delta \\ 0 & \text{if } S_n^r(k_i) \leq \delta \end{cases} \tag{6.60}$$

$I_n(k_i)$ に音声が含まれる場合 1，含まれない場合 0 となる．この判定結果をもとに，前フレーム定常ノイズと，現在のフレームのパワーとの混合比 $\alpha_{d,n}^C(k_i)$ を決める．

$$\alpha_{d,n}^C(k_i) = (\alpha_d - 1)I_n(k_i) + 1. \tag{6.61}$$

次に，分離音のパワースペクトルに含まれる漏れノイズを除去する．

$$S_n^{leak}(k_i) = \sum_{p=1}^{N} |Y_p(k_i)|^2 - |Y_n(k_i)|^2, \tag{6.62}$$

$$S_n^0(k_i) = |Y_n(k_i)|^2 - qS_n^{leak}(k_i), \tag{6.63}$$

ただし，$S_n^0(k_i) < S_{floor}$ のとき，

$$S_n^0(k_i) = S_{floor} \tag{6.64}$$

に値が変更される．

漏れノイズを除いたパワースペクトル $S_n^0(f,k_i)$ と，前フレームの推定定常ノイズ $\lambda^{sta}(f-1,k_i)$ または，BGNEstimator からの出力である $bf\lambda^{init}(f,k_i)$ を混合することで，現在のフレームの定常ノイズを求める．

$$\lambda_n^{sta}(f,k_i) = \begin{cases} \alpha_{d,n}^C(k_i)\lambda_n^{sta}(f-1,k_i) + (1-\alpha_{d,n}^C(k_i)rS_n^0(f,k_i) & \text{if音源位置に変更なし} \\ \alpha_{d,n}^C(k_i)\lambda_n^{init}(f,k_i) + (1-\alpha_{d,n}^C(k_i)rS_n^0(f,k_i) & \text{if音源位置に変更あり} \end{cases} \tag{6.65}$$

1-b) 漏れノイズ推定　1-b) で用いる変数は 表 6.42 に基づく．

いくつかのパラメータを次のように計算する．

$$\beta = -\frac{\alpha^{leak}}{1-(\alpha^{leak})^2 + \alpha^{leak}(1-\alpha^{leak})(N-2)} \tag{6.66}$$

$$\alpha = 1-(N-1)\alpha^{leak}\beta \tag{6.67}$$

このパラメータを用いて，平滑化されたスペクトル $S(k_i)$ と，式 (6.62) で求められた，他の分離音のパワーから自分の分離音のパワーを除いたパワースペクトル $S_n^{leak}(k_i)$ を混合する．

$$Z_n(k_i) \quad = \quad \alpha S_n(k_i) + \beta S_n^{leak}(k_i), \tag{6.68}$$

ただし，$Z_n(k_i) < 1$ になる場合は，$Z_n(k_i) = 1$ とする．

最終的な漏れノイズのパワースペクトル $\lambda^{leak}(k_i)$ は，

$$\lambda_n^{leak} \quad = \quad \alpha^{leak}\left(\sum_{n' \neq n} Z_{n'}(k_i)\right) \tag{6.69}$$

として求める．

1-c) 残響推定　1-c) で用いる変数は 表 6.43 に基づく．

残響のパワーは，前フレームの推定残響パワー $\lambda^{rev}(f-1,k_i) = [\lambda_1^{rev}(f-1,k_i),\ldots,\lambda_N^{rev}(f-1,k_i)]^T$ と，前フレームの分離スペクトル $\hat{S}(f-1,k_i) = [\hat{S}_1(f-1,k_i),\ldots,\hat{S}_N(f-1,k_i)]^T$ から次のように計算される．$\hat{S}_n(f-1,k_i)$ は複素数であることに注意．

$$\lambda_n^{rev}(f,k_i) \quad = \quad \gamma\left(\lambda_n^{rev}(f-1,k_i) + \Delta|\hat{S}_n(f-1,k_i)|^2\right) \tag{6.70}$$

2) SNR 推定:

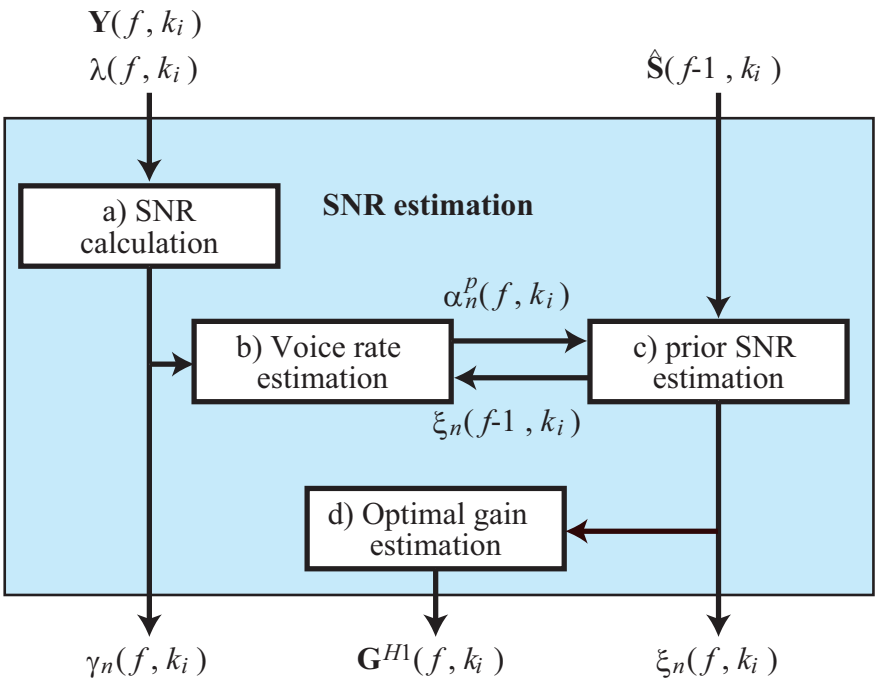


図 6.52: SNR 推定の手順

SNR 推定の流れを図 6.52 に示す．SNR 推定は，

a) SNR の計算

b) ノイズ混入前の事前 SNR 推定

c) 音声含有率の推定

d) 最適ゲインの推定
から成る．
表 6.44 のベクトルの要素は，各分離音の値に対応する．

2-a) SNR の計算　2-a) で用いる変数は，表 6.44 に従う．ここでは，入力の複素スペクトル $Y(k_i)$ と，前段で推定されたノイズのパワースペクトル $\lambda(k_i)$ を元に，SNR $\gamma_n(k_i)$ が計算される．

$$\gamma_n(k_i) = \frac{|Y_n(k_i)|^2}{\lambda_n(k_i)} \tag{6.71}$$

$$\gamma_n^C(k_i) = \begin{cases} \gamma_n(k_i) & \text{if } \gamma_n(k_i) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.72}$$

2-b) 音声含有率の推定　2-b) で用いる変数は，表 6.45 に従う．
音声含有率 $\alpha_n^p(f, k_i)$ は，前フレームの事前 SNR $\xi_n(f-1, k_i)$ を用いて次のように計算される．

$$\alpha_n^p(f, k_i) = \alpha_{mag}^p \left(\frac{\xi_n(f-1, k_i)}{\xi_n(f-1, k_i) + 1} \right)^2 + \alpha_{min}^p \tag{6.73}$$

2-c) ノイズ混入前の事前 SNR 推定　2-c) で用いる変数は，表 6.46 に従う．
事前 SNR $\xi_n(k_i)$ は，次のようにして計算する．

$$\xi_n(k_i) = \left(1 - \alpha_n^p(k_i)\right)\xi_{tmp} + \alpha_n^p(k_i)\gamma_n^C(k_i) \tag{6.74}$$

$$\xi_{tmp} = a\frac{|\hat{S}_n(f-1, k_i)|^2}{\lambda_n(f-1, k_i)} + (1-a)\xi_n(f-1, k_i) \tag{6.75}$$

ただし，ξ_{tmp} は計算上の一時的な変数で，前フレームの推定 SNR$\gamma_n(k_i)$ と，事前 SNR $\xi_n(k_i)$ の内分値である．また，$\xi_n(k_i) > \xi^{max}$ となる場合，$\xi_n(k_i) = \xi^{max}$ と値を変更する．

2-d) 最適ゲインの推定　2-d) で用いる変数は，表 6.47 に従う．
最適ゲイン計算の前に，上で求めた事前 SNR$\xi_n(k_i)$ と，推定 SNR$\gamma_n(k_i)$ を用いて，以下の中間変数 $v_n(k_i)$ を計算する．

$$v_n(k_i) = \frac{\xi_n(k_i)}{1 + \xi_n(k_i)}\gamma_n(k_i) \tag{6.76}$$

$v_n(k_i) > \theta^{max}$ の場合，$v_n(k_i) = \theta^{max}$ とする．
音声がある場合の最適ゲイン $G^{H1}(k_i) = [G_1^{H1}(k_i), \ldots, G_N^{H1}(k_i)]$ は，

$$G_n^{H1}(k_i) = \frac{\xi_n(k_i)}{1 + \xi_n(k_i)} \exp\left\{ \frac{1}{2} \mathtt{int}_{v_n(k_i)}^{\inf} \frac{e^{-t}}{t} \mathrm{d}t \right\} \tag{6.77}$$

として求める．ただし，

$$\begin{aligned} G_n^{H1}(k_i) &= 1 \quad \text{if} v_n(k_i) < \theta^{min} \\ G_n^{H1}(k_i) &= 1 \quad \text{if} G_n^{H1}(k_i) > 1. \end{aligned} \tag{6.78}$$

3) 音声存在確率推定:
音声存在確率推定の流れを図 6.53 に示す．音声存在確率推定は，
a) 3 種類の帯域ごとに事前 SNR の平滑化

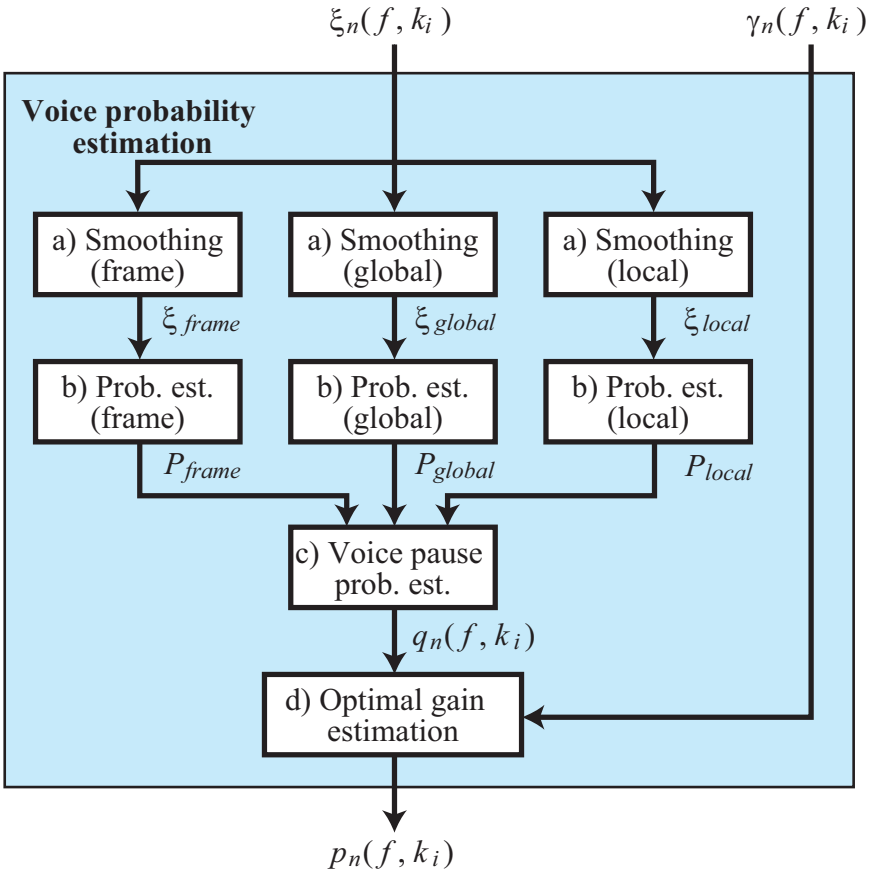


図 6.53: 音声存在確率推定の手順

b) 各帯域で，平滑化した SNR を元に，暫定的な音声確率を推定
c) 3 つの暫定確率をもとに音声休止確率を推定
d) 最終的な音声存在確率を推定
から成る．

3-a) 事前 SNR の平滑化　3-a) で用いる変数を表 6.48 にまとめる．

まず，式 (6.74) で計算された事前 SNR $\xi_n(f, k_i)$ と，前フレームの時間平滑化事前 SNR $\zeta_n(f-1, k_i)$ で，時間平滑化を行う．

$$\zeta_n(f, k_i) \quad = \quad b\zeta_n(f-1, k_i) + (1-b)\xi_n(f, k_i) \tag{6.79}$$

周波数方向の平滑化は，その窓の大きさによって，frame，global，local の順に小さくなっていく．

- frame での周波数平滑化
  周波数ビン $F_{st} \sim F_{en}$ の範囲で加算平均による平滑化が行われる．

$$\zeta_n^f(k_i) \quad = \quad \frac{1}{F_{en} - F_{st} + 1} \sum_{k_j = F_{st}}^{F_{en}} \zeta_n(k_j) \tag{6.80}$$

- global での周波数平滑化

global では，幅 G での hanning 窓を用いた周波数平滑化が行われる．

$$\zeta_n^g(k_i) = \sum_{j=-(G-1)/2}^{(G-1)/2} w_{han}(j+(G-1)/2)\zeta_n(k_{i+j}), \tag{6.81}$$

$$w_{han}(j) = \frac{1}{C}\left(0.5 - 0.5\cos\left(\frac{2\pi j}{G}\right)\right), \tag{6.82}$$

ただし，C は $\sum_{j=0}^{G-1} w_{han}(j) = 1$ にするための正規化係数.

- local での周波数平滑化
  local では，幅 F での三角窓を用いた周波数平滑化が行われる．

$$\zeta_n^l(k_i) = 0.25\zeta_n(k_i-1) + 0.5\zeta_n(k_i) + 0.25(k_i+1) \tag{6.83}$$

3-b) 暫定音声確率を推定　3-b) で用いる変数を表 6.49 に示す.

- $P_n^f(k_i)$ と $\zeta_n^{peak}(k_i)$ の計算
  まず，$\zeta_n^{peak}(f,k_i)$ を以下のように求める．

$$\zeta_n^{peak}(f,k_i) = \begin{cases} \zeta_n^f(f,k_i), & \text{if } \zeta_n^f(f,k_i) > Z_{thres}\zeta_n^f(f-1,k_i) \\ \zeta_n^{peak}(f-1,k_i), & \text{if otherwise.} \end{cases} \tag{6.84}$$

ただし，$\zeta_n^{peak}(k_i)$ の値はパラメータ $Z_{min}^{peak}, Z_{max}^{peak}$ の範囲に入るようにする．すなわち，

$$\zeta_n^{peak}(k_i) = \begin{cases} Z_{min}^{peak}, & \text{if } \zeta_n^{peak}(k_i) < Z_{min}^{peak} \\ Z_{max}^{peak}, & \text{if } \zeta_n^{peak}(k_i) > Z_{max}^{peak} \end{cases} \tag{6.85}$$

次に，$P_n^f(k_i)$ を次のように求める．

$$P_n^f(k_i) = \begin{cases} 0, & \text{if } \zeta_n^f(k_i) < \zeta_n^{peak}(k_i)Z_{min}^f \\ 1, & \text{if } \zeta_n^f(k_i) > \zeta_n^{peak}(k_i)Z_{max}^f \\ \frac{\log\left(\zeta_n^f(k_i)/\zeta_n^{peak}(k_i)Z_{min}^f\right)}{\log\left(Z_{max}^f/Z_{min}^f\right)}, & \text{otherwise} \end{cases} \tag{6.86}$$

- $P_n^g(k_i)$ の計算
  次の通りに計算する．

$$P_n^g(k_i) = \begin{cases} 0, & \text{if } \zeta_n^g(k_i) < Z_{min}^g \\ 1, & \text{if } \zeta_n^g(k_i) > Z_{max}^g \\ \frac{\log\left(\zeta_n^g(k_i)/Z_{min}^g\right)}{\log\left(Z_{max}^g/Z_{min}^g\right)}, & \text{otherwise} \end{cases} \tag{6.87}$$

- $P_n^l(k_i)$ の計算
  次の通りに計算する．

$$P_n^l(k_i) = \begin{cases} 0, & \text{if } \zeta_n^l(k_i) < Z_{min}^l \\ 1, & \text{if } \zeta_n^l(k_i) > Z_{max}^l \\ \frac{\log\left(\zeta_n^l(k_i)/Z_{min}^l\right)}{\log\left(Z_{max}^l/Z_{min}^l\right)}, & \text{otherwise} \end{cases} \tag{6.88}$$

3-c) 音声休止確率推定　3-c) で用いる変数を表 6.50 に示す．

音声休止確率 $q_n(k_i)$ は，3 つの周波数帯域の平滑化結果を元にして計算した暫定の音声確率 $P_n^{f,g,l}(k_i)$ を次のように統合して得られる．

$$q_n(k_i) \quad = \quad 1 - \left(1 - a^l + a^l P_n^l(k_i)\right)\left(1 - a^g + a^g P_n^g(k_i)\right)\left(1 - a^f + a^f P_n^f(k_i)\right), \tag{6.89}$$

ただし，$q_n(k_i) < q_{min}$ のとき，$q_n(k_i) = q_{min}$ とし，$q_n(k_i) > q_{max}$ のとき，$q_n(k_i) = q_{max}$ とする．

3-d) 音声存在確率推定　音声存在確率 $p_n(k_i)$ は，音声休止確率 $q_n(k_i)$，事前 SNR $\zeta_n(k_i)$，式 (6.76) により導出された中間変数 $v_n(k_i)$ を用いて次のように導出する．

$$p_n(k_i) \quad = \quad \left\{1 + \frac{q_n(k_i)}{1 - q_n(k_i)}\,(1 + \zeta_n(k_i))\exp\left(-v_n(k_i)\right)\right\}^{-1} \tag{6.90}$$

4) ノイズ除去: 出力である音声強調された分離音 $\hat{S}_n(k_i)$ は，入力である分離音スペクトル $Y_n(k_i)$ に対して，最適ゲイン $G_n^{H1}(k_i)$，音声存在確率 $p_n(k_i)$ を次のように作用させることで導出する．

$$\hat{S}_n(k_i) \quad = \quad Y_n(k_i) G_n^{H1}(k_i) p_n(k_i) \tag{6.91}$$

表 6.40: PostFilter のパラメータ表 (後半)

パラメータ名	型	デフォルト値	単位	説明
EST_VOICEP_SETTING	bool	false		音声確率推定に関するパラメータを設定する時，true にする．
EST_VOICEP_SETTING				以下，EST_VOICEP_SETTING が true の時に有効．
PRIOR_SNR_SMOOTH_FACTOR	float	0.7		時間平滑化係数．
MIN_FRAME_SMOOTH_SNR	float	0.1		周波数平滑化 SNR の最小値 (frame)．
MAX_FRAME_SMOOTH_SNR	float	0.316		周波数平滑化 SNR の最大値 (frame)．
MIN_GLOBAL_SMOOTH_SNR	float	0.1		周波数平滑化 SNR の最小値 (global)．
MAX_GLOBAL_SMOOTH_SNR	float	0.316		周波数平滑化 SNR の最大値 (global)．
MIN_LOCAL_SMOOTH_SNR	float	0.1		周波数平滑化 SNR の最小値 (local)．
MAX_LOCAL_SMOOTH_SNR	float	0.316		周波数平滑化 SNR の最大値 (local)．
UPPER_SMOOTH_FREQ_INDEX	int	99		周波数平滑化上限ビンインデックス．
LOWER_SMOOTH_FREQ_INDEX	int	8		周波数平滑化下限ビンインデックス．
GLOBAL_SMOOTH_BANDWIDTH	int	29		周波数平滑化バンド幅 (global)．
LOCAL_SMOOTH_BANDWIDTH	int	5		周波数平滑化バンド幅 (local)．
FRAME_SMOOTH_SNR_THRESH	float	1.5		周波数平滑化 SNR の閾値．
MIN_SMOOTH_PEAK_SNR	float	1.0		周波数平滑化 SNR ピークの最小値．
MAX_SMOOTH_PEAK_SNR	float	10.0		周波数平滑化 SNR ピークの最大値．
FRAME_VOICEP_PROB_FACTOR	float	0.7		音声確率平滑化係数 (frame)．
GLOBAL_VOICEP_PROB_FACTOR	float	0.9		音声確率平滑化係数 (global)．
LOCAL_VOICEP_PROB_FACTOR	float	0.9		音声確率平滑化係数 (local)．
MIN_VOICE_PAUSE_PROB	float	0.02		音声休止確率の最小値．
MAX_VOICE_PAUSE_PROB	float	0.98		音声休止確率の最大値．

表 6.41: 変数の定義

変数	説明，対応するパラメータ
$Y(k_i) = [Y_1(k_i), \ldots, Y_N(k_i)]^T$	周波数ビン k_i に対応する分離音複素スペクトル
$\lambda^{init}(k_i) = \left[\lambda_1^{init}(k_i), \ldots, \lambda_N^{init}(k_i)\right]^T$	定常ノイズ推定に用いる初期値パワースペクトル
$\lambda^{sta}(k_i) = \left[\lambda_1^{sta}(k_i), \ldots, \lambda_N^{sta}(k_i)\right]^T$	推定された定常ノイズパワースペクトル.
α_s	入力パワースペクトルの平滑化係数．パラメータ SPEC_SMOOTH_FACTOR，デフォルト 0.5
$S^{tmp}(k_i) = \left[S_1^{tmp}(k_i), \ldots, S_N^{tmp}(k_i)\right]$	最小パワー計算用のテンポラリ変数.
$S^{min}(k_i) = \left[S_1^{min}(k_i), \ldots, S_N^{min}(k_i)\right]$	最小パワーを保持する変数.
L	S_{tmp} の保持フレーム数．パラメータ BLOCK_LENGTH，デフォルト 80
δ	音声存在判定の閾値．パラメータ VOICEP_THRESHOLD，デフォルト 3.0
α_d	推定定常ノイズの混合比．パラメータ STATIONARY_NOISE_MIXTURE_FACTOR，デフォルト 0.98
$Y^{leak}(k_i)$	分離音に含まれると推定される漏れノイズのパワースペクトル
q	入力分離音パワーから漏れノイズを除くときの係数．パラメータ AMP_LEAK_FACTOR, デフォルト 1.5
S_{floor}	漏れノイズ最小値．パラメータ LEAK_FLOOR, デフォルト 0.1
r	定常ノイズ推定時の係数．パラメータ STATIONARY_NOISE_FACTOR, デフォルト 1.2

表 6.42: 変数の定義

変数	説明，対応するパラメータ
$\lambda^{leak}(k_i)$	漏れノイズのパワースペクトル，各分離音の要素から成るベクトル.
α^{leak}	全分離音パワーの合計に対する漏れ率． LEAK_FACTOR × OVER_CANCEL_FACTOR
$S_n(f, k_i)$	式 (6.56) で求める平滑化パワースペクトル

表 6.43: 変数の定義

変数	説明，対応するパラメータ
$\lambda^{rev}(f, k_i)$	時間フレーム f での残響のパワースペクトル
$\hat{S}(f-1, k_i)$	前フレームの PostFilter の出力したノイズ除去後分離音スペクトル
γ	前フレーム残響パワーの減衰係数．パラメータ REVERB_DECAY_FACTOR，デフォルト 0.5
Δ	前フレーム分離音の減衰係数．パラメータ DIRECT_DECAY_FACTOR，デフォルト 0.2

表 6.44: 主な変数の定義

変数	説明，対応するパラメータ
$Y(k_i)$	PostFilter ノードの入力である分離音の複素スペクトル
$\hat{S}(k_i)$	PostFilter ノードの出力となる，整形された分離音複素スペクトル
$\lambda(k_i)$	前段で推定されたノイズのパワースペクトル
$\gamma_n(k_i)$	分離音 n の SNR
$\alpha_n^p(k_i)$	音声含有率
$\xi_n(k_i)$	事前 SNR
$G^{H1}(k_i)$	分離音の SNR を向上させるための最適ゲイン

表 6.45: 変数の定義

変数	説明，対応するパラメータ
α_{mag}^p	事前 SNR 係数．パラメータ VOICEP_PROB_FACTOR，デフォルト 0.9
α_{min}^p	最小音声含有率．パラメータ MIN_VOICEP_PROB，デフォルト 0.05

表 6.46: 変数の定義

変数	説明，対応するパラメータ
a	前フレーム SNR の内分比．パラメータ PRIOR_SNR_FACTOR，デフォルト 0.8
ξ^{max}	事前 SNR の上限．パラメータ MAX_PRIOR_SNR，デフォルト 100

表 6.47: 変数の定義

変数	説明，対応するパラメータ
θ^{max}	中間変数 $v_n(k_i)$ 最大値．パラメータ MAX_OPT_GAIN，デフォルト 20
θ^{min}	中間変数 $v_n(k_i)$ 最小値．パラメータ MIN_OPT_GAIN，デフォルト 6

表 6.48: 変数の定義

変数	説明，対応するパラメータ
$\zeta_n(k_i)$	時間平滑化した事前 SNR
$\xi_n(k_i)$	事前 SNR
$\zeta_n^f(k_i)$	周波数平滑化 SNR (frame)
$\zeta_n^g(k_i)$	周波数平滑化 SNR (global)
$\zeta_n^l(k_i)$	周波数平滑化 SNR (local)
b	パラメータ PRIOR_SNR_SMOOTH_FACTOR，デフォルト 0.7
F_{st}	パラメータ LOWER_SMOOTH_FREQ_INDEX，デフォルト 8
F_{en}	パラメータ UPPER_SMOOTH_FREQ_INDEX，デフォルト 99
G	パラメータ GLOBAL_SMOOTH_BANDWIDTH，デフォルト 29
L	パラメータ LOCAL_SMOOTH_BANDWIDTH，デフォルト 5

表 6.49: 変数の定義

変数	説明，対応するパラメータ
$\zeta_n^{f,g,l}(k_i)$	各帯域で平滑化された SNR
$P_n^{f,g,l}(k_i)$	各帯域での暫定音声確率
$\zeta_n^{peak}(k_i)$	平滑化 SNR のピーク
Z_{min}^{peak}	パラメータ MIN_SMOOTH_PEAK_SNR，デフォルト値 1
Z_{max}^{peak}	パラメータ MAX_SMOOTH_PEAK_SNR，デフォルト値 10
Z_{thres}	FRAME_SMOOTH_SNR_THRESH，デフォルト値 1.5
$Z_{min}^{f,g,l}$	パラメータ MIN_FRAME_SMOOTH_SNR, MIN_GLOBAL_SMOOTH_SNR, MIN_LOCAL_SMOOTH_SNR，デフォルト値 0.1
$Z_{max}^{f,g,l}$	パラメータ MAX_FRAME_SMOOTH_SNRF, MAX_GLOBAL_SMOOTH_SNR, MAX_LOCAL_SMOOTH_SNR，デフォルト値 0.316

表 6.50: 変数の定義

変数	説明，対応するパラメータ
$q_n(k_i)$	音声休止確率
a^f	FRAME_VOICEP_PROB_FACTOR，デフォルト，0.7
a^g	GLOBAL_VOICEP_PROB_FACTOR，デフォルト，0.9
a^l	LOCAL_VOICEP_PROB_FACTOR，デフォルト，0.9
q_{min}	MIN_VOICE_PAUSE_PROB，デフォルト，0.02
q_{max}	MAX_VOICE_PAUSE_PROB，デフォルト，0.98

6.3.8 SpectralGainFilter

ノードの概要

本ノードは，入力された分離音スペクトルに最適ゲイン，音声存在確率（PostFilter を参照）を乗じ，出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

HRLE，CalcSpecSubGain を用いて分離音スペクトルからノイズを抑制した音声スペクトを得る際に用いる．

典型的な接続例

SpectralGainFilter の接続例は図 6.54 の通り．入力は，GHDSS から出力された分離スペクトルおよび CalcSpecSubGain などから出力される，最適ゲイン，音声存在確率である．図では出力の例として，Synthesize ,SaveRawPCM に接続し，音声ファイルを作成している．

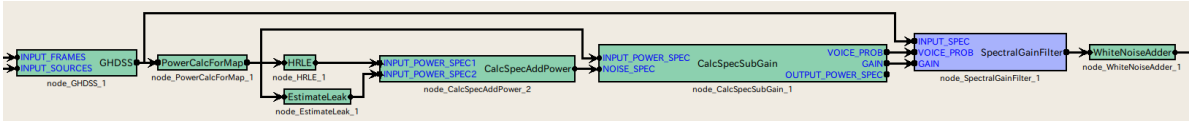


図 6.54: SpectralGainFilter の接続例

ノードの入出力とプロパティ

入力

INPUT_SPEC ：`Map<int, ObjectRef>` 型．GHDSS の出力と同じ型．音源 ID と分離音の複素スペクトルである，`Vector<complex<float> >` 型データのペア．

VOICE_PROB ：`Map<int, ObjectRef>` 型．音源 ID と音声存在確率の `Vector<float>` 型データのペア．

GAIN ：`Map<int, ObjectRef>` 型．音源 ID と最適ゲインの `Vector<float>` 型データのペア．

出力

OUTPUT_SPEC ：`Map<int, ObjectRef>` 型．GHDSS の出力と同じ型．音源 ID と分離音の複素スペクトルである，`Vector<complex<float> >` 型データのペア．

パラメータ

なし

ノードの詳細

本ノードは，入力された音声スペクトルに最適ゲイン，音声存在確率を乗じ，音声を強調した分離音スペクトルを出力する．

出力である音声が強調された分離音 $Y_n(k_i)$ は，入力である分離音スペクトルを $X_n(k_i)$ ，最適ゲインを $G_n(k_i)$, 音声存在確率を $p_n(k_i)$ とすると次のように表される．

$$Y_n(k_i) \quad = \quad X_n(k_i)G_n(k_i)p_n(k_i) \tag{6.92}$$

6.4 FeatureExtraction カテゴリ

6.4.1 Delta

ノードの概要

本ノードは，静的特徴ベクトルから動的特徴量ベクトルを求める．特徴抽出ノードである MSLSExtraction や MFCCExtraction の後段に接続するのが一般的な使い方である．これらの特徴量抽出ノードは，静的特徴ベクトルを求めると共に，動的特徴量を保存する領域を確保している．この時の動的特徴量は，0 に設定されている．Delta ノードでは，静的特徴ベクトル値を使って動的特徴量ベクトル値を計算し，値を設定する．従って，入出力でベクトルの次元数は変らない．

必要なファイル

無し．

使用方法

どんなときに使うのか

静的特徴から動的特徴量を求める場合に本ノードを使う．通常，MFCCExtraction や MSLSExtraction の後に用いる．

典型的な接続例

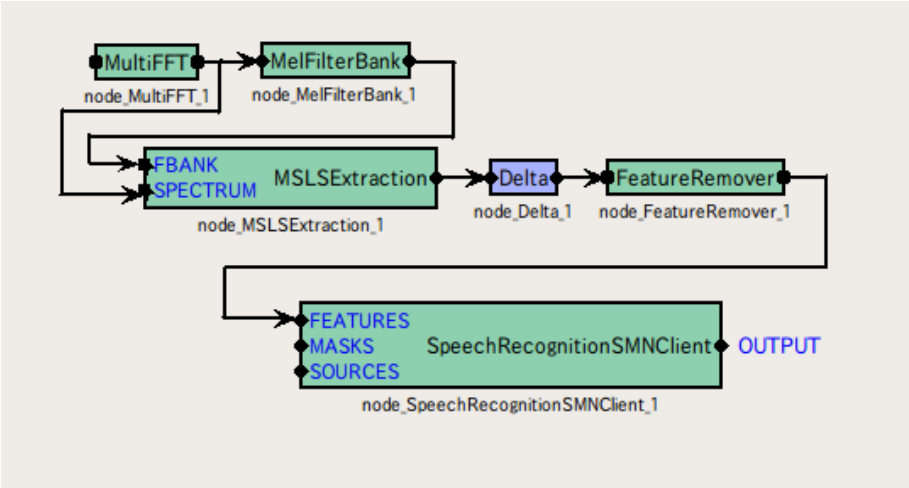


図 6.55: Delta の典型的な接続例

表 6.51: Delta パラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int	13		静的特徴の次元数

ノードの入出力とプロパティ

入力

INPUT : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

出力

OUTPUT : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

パラメータ

FBANK_COUNT : int 型である．処理する特徴量の次元数．値域は，正の整数．特徴量抽出ノードの直後に接続する場合は，特徴量抽出で指定した FBANK_COUNT を指定．ただし，特徴量抽出でパワー項を使用するオプションを true にしている場合は， FBANK_COUNT + 1 を指定する．

ノードの詳細

本ノードは，静的特徴ベクトルから動的特徴ベクトルを求める．入力の次元数は，静的特徴と動的特徴の次元数を合せた次元数である．FBANK_COUNT 以下の次元要素を静的特徴とみなし，動的特徴量を計算する．FBANK_COUNT より高次の次元要素に動的特徴量を入れる．

フレーム時刻 f における，入力特徴ベクトルを，

$$x(f) = [x(f,0), x(f,1), \ldots, x(f,P-1)]^T \tag{6.93}$$

と表す．ただし，P は，FBANK_COUNT である．

$$y(f) = [x(f,0), x(f,1), \ldots, x(f,2P-1)]^T \tag{6.94}$$

出力ベクトルの各要素は，

$$y(f,p) = \begin{cases} x(f,p), & \text{if } p = 0, \ldots, P-1, \\ w \sum_{\tau=-2}^{2} \tau \cdot x(f+\tau, p), & \text{if } p = P, \ldots, 2P-1, \end{cases} \tag{6.95}$$

である．ただし，$w = \frac{1}{\sum_{\tau=-2}^{\tau=2} \tau^2}$ である．図 6.56 に Delta の入出力フローを示す．

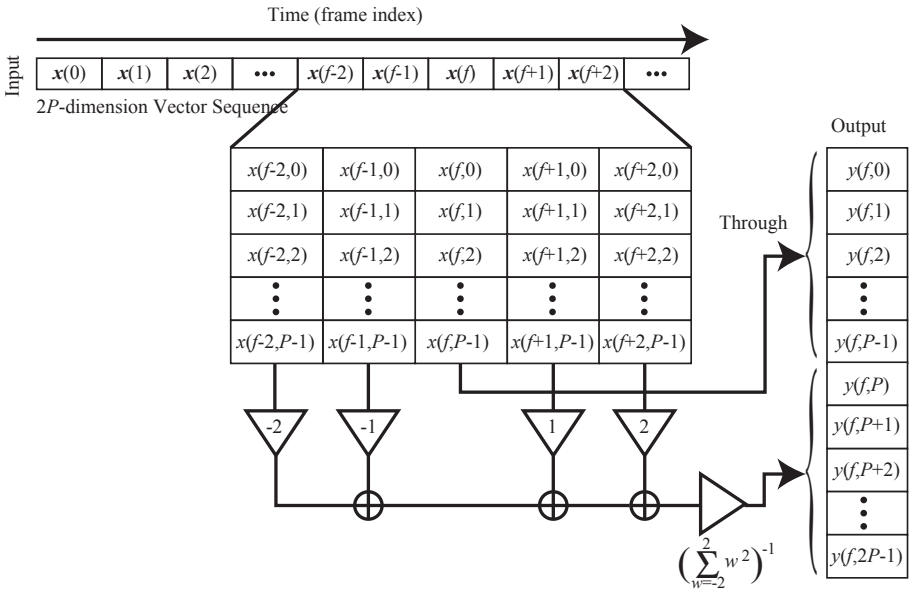


図 6.56: Delta の入出力フロー.

6.4.2 FeatureRemover

ノードの概要

本ノードは，入力ベクトル中から指定した次元要素を削除し，ベクトルの次元を減らしたベクトルを出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

音響特徴量やミッシングフィーチャーマスクベクトルなどのベクトル型の要素の中から不要な要素を削除し，次元数を減らすときに使う．

通常，特徴量の抽出処理は，静的特徴に続いて，動的特徴量を抽出する．その際に，静的特徴が不要となる場合がある．不要な特徴量を削除するには，本ノードを使用する．特に対数パワー項を削除することが多い．

典型的な接続例

MSLSExtraction や MFCCExtraction で対数パワー項を計算し，その後，Delta を用いると，デルタ対数パワー項を計算できる．対数パワー項を計算しなければデルタ対数パワーを計算できないため，一度対数パワー項を含めて音響特徴量を計算してから，対数パワー項を除去する．本ノードは，通常 Delta の後段に接続し，対数パワー項を削除するために用いる．

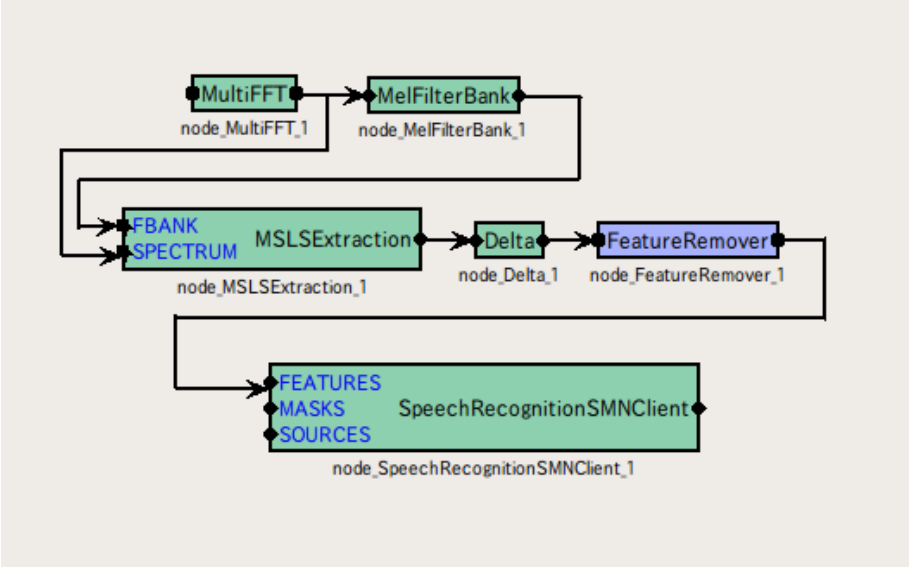


図 6.57: FeatureRemover の典型的な接続例

ノードの入出力とプロパティ

表 6.52: FeatureRemover パラメータ表

パラメータ名	型	デフォルト値	単位	説明
SELECTOR	Object	<Vector<int> >		次元インデックスからなるベクトル（複数指定可）

入力

INPUT : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

出力

OUTPUT : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

パラメータ

SELECTOR : Vector<int> 型．値域は，0 以上入力特徴量の次数未満．いくつ指定してもよい．1 次元目と 3 次元目の要素を削除し，入力ベクトルを 2 次元減らす場合は， <Vector<int> 0 2> とする．次元指定のインデクスが 0 から始まっていることに注意．

ノードの詳細

本ノードは，入力ベクトル中から不要な次元要素を削除し，ベクトルの次元を減らす．

音声信号を分析すると，分析フレームの対数パワーは，発話区間で大きい傾向がある．特に有声部分で大きい．従って，音声認識において対数パワー項を音響特徴量に取り入れることで認識精度の向上が見込める．しかしながら，対数パワー項を直接特徴量として用いると，収音レベルの違いが音響特徴に直接反映される．音響モデルの作成に用いた対数パワーレベルと収音レベルに差が生じると音声認識精度が低下する．機器の設定を固定しても，発話者が常に同一レベルで発話するとは限らない．そこで，対数パワー項の動的特徴量であるデルタ対数パワーを用いる．これにより，収音レベルの違いに頑健で，かつ発話区間や有声部分を表す特徴を捉えることが可能となる．

6.4.3 MelFilterBank

ノードの概要

入力スペクトルにメルフィルタバンク処理を行ない，各フィルタチャネルのエネルギーを出力する．入力スペクトルは，2 種類あり，入力によって出力結果が異なる点に留意．

必要なファイル

無し．

使用方法

どんなときに使うのか

音響特徴量を求める前処理として使用する．MultiFFT，PowerCalcForMap，PreEmphasis の直後に使用する．MFCCExtraction，MSLSExtraction の前段で使用する．

典型的な接続例

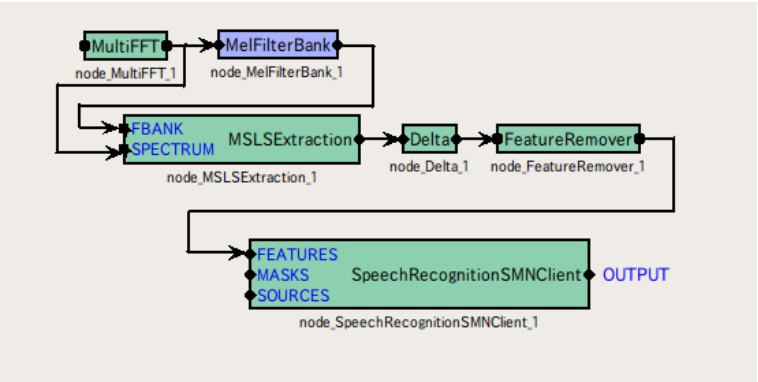


図 6.58: MelFilterBank の接続例

ノードの入出力とプロパティ

入力

INPUT : `Map<int, ObjectRef>` 型．音源 ID とパワースペクトル `Vector<float>` 型または，複素スペクトル `Vector<complex<float> >` 型のデータのペア．ただし，パワースペクトルを選択した場合，複素スペクトルを選択した場合と比較して出力エネルギーが 2 倍になる．

出力

表 6.53: MelFilterBank のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	分析フレーム長
SAMPLING_RATE	int	16000	[Hz]	サンプリング周波数
CUT_OFF	int	8000	[Hz]	ローパスフィルタのカットオフ周波数
MIN_FREQUENCY	int	63	[Hz]	フィルタバンクの下限周波数
MAX_FREQUENCY	int	8000	[Hz]	フィルタバンクの上限周波数
FBANK_COUNT	int	13		フィルタバンク数

OUTPUT : Map<int, ObjectRef> 型．音源 ID とフィルタバンクの出力エネルギーから構成されるベクトルの Vector<float> 型のデータのペア．出力ベクトルの次元数は，FBANK_COUNT の 2 倍である．0 から FBANK_COUNT-1 までに，フィルタバンクの出力エネルギーが入り，FBANK_COUNT から 2 *FBANK_COUNT-1 までには，0 が入る．0 が入れられる部分は，動的特徴量用のプレースホルダーである．動的特徴量が不要な場合は FeatureRemover を用いて削除する必要がある．

パラメータ

LENGTH : int 型．分析フレーム長である．入力スペクトルの周波数ビン数に等しい．値域は正の整数である．

SAMPLING_RATE : int 型．サンプリング周波数である．値域は正の整数である．

CUT_OFF : int 型．離散フーリエ変換時のアンチエーリアシングフィルタのカットオフ周波数．SAMPLING_RATE の 1/2 以下である．

MIN_FREQUENCY : int 型．フィルタバンクの下限周波数．値域は正の整数でかつ CUT_OFF 以下．

MAX_FREQUENCY : int 型．フィルタバンクの上限周波数．値域は正の整数でかつ CUT_OFF 以下．

FBANK_COUNT : int 型．フィルタバンク数である．値域は正の整数である．

ノードの詳細

メルフィルタバンク処理を行ない，各チャネルのエネルギーを出力する．各バンクの中心周波数は，メルスケール [1] 上で等間隔に配置する．チャネル毎の中心周波数は，最小周波数ビン SAMPLING_RATE/LENGTH から SAMPLING_RATECUT_OFF/LENGTH までを FBANK_COUNT 分割し決定する．

リニアスケールとメルスケールの変換式は，

$$m = 1127.01048 \log(1.0 + \frac{\lambda}{700.0}) \tag{6.96}$$

である．ただし，リニアスケール上での表現を λ (Hz)，メルスケール上での表現を m とする．図 6.59 に 8000 Hz までの変換例を示す．赤点は，SAMPLING_RATE が 16000 Hz，CUT_OFF が 8000 Hz，かつ FBANK_COUNT が 13 の場合の，各バンクの中心周波数を表す．各バンクの中心周波数が，メルスケール上で等間隔なことを確認できる．

図 6.60 にメルスケール上の各フィルタバンクの窓関数を示す．中心周波数部分で 1.0 となり，隣接チャネルの中心周波数部分で 0.0 となる三角窓である．中心周波数がチャネル毎にメルスケール上で等間隔で，対象な形状である．これらの窓関数は，リニアスケール上では図 6.61 のように表現される．高域のチャネルでは，広い帯域をカバーしている．

入力するリニアスケール上で表現されたパワースペクトルに図 6.61 に示す窓関数で重み付けし，各チャネル毎にエネルギーを求め，出力する．

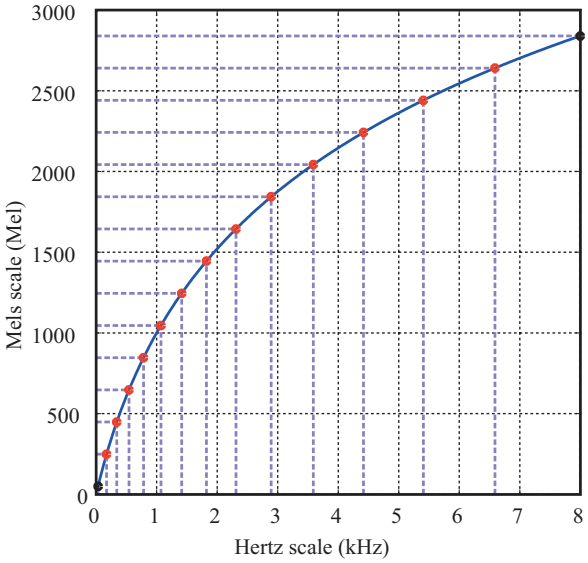


図 6.59: リニアスケールとメルスケールの対応

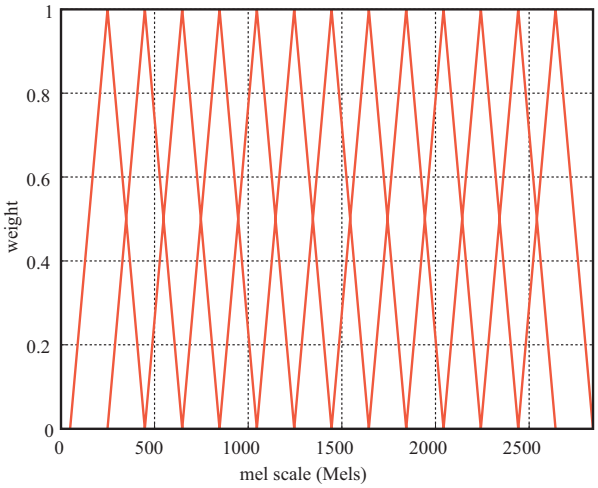


図 6.60: メルスケール上での窓関数

参考文献:

(1) Stanley Smith Stevens, John Volkman, Edwin Newman: “A Scale for the Measurement of the Psychological Magnitude Pitch”, Journal of the Acoustical Society of America 8(3), pp.185–190, 1937.

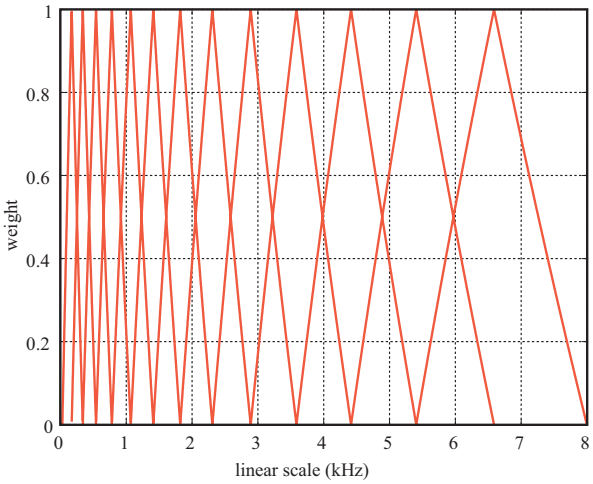


図 6.61: リニアスケール上での窓関数

6.4.4 MFCCExtraction

ノードの概要

本ノードは，音響特徴量の１つであるメルケプストラム係数 (MFCC : Mel-Frequency Cepstrum Coefficients) を求める．メルケプストラム係数と対数スペクトルパワーを要素とする音響特徴量ベクトルを生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

メルケプストラム係数を要素とする音響特徴量を生成するために用いる．音響特徴量ベクトルを生成するために用いる．例えば，音響特徴量ベクトルを音声認識ノードに入力し，音韻や話者を識別する．

典型的な接続例

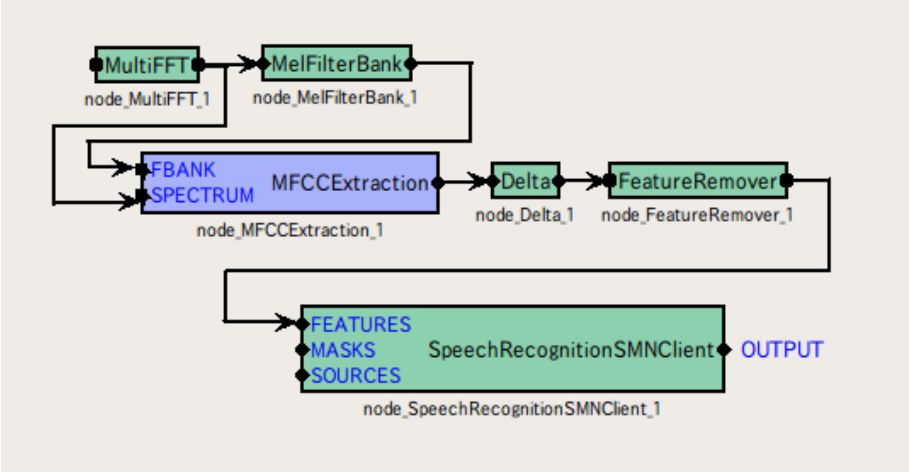


図 6.62: MFCCExtraction の典型的な接続例

ノードの入出力とプロパティ

表 6.54: MFCCExtraction のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int	24		入力スペクトルにかけるフィルタバンク数
NUM_CEPS	int	12		リフタリングで残すケプストラム係数の数
USE_POWER	bool	false		対数パワーを特徴量に含めるか含めないかの選択

入力

FBANK : Map<int, ObjectRef> 型．音源 ID とフィルタバンクの出力エネルギーから構成されるベクトルの Vector<float> 型のデータのペア．

SPECTRUM : Map<int, ObjectRef> 型．音源 ID と複素スペクトルから構成されるベクトルの Vector<complex<float> > 型のデータのペア．

出力

OUTPUT : Map<int, ObjectRef> 型．音源 ID と MFCC と対数パワー項から構成されるベクトルの Vector<float> 型のデータのペア．

パラメータ

FBANK_COUNT : int 型．入力スペクトルにかけるフィルタバンク数．デフォルト値は 24 である．値域は，正の整数．値を大きくすると 1 バンク当りの担当周波数帯域が狭くなり，周波数分解能の高い音響特徴量が求まる．より大きな FBANK_COUNT を設定すると，音響特徴をより精細に表現する．音声認識には，必ずしも精細な表現が最適ではなく，発声する音響環境に依存する．

NUM_CEP : int 型．リフタリングで残すケプストラム係数の数．デフォルト値は 12．値域は，正の整数．値を大きくすると音響特徴量の次元数が増える．より細かなスペクトル変化を表現する音響特徴量が求まる．

USE_POWER : bool 型．対数パワーを特徴量に含めて出力する場合は true 指定．

ノードの詳細

音響特徴量の 1 つであるメルケプストラム係数 (MFCC : Mel-Frequency Cepstrum Coefficients) と対数パワーを求める．MFCC と対数スペクトルパワーを次元要素とする音響特徴量を生成する．

対数スペクトルに，三角窓のフィルタバンクをかける．三角窓の中心周波数は，メルスケール上で等間隔になるように配置する．各フィルタバンクの出力対数エネルギーをとり，離散コサイン変換 (Discrate Cosine Transform) する．得られた係数をリフタリングした係数が MFCC である．

本ノードの入力部の FBANK には， 各フィルタバンクの出力対数エネルギーが入力されることが前提である．

フレーム時刻 f における FBANK への入力ベクトルを，

$$x(f) = [x(f,0), x(f,1), \cdots, x(f,P-1)]^T \tag{6.97}$$

と表す．ただし，P は，入力特徴ベクトルの次元数で，FBANK_COUNT である．出力されるベクトルは，$P+1$ 次元ベクトルで， メルケプストラム係数とパワー項から構成される．1 次元目から P 次元目までは，メルケプストラム係数で，$P+1$ 次元目は，パワー項である．本ノードの出力ベクトルは，

$$y(f) = [y(f,0), y(f,1), \ldots, y(f,P-1), E]^T \tag{6.98}$$

$$y(f,p) = L(p) \cdot \sqrt{\frac{2}{P}} \cdot \sum_{q=0}^{P-1} \left\{ \log(x(q)) \cos\left(\frac{\pi(p+1)(q+0.5)}{P} \right) \right\} \tag{6.99}$$

ただし，E は，パワー項（後述）で，リフタリング係数は，

$$L(p) = 1.0 + \frac{Q}{2} \sin\left(\frac{\pi(p+1)}{Q} \right), \tag{6.100}$$

である．ただし，$Q = 22$ である．

パワー項は，SPECTRUM 部の入力ベクトルから求める．入力ベクトルを

$$s = [s(0), \ldots, s(K-1)]^T, \tag{6.101}$$

と表す．ただし，K は，FFT 長である．K は，SPECTRUM に接続された `Map` の次元数によって決る．対数パワー項は，

$$E = \log\left(\frac{1}{K}\sum_{k=0}^{K-1} s(k)\right) \tag{6.102}$$

6.4.5 MSLSExtraction

ノードの概要

本ノードは，音響特徴量の１つであるメルスケール対数スペクトル (MSLS : Mel-Scale Log-Spectrum) と対数パワーを求める．メルスケール対数スペクトル係数と対数スペクトルパワーを要素とする音響特徴量ベクトルを生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

メルスケール対数スペクトル係数と対数パワーを次元要素とする．音響特徴量ベクトルを生成するために用いる．例えば，音響特徴量ベクトルを音声認識ノードに入力し，音韻や話者を識別する．

典型的な接続例

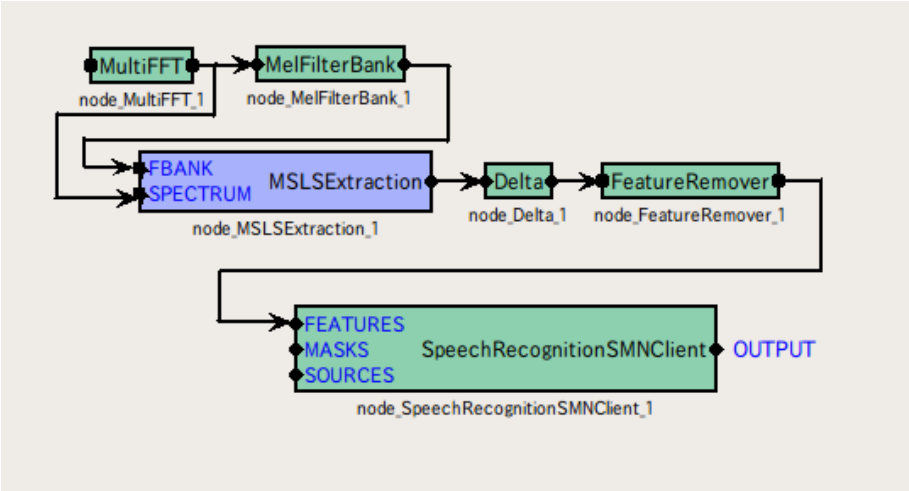


図 6.63: MSLSExtraction の典型的な接続例

ノードの入出力とプロパティ

入力

FBANK : `Map<int, ObjectRef>` 型．音源 ID とフィルタバンクの出力エネルギーから構成されるベクトルの `Vector<float>` 型のデータのペア．

表 6.55: MSLSExtraction のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int	13		入力スペクトルにかけるフィルタバンク数．実装は 13 に最適
NORMALIZATION_MODE	string	CEPSTRAL		特徴量の正規化手法
USE_POWER	bool	false		対数パワーを特徴量に含めるか含めないかの選択

SPECTRUM ：Map<int, ObjectRef> 型．音源 ID と複素スペクトルから構成されるベクトルの Vector<complex<float> > 型のデータのペア．

出力

OUTPUT ：Map<int, ObjectRef> 型．である．音源 ID と MSLS と対数パワー項から構成されるベクトルの Vector<float> 型のデータのペア．本ノードは， MSLS の静的特徴を求めるノードであるが，出力には，動的特徴量部を含んだベクトルを出力する．動的特徴量部分には， 0 が設定される．その様子を図 6.64 に示す．

パラメータ

FBANK_COUNT ：int 型．入力スペクトルにかけるフィルタバンク数．値域は，正の整数．値を大きくすると 1 バンク当りの担当周波数帯域が狭くなり，周波数分解能の高い音響特徴量が求まる．典型的な設定値は，13 から 24 である．ただし，現在の実装では，この値が 13 に固定されるよう最適化されているので，デフォルト値の利用を強く推奨する．より大きな FBANK_COUNT を設定すると，音響特徴をより精細に表現する．音声認識には，必ずしも精細な表現が最適ではなく，発声する音響環境に依存する．

NORMALIZATION_MODE ：string 型．CEPSTRAL または SPECTRAL を指定可能．正規化をケプストラムドメイン／スペクトラムドメインで行うかを選択．

USE_POWER ：true にすると音響特徴量に対数パワー項を追加．false にすると省略．音響特徴量にパワー項を利用することは稀であるが，音声認識には，デルタ対数パワーが有効であるとされる．true にし，後段でデルタ対数パワーを計算し，それを音響特徴量として用いる．

ノードの詳細

本ノードは，音響特徴量の１つであるメルスケール対数スペクトル (MSLS : Mel-Scale Log-Spectrum) と対数パワーを求める．メルスケール対数スペクトル係数と対数スペクトルパワーを次元要素とする音響特徴量を生成する．

本ノードの FBANK 入力端子には， 各フィルタバンクの出力対数エネルギーを入力する．指定する正規化手法によって，出力する MSLS の計算方法が異なる．

以下で，正規化手法ごとに本ノードの出力ベクトルの計算方法を示す．

CEPSTRAL ：FBANK 端子への入力を，

$$x = [x(0), x(1), \cdots, x(P-1)]^T \tag{6.103}$$

と表す．ただし，P は，入力特徴ベクトルの次元数で，FBANK_COUNT である．出力されるベクトルは，$P+1$ 次元ベクトルで， MSLS 係数とパワー項から構成される．1 次元目から P 次元目までは，MSLS で，

$P+1$ 次元目は，パワー項である．本ノードの出力ベクトルは，

$$y = [y(0), y(1), \ldots, y(P-1), E]^T \tag{6.104}$$

$$y(p) = \frac{1}{P}\sum_{q=0}^{P-1}\left\{L(q)\cdot\sum_{r=0}^{P-1}\left\{\log(x(r))\cos\left(\frac{\pi q(r+0.5)}{P}\right)\right\}\cos\left(\frac{\pi q(p+0.5)}{P}\right)\right\} \tag{6.105}$$

である．ただし，リフタリング係数は，

$$L(p) = \begin{cases} 1.0, & (p = 0, \ldots, P-1), \\ 0.0, & (p = P, \ldots, 2P-1), \end{cases} \tag{6.106}$$

とする．ただし，$Q = 22$ である．

SPECTRAL ：FBANK 部の入力を

$$x = [x(0), x(1), \cdots, x(P-1)]^T \tag{6.107}$$

と表す．ただし，P は，入力特徴ベクトルの次元数で，FBANK_COUNT である．出力されるベクトルは，$P+1$ 次元ベクトルで， MSLS 係数とパワー項から構成される．1 次元目から P 次元目までは，MSLS で，$P+1$ 次元目は，パワー項である．本ノードの出力ベクトルは，

$$y = [y(0), y(1), \ldots, y(P-1), E]^T \tag{6.108}$$

$$y(p) = \begin{cases} (\log(x(p)) - \mu) - 0.9(\log(x(p-1)) - \mu), & if \;\; p = 1, \ldots, P-1 \\ \log(x(p), & if \;\; p = 0, \end{cases} \tag{6.109}$$

$$\mu = \frac{1}{P}\sum_{q=0}^{P-1}\log(x(q)), \tag{6.110}$$

である．周波数方向の平均除去と，ピーク強調処理を適用ている．

対数パワー項は，SPECTRUM 端子の入力を

$$s = [s(0), s(1), \ldots, s(N-1)]^T \tag{6.111}$$

と表す．ただし，N は，SPECTRUM 端子に接続された `Map` のサイズによって決る．`Map` は，0 から π までのスペクトル表現を B 個のビンに格納しているとすると，$N = 2(B-1)$ である．この時，パワー項は，

$$p = \log\left(\frac{1}{N}\sum_{n=0}^{N-1}s(n)\right) \tag{6.112}$$

である．

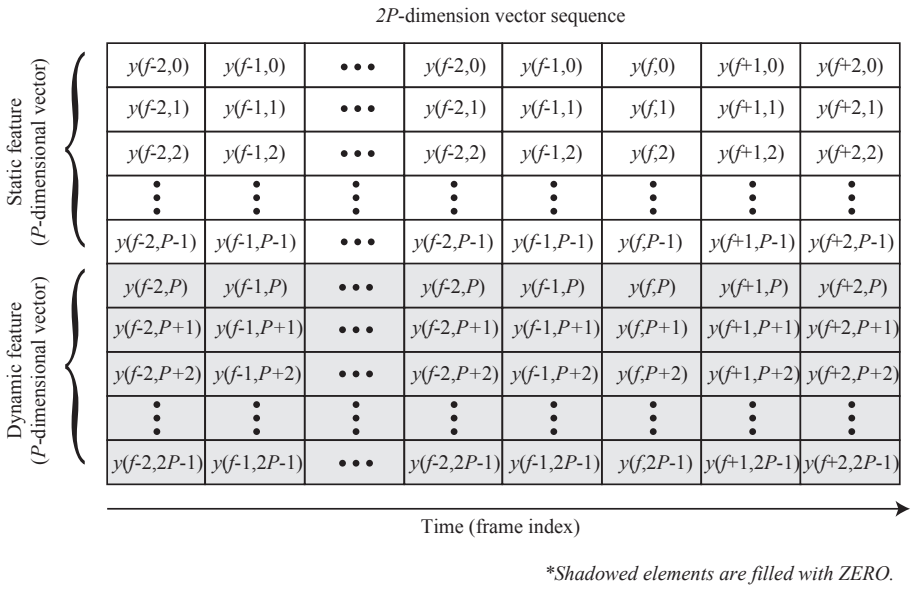


図 6.64: MSLSExtraction の出力パラメータ

6.4.6 PreEmphasis

ノードの概要

音声認識用の音響特徴量抽出の際に高域の周波数を強調する処理（プリエンファシス）を行い，ノイズへの頑健性を高める．

必要なファイル

無し．

使用方法

一般的に，MFCC 特徴量抽出の前に用いる．また，HARK で一般的に用いている MSLS 特徴量抽出の際にも前処理として用いることができる．

典型的な接続例

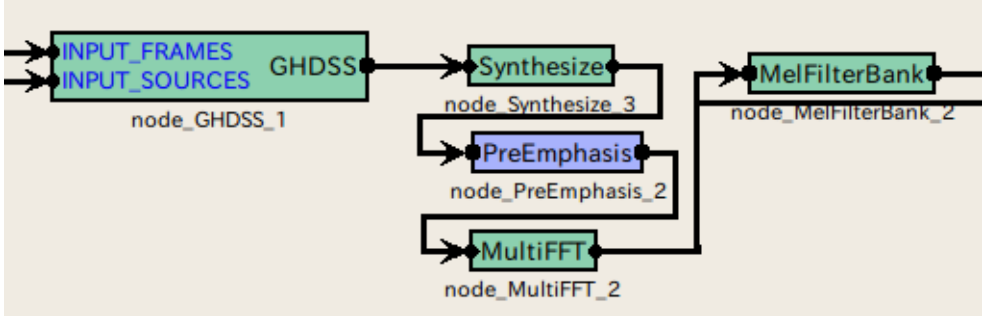


図 6.65: PreEmphasis の接続例

ノードの入出力とプロパティ

表 6.56: PreEmphasis のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	信号長もしくは FFT の窓長
SAMPLING_RATE	int	16000	[Hz]	サンプリングレート
PREEMCOEF	float	0.97		プリエンファシス係数
INPUT_TYPE	string	WAV		入力信号タイプ

入力

INPUT : Map<int, ObjectRef>，入力信号が時間領域波形の場合は，ObjectRef は，Vector<float> として扱われる．また，周波数領域の信号の場合は，Vector<complex<float> > として扱われる．

出力

OUTPUT : `Map<int, ObjectRef>`，高域強調された信号が出力される．`ObjectRef` は，入力の種類に対応して，時間領域波形では `Vector<float>`，周波数領域信号では `Vector<complex<float> >` となる．

パラメータ

LENGTH INPUT_TYPE が SPECTRUM の場合は FFT 長であり，前段のモジュールと値を合わせる必要がある．INPUT_TYPE が WAV の場合は，１フレームに含まれる信号の信号長を表し，同様に前段のノードと値を合わせる必要がある．通常の構成では，FFT 長と信号長は一致する．

SAMPLING_RATE LENGTH と同様，他のノードと値を合わせる必要がある．

PREEMCOEF 以下で c_p として表わされるプリエンファシス係数．音声認識では，0.97 が一般的に用いられる．

INPUT_TYPE 入力のタイプは WAV, SPECTRUM の２種類が用意されている．WAV は時間領域波形入力の際に用いる．また，SPECTRUM は周波数領域信号入力の際に用いる．

ノードの詳細

プリエンファシスの必要性や一般的な音声認識における効果に関しては，様々な書籍や論文で述べられているので，それらを参考にしてほしい．一般的には，この処理を行った方がノイズに頑健になると言われているが，HARK では，マイクロホンマイクロホンアレイ処理を行っているためか，この処理の有無による性能差はそれほど大きくない．ただし，音声認識で用いる音響モデルを学習する際に用いた音声データとパラメータを合わせる必要性がある．つまり，音響モデル学習で用いたデータにプリエンファシスを行っていれば，入力データに対してもプリエンファシスを行った方が性能が高くなる．

具体的には PreEmphasis は，入力信号の種類に対応して，２種類の処理からなっている．

時間領域での高域強調:

時間領域の場合は，t をフレーム内でのサンプルを表すインデックスとし，入力信号を $s[t]$，高域強調した信号を $p[t]$，プリエンファシス係数を c_p とすれば，以下の式によって表すことができる．

$$p[t] = \begin{cases} s[t] - c_p \cdot s[t-1] & t > 0 \\ (1 - c_p) \cdot s[0] & t = 0 \end{cases} \tag{6.113}$$

周波数領域での高域強調:

時間領域のフィルタと等価なフィルタを周波数領域で実現するため，上記のフィルタのインパルス応答に対して，周波数解析を行ったスペクトルフィルタを用いている．また，低域（下から 4 バンド分）と高域（$fs/2$ - 100 Hz 以上）は，誤差を考慮して，強制的に 0 としている．ただし，fs は，サンプリング周波数を表す．

6.4.7 SaveFeatures

ノードの概要

特徴量ベクトルをファイルに保存する．

必要なファイル

無し．

使用方法

どんなときに使うのか

MFCC，MSLS などの音響特徴量を保存する時に使用する．

典型的な接続例

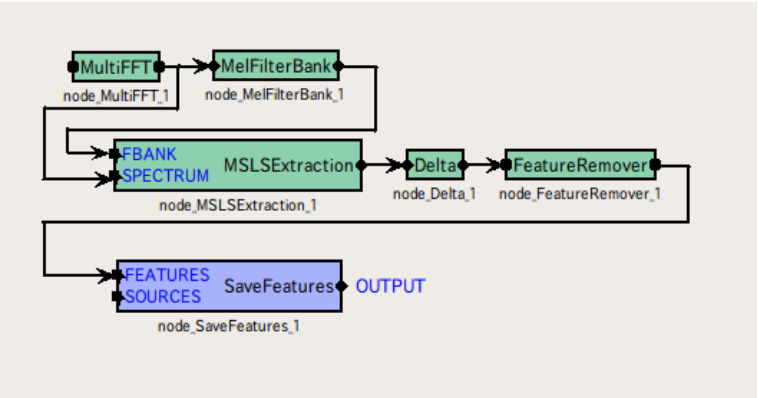


図 6.66: SaveFeatures の接続例

ノードの入出力とプロパティ

表 6.57: SaveFeatures のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
BASENAME	string			保存する時のファイル名の Prefix

入力

FEATURES ：Map<int, ObjectRef> 型．特徴量ベクトルは Vector<float> で示される．

SOURCES ：Vector<ObjectRef> 型である．この入力は，オプションである．

出力

OUTPUT : `Map<int, ObjectRef>` 型である．

パラメータ

BASENAME : `string` 型である．保存する時のファイル名の Prefix で，保存時は，Prifix の後に SOURCES の ID が付与されて特徴量が保存される．

ノードの詳細

特徴量ベクトルを保存する．保存するファイルの形式は，ベクトル要素を IEEE 758 の 32 ビット浮動小数点数形式，リトルエンディアンで保存する．名前付ルールは， BASENAME プロパティで与えた Prifix の後に ID 番号が付与される．

6.4.8 SaveHTKFeatures

ノードの概要

特徴量ベクトルを HTK (The Hidden Markov Model Toolkit) で扱えるファイル形式で保存する．

必要なファイル

無し．

使用方法

どんなときに使うのか

MFCC，MSLS などの音響特徴量を保存する時に使用する．SaveFeatures と異なり，HTK で扱えるように専用のヘッダが追加されて保存できる．

典型的な接続例

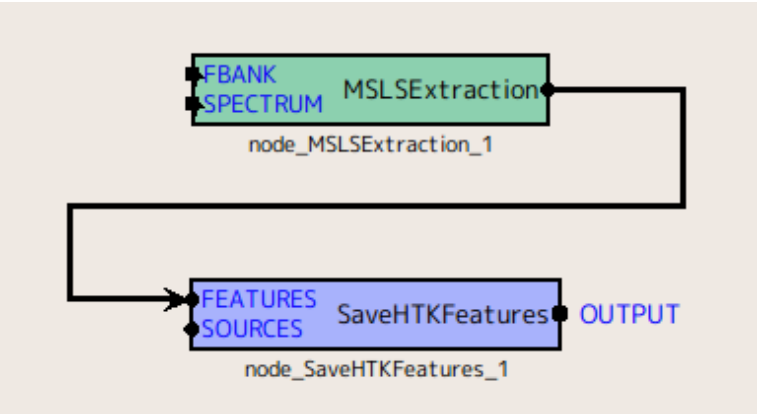


図 6.67: SaveHTKFeatures の接続例

ノードの入出力とプロパティ

表 6.58: SaveHTKFeatures のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
BASENAME	`string`			保存する時のファイル名の Prefix
HTK_PERIOD	`int`	100000	[100 nsec]	フレーム周期
FEATURE_TYPE	`string`	USER		特徴量の型

入力

FEATURES : `Map<int, ObjectRef>` 型である．

SOURCES : `Vector<ObjectRef>` 型である．この入力は，オプションである．

出力

OUTPUT : `Map<int, ObjectRef>` 型である．

パラメータ

BASENAME : `string` 型である．保存する時のファイル名の Prefix で，保存時は，Prifix の後に SOURCES の ID が付与されて特徴量が保存される．

HTK_PERIOD : フレーム周期の設定で単位は [100 nsec] である．サンプリング周波数 16000[Hz] でシフト長 160 の場合，フレーム周期は 10[msec] となるので 10[ms] = 100000 * 100[nsec] で，100000 が適当な設定値となる．

FEATURE_TYPE : HTK で扱う特徴量の形式の指定．HTK 独自の型に従う．例えば，`MFCC_E_D` の場合は「(MFCC+パワー) + デルタ (MFCC+パワー)」となる．HARK で計算した特徴量の中身と合わせるように設定する．詳細は HTKbook を参照されたい．

ノードの詳細

特徴量ベクトルを HTK で扱われる形式で保存する．保存するファイルの形式は，HTK 固有のヘッダを付与した後，ベクトル要素を IEEE 758 の 32 ビット浮動小数点数形式，ビッグエンディアンで保存する．名前付ルールは， BASENAME プロパティで与えた Prifix の後に ID 番号が付与される．HTK のヘッダの設定はプロパティで変更可能.

6.4.9 SpectralMeanNormalization

ノードの概要

入力音響特徴量から特徴量の平均を除去することを意図したノードである．ただし，実時間処理を実現するためには，当該発話の平均を除去することができない問題がある．当該発話の平均値をなんらかの値を用いてい推定あるいは，近似する必要がある．

必要なファイル

無し．

使用方法

どんなときに使うのか

音響特徴量の平均を除去したい時に使用する．音響モデル学習用音声データと認識用音声データの収録環境の平均値のミスマッチを除去できる．

音声収録環境においてマイクロホンの特性は，統一できないことが多い．特に，音響モデル学習時と認識時の音声収録環境は，必ずしも等しくない．通常，学習用の音声コーパス作成者と，認識用音声データの収録者が異なるから環境を揃えることは困難である．従って，音声の収録環境に依存しない特徴量を用いる必要がある．

例えば，学習データ収録に使用するマイクロホンと認識時に使用するマイクロフォンは通常異なる．マイクロホンの特性の違いが，収録音の音響特徴のミスマッチとして現れ，認識性能の低下を招く．マイクロホンの特性の違いは，時不変であり，平均スペクトルの差となって現れる．従って，平均スペクトルを除去することで，簡易的に収録環境に依存した成分を特徴量から除去できる．

典型的な接続例

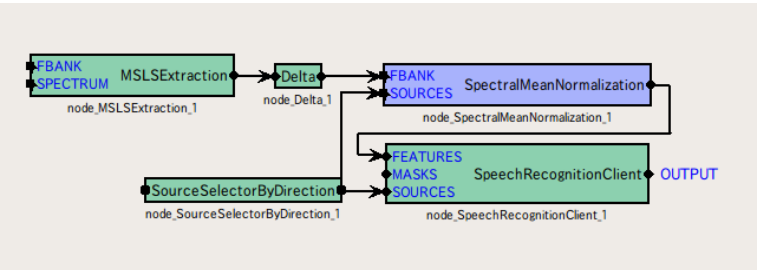


図 6.68: SpectralMeanNormalization の接続例

ノードの入出力とプロパティ

入力

FBANK ：`Map<int, ObjectRef>` 型．音源 ID と特徴量ベクトルの `Vector<float>` 型のデータのペア．

SOURCES ：`Vector<ObjectRef>` 型である．音源位置．

表 6.59: SpectralMeanNormalization のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int	13		入力特徴パラメータの次元数

出力

OUTPUT : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

パラメータ

FBANK_COUNT : int 型である．値域は 0 または正の整数である．

ノードの詳細

入力音響特徴量から特徴量の平均を除去することを意図したノードである．ただし，実時間処理を実現するためには，当該発話の平均を除去することができない問題がある．当該発話の平均値をなんらかの値を用いてい推定あるいは，近似する必要がある．

当該発話の平均を除去する替りに，前発話の平均を近似値とし，除去することで実時間平均除去を実現する．この方法では，更に音源方向を考慮しなければならない．音源方向によって伝達関数が異なるため，当該発話と前発話が異なる方向から受音された場合には，前発話の平均は，当該発話の平均の近似として不適当である．この場合，当該発話の平均の近似として，当該発話よりも前に発話されかつ，同方向からの発話の平均を用いる．

最後に以後の平均除去に備え，当該発話の平均を計算し，当該発話方向の平均値としてメモリ内に保持する．発話中に音源が 10 [deg] 以上移動する場合は，別音源として，平均を計算する．

6.5 MFM カテゴリ

6.5.1 DeltaMask

ノードの概要

本ノードは，静的特徴のミッシングフィーチャーマスクベクトルから動的特徴量のミッシングフィーチャーマスクベクトルを求め，静的特徴と動的特徴のミッシングフィーチャーマスクから構成されるマスクベクトルを生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

ミッシングフィーチャー理論に基づき，特徴量を信頼度に応じてマスクして音声認識を行うために用いる．通常，MFMGeneration の後段に用いる．

典型的な接続例

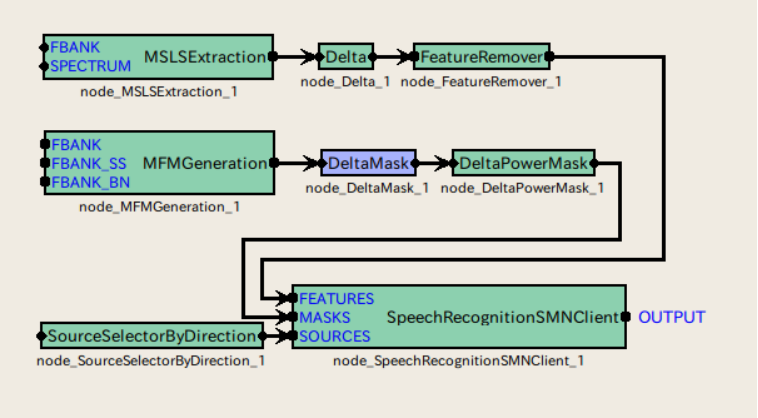


図 6.69: DeltaMask の典型的な接続例

ノードの入出力とプロパティ

表 6.60: DeltaMask パラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int			静的特徴の次元数

入力

INPUT : `Map<int, ObjectRef>` 型．音源 ID と特徴量のマスクベクトルの `Vector<float>` 型のデータのペア．マスク値は，0.0 から 1.0 の実数で，0.0 が特徴量を信頼しない， 1.0 が信頼する状態を表す．

出力

OUTPUT : `Map<int, ObjectRef>` 型．音源 ID と特徴量のマスクベクトルの `Vector<float>` 型のデータのペア．マスク値は，0.0 から 1.0 の実数で，0.0 が特徴量を信頼しない状態， 1.0 が信頼する状態を表す．

パラメータ

FBANK_COUNT : `int` 型である．処理する特徴量の次元数．値域は，正の整数．

ノードの詳細

本ノードは，静的特徴のマスクベクトルから動的特徴量のマスクベクトルを求め，静的特徴と動的特徴のミッシングフィーチャーマスクから構成されるマスクベクトルを生成する．

フレーム時刻 f における，入力マスクベクトルを，

$$m(f) = [m(f,0), m(f,1), \ldots, m(f,2P-1)]^T \tag{6.114}$$

と表す．ただし，P は，入力マスクベクトルのうち，静的特徴を表わす次元数を表わし，FBANK_COUNT で与える．静的特徴のマスク値を用い，動的特徴のマスク値を求め， P から $2P-1$ 次元の要素に入れて出力ベクトルを生成する．出力ベクトル $m'(f)$ は，

$$y'(f) = [m'(f,0), m'(f,1), \ldots, m'(f,2P-1)]^T \tag{6.115}$$

$$m'(f,p) = \begin{cases} m(f,p), & \text{if } p = 0, \ldots, P-1, \\ \prod_{\tau=-2}^{2} m(f+\tau, p), & \text{if } p = P, \ldots, 2P-1, \end{cases} \tag{6.116}$$

である．図 6.70 に DeltaMask の入出力フローを示す．

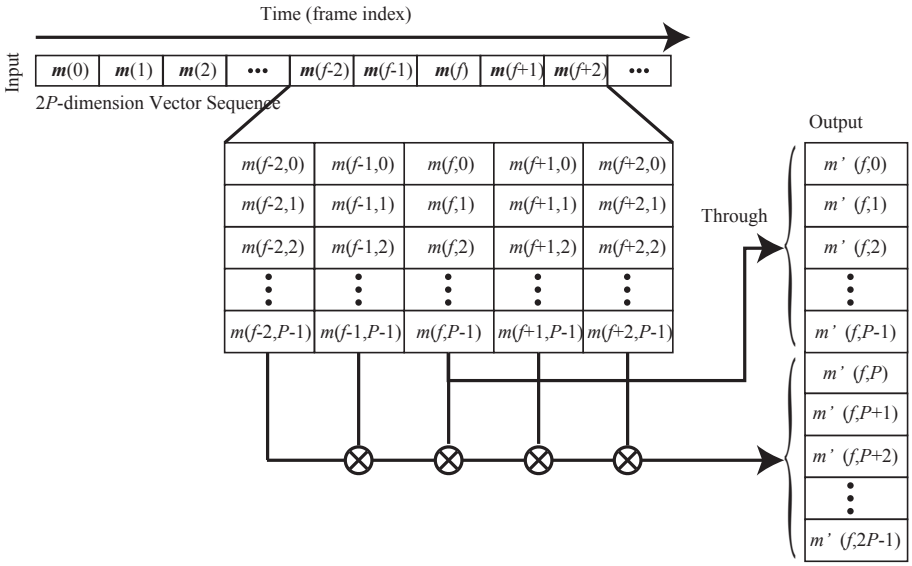


図 6.70: DeltaMask の入出力フロー.

6.5.2 DeltaPowerMask

ノードの概要

本ノードは，音響特徴量の１つである動的対数パワーのマスク値を生成する．生成したマスク値を入力のマスクベクトルの要素に追加する．

必要なファイル

無し．

使用方法

どんなときに使うのか

ミッシングフィーチャー理論に基づき，特徴量を信頼度に応じてマスクして音声認識を行う．通常，DeltaMask の後段に用いる．

典型的な接続例

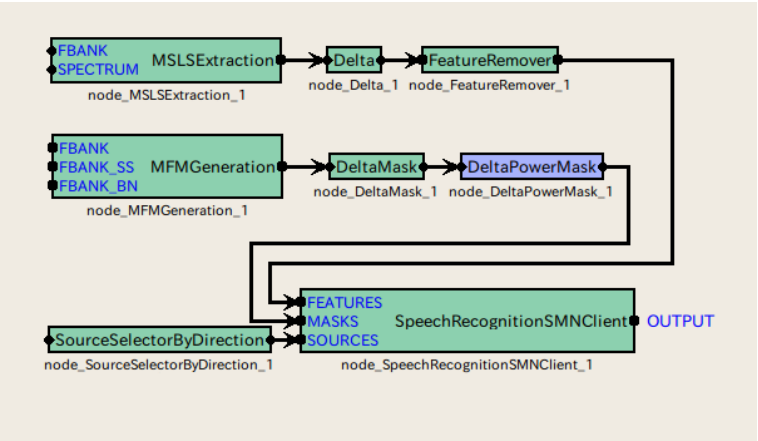


図 6.71: DeltaPowerMask の典型的な接続例

ノードの入出力とプロパティ

INPUT ：`Map<int, ObjectRef>` 型．音源 ID と特徴量のマスクベクトルの `Vector<float>` 型のデータのペア．マスク値は，0.0 から 1.0 の実数で，0.0 が特徴量を信頼しない状態， 1.0 が信頼する状態を表す．

出力

OUTPUT ：`Map<int, ObjectRef>` 型．音源 ID と特徴量のマスクベクトルの `Vector<float>` 型のデータのペア．マスク値は，0.0 から 1.0 の実数で，0.0 が特徴量を信頼しない状態， 1.0 が信頼する状態を表す．

パラメータ

ノードの詳細

本ノードは，音響特徴量の１つである動的対数パワーのマスク値を生成する．生成するマスク値は，常に 1.0 である．出力マスクの次元数は，入力マスクの次元数 + 1 次元である．

6.5.3 MFMGeneration

ノードの概要

本ノードは，ミッシングフィーチャー理論に基く音声認識のためのミッシングフィーチャーマスク (Missing-Feature-Mask:MFM) を生成する．

必要なファイル

無し．

使用方法

どんなときに使うのか

ミッシングフィーチャー理論に基く音声認識するために使用する．MFMGeneration は，PostFilter と GHDSS の出力からミッシングフィーチャーマスクを生成する．そのため PostFilter と GHDSS の利用が前提条件である．

典型的な接続例

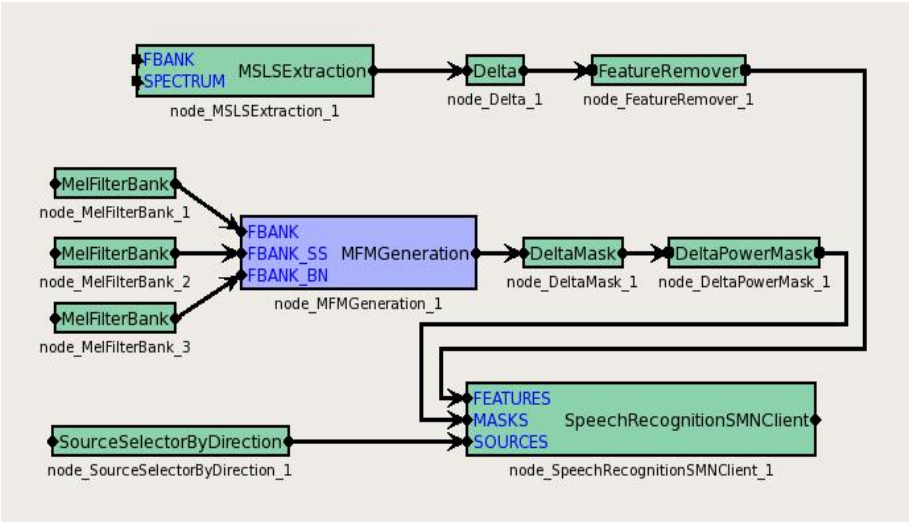


図 6.72: MFMGeneration の接続例

ノードの入出力とプロパティ

入力

FBANK : `Map<int, ObjectRef>` 型．音源 ID と PostFilter の出力から求めたメルフィルタバンク出力エネルギーから構成されるベクトルの `Vector<float>` 型のデータのペア．

FBANK_SS : `Map<int, ObjectRef>` 型．音源 ID と GHDSS の出力から求めたメルフィルタバンク出力エネルギーから構成されるベクトルの `Vector<float>` 型のデータのペア．

表 6.61: MFMGeneration のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
FBANK_COUNT	int	13		音響特徴量の次元数
THRESHOLD	float	0.2		0.0 から 1.0 の間の連続値を 0.0 (信頼しない) または 1.0 (信頼する) に量子化するためのしきい値

FBANK_BN : Map<int, ObjectRef> 型．音源 ID と BGNEstimator の出力から求めたメルフィルタバンク出力エネルギーから構成されるベクトルの Vector<float> 型のデータのペア．

出力

OUTPUT : Map<int, ObjectRef> 型．音源 ID と ミッシングフィーチャーマスクベクトルから構成されるベクトルの Vector<float> 型のデータのペア．ベクトルの要素は 0.0 (信頼しない) または 1.0 (信頼する) である．出力ベクトルは，2*FBANK_COUNT 次元ベクトルで， FBANK_COUNT 以上の次元要素は，全て 0 である．動的特徴量用のミッシングフィーチャーマスクのプレースホルダである．

パラメータ

FBANK_COUNT : int 型である．音響特徴量の次元数である．

THRESHOLD : float 型である．ノード内部で計算する 0.0(信頼しない) から 1.0(信頼する) までの信頼度を量子化するためのしきい値である．しきい値に 0.0 を設定すると，すべての信頼度がしきい値以上になり，すべてのマスク値が 1.0 になる．このときの処理は，通常の音声認識処理と等価になる．

ノードの詳細

ミッシングフィーチャー理論に基く音声認識のためのミッシングフィーチャーマスクを生成する．

信頼度 $r(p)$ をしきい値 THRESHOLD でしきい値処理し，マスク値を 0.0 (信頼しない) また 1.0 (信頼する) に量子化する．信頼度は，PostFilter , GHDSS , BGNEstimator の出力から求めたメルフィルタバンクの出力エネルギー $f(p), b(p), g(p)$, から求める．このときフレーム番号 f のマスクベクトルは，

$$m(f) = [m(f,0), m(f,1), \ldots, m(f,P-1)]^T \tag{6.117}$$

$$m(f,p) = \begin{cases} 1.0, & r(p) > \mathrm{THRESHOLD} \\ 0.0, & r(p) \leq \mathrm{THRESHOLD} \end{cases}, \tag{6.118}$$

$$r(p) = \min(1.0, (f(p) + 1.4 * b(p))/(fg(p) + 1.0)), \tag{6.119}$$

$$\tag{6.120}$$

である．ただし，P は，入力特徴ベクトルの次元数で，FBANK_COUNT で指定する正の整数である．実際に出力するベクトルの次元数は， 2*FBANK_COUNT 次のベクトルである．FBANK_COUNT 以上の次元要素は，0 で埋められる．これは，動的特徴量マスク値を入れるためのプレースホルダである．図 6.73 に出力ベクトル列の模式図を示す．

2P-dimension vector sequence

Static feature (*P*-dimensional vector)	$m(f0,0)$	$m(1,0)$	• • •	$m(f\text{-}2,0)$	$m(f\text{-}1,0)$	$m(f,0)$	$m(f\text{+}1,0)$	$m(f\text{+}2,0)$
	$m(f0,1)$	$m(1,1)$	• • •	$m(f\text{-}2,1)$	$m(f\text{-}1,1)$	$m(f,1)$	$m(f\text{+}1,1)$	$m(f\text{+}2,1)$
	$m(0,2)$	$m(1,2)$	• • •	$m(f\text{-}2,2)$	$m(f\text{-}1,2)$	$m(f,2)$	$m(f\text{+}1,2)$	$m(f\text{+}2,2)$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$m(0,P\text{-}1)$	$m(1,P\text{-}1)$	• • •	$m(f\text{-}2,P\text{-}1)$	$m(f\text{-}1,P\text{-}1)$	$m(f,P\text{-}1)$	$m(f\text{+}1,P\text{-}1)$	$m(f\text{+}2,P\text{-}1)$
Dynamic feature (*P*-dimensional vector)	$m(0,P)$	$m(1,P)$	• • •	$m(f\text{-}2,P)$	$m(f\text{-}1,P)$	$m(f,P)$	$m(f\text{+}1,P)$	$m(f\text{+}2,P)$
	$m(0,P\text{+}1)$	$m(1,P\text{+}1)$	• • •	$m(f\text{-}2,P\text{+}1)$	$m(f\text{-}1,P\text{+}1)$	$m(f,P\text{+}1)$	$m(f\text{+}1,P\text{+}1)$	$m(f\text{+}2,P\text{+}1)$
	$m(0,P\text{+}2)$	$m(1,P\text{+}2)$	• • •	$m(f\text{-}2,P\text{+}2)$	$m(f\text{-}1,P\text{+}2)$	$m(f,P\text{+}2)$	$m(f\text{+}1,P\text{+}2)$	$m(f\text{+}2,P\text{+}2)$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	$m(0,2P\text{-}1)$	$m(1,2P\text{-}1)$	• • •	$m(f\text{-}2,2P\text{-}1)$	$m(f\text{-}1,2P\text{-}1)$	$m(f,2P\text{-}1)$	$m(f\text{+}1,2P\text{-}1)$	$m(f\text{+}2,2P\text{-}1)$

Time (frame index)

**Shadowed elements are filled with ZERO.*

図 6.73: MFMGeneration の出力ベクトル列

6.6 ASRIF カテゴリ

6.6.1 SpeechRecognitionClient

ノードの概要

音響特徴量をネットワーク経由で音声認識ノードに送信するノードである．

必要なファイル

無し．

使用方法

どんなときに使うのか

音響特徴量を HARK 外のソフトウェアに送信するために用いる．例えば，大語彙連続音声認識ソフトウェア Julius [(1)] に送信し，音声認識を行う．

典型的な接続例

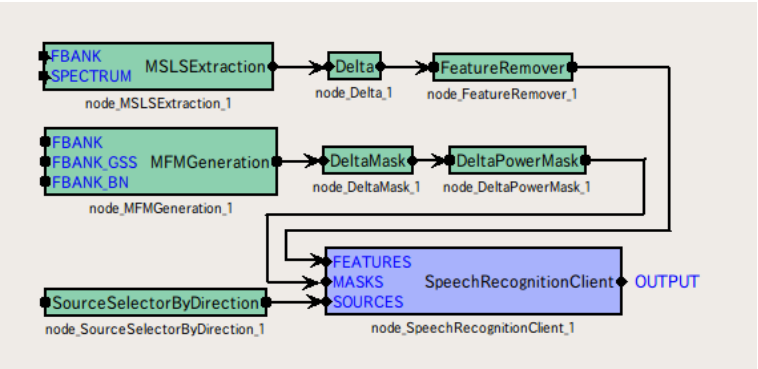


図 6.74: SpeechRecognitionClient の接続例

ノードの入出力とプロパティ

表 6.62: SpeechRecognitionClient のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
MFM_ENABLED	bool	true		ミッシングフィーチャーマスクを送出するかしないかの選択
HOST	string	127.0.0.1		Julius/Julian が動いているサーバのホスト名/IP アドレス
PORT	int	5530		ネットワーク送出用ポート番号
SOCKET_ENABLED	bool	true		ソケット出力をするかどうかを決めるフラグ

入力

FEATURES : Map<int, ObjectRef> 型．音源 ID と特徴量ベクトルの Vector<float> 型のデータのペア．

MASKS : Map<int, ObjectRef> 型．音源 ID とマスクベクトルの Vector<float> 型のデータのペア．

SOURCES : Vector<ObjectRef> 型．

出力

OUTPUT : Vector<ObjectRef> 型．

パラメータ

MFM_ENABLED : bool 型．true の場合， MASKS を転送する．false の場合は，入力の MASKS を無視し，すべて 1 のマスクを転送する．

HOST : string 型．音響パラメータを転送するホストの IP アドレスでる．SOCKET_ENABLED が false の場合は，無効である．

PORT : int 型．音響パラメータ転送するソケット番号である． SOCKET_ENABLED が false の場合は，無効である．

SOCKET_ENABLED : bool 型．true で音響パラメータをソケットに転送し，false で転送しない．

ノードの詳細

MFM_ENABLED が true かつ SOCKET_ENABLED のとき，音響特徴量ベクトルとマスクベクトルをネットワークポートを経由で音声認識ノードに送信するノードである．MFM_ENABLED が false のとき，ミッシングフィーチャー理論を使わない音声認識になる．実際には，マスクベクトルの値をすべて 1 ，つまりすべての音響特徴量を信頼する状態にしてマスクベクトルを送り出す．SOCKET_ENABLED が false のときは，特徴量を音声認識ノードに送信しない．これは，音声認識エンジンが外部プログラムに依存しているため，外部プログラムを動かさずに HARK のネットワーク動作チェックを行うために使用する．HOST は，ベクトルを送信する外部プログラムが動作する HOST の IP アドレスを指定する．PORT は，ベクトルを送信するネットワークポート番号を指定する．

参考文献:

(1) http://julius.sourceforge.jp/en_index.php

6.6.2 SpeechRecognitionSMNClient

ノードの概要

音響特徴量をネットワーク経由で音声認識ノードに送信するノードである．SpeechRecognitionClient との違いは，入力特徴ベクトルの平均除去 (Spectral Mean Normalization: SMN) を行う点である．ただし，実時間処理を実現するためには，当該発話の平均を除去することができない問題がある．当該発話の平均値をなんらかの値を用いてい推定あるいは，近似する必要がある．近似処理の詳細は，ノードの詳細部分を参照のこと．

必要なファイル

無し．

使用方法

どんなときに使うのか

音響特徴量を HARK 外のソフトウェアに送信するために用いる．例えば，大語彙連続音声認識ソフトウェア Julius [1] に送信し，音声認識を行う．

典型的な接続例

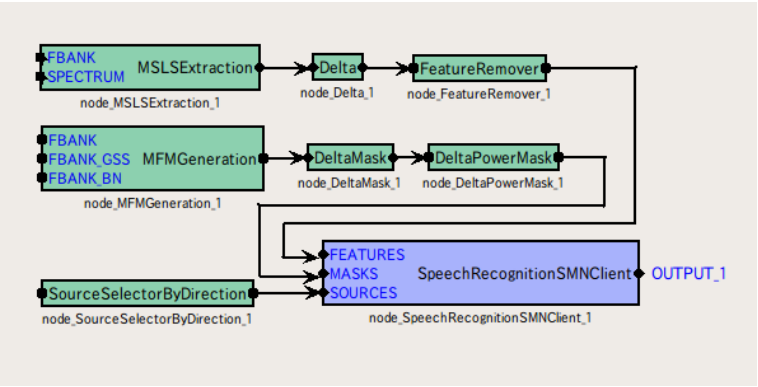


図 6.75: SpeechRecognitionSMNClient の接続例

ノードの入出力とプロパティ

入力

FEATURES ：`Map<int, ObjectRef>` 型．音源 ID と特徴量ベクトルの `Vector<float>` 型のデータのペア．

MASKS ：`Map<int, ObjectRef>` 型．音源 ID とマスクベクトルの `Vector<float>` 型のデータのペア．

SOURCES ：`Vector<ObjectRef>` 型．

出力

表 6.63: SpeechRecognitionSMNClient のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
MFM_ENABLED	bool	true		ミッシングフィーチャーマスクを送出するかしないかの選択
HOST	string	127.0.0.1		Julius/Julian が動いているサーバのホスト名/IP アドレス
PORT	int	5530		ネットワーク送出用ポート番号
SOCKET_ENABLED	bool	true		ソケット出力をするかどうかを決めるフラグ

OUTPUT : Vector<ObjectRef> 型.

パラメータ

MFM_ENABLED : bool 型. true の場合, MASKS を転送する. false の場合は, 入力の MASKS を無視し, すべて 1 のマスクを転送する.

HOST : string 型. 音響パラメータを転送するホストの IP アドレスでる. SOCKET_ENABLED が false の場合は, 無効である.

PORT : int 型. 音響パラメータを転送するソケット番号である. SOCKET_ENABLED が false の場合は, 無効である.

SOCKET_ENABLED : bool 型. true で音響パラメータをソケットに転送し, false で転送しない.

ノードの詳細

MFM_ENABLED が true かつ SOCKET_ENABLED のとき, 音響特徴量ベクトルとマスクベクトルをネットワークポートを経由で音声認識ノードに送信するノードである. MFM_ENABLED が false のとき, ミッシングフィーチャー理論を使わない音声認識になる. 実際には, マスクベクトルの値をすべて 1 , つまりすべての音響特徴量を信頼する状態にしてマスクベクトルを送り出す. SOCKET_ENABLED が false のときは, 特徴量を音声認識ノードに送信しない. これは, 音声認識エンジンが外部プログラムに依存しているため, 外部プログラムを動かさずに HARK のネットワーク動作チェックを行うために使用する. HOST は, ベクトルを送信する外部プログラムが動作する HOST の IP アドレスを指定する. PORT は, ベクトルを送信するネットワークポート番号を指定する.

参考文献:

(1) http://julius.sourceforge.jp/en_index.php

6.7 MISC カテゴリ

6.7.1 ChannelSelector

ノードの概要

マルチチャネルの音声波形や複素スペクトルのデータから，指定したチャネルのデータだけを指定した順番に取り出す．

必要なファイル

無し．

使用方法

どんなときに使うのか

入力されたマルチチャネルの音声波形や複素スペクトルのデータの中から，必要のないチャネルを削除したいとき，あるいは，チャネルの並びを入れ替えたいとき．あるいは，チャネルを複製したいとき．

典型的な接続例

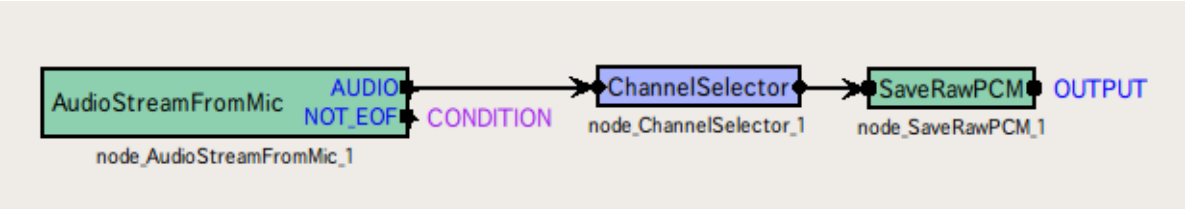


図 6.76: ChannelSelector の典型的な接続例

図 6.76 に典型的な接続例を示す．このネットワークファイルによって，マルチチャネルの音声ファイルのいくつかのチャネルだけを抽出できる．主な入力元は AudioStreamFromMic，AudioStreamFromWave，MultiFFT，主な出力先は SaveRawPCM，MultiFFT などである．

ノードの入出力とパラメータ

入力

INPUT : `Matrix<float>` もしくは `Matrix<complex<float> >` 型．マルチチャネルの音声波形または複素スペクトルのデータ．

出力

OUTPUT : `Matrix<float>` もしくは `Matrix<complex<float> >` 型．マルチチャネルの音声波形または複素スペクトルのデータ．

表 6.64: ChannelSelector パラメータ表

パラメータ名	型	デフォルト値	単位	説明
SELECTOR	`Object`	`Vector< int >`		出力するチャネルの番号を指定

パラメータ

SELECTOR : 型，デフォルト値は無し．使用するチャネルの，チャネル番号を指定する．チャネル番号は 0 からはじまる．

例: 5 チャネル (0-4) のうち 2，3，4 チャネルだけを使うときは (1) のように，さらに 3 チャネルと 4 チャネルを入れ替えたい時は (2) のように指定する．

(1) `<Vector<int> 2 3 4>`

(2) `<Vector<int> 2 4 3>`

ノードの詳細

入力の $N \times M$ 型行列 (`Matrix`) から指定したチャネルの音声波形（もしくは複素スペクトル）データだけを抽出し，新たな $N' \times M$ 型行列のデータを出力する．ただし，N は入力チャネル数，N' は出力チャネル数．

6.7.2 CombineSource

ノードの概要

LocalizeMUSIC や ConstantLocalization 等から出力された 2 つの音源定位結果を結合し，一つにまとめて出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

複数の音源定位結果を GHDSS 等の後段処理で使用したい時に使える．以下の使用事例が考えられる．

- LocalizeMUSIC を複数使用する場合
- ConstantLocalization と LocalizeMUSIC の両方を組み合わせる場合

典型的な接続例

主に，ConstantLocalization，LoadSourceLocation，LocalizeMUSIC などの音源定位結果を入力として接続し，GHDSS や SpeechRecognitionClient などの音源定位結果が必要なモジュールを出力側に接続する．
図 6.77 は，LocalizeMUSIC と ConstantLocalization を組み合わせる例である．

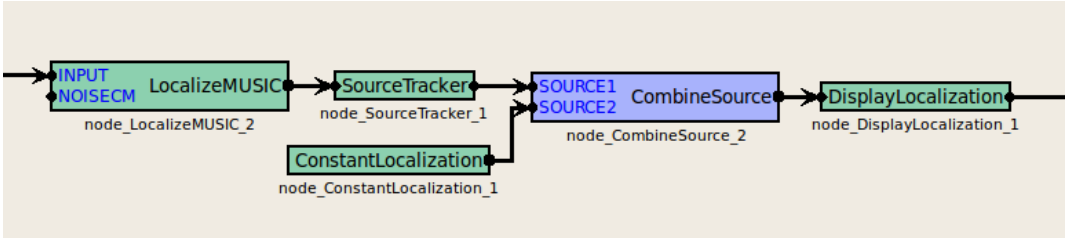


図 6.77: CombineSource の接続例

ノードの入出力とプロパティ

入力

SOURCES1 ： `Vector<ObjectRef>` 型．結合したい音源定位結果を接続する．`ObjectRef` が参照するのは，`Source` 型のデータである．

SOURCES2 ： `Vector<ObjectRef>` 型．結合したい音源定位結果を接続する．`ObjectRef` が参照するのは，`Source` 型のデータである．

出力

SRCOUT : Vector<ObjectRef> 型．結合後の音源定位結果を出力する．ObjectRef が参照するのは，Source 型のデータである．

パラメータ

なし

ノードの詳細

本ノードは二つの音源定位結果を一つにまとめて出力する．音源の方位角，仰角，パワー，ID は引き継がれる．

6.7.3 DataLogger

ノードの概要

入力されたデータにパラメータで指定したラベルを付与して標準出力またはファイルに出力する．

必要なファイル

無し．

使用方法

どんなときに使うのか

ノードのデバッグの際や，ノードの出力を保存して実験や解析に利用したい場合に使う．

典型的な接続例

例えば，各音源の特徴量をテキストで出力して解析したい場合には，以下のような接続すればよい．

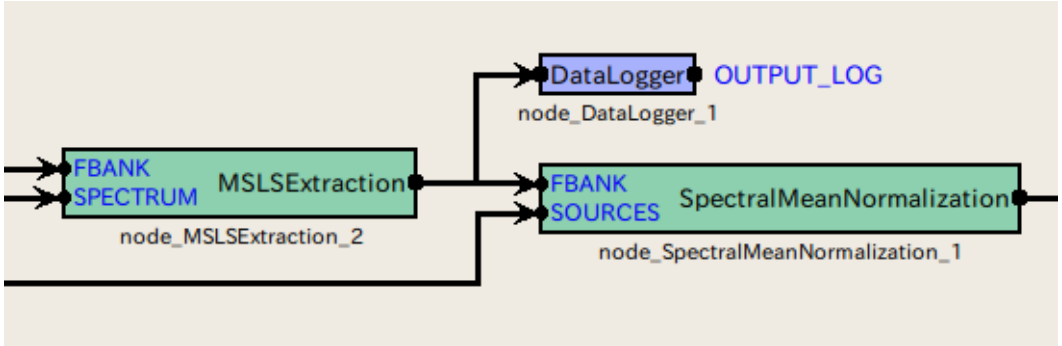


図 6.78: DataLogger の接続例

ノードの入出力とプロパティ

表 6.65: ModuleName のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LABEL	string			出力するデータに付与するラベル

入力

INPUT : any．ただし，サポートする型は，Map<int, float>，Map<int, int> または Map<int, ObjectRef> である．Map<int, ObjectRef> の ObjectRef は， Vector<float> または Vector<complex<float> > のみをサポートしている．

出力

OUTPUT ：入力と同じ．

パラメータ

LABEL ：複数の DataLogger を利用したときにどの DataLogger が出力した結果か分かるように，出力するデータに付与する文字列を指定する．

ノードの詳細

入力されたデータにパラメータで指定したラベルを付与して標準出力またはファイルに出力する．サポートしている型は HARK でよく利用する音源 ID を キーとした `Map` 型のみである．出力される形式は以下の通りである．

ラベル フレームカウント キー 1 値 1 キー 2 値 2 . . .

本ノードの 1 フレームカウントの出力は 1 行で，上記のように最初にパラメータで指定したラベル，次にフレームカウント，その後に `Map` 型のキーと値を全てスペース区切りで出力する．値が `Vector` の時は，すべての要素がスペース区切りで出力される．

6.7.4 MatrixToMap

ノードの概要

Matrix<float> 型や，Matrix<complex<float> > 型のデータを Map<int, ObjectRef> 型に変換する．

必要なファイル

無し．

使用方法

どんなときに使うのか

入力が Map<int, ObjectRef> 型しか受け付けないノード，例えば PreEmphasis，MelFilterBank や SaveRawPCM など，に接続する際に使用する．

典型的な接続例

MatrixToMap ノードの接続例を図 6.79，6.80 に示す．

図 6.79 は，AudioStreamFromMic ノードでマイクロホンから音声波形データを取り込み，ChannelSelector ノードにて必要なチャネルを選別し，MatrixToMap ノードによって Matrix<float> 型データを Map<int, ObjectRef> 型に変換する．その出力を SaveRawPCM ノードに接続し，波形をファイルとして保存する．

図 6.80 は，波形のスペクトルを Map<int, ObjectRef> 型で得たいときの MatrixToMap ノードの使い方である．図のように，MultiFFT ノードを通すのは，MatrixToMap の前でも後でも良い．

ノードの入出力とプロパティ

入力

INPUT : Matrix<float> または Matrix<complex<float> > 型．

出力

OUTPUT : Map<int, ObjectRef> 型．入力データに ID が付与された構造体．

パラメータ

無し．

ノードの詳細

ID の付き方: ID の値は常に 0 となる．

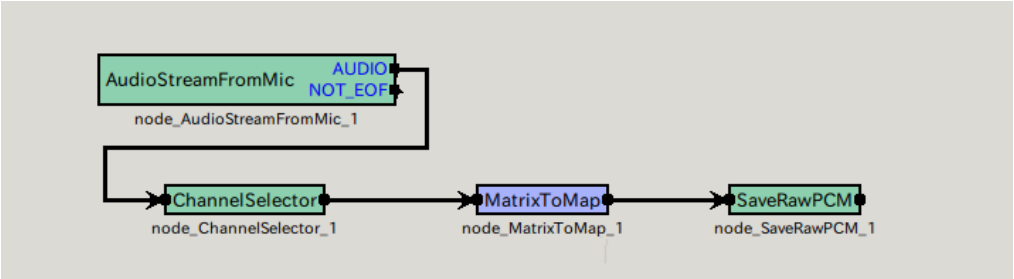


図 6.79: MatrixToMap の接続例 – SaveRawPCM への接続

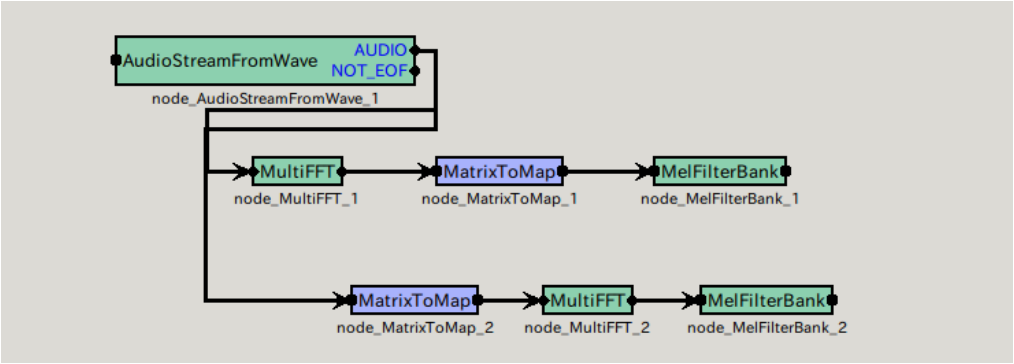


図 6.80: MatrixToMap の接続例 – MultiFFT との接続

6.7.5 MultiDownSampler

ノードの概要

入力信号をダウンサンプリングして出力する．ローパスフィルタは窓関数法を用いており，窓関数はカイザー窓である．

必要なファイル

無い．

使用方法

どんなときに使うのか 入力信号のサンプリング周波数が 16kHz でない場合等に用いる．HARK ノードは基本的にサンプリング周波数を 16kHz と仮定している．そのため，入力信号が 48kHz であったりする場合には，ダウンサンプリングを行ない，サンプリング周波数を 16kHz まで下げる必要がある．

注意 1（ADVANCE の値域）： 処理の都合上，前段の入力ノード，例えば，AudioStreamFromMic や AudioStreamFromWave のパラメータ設定に制限を設ける．それらのパラメータ，LENGTH と ADVANCE の差: OVERLAP = LENGTH - ADVANCE，は十分大きな値でなければならない．より具体的には，このノードのローパスフィルタ長 N より大きな値でなけらばならない．このノードのデフォルトの設定では，おおよそ 120 以上あれば十分であるので，ADVANCE が LENGTH の 4 分の 1 以上なら問題は起きないであろう．また，次の**注意 2** の要求も満す必要がある．

注意 2（ADVANCE 値の設定）： このノードの ADVANCE は，後段ノード（GHDSS など）における ADVANCE 値の SAMPLING_RATE_IN / SAMPLING_RATE_OUT 倍に設定する必要がある．これは仕様であり，これ以外の値に設定したときの動作は保証しない．例えば，後段ノードで ADVANCE = 160 に設定されている場合かつ SAMPLING_RATE_IN / SAMPLING_RATE_OUT = 3 である場合，このノードや前段ノードの ADVANCE は 480 に設定する必要がある．

注意 3（前段ノードでの LENGTH 値の要求）： このノード以前（AudioStreamFromMic など）における LENGTH 値も，後段ノード（GHDSS など）での値の SAMPLING_RATE_IN / SAMPLING_RATE_OUT 倍に設定しておくことを要求する．例えば，SAMPLING_RATE_IN / SAMPLING_RATE_OUT = 3 なら，GHDSS で LENGTH = 512，ADVANCE = 160 に設定されているなら，AudioStreamFromMic では LENGTH = 1536，ADVANCE = 480 に設定するのが望ましい．

典型的な接続例 下記に典型的な接続例を示す．このネットワークファイルは，Wave ファイル入力を読み込み，ダウンサンプルを行ない，Raw ファイルで保存を行なう．Wave ファイル入力は Constant，InputStream，AudioStreamFromMic，を繋ぐことで実現される．その後，MultiDownSampler によって，ダウンサンプリングを実行し，SaveRawPCM で出力波形を保存する．

ノードの入出力とプロパティ

入力

INPUT ：`Matrix<float>` 型. マルチチャネル音声波形データ（時間領域波形）.

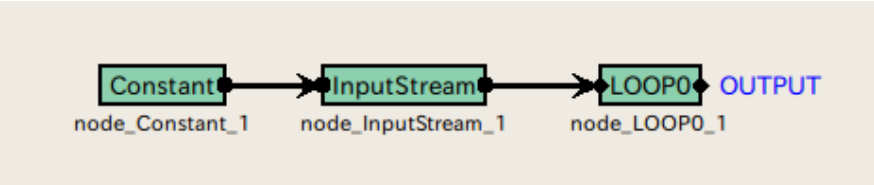


図 6.81: MultiDownSampler の接続例: Main ネットワーク

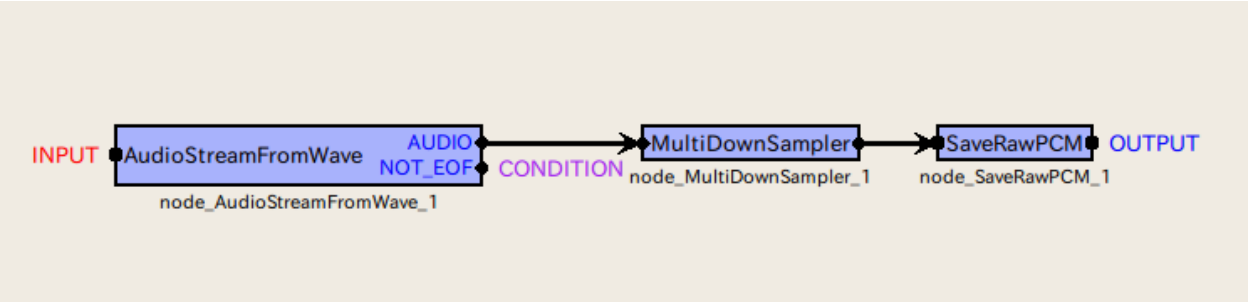


図 6.82: MultiDownSampler の接続例: Iteration(LOOP0) ネットワーク

出力

OUTPUT ：`Matrix<float>` 型．ダウンサンプルされたマルチチャネル音声波形データ（時間領域波形）．

表 6.66: MultiDownSampler のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
ADVANCE	`int`	480	[pt]	INPUT 信号でのイタレーション毎にフレームをシフトさせる長さ．特殊な設定が必要であるので，パラメータ説明を参考にすること．
SAMPLING_RATE_IN	`int`	48000	[Hz]	INPUT 信号のサンプリング周波数．
SAMPLING_RATE_OUT	`int`	16000	[Hz]	OUTPUT 信号のサンプリング周波数．
Wp	`float`	0.28	$[\frac{\omega}{2\pi}]$	ローパスフィルタ通過域端．INPUT を基準とした正規化周波数 [0.0 – 1.0] の値で指定．
Ws	`float`	0.34	$[\frac{\omega}{2\pi}]$	ローパスフィルタ阻止域端．INPUT を基準とした正規化周波数 [0.0 – 1.0] の値で指定．
As	`float`	50	[dB]	阻止域最小減衰量．

パラメータ 各パラメータはローパスフィルタ，ここではカイザー窓の周波数特性を定めるものが多い．図 6.83 に記号とフィルタ特性の関係を示すので，対応関係に注意して読み進めること．

ADVANCE ：`int` 型．480 がデフォルト値．音声波形に対する処理のフレームを，波形の上でシフトする幅をサンプル数で指定する．ただし，INPUT 以前のノードで設定されている値を用いる．**注意:** OUTPUT 以降で設定されている値の SAMPLING_RATE_IN / SAMPLING_RATE_OUT 倍の値に設定する必要がある．

SAMPLING_RATE_IN ：`int` 型．48000 がデフォルト値．入力波形のサンプリング周波数を指定する．

SAMPLING_RATE_OUT ：`int` 型．16000 がデフォルト値．出力波形のサンプリング周波数を指定する．この時，SAMPLING_RATE_IN の整数分の一の値しか対応できないことに注意が必要．

Wp : `float` 型．デフォルト値は 0.28．ローパスフィルタ通過域端周波数を INPUT を基準とした正規化周波数 [0.0 – 1.0] の値によって指定する．入力のサンプリング周波数が 48000 [Hz] で，0.28 の値に設定した場合，約 48000 ∗ 0.28 = 13440 [Hz] から，ローパスフィルタのゲインが減少しはじめる．

Ws : `float` 型．デフォルト値は 0.34．ローパスフィルタ阻止域端周波数を INPUT を基準とした正規化周波数 [0.0 – 1.0] の値によって指定する．入力のサンプリング周波数が 48000 [Hz] で，0.34 の値に設定した場合，約 48000 ∗ 0.34 = 16320 [Hz] から，ローパスフィルタのゲインが安定しはじめる．

As : `float` 型．デフォルト値は 50．阻止域最小減衰量を [dB] で表現した値．デフォルト値を用いた場合，阻止帯域のゲインは通過帯域を 0 とした場合，約 -50 [dB] となる．

ここで，Wp，Ws，As の値をシビアに設定，例えば，Wp，Ws を遮断周波数 Ws 近くに設定するとカイザー窓の周波数特性精度が向上する．しかし，ローパスフィルタの次元が増大し，処理時間の増大を招く．この関係はトレードオフである．

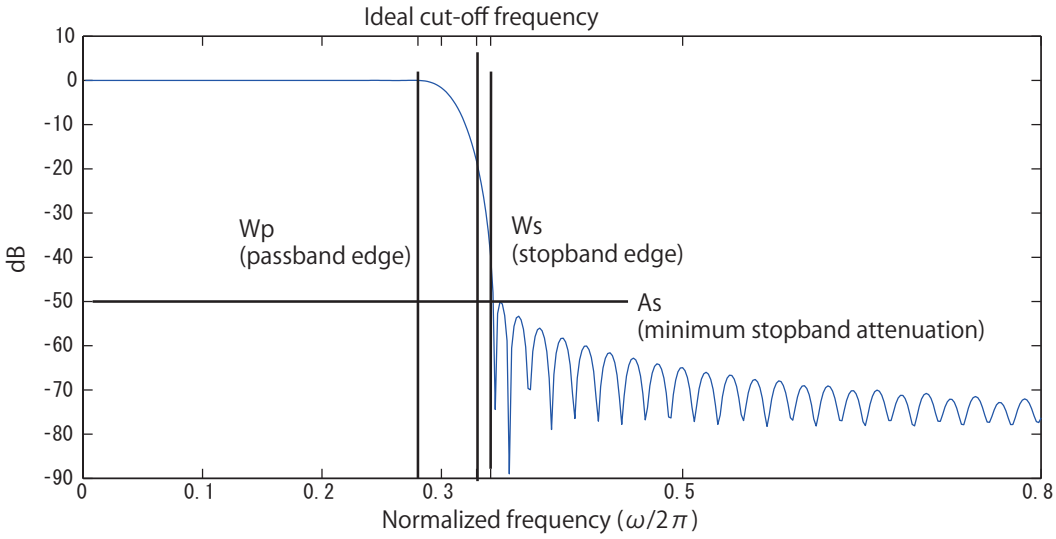


図 6.83: デフォルト設定のフィルタ特性: 横軸 正規化周波数 [$\omega/2\pi$]，縦軸 ゲイン [dB]

ノードの詳細

MultiDownSampler は，マルチチャネル信号をカイザー窓を用いたローパスフィルタによって帯域制限を行い，ダウンサンプリングを行なうノードである．具体的には 1) カイザー窓，2) 理想低域応答，の合成によって FIR ローパスフィルタを作成・実行した後，SAMPLING_RATE_OUT/SAMPLING_RATE_IN のダウンサンプルを行なう．

FIR フィルタ: 有限インパルス応答 $h(n)$ を用いたフィルタリングは次式によって行なわれる．

$$s_{\mathrm{out}}(t) = \sum_{i=0}^{N} h(n) s_{\mathrm{in}}(t-n) \tag{6.121}$$

ここで，$s_{\mathrm{out}}(t)$ は出力信号，$s_{\mathrm{in}}(t)$ は入力信号である．

マルチチャネル信号の場合は，各チャネルの信号に対して独立にフィルタリングが行なわれる．この時，用いる有限インパルス応答 $h(n)$ は同一のものである．

理想低域応答: 遮断周波数が ω_c である理想低域応答は以下の式によって定められる．

$$H_i(e^{j\omega}) = \begin{cases} 1, & |\omega| < \omega_c \\ 0, & \text{otherwise} \end{cases} \tag{6.122}$$

このインパルス応答は,

$$h_i(n) = \frac{\omega_c}{\pi}\left(\frac{sin(\omega n)}{\omega n}\right), \quad -\infty \leq n \leq \infty \tag{6.123}$$

となる．このインパルス応答は非因果的かつ有界入力有界出力（BIBO: bounded input bounded output）安定条件を満さない．

この理想フィルタから FIR フィルタを得るには，インパルス応答を途中で打ち切る．

$$h(n) = \begin{cases} h_i(n), & |n| \leq \frac{N}{2} \\ 0, & \text{otherwise} \end{cases} \tag{6.124}$$

ここで N はフィルタの次数である．このフィルタはインパルス応答の打ち切りによって，通過域と阻止域にはリプルが発生する．また，阻止域最小減衰量 As も約 21 dB に止まり，十分な減衰量を得ることができない．

カイザー窓を用いた窓関数法によるローパスフィルタ: 上述の打ち切り法による特性を改善するため，理想インパルス応答 $h_i(n)$ に窓関数 $v(n)$ を掛けた，次式のインパルス応答を代りに用いる．

$$h(n) = h_i(n)v(n) \tag{6.125}$$

ここではカイザー窓を用いて，ローパスフィルタを設計する．カイザー窓は次式によって定義される．

$$v(n) = \begin{cases} \frac{I_0\left(\beta\sqrt{1-(nN/2)^2}\right)}{I_0(\beta)}, & -\frac{N}{2} \leq n \leq \frac{N}{2} \\ 0, & \text{otherwise} \end{cases} \tag{6.126}$$

ここで，β は窓の形状を定めるパラメータ，$I_0(x)$ は 0 次の変形ベッセル関数であり，

$$I_0(x) = 1 + \sum_{k=1}^{\infty}\left(\frac{(0.5x)^k}{k!}\right) \tag{6.127}$$

から得られる．

パラメータ β は低域通過フィルタで求められる減衰量に応じて決まる．ここでは下記の指標によって定める．

$$\beta = \begin{cases} 0.1102(As - 8.7) & As > 50, \\ 0,5842(As - 21)^{0.4} + 0.07886(As - 21) & 21 < As < 50, \\ 0 & As < 21 \end{cases} \tag{6.128}$$

残りはフィルタ次数と遮断周波数 ω_c を定めれば，窓関数法によってローパスフィルタを実現できる．フィルタ次数 N は，阻止域最小減衰量 As と遷移域 $\Delta f = (Ws - Wp)/(2\pi)$ を用いて,

$$N \approx \frac{As - 7.95}{14.36\Delta f} \tag{6.129}$$

と見積もる．また，遮断周波数 ω_c を $0.5(Wp + Ws)$ と設定する．

ダウンサンプリング: ダウンサンプリングは，ローパスフィルタを通過させた信号から SAMPLING_RATE_IN / SAMPLING_RATE_OUT のサンプル点を間引くことによって実現される．例えば，デフォルトの設定では 48000/16000 = 3 であるから，入力サンプルを 3 回に 1 回 取り出し，出力サンプルとすれば良い．

参考文献:

(1) 著: P. Vaidyanathan, 訳: 西原 明法, 渡部 英二, 吉田 俊之, 杉野 暢彦: “マルチレート信号処理とフィルタバンク”, 科学技術出版, 2001.

6.7.6 MultiFFT

ノードの概要

マルチチャネル音声波形データに対し，高速フーリエ変換 (Fast Fourier Transformation: FFT) を行う．

必要なファイル

無し．

使用方法

どんなときに使うのか

このノードは，マルチチャネル音声波形データを，スペクトルに変換して時間周波数領域で解析を行いたいときに用いる．音声認識に用いる特徴抽出の前処理として用いられることが多い．

典型的な接続例

図 6.84 で，MultiFFT ノードに `Matrix<float>`，`Map<int, ObjectRef>` 型の入力を与える例を示す．

図 6.84 の上のパスは AudioStreamFromWave ノードから `Matrix<float>` 型の多チャネル音響信号を受け取り，MultiFFT ノードで `Matrix<complex<float> >` 型の複素スペクトルに変換したのち，LocalizeMUSIC ノードに入力される．

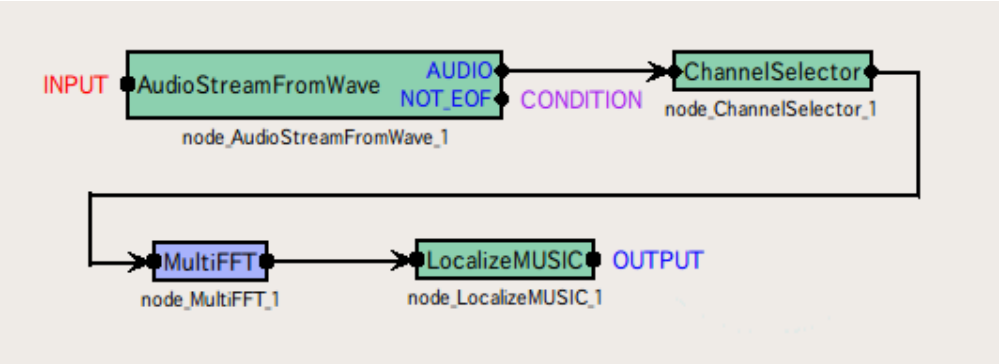


図 6.84: MultiFFT の接続例

ノードの入出力とプロパティ

入力

INPUT ：型は `Matrix<float>` または `Map<int, ObjectRef>`．マルチチャネル音声波形データ．`Map<int, ObjectRef>` の 部 分は `Vector<float>` 型．行列のサイズが $M \times L$ のとき，M がチャネル数，L が波形のサンプル数を表す．L は，パラメータ LENGTH と値が等しい必要がある．

出力

表 6.67: MultiFFT のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	フーリエ変換を適用する信号の長さ
WINDOW	string	CONJ		フーリエ変換を行う際の窓関数の種類．CONJ，HAMMING，RECTANGLE から選択する．それぞれ，複素窓，ハミング窓，矩形窓を示す．
WINDOW_LENGTH	int	512	[pt]	フーリエ変換を行う際の窓関数の長さ．

OUTPUT ：型は Matrix<complex<float> > または Map<int, ObjectRef>．入力に対応したマルチチャネル複素スペクトル．入力が Matrix<float> 型のとき，出力は Matrix<complex<float> > 型となり，入力が Map<int, ObjectRef> 型のとき，出力は Map<int, ObjectRef> 型となる．部 分は Vector<complex<float> > 型となる．入力行列のサイズが $M \times L$ のとき，出力行列のサイズは，$M \times L/2 + 1$ となる．

パラメータ

LENGTH ：型は int．デフォルト値は 512．フーリエ変換を適用する信号の長さを指定する．アルゴリズムの性質上，2 のべき乗の値をとる．また，WINDOW_LENGTH より大きい値にする必要がある．

WINDOW ：型は string．デフォルト値は CONJ．CONJ，HAMMING，RECTANGLE から選択する．それぞれ，複素窓，ハミング窓，矩形窓を意味する．音声信号の解析には，HAMMING 窓がよく用いられる．

WINDOW_LENGTH ：型は int．デフォルト値は 512．窓関数の長さを指定する．値を大きくすると，周波数解像度は増す半面，時間解像度は減る．直感的には，この値を増やすと，音の高さの違いに敏感になるが，音の高さの変化に鈍感になる．

ノードの詳細

LENGTH，WINDOW_LENGTH の目安: 音声信号の解析には，20 ~ 40 [ms] に相当する長さのフレームで分析するのが適当である．サンプリング周波数を f_s [Hz]，窓の時間長を x [ms] とすると，フレーム長 L [pt] は，

$$L = \frac{f_s x}{1000}$$

で求められる．
例えば，サンプリング周波数が 16 [kHz] のとき，デフォルト値の 512 [pt] は，32 [ms] に相当する．パラメータ LENGTH は，高速フーリエ変換の性質上，2 の累乗の値が適しているため，512 を選ぶ．
　より音声の解析に適したフレームの長さを指定するため，窓関数の長さ WINDOW_LENGTH は，400 [pt] (サンプリング周波数が 16 [kHz] のとき，25 [ms] に相当) に設定することもある．

各窓関数の形: 各窓関数 $w(k)$ の形は次の通り．k はサンプルのインデックス，L は窓関数の長さ，FFT 長を $NFFT$ とし，k は $0 \leq k < L$ の範囲を動く．FFT 長が窓の長さよりも大きいとき，$NFFT \leq k < L$ における窓関数の値には，0 が埋められる．

CONJ，複素窓:

$$w(k) = \begin{cases} 0.5 - 0.5\cos\left(\frac{4k}{L}C\right), & \text{if} 0 \le k < L/4 \\ \sqrt{1 - \left\{0.5 - 0.5\cos\left(\frac{2L-4k}{L}C\right)\right\}^2}, & \text{if} L/4 \le k < 2L/4 \\ \sqrt{1 - \left\{0.5 - 0.5\cos\left(\frac{4k-2L}{L}C\right)\right\}^2}, & \text{if} 2L/4 \le k < 3L/4 \\ 0.5 - 0.5\cos\left(\frac{4L-4k}{L}C\right), & \text{if} 3L/4 \le k < L \\ 0, & \text{if} NFFT \le k < L \end{cases},$$

ただし，$C = 1.9979$ である．

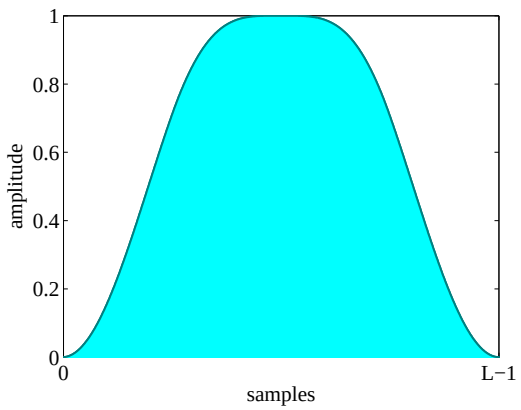


図 6.85: 複素窓関数の形状

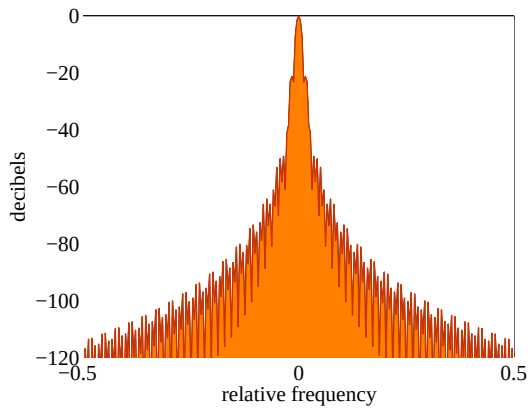


図 6.86: 複素窓関数の周波数応答

図 6.85 と図 6.86 はそれぞれ，複素窓関数の形状および周波数応答である．図 6.86 における横軸はサンプリング周波数に対して，相対的な周波数の値を意味する．一般に，窓関数の周波数応答は，横軸が 0 におけるピークが鋭い方が性能が良いとされる．縦軸の値は，フーリエ変換などの周波数解析を行ったとき，ある周波数ビンに他の周波数成分のパワーが漏れてくる量を表す．

HAMMING，ハミング窓:

$$w(k) = \begin{cases} 0.54 - 0.46\cos\frac{2\pi k}{L-1}, & \text{if } 0 \le k < L, \\ 0, & \text{if } L \le k < NFFT \end{cases}$$

ただし，π は円周率を表す．

図 6.87，6.87 はそれぞれ，ハミング窓関数の形状と周波数応答である．

RECTANGLE，矩形窓:

$$w(k) = \begin{cases} 1, & \text{if } 0 \le k < L \\ 0, & \text{if } L \le k < NFFT \end{cases}$$

図 6.89，6.89 はそれぞれ，矩形窓関数の形状と周波数応答である．

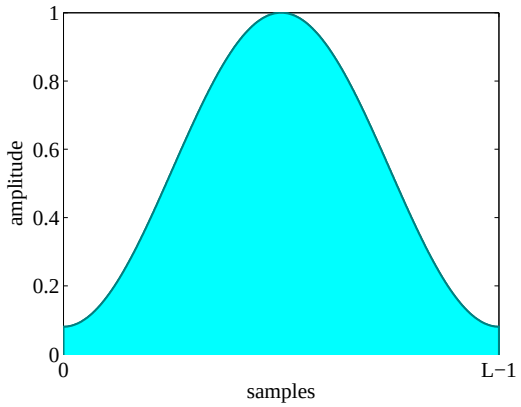


図 6.87: ハミング窓関数の形状

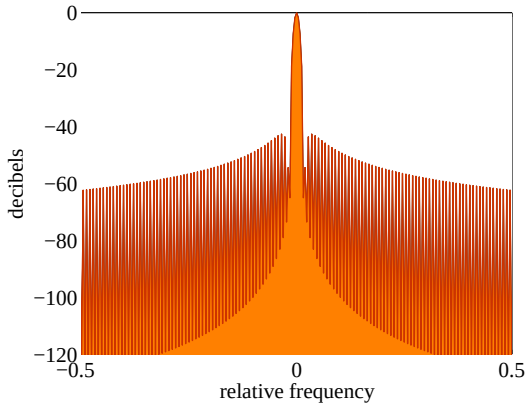


図 6.88: ハミング窓関数の周波数応答

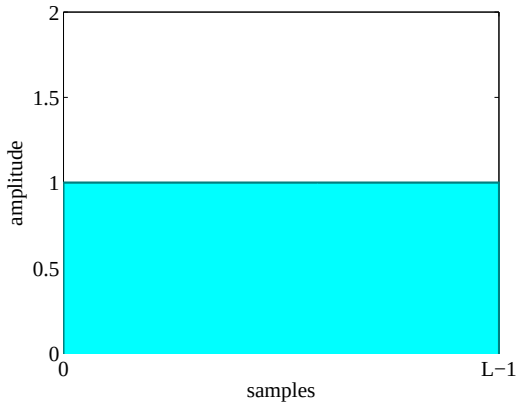


図 6.89: 矩形窓関数の形状

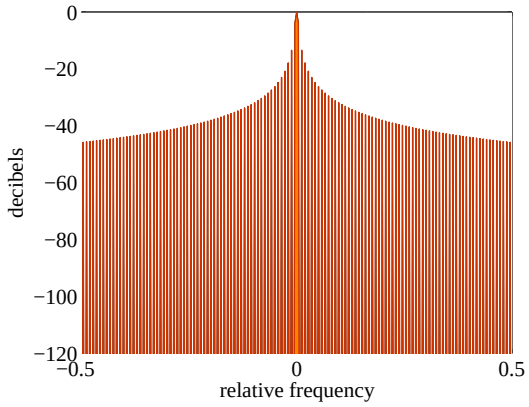


図 6.90: 矩形窓関数の周波数応答

6.7.7 MultiGain

ノードの概要

入力信号のゲインを調節する．

必要なファイル

無し．

使用方法

どんなときに使うのか

主に入力信号をクリップしないよう，もしくは，増幅する場合に使用する．例えば，入力に 24 [bit] で量子化された音声波形データを用いる場合，16 [bit] を仮定して構築したシステムを用いる場合には，このノードを利用して，8 [bit] 分，ゲインを落とすなどといった用途に用いる．

典型的な接続例

AudioStreamFromMic や AudioStreamFromWave の直後に直接配置するか，ChannelSelector を間に挟んで配置することが多い．

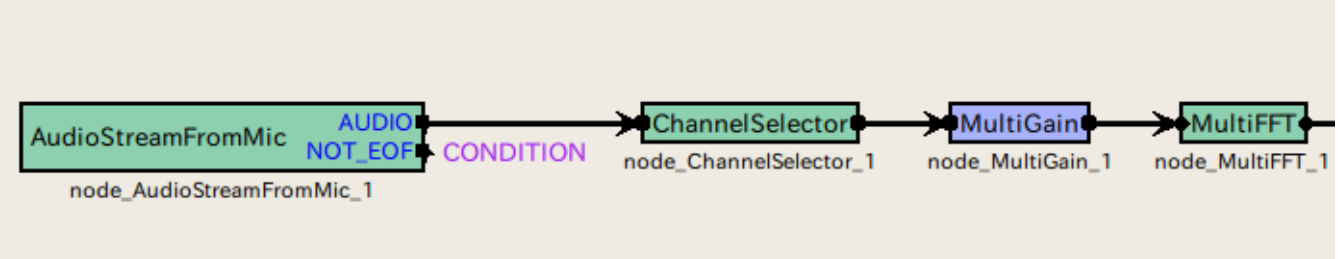


図 6.91: MultiGain の接続例

ノードの入出力とプロパティ

表 6.68: MultiGain のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
GAIN	`float`	1.0		ゲイン値

入力

INPUT : `Matrix<float>` 型．マルチチャネル音声波形データ（時間領域波形）．

出力

OUTPUT ：`Matrix<float>` 型．ゲイン調節されたマルチチャネル音声波形データ（時間領域波形）．

パラメータ

GAIN ：`float` 型．ゲインパラメータ．1.0 で，入力をそのまま出力することに相当する．

ノードの詳細

入力の各チャネルが GAIN パラメータで指定した値を乗じた値となって出力される．使用時は時間領域波形の入力を仮定していることに注意．

例えば，40 dB ゲインを落としたい場合には，下記のような計算を行い，0.01 を指定すればよい．

$$20 \log x = -40 \tag{6.130}$$

$$x = 0.01 \tag{6.131}$$

6.7.8 PowerCalcForMap

ノードの概要

Map<int, ObjectRef> 型の ID 付きマルチチャネル複素スペクトルを，実パワー（または振幅）スペクトルに変換する．

必要なファイル

無し．

使用方法

どんなときに使うのか

複素スペクトルを実パワー（または振幅）スペクトルに変換したいときに用いる．入力が Map<int, ObjectRef> 型のときはこのノードを用いる．入力が Matrix<complex<float> > 型の時は，PowerCalcForMatrix ノードを用いる．

典型的な接続例

図 6.92 に PowerCalcForMap ノードの使用例を示す．MultiFFT ノードから得られた Map<int, ObjectRef> 型複素スペクトルを，Map<int, ObjectRef> 型のパワースペクトルに変換したのち，MelFilterBank ノードに入力している．

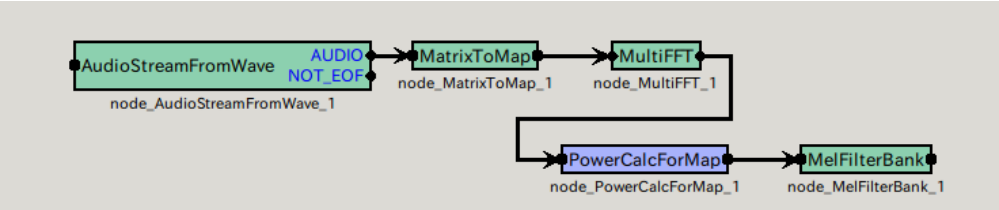


図 6.92: PowerCalcForMap の接続例

ノードの入出力とプロパティ

表 6.69: PowerCalcForMap のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
POWER_TYPE	string	POW		パワーか振幅かの選択

入力

INPUT : Map<int, ObjectRef> 型．ObjectRef 部分に，Matrix<complex<float> > 型の複素行列が格納されている．

出力

OUTPUT : `Map<int, ObjectRef>` 型．`ObjectRef` 部分に，入力の複素行列の各要素について，パワー（または絶対値）を取った実行列が格納されている．

パラメータ

POWER_TYPE : パワースペクトル (POW) か振幅スペクトル (MAG) かの選択．

ノードの詳細

入力の複素行列 $M_{i,j}$ (i, j はそれぞれ，行，列のインデックス) に対して，出力の実行列 $N_{i,j}$ は次のように求める．

$$
\begin{aligned}
N_{i,j} &= M_{i,j}M^*_{i,j} \ \text{(if POWER_TYPE=POW)}, \\
N_{i,j} &= abs(M_{i,j}) \ \text{(if POWER_TYPE=MAG)},
\end{aligned}
$$

ただし，$M^*_{i,j}$ は，$M_{i,j}$ の複素共役を表す．

6.7.9 PowerCalcForMatrix

ノードの概要

Matrix<complex<float> > 型のマルチチャネル複素スペクトルを，実パワー（または振幅）スペクトルに変換する．

必要なファイル

無し．

使用方法

どんなときに使うのか

複素スペクトルを実パワー（または振幅）スペクトルに変換したいときに用いる．入力が Matrix<complex<float> > 型のときはこのノードを用いる．入力が Map<int, ObjectRef> 型のときは PowerCalcForMap ノードを用いる．

典型的な接続例

図 6.93 に PowerCalcForMatrix ノードの使用例を示す．MultiFFT ノードから得られた Matrix<complex<float> > 型複素スペクトルを，Matrix<float> 型のパワースペクトルに変換したのち，BGNEstimator ノードに入力している．

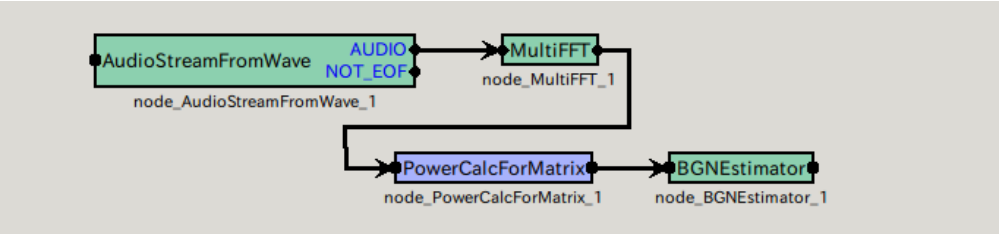


図 6.93: PowerCalcForMatrix の接続例

ノードの入出力とプロパティ

表 6.70: PowerCalcForMatrix のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
POWER_TYPE	string	POW		パワーか振幅かの選択

入力

INPUT : Matrix<complex<float> > 型．各要素が複素数の行列．

出力

OUTPUT ：`Matrix<float>` 型．入力の各要素のパワー（または絶対値）を取った実行列．

パラメータ

POWER_TYPE ：パワースペクトル (POW) か振幅スペクトル (MAG) かの選択

ノードの詳細

入力の複素行列 $M_{i,j}$ (i, j はそれぞれ行，列のインデックス) に対して，出力の実行列 $N_{i,j}$ は次のように求める．

$$
\begin{aligned}
N_{i,j} &= M_{i,j}M^*_{i,j} \ \text{(if POWER_TYPE=POW)}, \\
N_{i,j} &= abs(M_{i,j}) \ \text{(if POWER_TYPE=MAG)},
\end{aligned}
$$

ただし，$M^*_{i,j}$ は，$M_{i,j}$ の複素共役を表す．

6.7.10 SegmentAudioStreamByID

ノードの概要

ID 情報を利用した音響ストリームを切り出し，ID 情報を付加した出力を行う．

必要なファイル

なし

使用方法

どんなときに使うのか

音響信号全体を一つのストリームとして処理するのではなく，音声部分だけなど，ある部分のみ切り出して処理を際に有用なノードである．ID をキーとして切り出しを行うので，入力には ID 情報が必須である．同じ ID が続く区間を信号の区間として切り出し，ID 情報を付加して，一チャネルの `Map` データとして出力する．

典型的な接続例

入力ストリームが 2 つの音声信号の混合音であると仮定する．ユーザが GHDSS などを用いて分離した信号と時間的に同じ区間のオリジナルの混合音を比較したい場合，1 ch の音響ストリームと音源定位によって検出した音源をこのノードに入力する．この際に，出力は， GHDSS や PostFilter 分離音と完全に同じフォーマット (`Map`) で出力される．

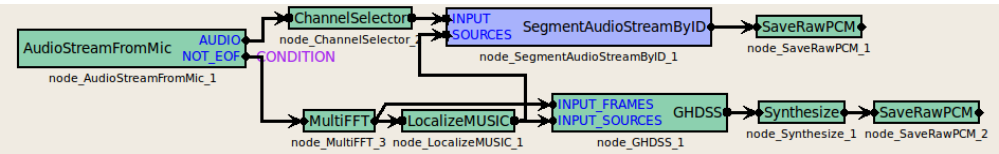


図 6.94: SegmentAudioStreamByID の接続例

ノードの入出力とプロパティ

入力

INPUT : `any` , `Matrix<complex<float> >` や `Matrix<float>` など．

SOURCES : `Vector<ObjectRef>` , ID 付きの音源方向．各 Vector の中身は，ID 付きの音源情報を示す Source 型になっている．特徴量ベクトルが各音源毎に格納される．このパラメータの指定は必須である．

出力

OUTPUT : `Map<int, ObjectRef>` , `ObjectRef` は，`Vector<float>` , `Vector<complex<float> >` へのスマートポインタである．

ノードの詳細

ID 情報を利用した音響ストリームを切り出し，ID 情報を付加した出力を行う．このノードは，Matrix<complex<float>> や Matrix<float> を入力で与えられた ID を用いて Map<int, ObjectRef> に変換する現状では入力として 1 ch データしかサポートしていないことに注意.

6.7.11 SourceSelectorByDirection

ノードの概要

入力された音源定位結果のうち，指定した水平方向の角度の範囲にあるもののみを通過させる，フィルタリングノード．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源の方向に関する事前情報があるとき (前方にしか音源は無いと分かっている場合など) にその方向のみの定位結果を得る場合に使う．あるいは，ノイズ源の方向が分かっているときに，その方向以外を指定すれば，ノイズの定位結果を除去することも可能である．

典型的な接続例

主に，ConstantLocalization，LoadSourceLocation，LocalizeMUSIC などの音源定位結果を接続する．

図 6.95 に示す接続例は，音源定位結果のログファイルのうち，指定した範囲の方向のみを取り出すネットワークである．

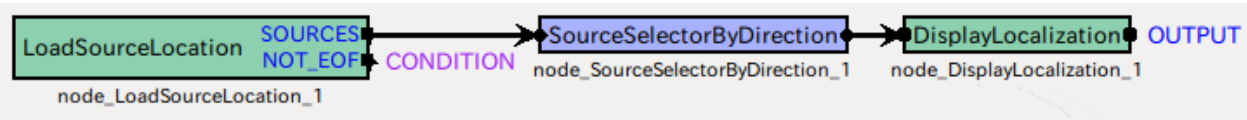


図 6.95: SourceSelectorByDirection の接続例: これは Iterator サブネットワーク

ノードの入出力とプロパティ

入力

SOURCES ： `Vector<ObjectRef>` 型．入力となる音源定位結果を接続する．`ObjectRef` が参照するのは，`Source` 型のデータである．

出力

OUTPUT ：`Vector<ObjectRef>` 型． フィルタリングされた後の音源定位結果を意味する．`ObjectRef` が参照するのは，`Source` 型のデータである．

パラメータ

表 6.71: SourceSelectorByDirection のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
MIN_AZIMUTH	`float`	-20.0	[deg]	通過させる音源方向の最小値
MAX_AZIMUTH	`float`	20.0	[deg]	通過させる音源方向の最大値

MIN_AZIMUTH，MAX_AZIMUTH ： `float` 型．角度は通過させる音源の左右方向 (方位角) を表す．角度 θ[deg] が $\theta \in [-180, 180]$ の範囲に無ければ, 範囲に入るように 360 を加算/減算される。

ノードの詳細

ビームフォーマなどのマイクロホンアレイ信号処理による空間フィルタリングをするわけではなく，あくまで定位結果の音源方向の情報を元にフィルタリングを行う．

6.7.12 SourceSelectorByID

ノードの概要

複数の音源分離結果のうち，ID が指定した値以上のものだけを出力させたいときに用いる．特に，GHDSS ノードの FIXED_NOISE プロパティを `true` にした場合は，定常ノイズ分離結果に負の ID が振られるので，それ以外の音を処理するためのフィルタとして使用する．

必要なファイル

無し．

使用方法

どんなときに使うのか

音源分離ノード GHDSS は，ロボットの電源を入れるが移動はさせない．すなわちノイズ (ファンの音など) が定常かつ既知，という条件下で音源分離をする場合，そのノイズの分離結果の ID を-1 として出力する．このとき，GHDSS ノードの後に SourceSelectorByID を接続し，MIN_ID を 0 に設定すると，定常ノイズの分離音を以後の処理で無視することができる．

典型的な接続例

図 6.96 に接続例を示す．図に示すように，このノードは，典型的には GHDSS の後段に接続される．

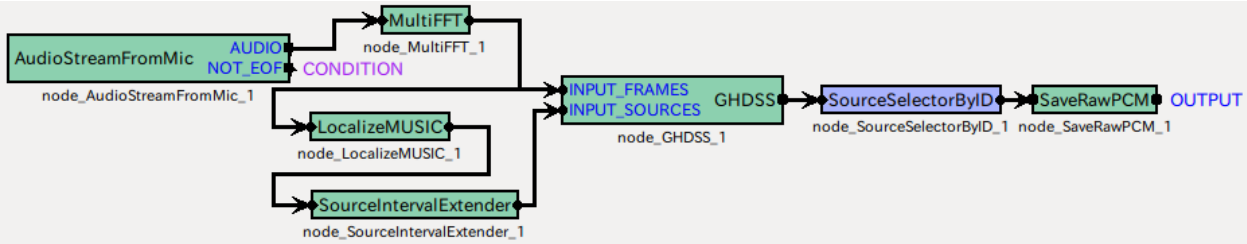


図 6.96: SourceSelectorByID の接続例: これは Iterator サブネットワーク

ノードの入出力とプロパティ

入力

INPUT : `Map<int, ObjectRef>` 型．通常は音源分離ノードの後段に接続されるので，`Map` のキーになる `int` には音源 ID が対応する．`ObjectRef` は分離を表す `Vector<float>` 型 (パワースペクトル) か `Vector<complex<float>>` 型 (複素スペクトル) である．

出力

OUTPUT : `Map<int, ObjectRef>` 型．音源 ID が MIN_ID より大きいデータだけを抽出したデータが出力される．`Map` の内容は INPUT と同じになる．

パラメータ

表 6.72: SourceSelectorByID のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
MIN_ID	`int`	0		この値より大きい ID の音源は通す．

MIN_ID : `int` 型．このパラメータ値以上の音源 ID を持つ分離音を通過．デフォルト値は 0 ．GHDSS の後段に接続するのであれば変更不要．

6.7.13 Synthesize

ノードの概要

周波数領域の信号を時間領域の波形に変換する.

必要なファイル

無し.

使用方法

周波数領域の信号を時間領域の波形に変換する際に用いる.

典型的な接続例

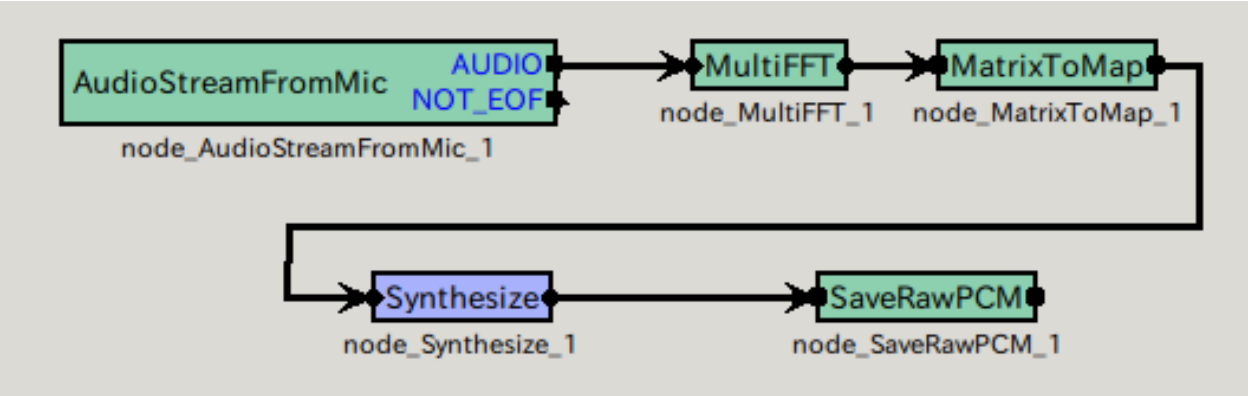


図 6.97: Synthesize の接続例

ノードの入出力とプロパティ

表 6.73: Synthesize のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	FFT 長
ADVANCE	int	160	[pt]	シフト長
SAMPLING_RATE	int	16000	[Hz]	サンプリングレート
MIN_FREQUENCY	int	125	[Hz]	最小周波数
MAX_FREQUENCY	int	7900	[Hz]	最大周波数
WINDOW	string	HAMMING		窓関数
OUTPUT_GAIN	float	1.0		出力ゲイン

入力

INPUT : `Map<int, ObjectRef>` 型． `ObjectRef` は `Vector<complex<float> >` .

出力

OUTPUT : `Map<int, ObjectRef>` 型．`ObjectRef` は `Vector<float>` .

パラメータ

LENGTH FFT 長，他のノード（ MultiFFT ）と値を合わせる必要がある．

ADVANCE シフト長，他のノード（ MultiFFT ）と値を合わせる必要がある．

SAMPLING_RATE サンプリングレート，他のノードと値を合わせる必要がある．

MIN_FREQUENCY 波形生成時に用いる最小周波数値

MAX_FREQUENCY 波形生成時に用いる最大周波数値

WINDOW 窓関数，HAMMING，RECTANGLE, CONJ から選択

OUTPUT_GAIN 出力ゲイン

ノードの詳細

入力された周波数領域の信号に対して，低域 4 バンド分，および，$\omega_s/2 - 100$ [Hz] 以上の周波数ビンについては 0 を代入したのち逆 FFT を適用する．次に，指定された窓をかけ，overlap-add 処理を行う．overlap-add 処理は，フレーム毎に逆変換を行い，時間領域の戻した信号をずらしながら加算することにより，窓の影響を軽減する手法である．詳細は，参考文献で挙げている web ページを参照すること．最後に，得られた時間波形に出力ゲインを乗じて，出力する．

なお，overlap-add 処理を行うために，フレームの先読みをする必要があり，結果として，このノードは処理系全体に遅延をもたらす．遅延の大きさは，下記で計算できる．

$$delay = \begin{cases} |\mathrm{LENGTH}/\mathrm{ADVANCE}| - 1, & \text{if } \mathrm{LENGTH}\ mod\ \mathrm{ADVANCE} \neq 0, \\ \mathrm{LENGTH}/\mathrm{ADVANCE}, & \text{otherwise.} \end{cases} \tag{6.132}$$

HARK のデフォルトの設定では，LENGTH = 512, ADVANCE = 160 であるので，遅延は 3 [frame]，つまり，システム全体に与える遅延は 30 [ms] となる．

参考文献

(1) http://en.wikipedia.org/wiki/Overlap-add

6.7.14 WhiteNoiseAdder

ノードの概要

入力信号に白色ノイズを付加する．

必要なファイル

無し．

使用方法

分離後の非線形歪みの影響を緩和するために敢えてノイズを付加する場合に用いる．例えば，PostFilter は，非線形処理を行うため，musical ノイズの発生を避けることは難しい．このようなノイズは，音声認識性能に大きく影響する場合がある．適量の既知ノイズを加えることにより，このようなノイズの影響を低減できることが知られている．

典型的な接続例

例を図示．具体的なノード名をあげる．

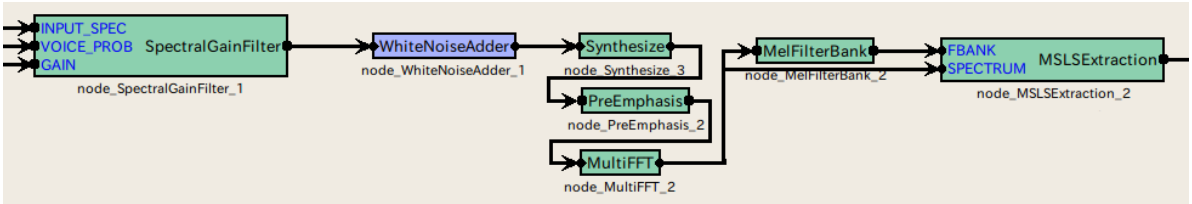


図 6.98: WhiteNoiseAdder の接続例

ノードの入出力とプロパティ

表 6.74: WhiteNoiseAdder のパラメータ表

パラメータ名	型	デフォルト値	単位	説明
LENGTH	int	512	[pt]	FFT 長
WN_LEVEL	float	0		付加ノイズレベル

入力

INPUT : Map<int, ObjectRef> 型． ObjectRef は Vector<complex<float> > であるため，周波数領域の信号を入力することが前提である．

出力

OUTPUT : Map<int, ObjectRef> 型．ObjectRef は Vector<complex<float> > である．白色ノイズが付加された信号が出力される．

パラメータ

LENGTH ：FFT 長，他のノードと値を合わせる必要がある．

WN_LEVEL ：付加ノイズレベル，時間領域での最大振幅値を指定．

ノードの詳細

入力信号の各周波数ビンに対して，

$$\frac{\sqrt{\mathrm{LENGTH}}}{2} \cdot \mathrm{WN_LEVEL} \cdot e^{2\pi jR} \tag{6.133}$$

を加算する．R は，$0 \leq R \leq 1$ となる乱数である (各周波数ビンごとに異なる値となる)． $\sqrt{\mathrm{LENGTH}}/2$ は，FFT による周波数解析の際に生じる時間領域と周波数領域のスケーリングのずれを補正するための項である．

6.8 Flow Designer に依存しないモジュール

6.8.1 JuliusMFT

概要

JuliusMFT は，大語彙音声認識システム Julius を HARK 用に改造を行った音声認識モジュールである．HARK 0.1.x 系では，大語彙音声認識システム Julius 3.5 をもとに改良されたマルチバンド版 Julius[1] に対するパッチとして，提供していたが，HARK 1.0 では，Julius 4.1 系をベースに実装および機能を大きく見直した．HARK 1.0 の JuliusMFT では，オリジナルの Julius と比較して，下記の 4 点に対応するための変更を行っている．

- ミッシングフィーチャー理論の導入
- MSLS 特徴量のネットワーク入力 (mfcnet) 対応
- 音源情報 (SrcInfo) の追加に対応
- 同時発話への対応（排他処理）

実装に関しては，Julius 4.0 から導入されたプラグイン機能を用いて極力 Julius 本体に変更を加えない形で，実装を行った，

インストール方法，使用方法を説明するとともに，Julius との違い，FlowDesigner 上の HARK モジュールとの接続について，解説する．

起動と設定

JuliusMFT の実行は，例えば設定ファイル名を julius.conf とすれば，以下のように行う．

```
> julius_mft -C julius.jconf
> julius_mft.exe -C julius.jconf (Windows 版)
```

HARK では，JuliusMFT を起動したのち，IP アドレスやポート番号が正しく設定された SpeechRecognition-Client (または SpeechRecognitionSMNClient) を含んだネットワークを起動することにより，JuliusMFT とのソケット接続が行われ，音声認識が可能な状態となる．

上述の julius.jconf は JuliusMFT の設定を記述したテキストファイルである．設定ファイルの中身は，基本的に “-” で始まる引数オプションからなっており，起動時に直接，julius のオプションとして引数指定することも可能である．また，# 以降はコメントとして扱われる．Julius で用いるオプションは，http://julius.sourceforge.jp/juliusbook/ja/desc_option.html にまとめられているので，そちらを参照していただきたいが，最低限必要な設定は，以下の 7 種類である．

- -notypecheck
- -plugindir /usr/lib/julius_plugin
- -input mfcnet
- -gprune add_mask_to_safe
- -gram grammar
- -h hmmdefs
- -hlist allTriphones

[1] http://www.furui.cs.titech.ac.jp/mband_julius/

- **-notypecheck**
  特徴パラメータの型チェックをスキップする設定．オリジナルの Julius では任意で指定可能なオプションであるが，JuliusMFT では指定が必須のオプションとなっている．このオプションを指定しないとタイプチェックが行われるが，JuliusMFT のプラグインでは，特徴量と共にマスクデータを計算しているため (マスクなしの場合でも 1.0 を出力する)，タイプチェックでサイズ不一致となり，認識が行われない．
- **-plugindir プラグインディレクトリ名**
  プラグイン (*.jpi) が存在するディレクトリを指定する．引数にカレントディレクトリからの相対パス，もしくはプラグインの絶対パスを指定する．このパスは，apt-get でインストールした場合には，/usr/lib/julius_plugin，ソースコードをパス指定せずコンパイルおよびインストールした場合には/usr/local/lib/julius_plugin がデフォルトとなる．なお，このパスは，-input mfcnet や，-gprune add_mask_to_safe などプラグインで実現されている機能の指定よりも前に指定する必要がある．このパス下にある拡張プラグインファイルは，実行時に全て読み込まれるので注意．尚、Windows 版においては、プラグインを使用しないためプラグインディレクトリ名は任意となるが、本オプションは必須となる．
- **-input mfcnet**
  -input 自体はオリジナル Julius で実装されているオプションで，マイクロホン，ファイル，ネットワーク経由の入力などがサポートされている．JuliusMFT では，SpeechRecognitionClient (または，SpeechRecognitionSMNClient) から送信される音響特徴量とマスクをネットワーク経由で受信できるようにこのオプションを拡張し，音声入力ソースとして mfcnet を指定できるようにした．この機能は -input mfcnet と指定することにより，有効にすることができる．また，mfcnet 指定時のポート番号は，オリジナルの Julius で，音声入力ソース adinnet 用ポート番号を指定するために使用される-adport を用いて “-adport ポート番号” のように指定することができる．
- **-gprune**
  既存の出力確率計算にマスクを利用する場合に使用する枝刈りアルゴリズムを指定する．基本的に，HARK 0.1.x で提供していた julius_mft(ver3.5) に搭載された機能を移植したもので，{*add_mask_to_safe*||*add_mask_to_heu*||*add_m*, *ask_to_beam*||*add_mask_to_none*} の 4 種類からアルゴリズムを選択する (指定しない場合はデフォルトの計算方法となる)．それぞれオリジナル Julius の {*safe*||*heuristic*||*beam*||*none*} に対応している．なお，julius_mft(ver3.5) の eachgconst を用いた計算方法は厳密には正確ではないため，オリジナルと比較すると計算結果 (score) に誤差が出てしまっていた．今回，オリジナルと同様の計算方法を取り入れ，この誤差問題を解決している．
- **-gram grammar**
  言語モデルを指定する．オリジナル Julius と同様.
- **-h hmmdefs**
  音響モデル (HMM) を指定する．オリジナル Julius と同様.
- **-hlist allTriphones**
  HMMList ファイルを指定する．オリジナル Julius と同様.

なお，後述のモジュールモードで利用する際には，オリジナル Julius と同様に-module オプションを指定する必要がある．

詳細説明

mfcnet 通信仕様

mfcnet を音声入力ソースとして利用するには，上述のように，JuliusMFT 起動時に “-input mfcnet” を引数とし

て指定する．この際，JuliusMFT は TCP/IP 通信サーバとなり，特徴量を受信する．また，HARK のモジュールである SpeechRecognitionClient や，SpeechRecognitionSMNClient は，音響特徴量とミッシングフィーチャーマスクを JuliusMFT に送出するためのクライアントして動作する．クライアントは，1 発話ごとに JuliusMFT に接続し，送信終了後ただちに接続を切断する．送信されるデータはリトルエンディアンである必要がある（ネットワークバイトオーダーでないことに注意）．具体的には，1 発話に対して以下の流れで通信を行う．

1. ソケット接続

   ソケットを開き，JuliusMFT に接続.

2. **通信初期化（最初に 1 回だけ送信するデータ）**

   クライアントから，ソケット接続直後に 1 回だけ，表 6.75 に示すこれから送信する音源に関する情報を送信する．音源情報は SourceInfo 構造体（表 6.76）で表され，音源 ID，音源方向，送信を開始した時刻を持つ．時刻は，<sys/time.h> で定義されている timeval 構造体で表し，システムのタイムゾーンにおける紀元（1970 年 1 月 1 日 00:00:00）からの経過時間である．以後，時刻は紀元からの経過時間を指すものとする．

表 6.75: 最初に 1 回だけ送信するデータ

サイズ [byte]	型	送信するデータ
4	`int`	28（= sizeof(SourceInfo))
28	SourceInfo	これから送信する特徴量の音源情報

表 6.76: SourceInfo 構造体

メンバ変数名	型	説明
source_id	`int`	音源 ID
azimuth	`float`	水平方向 [deg]
elevation	`float`	垂直方向 [deg]
time	timeval	時刻 (64 bit 処理系に統一，サイズは 16 バイト)

3. **データ送信（毎フレーム)**

   音響特徴量とミッシングフィーチャーマスクを送信する．表 6.77 に示すデータを 1 フレームとし，1 発話の特徴量を音声区間が終了するまで繰り返し送信する．特徴量ベクトルとマスクベクトルの次元数は同じ大きさであることが JuliusMFT 内部で仮定されている．

表 6.77: 毎フレーム送信するデータ

サイズ [byte]	型	送信するデータ
4	`int`	N1=（特徴量ベクトルの次元数）× sizeof(`float`)
N1	`float` [N1]	特徴量ベクトルの配列
4	`int`	N2=(マスクベクトルの次元数) × sizeof(`float`)
N2	`float` [N2]	マスクベクトルの配列

4. **終了処理**

   1 音源分の特徴量を送信し終えたら，終了を示すデータ（表 6.78）を送信してソケットを閉じる．

表 6.78: 終了を示すデータ

サイズ [byte]	型	送信するデータ
4	int	0

モジュールモード通信仕様　-module を指定するとオリジナル Julius と同様にモジュールモードで動作させることができる．モジュールモードでは， JuliusMFT が TCP/IP 通信のサーバとして機能し， JuliusMFT の状態や認識結果を jcontrol などのクライアントに提供する．また，コマンドを送信することにより動作を変更することができる．日本語文字列の文字コードは，通常 EUC-JP を利用しており，引数によって変更可能である．データ表現には XML ライクな形式が用いられており，一つのメッセージごとにデータの終了を表す目印として ”.”（ピリオド）が送信される．JuliusMFT で送信される代表的なタグの意味は以下の通りである．

- INPUT タグ入力に関する情報を表すタグで，属性として STATUS と TIME がある．STATUS の値は LISTEN，STARTREC，ENDREC のいずれかの状態をとる． LISTEN のときは Julius が音声を受信する準備が整ったことを表す．STARTREC は特徴量の受信を開始したことを表す．ENDREC は受信中の音源の最後の特徴量を受信したことを表す．TIME はそのときの時刻を表す．
- SOURCEINFO タグ音源に関する情報を表す，JuliusMFT オリジナルのタグである．属性として ID，AZIMUTH，ELEVATION，SEC，USEC がある．SOURCEINFO タグは第 2 パス開始時に送信される．ID は HARK で付与した音源 ID（話者 ID ではなく，各音源に一意に振られた番号）を，AZIMUTH は音源の最初のフレームのときのマイクロホンアレー座標系からみた水平方向（度）を，ELEVATION は同垂直方向（度）を，SEC と USEC は音源の最初のフレームの時刻を表し SEC が秒の位，USEC がマイクロ秒の位を表す．
- RECOGOUT タグ認識結果を表すタグで，子要素は漸次出力，第 1 パス，第 2 パスのいずれかである．漸次出力の場合は，子要素として PHYPO タグを持つ．第 1 パスと第 2 パス出力の場合は，子要素として文候補の SHYPO タグを持つ．第 1 パスの場合は，最大スコアとなる結果のみが出力され，第 2 パスの場合はパラメータで指定した数だけ候補を出力するので，その候補数だけ SHYPO タグが出力される．
- PHYPO タグ漸次候補を表すタグで，子要素として単語候補 WHYPO タグの列が含まれる．属性として PASS，SCORE，FRAME，TIME がある．PASS は何番目のパスかを表し，必ず 1 である．SCORE はこの候補のこれまでの累積スコアを表す．FRAME はこの候補を出力するのにこれまでに処理したフレーム数を表す．TIME は，そのときの時刻（秒）を表す．
- SHYPO タグ文仮説を表すタグで，子要素として単語仮説 WHYPO タグの列が含まれる．属性として PASS，RANK，SCORE，AMSCORE，LMSCORE がある． PASS は何番目のパスかを表し，属性 PASS があるときは必ず 1 である．RANK は仮説の順位を表し，第 2 パスの場合にのみ存在する．SCORE はこの仮説の対数尤度，AMSCORE は対数音響尤度，LMSCORE は対数言語確率を表す．
- WHYPO タグ単語仮説を表すタグで，属性として WORD，CLASSID，PHONE，CM を含む．WORD は表記を，CLASSID は統計言語モデルのキーとなる単語名を，PHONE は音素列を，CM は単語信頼度を表す．単語信頼度は，第 2 パスの結果にしか含まれない．
- SYSINFO タグシステムの状態を表すタグで，属性として PROCESS がある．PROCESS が EXIT のときは正常終了を，ERREXIT のときは異常終了を，ACTIVE のときは音声認識が動作可能である状態を，SLEEP のときは音声認識が停止中である状態を表す．これらのタグや属性が出力されるかどうかは，Julius MFT の起動時に指定された引数によって変わる．SOURCEINFO タグは必ず出力され，それ以外はオリジナルの Julius と同じなので，オリジナルの Julius の引数ヘルプを参照のこと．

オリジナルの Julius と比較した場合，JuliusMFT における変更点は，以下の 2 点である．

- 上述の音源定位に関する情報用タグである SOURCEINFO タグ関連の追加，および，関連する下記のタグへの音源 ID(SOURCEID) の埋め込み．

  STARTRECOG, ENDRECOG, INPUTPARAM, GMM, RECOGOUT, REJECTED, RECOGFAIL, GRAPHOUT, SOURCEINFO

- 同時発話時の排他制御による処理遅れを改善するために，モジュールモードのフォーマット変更を行った．具体的には，これまで発話単位で排他制御を行っていたが，これをタグ単位で行うよう出力が複数回に分かれており，一度に出力する必要がある下記のタグの出力に改造を施した．

  《開始タグ・終了タグに分かれているもの》

  - <RECOGOUT>...</RECOGOUT>
  - <GRAPHOUT>...</GRAPHOUT>
  - <GRAMINFO>...</GRAMINFO>
  - <RECOGPROCESS>...</RECOGPROCESS>

  《1 行完結のタグであるが，内部では複数回に分けて出力されているもの》

  - <RECOGFAIL ... />
  - <REJECTED ... />
  - <SR ... />

JuliusMFT 出力例

1. 標準出力モードの出力例

```
Stat: server-client: connect from 127.0.0.1
forked process [6212] handles this request
waiting connection...
source_id = 0, azimuth = 5.000000, elevation = 16.700001, sec = 1268718777, usec = 474575
### Recognition: 1st pass (LR beam)
.....................................................................................
pass1_best: <s> 注文お願いします </s>
pass1_best_wordseq: 0 2 1
pass1_best_phonemeseq: silB | ch u: m o N o n e g a i sh i m a s u | silE
pass1_best_score: 403.611420
### Recognition: 2nd pass (RL heuristic best-first)
STAT: 00 _default: 19 generated, 19 pushed, 4 nodes popped in 202
sentence1: <s> 注文お願いします </s>
wseq1: 0 2 1
phseq1: silB | ch u: m o N o n e g a i sh i m a s u | silE
cmscore1: 1.000 1.000 1.000
score1: 403.611786
```

connection end
ERROR: an error occured while recognition, terminate stream ←このエラーログが出るのは仕様

2. モジュールモードの出力サンプル

   下記の XML ライクな形式でクライアント (jcontrol など) に出力される．行頭の “>” は，jcontrol を用いた場合に jcontrol によって出力される（出力情報には含まれない）．

```
      > <STARTPROC/>
> <STARTRECOG SOURCEID="0"/>
> <ENDRECOG SOURCEID="0"/>
> <INPUTPARAM SOURCEID="0" FRAMES="202" MSEC="2020"/>
> <SOURCEINFO SOURCEID="0" AZIMUTH="5.000000" ELEVATION="16.700001" SEC="1268718638" USEC="10929"/
> <RECOGOUT SOURCEID="0">
>   <SHYPO RANK="1" SCORE="403.611786" GRAM="0">
>     <WHYPO WORD="<s>" CLASSID="0" PHONE="silB" CM="1.000"/>
>     <WHYPO WORD="注文お願いします" CLASSID="2" PHONE="ch u: m o N o n e g a i sh i m a s u" CM="1.
>     <WHYPO WORD="</s>" CLASSID="1" PHONE="silE" CM="1.000"/>
>   </SHYPO>
> </RECOGOUT>
```

注意事項

- -outcode オプションの制約

  タグ出力をプラグイン機能を用いて実装したため，出力情報タイプを指定できる -outcode オプションもプラグイン機能を用いて実現するように変更した．このため，プラグインが読み込まれていない状態で -outcode オプションを指定すると，エラーとなってしまう．

- 標準出力モードの発話終了時のエラーメッセージ

  標準出力モードで出力されるエラー “ERROR: an error occured while recognition, terminate stream” （出力例を参照）は，作成した特徴量入力プラグイン (mfcnet) で生成した子プロセスを終了する際に，強制的にエラーコードを julius 本体側に返しているため出力される．Julius 本体に極力修正を加えないようこのエラーに対する対処を行わず，仕様としている．なお，モジュールモードでは，このエラーは出力されない．

インストール方法

- apt-get を用いる方法

  apt-get の設定ができていれば，下記でインストールが完了する．なお，Ubuntu では オリジナルの Julius もパッケージ化されているため，オリジナルの Julius がインストールされている場合には，これを削除してから，以下を実行すること．

```
> apt-get install julius-4.1.4-hark julius-4.1.3-hark-plugin
```

- ソースからインストールする方法

  1. julius-4.1.4-hark と julius_4.1.3_plugin をダウンロードし，適当なディレクトリに展開する．
  2. julius-4.1.4-hark ディレクトリに移動して以下のコマンドを実効する．デフォルトでは，/usr/local/bin にインストールされてしまうため，パッケージと同様に /usr/bin にインストールするためには，以下のように -prefix を指定する．

     ```
     ./configure --prefix=/usr --enable-mfcnet; make; sudo make install
     ```

  3. 実行して以下の表示が出力されれば Julius のインストールは正常に終了している．

     ```
     > /usr/bin/julius
     Julius rev.4.1.4 - based on JuliusLib? rev.4.1.4 (fast) built for
     i686-pc-linux
     Copyright (c) 1991-2009 Kawahara Lab., Kyoto University Copyright
     (c) 1997-2000 Information-technology Promotion Agency, Japan Copyright
     (c) 2000-2005 Shikano Lab., Nara Institute of Science and Technology
     Copyright (c) 2005-2009 Julius project team, Nagoya Institute of
     Technology
     Try '-setting' for built-in engine configuration.
     Try '-help' for run time options.
     >
     ```

  4. 次にプラグインをインストールする． julius_4.1.3_plugin ディレクトリに移動して以下のコマンドを実行する．

     ```
     > export JULIUS_SOURCE_DIR=../julius_4.1.4-hark; make; sudo make install
     ```

     JULIUS_SOURCE_DIR には julius_4.1.4-hark のソースのパスを指定する．今回は同じディレクトリに Julius と plugin のソースを展開した場合を想定した．

     以上でインストール完了である．

  5. /usr/lib/julius_plugin 下にプラグインファイルがあるかどうかを確認する．

     ```
     > ls /usr/lib/julius_plugin
     calcmix_beam.jpi calcmix_none.jpi mfcnet.jpi calcmix_heu.jpi calcmix_safe.jpi
     >
     ```

     以上のように 5 つのプラグインファイルが表示されれば，正常にインストールできている．

- Windows 版のインストール方法については、3.2 章のソフトウェアのインストール方法を参照.

第7章　サポートツール

7.1　HARKTOOL

7.1.1　概要

HARKTOOL は，GHDSS で用いる分離用伝達関数ファイルと LocalizeMUSIC で用いる定位用伝達関数ファイルを生成・可視化するツールである．なお，ファイル作成方法には，下記の GUI 画面から作成する方法と，7.1.11 のコマンドのみで作成する方法の 2 種類がある．

HARKTOOL の機能を下記に示す.

- インパルス応答リストファイル作成
- TSP 応答リストファイル作成
- マイクロホン位置ファイル作成
- ノイズ位置ファイル作成
- 定位用伝達関数ファイル作成
- 分離用伝達関数ファイル作成
- 作成ファイルのグラフ表示

伝達関数ファイルの生成に必要なファイルは，**インパルス応答を録音したファイルを使用する場合と，TSP 応答を録音したファイルを使用する場合とでは異なり，それぞれ以下の通りである．**

インパルス応答を録音したファイルを使用する場合：

1. インパルス応答リストファイル (7.1.5 インパルス応答リストファイル作成方法参照)
2. マイクロホン位置ファイル (7.1.7 マイクロホン位置情報ファイル作成方法参照)
3. インパルス応答を録音したファイル

TSP 応答を録音したファイルを使用する場合：

1. TSP 応答リストファイル (7.1.6 TSP 応答リストファイル作成方法参照)
2. マイクロホン位置ファイル (7.1.7 マイクロホン位置情報ファイル作成方法参照)
3. TSP 応答を録音したファイル

これらのファイルをウィンドウから指定することで，伝達関数を生成する．

伝達関数ファイルは HARKTOOL を用いてシミュレーションで作成可能である．この場合，上記のマイクロホン位置ファイルのみから伝達関数を生成する．ただし，この時伝達関数ファイルはマイクロホンが自由空間に置かれたと仮定して生成されるので，マイクロホンが実際に設置されている物体の反響など（e.g. ロボットの頭の反射など）は無視される．

7.1.2　インストール方法

HARK がサポートしているディストリビューション/バージョンであれば，apt-get でインストールが可能である．リポジトリの登録は HARK のウェブページを参照.

>sudo apt-get install harktool4

7.1.3　起動方法

HARKTOOL の起動は, 次のコマンドを実行する.

- 通常：

  >harktool4

- GUI を日本語表示で起動したい場合：

  >export LC_ALL=”ja_JP.utf-8”

  >harktool4

- GUI を英語表示で起動したい場合：

  >export LC_ALL=C

  >harktool4

起動後，図 7.1 のような初期画面が表示され，数秒後に作業画面が表示される.

図 7.1: 起動画面

7.1.4　作業画面説明

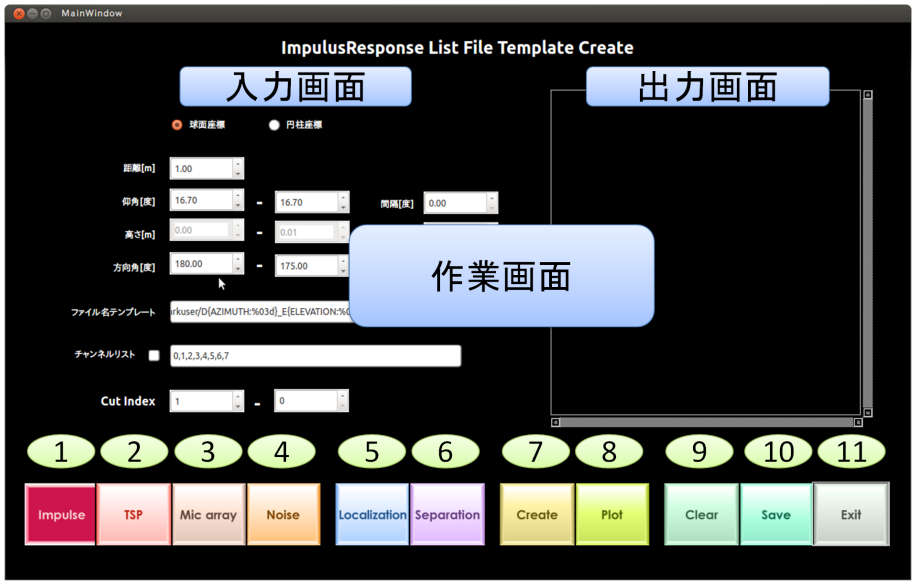


図 7.2: 作業画面

■作業画面のボタンの説明

1. Impluse
   インパルス応答リストファイルを作成する．
   作成したインパルス応答のグラフ表示を行う．

2. TSP
   TSP 応答リストファイルを作成する．
   作成した TSP 応答のグラフ表示を行う．

3. Mic array
   マイク位置情報ファイルを作成する．
   作成したマイク位置情報ファイルのグラフ表示を行う．

4. Noise
   ノイズ位置情報ファイルを作成する．
   作成したノイズ位置情報ファイルのグラフ表示を行う．

5. Localization
   定位用伝達関数ファイルを作成する．

6. Separation
   分離用伝達関数ファイルを作成する．

7. Create
   テンプレートファイル及び伝達関数を作成する．

8. Plot
   グラフを表示する．

9. Clear
入力項目を初期値に戻す.

10. Save
作成したファイルを保存する.

11. Exit
HARKTOOL4 を終了する.

■作業画面の上部の，メニューバーの「FILE」タブの各項目の説明

12. 「open」
作成したファイルを開く.

13. 「save」
作成したファイルを保存する.

■作業画面の上部の，メニューバーの「MENU」タブの各項目の説明

13. 「Impulse」
「Impulse」ボタンと同じ機能

14. 「tsp」
「TSP」ボタンと同じ機能

15. 「MicArray」
「Mic array」ボタンと同じ機能

16. 「noise」
「Noise」ボタンと同じ機能

17. 「localization」
「Localization」ボタンと同じ機能

18. 「separation」
「Separation」ボタンと同じ機能

19. 「create」
「Create」ボタンと同じ機能

20. 「plot」
「PLOT」ボタンと同じ機能

21. 「clear」
「Clear」ボタンと同じ機能

22. 「Exit」
「Exit」ボタンと同じ機能

7.1.5　インパルス応答リストファイル作成方法

- インパルス応答リストファイルとは，インパルス応答を録音したファイル群と測定位置の対応を表しているテキストファイルである．
- 「5.3.1 音源位置リスト情報 (srcinf) 形式」にもとづいて作成される必要がある．
- 下記に HARKTOOL のテンプレート作成機能を用いて作成する方法を示す．

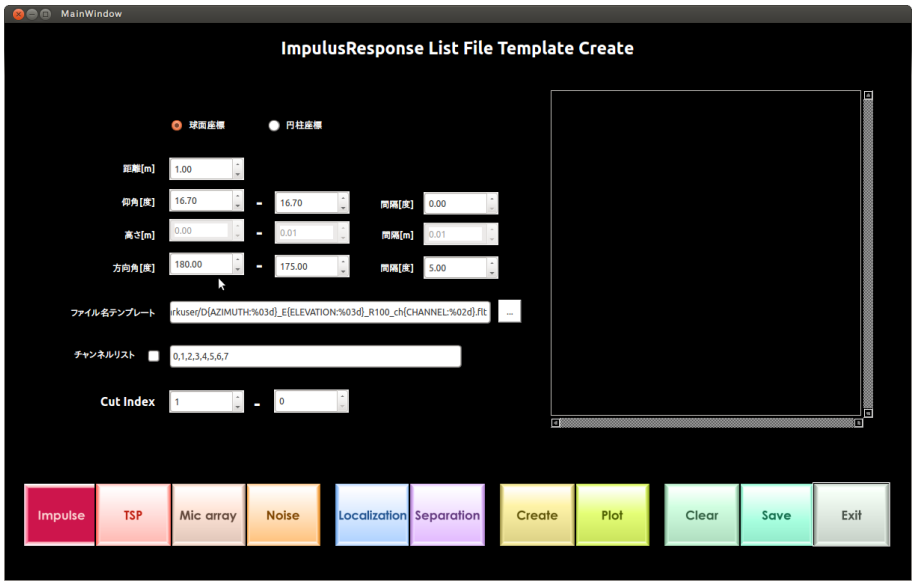


図 7.3: ImpulseResponse ListFile 作成画面

作成手順

1. 作業画面の「Impluse」ボタンをクリックする．
2. 左側の設定入力画面に設定値を入力する．

設定項目概要

- **球面座標　円柱座標**
  使用する座標系を球面座標または円柱座標を指定する．
  球面座標を選択した場合は，仰角が入力可能になり，円柱座標を選択した場合は，高さが入力可能になります．
- **距離 [m]**
  スピーカとマイクロホンとの水平方向距離
- **仰角 [度]**
  マイクロホンからみたスピーカ方向の仰角．
  次項目で設定する間隔 [度] ごとにテンプレートへ追加される．

- **高さ [m]**
  マイクロホンからみたスピーカ方向の高さ.
  次項目で設定する間隔 [m] ごとにテンプレートへ追加される.
- **方向角 [度]**
  左が開始角度, 右は終了角度.
  開始角度から反時計周りに, 次項目で設定する間隔 [度] ごとにテンプレートへ追加される.
- **ファイル名テンプレート**
  インパルス応答を録音したファイル（flt 形式）の格納場所を指定する.
  球面座標系の場合は文字列「{AZIMUTH:%03d}」,「{ELEVATION:%03d}」,「{CHANNEL:%02d}」を,
  円柱座標系の場合は文字列「{AZIMUTH:%03d}」,「{HEIGHT:%03d}」,「{CHANNEL:%02d}」を含む．実際のファイル名ではこれらは対応する角度や高さに置きかえられる.
- **チャンネルリスト**
  使用するチャンネル番号．使用するチャンネルを選択する場合は，チェックを入れて使用する
  チャンネル番号をカンマ区切りで記入する．全チャンネルを使用する場合は，チェック不要である.
- **Cut Index[サンプル]**
  伝達関数作成時に使用するサンプルの開始位置と終了位置.
  開始インデックスから終了インデックスまでのサンプルを使用して伝達関数が作成される.
  開始位置はデフォルトのままでよい.
  終了位置は直接音よりも反射音が大きい場合など，反射音を無視（削除）したい場合に使用する.

3. 画面下部の「Create」ボタンを押下するとインパルス応答リストファイルが作成され,
   そのファイルの内容が右側に表示される.
4. 右下の「Plot」ボタンを押下すると，入力パラメータに基づいたグラフが表示される.
5. 作成したファイルの保存は画面下部の「Save」ボタンから行う.

7.1.6 TSP 応答リストファイル作成方法

- TSP 応答リストファイルとは，TSP 信号を録音したファイル群と測定位置の対応を表すファイルである．
- 「5.3.1 音源位置リスト情報 (srcinf) 形式」にもとづいて作成される必要がある．
- 下記に，HARKTOOL のテンプレート作成機能を用いて作成する方法を示す．

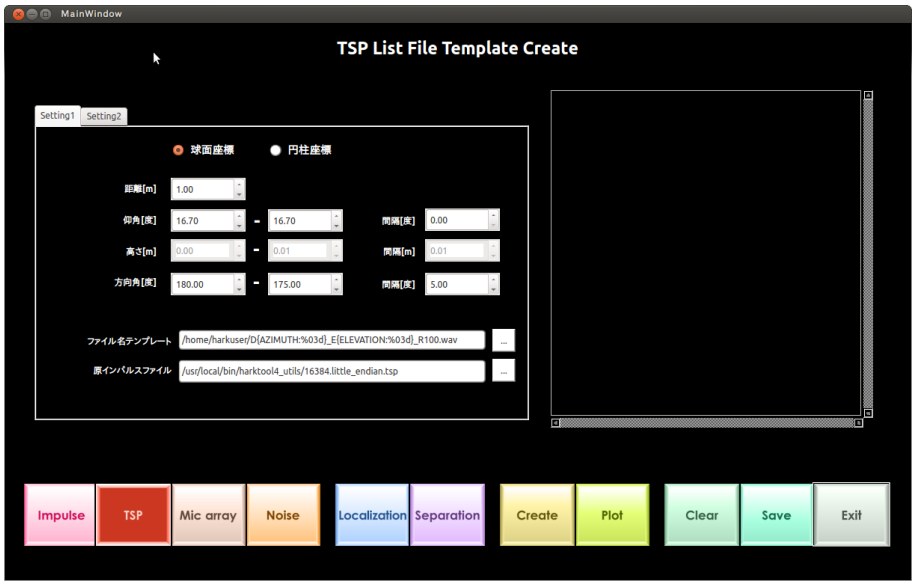


図 7.4: TSP ListFile 作成画面 (Setting1 タブ)

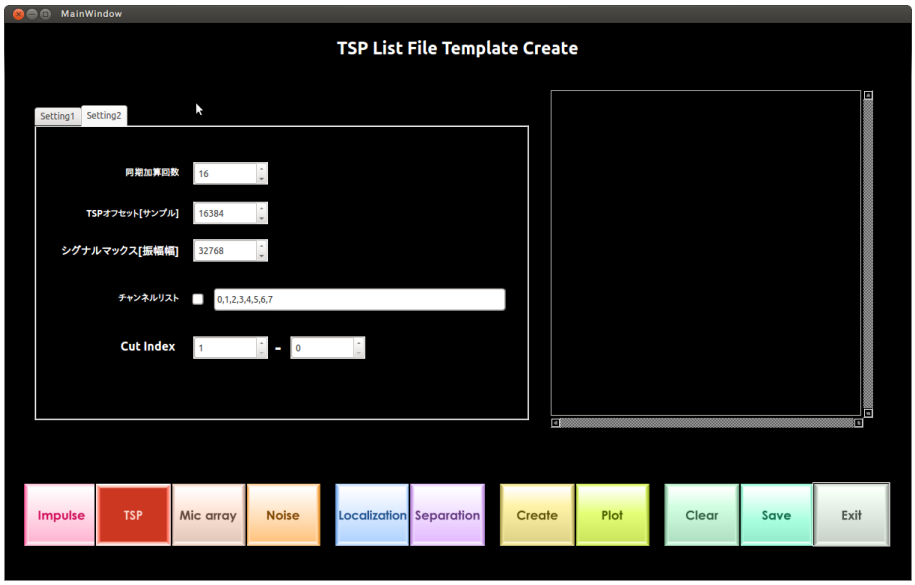


図 7.5: TSP ListFile 作成画面 (Setting2 タブ)

作成手順

1. 作業画面の「TSP」ボタンをクリックする．

2. 左側の設定入力画面に設定値を入力する.

設定項目概要

■ Setting1

- **球面座標　円柱座標**
  使用する座標系を球面座標または円柱座標を指定する.
  球面座標を選択した場合は，仰角が入力可能になり，円柱座標を選択した場合は，高さが入力可能になります.
- **距離 [m]**
  スピーカとマイクロホンとの水平方向距離.
- **仰角 [度]**
  マイクロホンからみたスピーカ方向の仰角.
  次項目で設定する間隔 [度] ごとにテンプレートへ追加される.
- **高さ [m]**
  マイクロホンからみたスピーカ方向の高さ.
  次項目で設定する間隔 [m] ごとにテンプレートへ追加される.
- **方向角 [度]**
  左が開始角度, 右は終了角度.
  開始角度から反時計周りに, 次項目で設定する間隔 [度] ごとにテンプレートへ追加される.
- **ファイル名テンプレート**
  TSP 応答を録音したファイル（flt 形式）の格納場所を指定する.
  球面座標系の場合は文字列「{AZIMUTH:%03d}」,「{ELEVATION:%03d}」を，円柱座標系の場合は文字列「{AZIMUTH:%03d}」,「{HEIGHT:%03d}」を含む．実際のファイル名ではこれらは対応する角度や高さに置きかえられる.
- **原インパルスファイル**
  録音に使用した **TSP 信号**, 1 周期分のファイル名

■ Setting2

- **同期加算回数**
  TSP 信号の録音時に TSP 信号を連続再生した回数
- **TSP オフセット [サンプル]**
  録音に使用した TSP 信号のサンプル数, デフォルトのままで良い.
- **シグナルマックス [振幅値]**
  最大振幅値とする数値. デフォルトのままで良い.
- **チャンネルリスト**
  使用するチャンネル番号．使用するチャンネルを選択する場合は，チェックを入れて使用する
  チャンネル番号をカンマ区切りで記入する．全チャンネルを使用する場合は，チェック不要である.
- **Cut Index[サンプル]**
  伝達関数作成時に使用するサンプルの開始位置と終了位置.
  開始インデックスから終了インデックスまでのサンプルを使用して伝達関数が作成される.
  開始位置はデフォルトのままでよい.
  終了位置は直接音よりも反射音が大きい場合など，反射音を無視（削除）したい場合に使用する.

3. 画面下部の「Create」ボタンを押下する．
   TSP 応答リストファイルが作成され，そのファイルの内容が右側に表示される．

4. 右下の「Plot」ボタンを押下すると，入力パラメータに基づいたグラフが表示される．

5. 作成したファイルの保存は画面下部の「Save」ボタンから行う．

7.1.7　マイクロホン位置情報ファイル作成方法

- マイクロホン位置情報ファイルは, マイクロホン位置を記述している. テキストファイルである.
- 5.2.1 マイクロホン位置テキスト形式 にもとづいて作成する必要がある.
- マイクロホン位置情報ファイルは, LocalizeMUSIC モジュールと GHDSS モジュールに使用する伝達関数ファイルの生成に必要である.
- 下記に HARKTOOL のテンプレート作成機能を用いて作成する方法を示す.

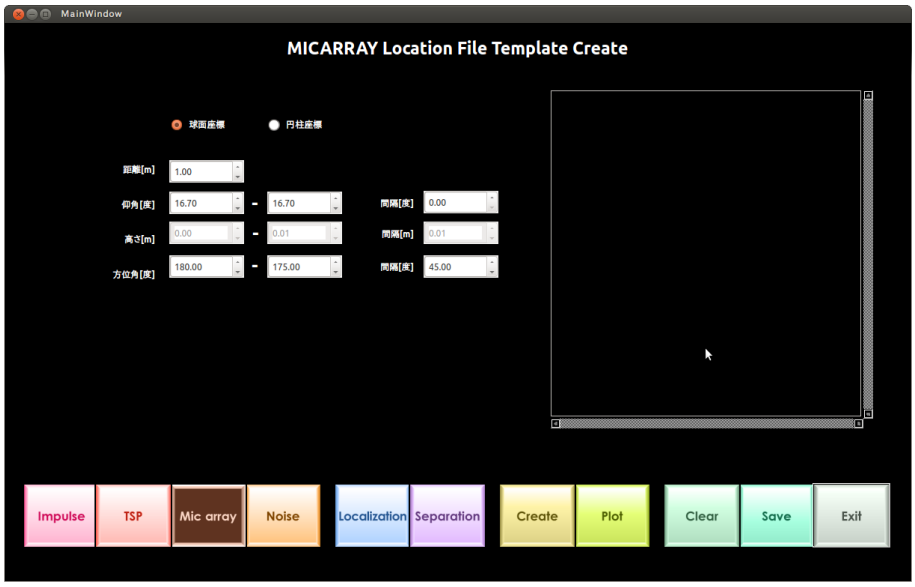


図 7.6: MICARY-LocationFile の編集

作成手順

1. 作業画面の「Mic array」ボタンをクリックする.
2. 左側の設定入力画面に設定値を入力する.

   設定項目概要

   - **球面座標　円柱座標**
     使用する座標系を球面座標または円柱座標を指定する．球面座標を選択した場合は，仰角が入力可能になり，円柱座標を選択した場合は，高さが入力可能になります.
   - **距離 [m]**
     スピーカとマイクロホンとの水平方向距離.
   - **仰角 [度]**
     マイクロホンからみたスピーカ方向の仰角．開始，終了，間隔を入力します.
   - **高さ [m]**
     マイクロホンからみたスピーカ方向の高さ．開始，終了，間隔を入力します.

- **方向角 [度]**
  左が終了角度, 右は開始角度.
  開始角度から反時計周りに，次項目で設定する間隔 [度] ごとにテンプレートへ追加される.

3. 画面下部の「Create」ボタンを押下するとマイクロホン位置情報ファイルが作成され，そのファイルの内容が右側に表示される.

4. 右下の「Plot」ボタンを押下すると，入力パラメータに基づいたグラフが表示される.

5. 作成したファイルの保存は画面下部の「Save」ボタンから行う.

7.1.8 ノイズ位置情報ファイル作成方法

- ノイズ位置情報ファイルは, ノイズ位置を記述している. テキストファイルである.
- 下記に，HARKTOOL のテンプレート作成機能を用いて作成する方法を示す.

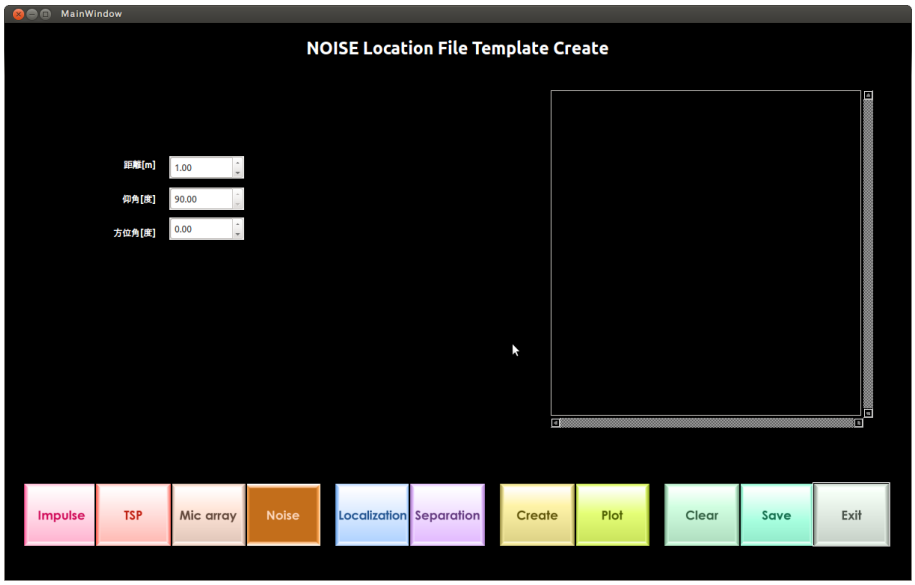


図 7.7: noise-LocationFile の編集

作成手順

1. 作業画面の「Noise」ボタンをクリックする.
2. 左側の設定入力画面に設定値を入力する.

   設定項目概要

   - **距離 [m]**
     スピーカとマイクロホンとの水平方向距離
   - **仰角 [度]**
     マイクロホンからみたスピーカ方向の仰角
   - **方向角 [度]**
     マイクロホンからみたスピーカの水平方向角
3. 画面下部の「Create」ボタンを押下するとノイズ位置情報ファイルが作成され，そのファイルの内容が右側に表示される.
4. 右下の「Plot」ボタンを押下すると，入力パラメータに基づいたグラフが表示される.
5. 作成したファイルの保存は画面下部の「Save」ボタンから行う.

7.1.9 定位用伝達関数ファイルの作成

- 定位用伝達関数ファイルは, LocalizeMUSIC モジュールで使用する設定ファイルである.
- ImpulseResponse-ListFile または，TSP-ListFile と MICARY-LocationFile から定位用伝達関数ファイルを生成する.

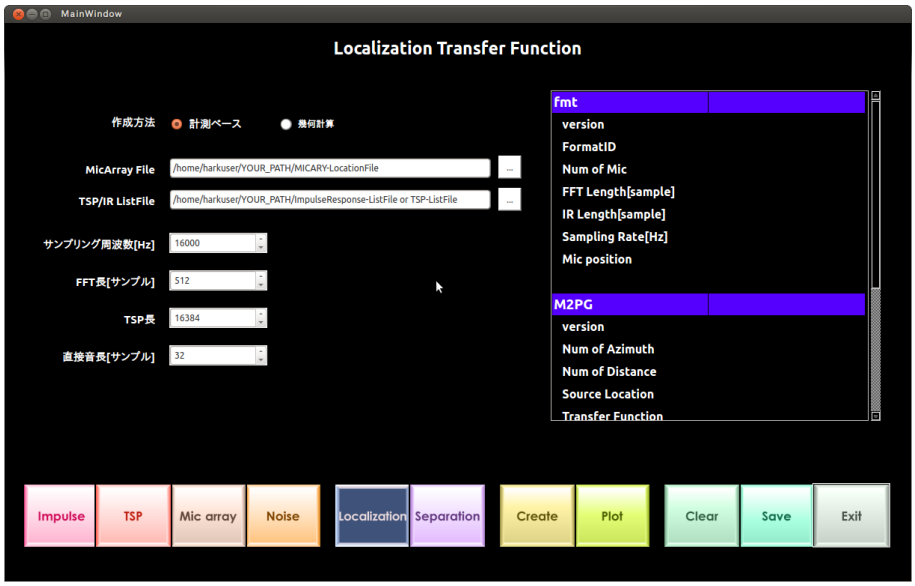


図 7.8: 定位用伝達関数ファイルの作成

1. 作業画面の「Localization」ボタンをクリックする.
2. 左側の設定入力画面に設定値を入力する.

   設定項目概要

   - **作成方法**
     計測ベースまたは, 幾何計算を選択できる.
     実際に計測した場合は計測ベースを選択する.
     幾何計算を選択すると伝達関数をシミュレーション生成する.
   - **MicArray File**
     マイクロホン位置情報ファイル.
     7.1.7 節で作成したファイルを指定する.
   - **TSP/IR ListFile**
     7.1.6 節の TSP 応答リストファイルまたは，7.1.5 節のインパルス応答リストファイルを指定する.
   - **サンプリング周波数 [Hz]**
     伝達関数のサンプリング周波数
   - **FFT 長 [サンプル]**
     伝達関数を離散周波数表現するときのビン数
   - **TSP 長 [サンプル]**
     録音した TSP 信号 1 個分の長さ. あるいはインパルス応答長を設定する.

- **直接音長 [サンプル]**
  伝達関数生成の際に使用するサンプル数

3. 画面下部の「Create」ボタンを押下すると，定位用伝達関数ファイルが作成され，そのファイルの内容が右側に表示される．

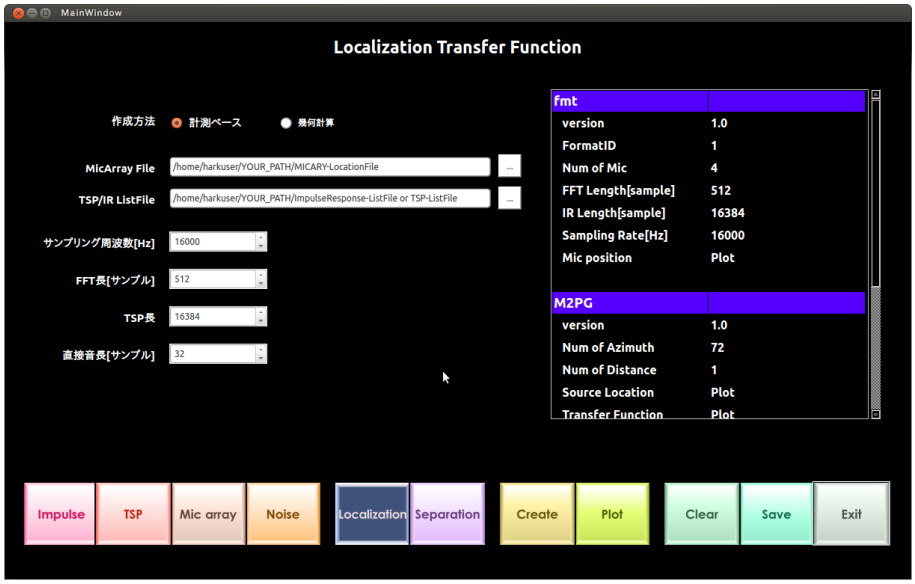


図 7.9: 定位用伝達関数表示

出力画面の説明

出力画面に「plot」と表示されている箇所は，ダブルクリックでグラフ表示可能である．

4. 「Mic Position」のグラフ表示
右側枠内の「Mic Position」横の「Plot」表示をダブルクリックする．

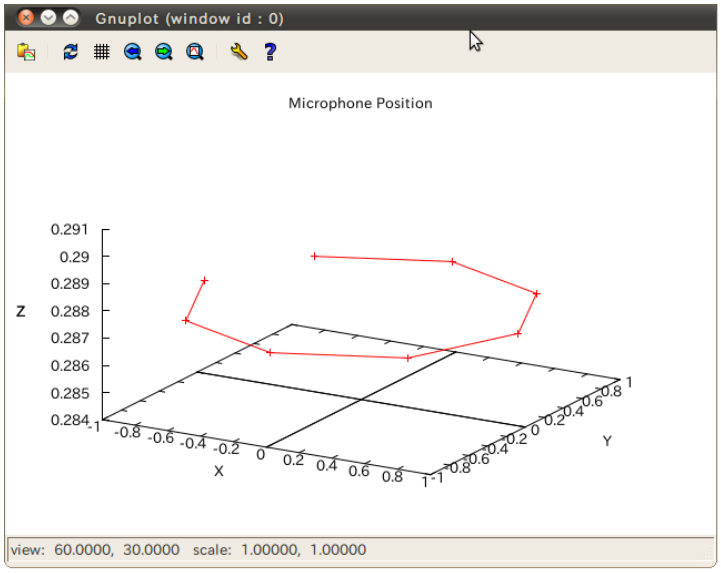


図 7.10: Mic Position のグラフ表示結果

5. 「Source Location」のグラフ表示
右側枠内の「Source Location」横の「Plot」表示をダブルクリックする．

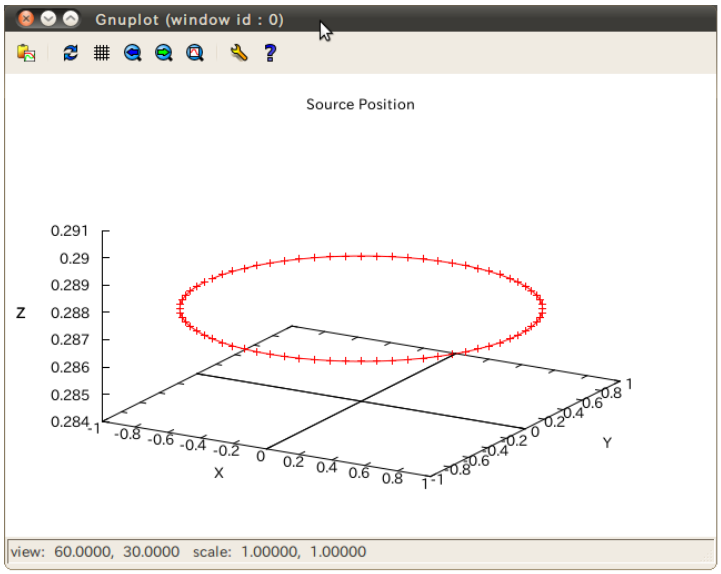


図 7.11: Source Location のグラフ表示結果

6. 「Transfer Function」のグラフ表示

右側枠内の「Transfer Function」の横の「Plot」をダブルクリックをすると, 図 7.12 のグラフ表示用設定画面が表示される.

下記の「設定項目概要」を参考に値を設定，右下の「Plot」ボタンをクリックすると，グラフが表示される

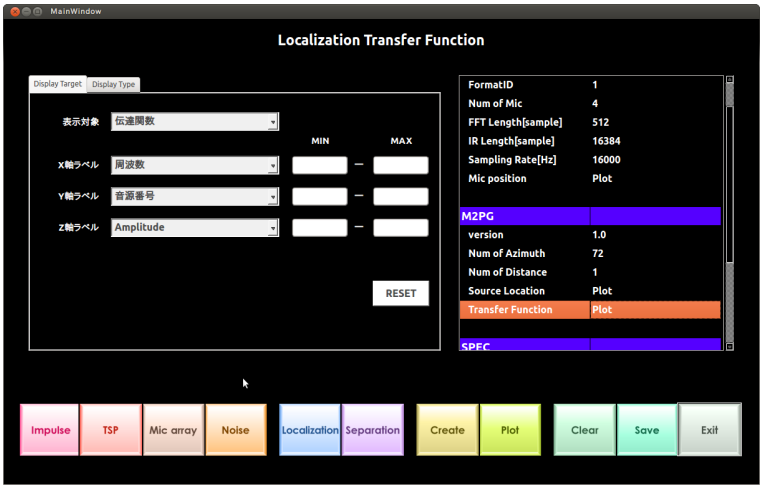


図 7.12: Transfer Function 設定画面 (Display Target タブ)

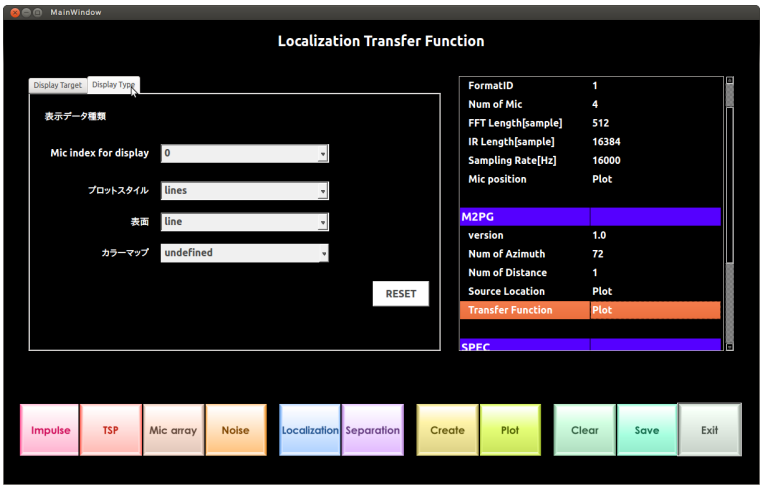


図 7.13: Transfer Function 設定画面 (Display Type タブ)

設定項目概要

■ Display Target タブ

- **表示対象**
  グラフの種別で，下記から選択する．
  - 伝達関数
  - 伝達関数の逆フーリエ変換
- **X 軸ラベル**
  X 軸のラベルを下記から選択する．
  - 周波数 （表示対象が「伝達関数」の場合）
  - 時間 （表示対象が「伝達関数の逆フーリエ変換」の場合）
  - 音源番号
  - マイク番号

  Min と Max に X 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : X 軸に表示する最小値
  Max : X 軸に表示する最大値
- **Y 軸ラベル**
  Y 軸のラベルを下記から選択する．
  - 周波数 （表示対象が「伝達関数」の場合）
  - 時間 （表示対象が「伝達関数の逆フーリエ変換」の場合）
  - 音源番号
  - マイク番号

  Min と Max に Y 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : Y 軸に表示する最小値
  Max : Y 軸に表示する最大値
- **Z 軸ラベル**
  Z 軸のラベルを下記から選択する．
  - Amplitude
  - dB
  - Phase

  Min と Max に Z 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : Z 軸に表示する最小値
  Max : Z 軸に表示する最大値

■ Display Type タブ

- **Mic (or SRC or Frec) index for display**
  表示するマイク (or ソース or 周波数) 番号を選択する

- **プロットパラメータ**
  表示するグラフのスタイルを選択する
  以下についてそれぞれ選択できる.
  - プロットスタイル (undefined, dots, points, lines, lines_points)
  - 表面 (undefined, line, mesh)
  - カラーマップ (undefined, pm3d, pm3d map, pm3d at b)

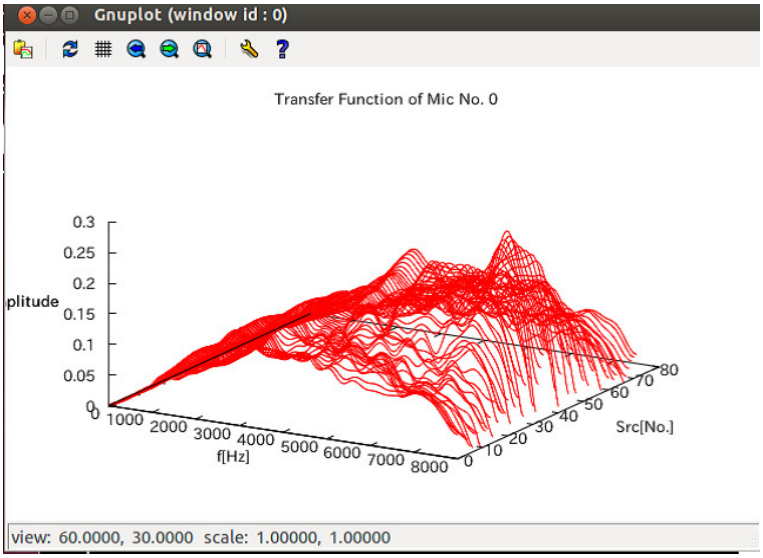


図 7.14: Transfer Function グラフ表示

7. 作成したファイルの保存は画面下部の「Save」ボタンから行う.

7.1.10　分離用伝達関数ファイルの作成

- 分離用伝達関数ファイルは, GHDSS モジュールで使用されるファイルである.
- ImpulseResponse-ListFile または，TSP-ListFile と MICARY-LocationFile から
- 分離用伝達関数ファイルを生成する.

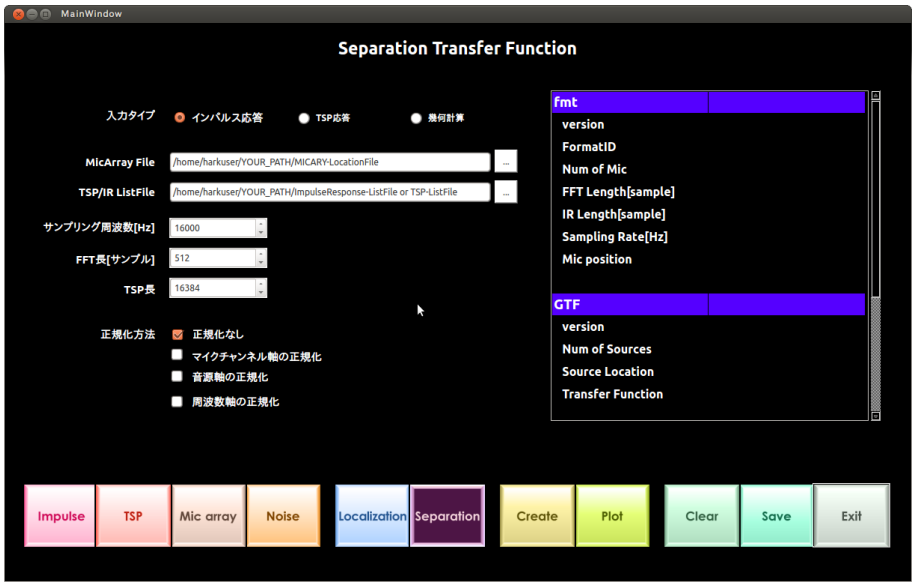


図 7.15: 分離用伝達関数ファイルの作成

1. 作業画面の「Separation」ボタンをクリックする.
2. 左側の設定入力画面に設定値を入力する.

   設定項目概要

   - **入力タイプ**
     インパルス応答，TSP 応答，幾何計算から選択できる.
     - 「インパルス応答」: 録音したインパルス応答から伝達関数を求める場合は選択する.
     - 「TSP 応答」: 録音した TSP 信号から伝達関数を求める場合は選択する.
     - 「幾何計算」: シミュレーションで伝達関数を求める場合は選択する.
   - **MicArray File**
     マイクロホン位置設定ファイル名.
     7.1.7 節で作成したファイルを指定する.
   - **TSP/IR ListFile**
     7.1.6 節の TSP 応答リストファイルまたは，7.1.5 節のインパルス応答リストファイルを指定する.
   - **サンプリング周波数 [Hz]**
     伝達関数のサンプリング周波数

- **FFT 長 [サンプル]**
  伝達関数を離散周波数表現する際のビン数
- **TSP 長 [サンプル]**
  録音した TSP 信号 1 個分の長さ. あるいはインパルス応答長を指定する.
- **正規化手法**
  正規化する軸を以下から選択できる.
  - 正規化なし
  - マイクチャネル軸の正規化
  - 音源軸の正規化
  - 周波数軸の正規化

3. 画面下部の「Create」ボタンを押下すると，分離用伝達関数ファイルが作成され
   そのファイルの内容が右側に表示される.

4. グラフ表示による確認
   出力画面に「Plot」と表示されているものに関しては，グラフ表示可能である.
   「Plot」表示をダブルクリックするとグラフ表示されるので，結果が正しいかを確認する.
   (図 7.16 参照)

5. 「Mic Position」のグラフ表示
   「Mic Position」横の「Plot」表示をダブルクリックする.

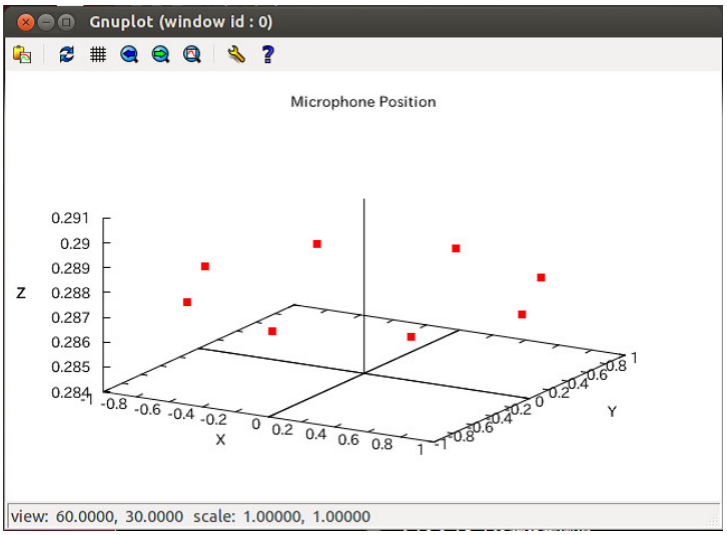


図 7.16: Mic Position のグラフ表示結果

6. 「Source Location」のグラフ表示
「Source Location」横の「Plot」表示をダブルクリックする．

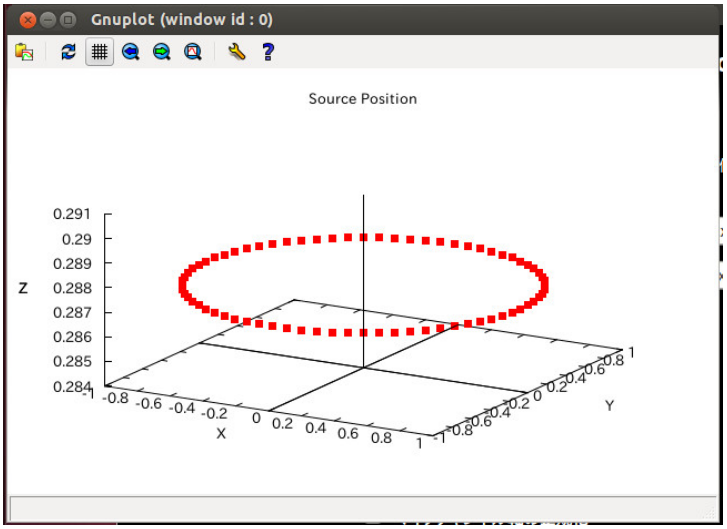


図 7.17: Source Location のグラフ表示結果

7. 「Transfer Function」のグラフ表示

「Transfer Function」の横の「Plot」をダブルクリックをすると，図 7.18 のグラフ表示用設定画面が表示される．

下記の「設定項目概要」を参考に値を設定，右下の「Plot」ボタンをクリックすると，グラフが表示される

(例：図 7.20 参照)

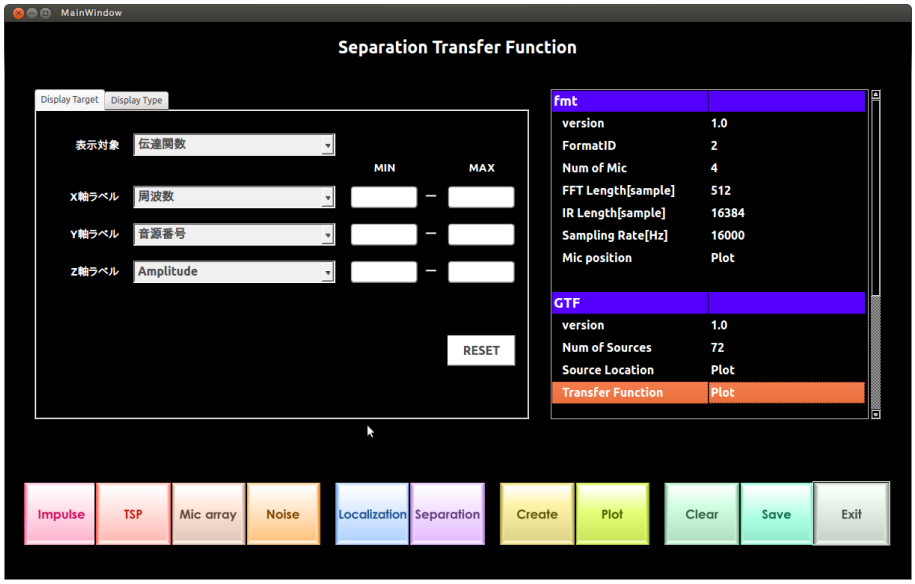


図 7.18: Transfer Function 設定画面 (Display Target タブ)

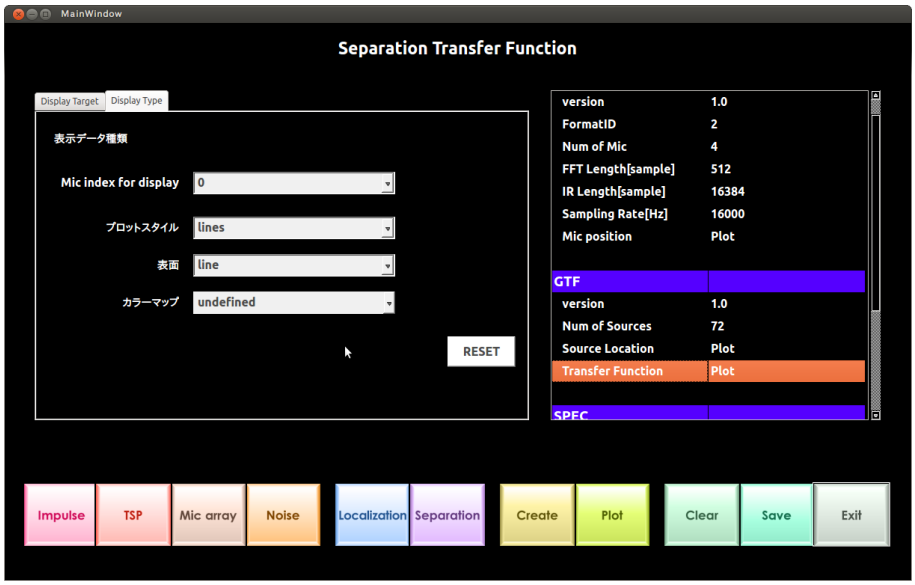


図 7.19: Transfer Function 設定画面 (Display Type タブ)

設定項目概要

■ Display Target タブ

- **表示対象**
  グラフの種別で，下記から選択する．
  - 伝達関数
  - 伝達関数の逆フーリエ変換
- **X 軸ラベル**
  X 軸のラベルを下記から選択する．
  - 周波数 （表示対象が「伝達関数」の場合）
  - 時間 （表示対象が「伝達関数の逆フーリエ変換」の場合）
  - 音源番号
  - マイク番号

  Min と Max に X 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : X 軸に表示する最小値
  Max : X 軸に表示する最大値
- **Y 軸ラベル**
  Y 軸のラベルを下記から選択する．
  - 周波数 （表示対象が「伝達関数」の場合）
  - 時間 （表示対象が「伝達関数の逆フーリエ変換」の場合）
  - 音源番号
  - マイク番号

  Min と Max に Y 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : Y 軸に表示する最小値
  Max : Y 軸に表示する最大値
- **Z 軸ラベル**
  Z 軸のラベルを下記から選択する．
  - Amplitude
  - dB
  - Phase

  Min と Max に Z 軸に表示する間隔を設定することが可能である．
  全表示する場合は，記入不要である．
  Min : Z 軸に表示する最小値
  Max : Z 軸に表示する最大値

■ Display Type タブ

- **Mic(or SRC or FREQ) index for display**
  表示するマイク (or ソース or 周波数) 番号を選択する

- **プロットパラメータ**
  表示するグラフのスタイルを選択する
  以下についてそれぞれ選択できる.
  - プロットスタイル (undefined, dots, points, lines, lines_points)
  - 表面 (undefined, line, mesh)
  - カラーマップ (undefined, pm3d, pm3d map, pm3d at b)

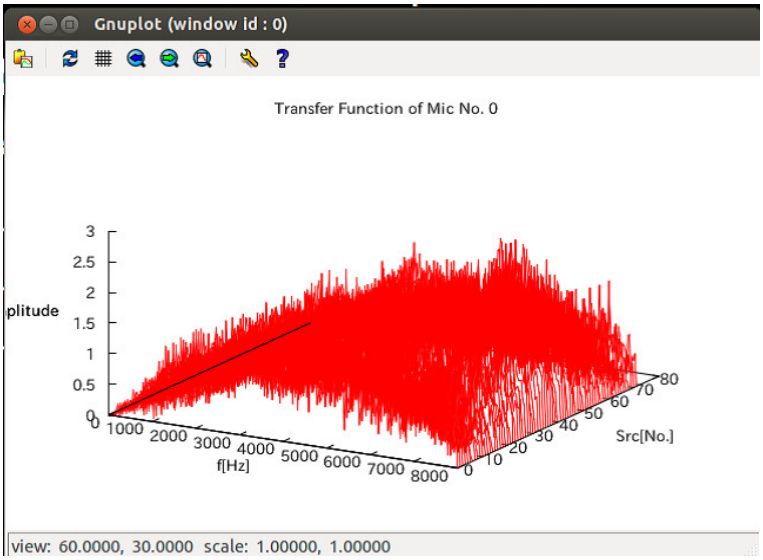


図 7.20: Transfer Function グラフ表示

8. 作成したファイルの保存は画面下部の「Save」ボタンから行う.

7.1.11 コマンド実行形式

下記に harktool4 をコマンドのみで実行する方法を記載する.
使用可能な機能を下記に示す.

- インパルス応答リストファイル作成
- TSP 応答リストファイル作成
- マイクロホン位置ファイル作成
- ノイズ位置ファイル作成
- 定位用伝達関数ファイル作成
- 分離用伝達関数ファイル作成

1. インストール方法

   HARK がサポートしているディストリビューション/バージョンであれば, `apt-get` でインストールが可能.

   ```
   sudo apt-get install harktool4_cui
   ```

2. 使用方法

下記のコマンドをターミナルから入力する

```
harktool4_cui [options] input-files
```

(例)ImpulseResponseListFile のテンプレートファイルを作成する場合

方位角範囲	-180〜180 1 度間隔
標高	16.7 度のみ
中心から音源までの半径	1m

```
harktool4_cui
--mode template           # テンプレートモード
--output-file ImpulseResponseListFile.xml # 出力ファイル名
--output-format xml       # XML フォーマット
--output-type ir          # ImpulseResponse タイプ
--azifrom -180.000000    # 方位角範囲開始 -180 度
--aziint 5.000000        # 方位角の間隔      1 度ずつ
--numazi 72              # 方位角の数  72
--elvfrom 16.700000      # 標高範囲の開始  16.7 度
--elvint 0.000000        # 間隔なし
--numelv 1               # 間隔数   1 つ
--radius 1.000000        # 半径    1m
--cut-start 1            # 開始インデックス  1 から
--cut-end 0              # 終了インデックス  最後まで
--filepath /home/username/D{AZIMUTH:%03d}_E{ELEVATION:%03d}_R100_ch{CHANNEL:%02d}.flt
```

[共通オプション:]

–help　　help 出力

-m [–mode] arg　　モード [template / tf]

-o [–output-file] arg　　出力 (default is STDOUT)

-f [–output-format] arg　出力フォーマット [xml / binary]

-t [–output-type] arg　　出力タイプ

(1) テンプレートモード [mic / noise / ir / tsp]

(2) 伝達関数モード [m2pg / gtf]

[オプション:]

※カッコ内は，初期値である．

- –nummic arg (=8)　　マイクロホン数

- –azifrom arg (=0)　方位角範囲開始
- –aziint arg (=5)　方位角の間隔
- –numazi arg (=72)　方位角の数
- –elvfrom arg (=0)　標高範囲の開始（音源の上下方向の角度：-90deg から 90deg の角度）
- –elvint arg (=5)　標高の間隔
- –numelv arg (=1)　間隔数
- –heightfrom arg (=0)　高さ (m) の範囲の開始
- –heightint arg (=0.01)　高さ (m) の間隔
- –numheight arg (=1)　高さ数
- –radius arg (=1)　中心からマイクや音源までの半径
- –synchronous-average arg (=16)　同期平均の数
- –original-impulse-file arg(=original-impulse-file.tsp)　オリジナルのインパルスファイル
- –tsp-offset arg (=16384)　サンプル中のインパルスファイル
- –tsp-length arg (=16384)　サンプルの１つの TSP の長さ
- –signal-max arg (=32768)　TSP 信号の最大値幅
- –cut-start arg (=1)　浮動小数点ベクトルのカット開始インデックス
- –cut-end arg (=0)　浮動小数点ベクトルのカット終了インデックス
- –mic-channels arg　マイクチャンネルのカンマ区切りリスト
- –filepath arg　ファイルパスのテンプレート文字列
- –nfft arg (=512)　FFT ポイント数
- –normalize-src　音源軸の正規化
- –normalize-mic　マイクチャンネル軸の正規化
- –normalize-freq　周波数軸の正規化
- –geometry-calculus arg (=0)　幾何計算フラグ（0:幾何計算しない 1:幾何計算する）
- –sampling-freq arg (=16000)　サンプリング周波数
- –direct-length arg (=32)　直接音の長さ

7.2 wios

7.2.1 概要

wios とは, HARK がサポートしている 3 種類のデバイス (1) ALSA がサポートするデバイス, (2) RASP シリーズ, (3) TD-BD-16ADUSB の録音, 再生, 同時録音再生を行うツールである. デバイスに関しては 8 を参照. 特に同時録音再生が可能なので, 音源定位、音源分離に必要なインパルス応答の測定に使える.

7.2.2 インストール方法

HARK がサポートしているディストリビューション/バージョンであれば, `apt-get` でインストールが可能. リポジトリの登録は HARK のウェブページを参照.

```
sudo apt-get install wios
```

7.2.3 使用方法

wios の詳細なオプションは、引数なしで実行すると見ることができる．重要なオプションは 3 種類で, 動作モード (録音, 再生, 同時録音再生) と, 使用デバイス (ALSA, RASP, TDBD)，そして動作指定オプションである. それぞれ自由な組み合わせた指定が可能である．たとえば デフォルトの ALSA デバイスから、44.1kHz で 2 秒録音し，voice.wav に保存したい時は次のようにすればよい.

```
wios -x 0 -f 44100 -t 2 -o voice.wav
```

他のコマンド例は, HARK cookbook の「多チャネル録音したい」を参照.

次に，オプションの種類ごとに説明する.

動作モード

録音 ：オプションに `-r` or `--rec` を与える．`-o` オプションで wave ファイル名を与えると，指定されたデバイスを通してマルチチャネルの音を録音できる．ファイル名を与えない場合のデフォルト名は `da.wav`.

再生 ：オプションに`-p` or `--play` を与える．`-i` オプションで wave ファイル名を与えると，指定されたデバイスを通して音声を再生できる．ファイル名を与えない場合のデフォルト名は `ad.wav`.

同時録音再生 ：オプションに `-s` or `--sync` を与える．`-i` オプションに再生するファイル名を，`-o` オプションに録音するファイル名を与える．すると wios はファイルの再生と録音を同時に開始する.

デバイスの選択

デバイスは, 2 つのオプション `-x` と `-d` を使って指定する.

ALSA ：オプションは `-x 0`．使用するデバイスを`-d` で指定する. デフォルトは `plughw:0,0`

TDBD ：オプションは `-x 1`．使用するデバイスファイルを`-d` で指定する．デフォルトは `/dev/sinichusb0`

RASP ：オプションは `-x 2`．使用するデバイスの IP アドレスを`-d` で指定する．デフォルトは `192.168.33.24`

上記のオプションに加えて, `-x` の値ごとに，デバイス専用の追加オプションを指定することも可能．詳しくは wios を引数なしで実行した際のオプションを参照.

動作指定

- `-t`: 録音/再生時間.
- `-e`: 量子化ビット数．デフォルトは 16bit.
- `-f`: サンプリング周波数．デフォルトは 16000Hz.
- `-c`: チャネル数．デフォルトは 1ch.

第8章　HARK 対応マルチチャネル A/D 装置の紹介と設定

HARK にマルチチャネル A/D 装置（以後，単に A/D 装置と呼ぶ）を接続することで，マイクロホンアレイ処理が可能になる．マイクロホンアレイ処理を必要とするユーザは，本章を参考にし，必要な A/D 装置の設定を行う．

HARK でサポートする A/D 装置は，以下の 3 機種である．

1. System In Frontier,　Inc.　RASP シリーズ,
2. ALSA ベースのデバイス (例，RME Hammerfall DSP Multiface シリーズ),
3. 東京エレクトロンデバイス TD-BD-16ADUSB.

各装置についてインストール方法と設定方法を説明する．

HARK 1.0 では，TD-BD-16ADUSB はサードパーティ版 HARK のみが対応していた．しかし，HARK 1.1 からはサードパーティ版は HARK と統一された．そのため，HARK のパッケージをインストールすれば TD-BD-16ADUSB は使用可能となる．

8.1 System In Frontier， Inc． RASP シリーズ

本節では， HARK がサポートしている マルチチャネル A/D 装置の RASP シリーズ (株式会社システムインフロンティア) のうち, 無線 RASP と RASP-24 の使用方法を述べる. サンプルファイルなどは HARK クックブックの録音ネットワークサンプルの節を参照.

8.1.1 無線 RASP

図 8.1: 無線 RASP の写真

図 8.1 に無線 RASP の概観を示す．無線 RASP は，単体で 16ch A/D 変換と 2ch D/A 変換が可能な A/D，D/A 変換装置である．TCP/IP 経由でコマンドを送受信し，ホスト PC から A/D，D/A 変換処理の設定とデータの送受信を行う．サンプリング周波数や，アナログ入力部分のフィルタをソフトウェアから容易に変更でき使い勝手が良い.

無線 RASP の ネットワーク設定

無線 RASP は，単体のボックスで構成されている．ホスト PC との接続は，無線 LAN で行うため，別途無線 LAN に対応したルーターが必要である．無線 RASP には，事前に IP アドレスを設定する他，無線ルーターの SSID を登録しておく必要がある．IP アドレスは出荷時設定をそのまま使用してもよいが，SSID の設定は必須である．添付マニュアルを参照して設定を完了させておく.

無線 RASP 用ソフトウェアのインストールと設定

無線 RASP を操作するためには，操作するホスト PC に，ライブラリのインストールとサンプルプログラムをインストールする必要がある．無線 RASP に同梱の CD からファイルをコピーし，インストール作業を行う．詳細は，添付資料を参考に行う.

HARK の使用には，少なくとも無線 RASP のアナログフィルタの設定を初期化できる環境を用意する必要がある．具体的には， `ws_fpaa_config` というプログラムのインストールが必須である．ライブラリをインストールし，サンプルプログラムをコンパイルすると `ws_fpaa_config` を生成できる． `ws_fpaa_config` は，無線 RASP の電源を切ると設定値が消えるため，電源を入れた直後に必ず実行する必要がある.

無線 RASP を用いた HARK での録音テスト

まず表 8.1 に示す使用機材を確認する．

表 8.1: 使用機材一覧

機材名	説明
無線 RASP	A/D，D/A 変換装置
無線 LAN 対応ルーター	RASP と ホスト PC 接続用
PC	無線 RASP 操作用 PC

HARK を用いた録音テストには，図 8.2 に示すネットワークファイル test．n を用いる．AudioStreamFromMic から ChannelSelector を経由し，MatrixToMap から SaveRawPCM への接続となっている．各プロパティの設定を表 8.2 に示す．”DEVICE” パラメータには，RASP に設定した IP アドレスを指定する (test.n では，192.68.0.1 と仮定．適宜読み替えて設定する)．

```
./test.n
```

とすることで，無線 RASP のアナログ入力 1 の音声が 3 秒間録音される．PCM 16 ビット量子化，16kHz サンプリングの音声がリトルエンディアン RAW 形式で保存される．

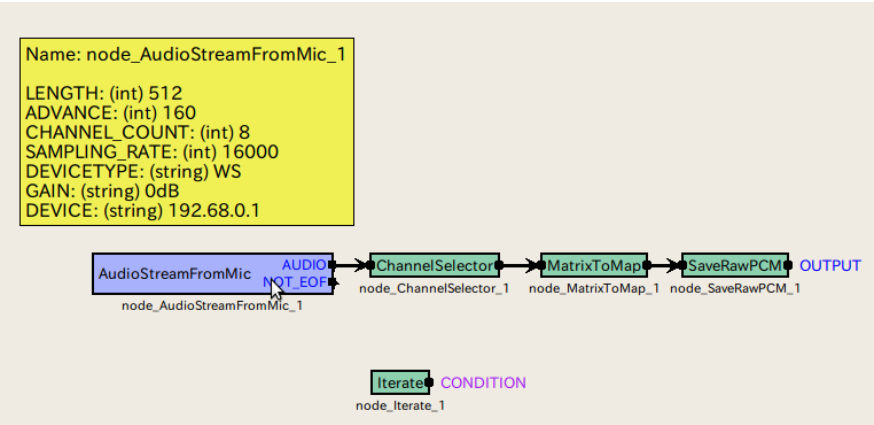


図 8.2: ネットワーク (test.n)

8.1.2 RASP-24

図 8.3 に RASP-24 の写真を示す。無線 RASP と同様に, RASP-24 も様々なサンプリング周波数で 8, 16 チャネルのマイクロホンを同時に録音できる A/D 装置である。また 2ch の音声出力 (D/A) 機能もあり, A/D と同期再生録音も可能である。無線 RASP との違いは、分解能である. 無線 RASP の分解能が 16bit であるのに大して, RASP-24 は 24bit なのでより量子化誤差の小さい A/D 変換が可能となる.

録音用計算機と RASP-24 との接続は TCP/IP を用いる．RASP-24 には LAN ポートが搭載されているのでそれを用いた有線接続で録音してもよいし，同じく搭載されている USB ポートに無線 LAN 子機を接続すれば無線で録音もできる．詳しくは製品に同梱されている説明書を参照.

基本的な使い方は 無線 RASP と同様である．ネットワークの設定を行い， AudioStreamFromMic を用いて録音を行う．無線 RASP との違いは，電源投入後にファームウェアの書き込みが不要ということである．した

表 8.2: 録音ネットワークのプロパティ

モジュール名	プロパティ名	値の型	設定値
AudioStreamFromMic	LENGTH	int	512
	ADVANCE	int	160
	CHANNEL_COUNT	int	16
	SAMPLING_RATE	int	16000
	DEVICETYPE	string	WS
	DEVICE	string	192．68．0．1
ChannelSelector	SELECTOR	Object	<Vector <int > 0 >
SaveRawPCM	BASENAME	string	sep_
	ADVANCE	int	160
	BITS	int	16
Iterate	MAX_ITER	int	300

がって, ws_fpaa_config を実行する必要はない．サンプルネットワークは HARK クックブックの録音サンプルネットワークの節を参照.

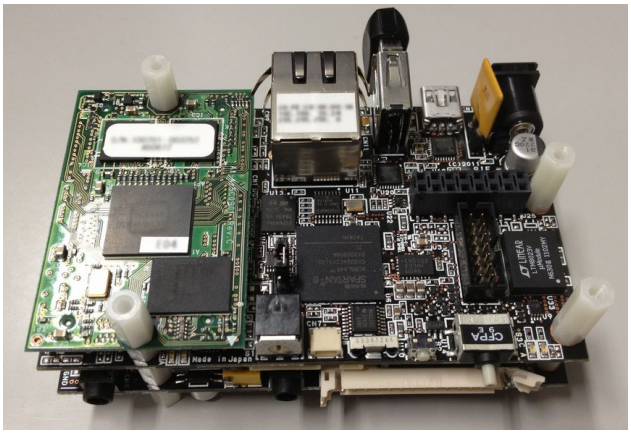


図 8.3: RASP-24 の写真

8.2 RME Hammerfall DSP シリーズ Multiface AE

本節では，HARK 開発チームが動作確認を行っているマルチチャネル A/D 装置の一つである RME 社の Hammerfall DSP Multiface シリーズ（以下，Hammerfall DSP Multiface シリーズ を単に Multiface と呼ぶ）を HARK で使用するための設定手順について説明する．

Multiface は，単体で 10 ch A/D 変換と 10 ch D/A 変換可能な A/D，D/A 装置である．10 ch のうちアナログ部が 8 ch で，SPDIF 部が 2 ch である．Linux で標準的なインタフェースである ALSA に準拠したデバイスであり，ここでは ALSA 準拠デバイスの例として設定方法を紹介する．

Multiface 以外の ALSA 準拠デバイスはサポート外であるが，他の ALSA 準拠デバイスを ALSA ドライバを通じて OS に認識させることができれば Multiface と同様に， HARK でマルチチャネル録音できる可能性がある．

Multiface のインストール方法，および，HARK 上での Multiface を接続したマイクロホンからの音声録音方法を取り上げる．説明と動作確認は，Ubuntu 9.04，10．04 で行っている．

8.2.1 Multiface の PC への接続

RME 社製の 24bit/96kHz マルチチャネル A/D 装置は，下記の 3 つの装置から構成される．

- **RME HDSP I/O ボックス**: RME HDSP Multiface AE,
- **RME HDSP インタフェース**: RME HDSP PCI カード，RME HDSP CardBus カードのいずれか
  RME HDSP PCI-Express カード，および RME HDSP Express カードは，動作未確認であり，現在サポート外である．
- **マイクロホン用プリアンプ**: RME OctaMic-II．RME OctaMic や YAMAHA HA-8 も使用実績があるが，これらのモデルは製造中止になっている．

ただし，マイクロホン用プリアンプは，必ずしも必要ではなく，録音レベルを確保できるならば，接続する必要はない．これらのハードウエアは，添付マニュアルを参照して，PC に接続する．

Multiface 用ソフトウェアのインストールと設定

Multiface を Linux にマルチチャネル A/D 機器として認識させるためには，ALSA デバイス用汎用ドライバと，Multiface 用のファームウエア及び，ソフトウェアをインストールし，パラメータを設定する必要がある．

以下のインストールは，パッケージからインストールする方法を強く推奨する．ソースからコンパイル，インストールという方法もあるが，他のソフトウェアとの競合が起り，コンパイルとインストールが見かけ上，成功しても，実際には音が再生・録音できないという現象に陥りやすい．特に pulseAudio との競合が起るので注意が必要である．各自のシステム環境に強く依存するため，ソースからのインストールの具体的な解説は割愛する．

以下パッケージから，ALSA デバイス用汎用ドライバのインストール方法を説明した後に，Multiface の初期化に必要な hdsp コマンドを使用可能にするためのインストール作業について述べ，hdsp コマンドによる設定について述べる．最後に，HARK を使った録音ネットワークの例題について解説する．

ALSA デバイス用汎用ドライバのインストール

ALSA デバイス用汎用ドライバ (alsa-lib, alsa-driver) は，インストールされていることが多い．インストールされているか確認するには，以下の太字の部分を入力する．> はコマンドプロンプトを表す．イタリック体部分のメッセージが表示されれば，alsa-lib，alsa-drive がインストールされているので，これらのインストールは必要ない．Version 番号は，使用環境によって異なるので作業例中の Version 1．0．24 は，各自の使用環境で異なるバージョン番号を示す場合がある．適宜読み換えて作業を進める．今回のテスト環境では， Version 1．0．24 を使用した．

```
> cat /proc/asound/version
 Advanced Linux Sound Architecture Drive Version 1. 0. 24.
```

インストールされていなければ，次の節の作業で自動的に自動インストールされるので，このまま次節の作業に入ればよい．

Multiface の初期化に必要なコマンドを使用可能にするためのインストール作業

必要なパッケージは， 以下の 2 つである．

- alsa-firmware-loaders
- alsa-tools-gui

これらのパッケージのインストールには，

- Synaptic パッケージマネージャ
- apt-get

のいずれかを使用する．以上のパッケージをインストールすることで，Multiface の設定に必要な以下の 3 つのコマンドが使用可能になる．

- hdsploader (Package : alsa-firmware-loaders)
- hdspconf (Package : alsa-tools-gui)
- hdspmixer (Package : alsa-tools-gui)

apt-get を使用したインストール例を示す．

```
> sudo apt-get update
> sudo apt-get install alsa-firmware-loaders
> sudo apt-get install alsa-tools-gui
```

最後に，multiface_firmware_rev11．bin を入手するために，alsa-firmware を alsa の Web サイトからソースをダウンロードする．(http://www．alsa-project．org/main/index．php/Main_Page)．2011/12/13 現在，利用可能な最新バージョンは， 1．0．24 である．alsa-firmware-1．0．24．tar．bz2 をダウンロードする．ダウンロード後，ファイルは，システムメニューの「場所」「ダウンロード」に保存される．これを,「場所」「ホーム」にコピーすると作業しやすい．

適当なディレクトリにダウンロードしたファイル alsa-firmware-1．0．24．tar.bz2 をコピーする．最初に bunzip2 と tar でファイルを展開する．その後，コンパイル作業に入る．具体的な作業手順を以下に示す．

```
> bunzip2 -d alsa-firmware-1.0.24.tar.bz2
> tar vfx alsa-firmware-1.0.24.tar
    alsa-firmware-1.0.24/
    alsa-firmware-1.0.24/multisound/
          中略
    config. status: creating aica/Makefile
    config. status: executing depfiles commands
> make
> sudo mkdir -p /lib/firmware/hdsploader
> sudo cp hdsploader/multiface_firmware_rev11.bin /lib/firmware/hdsploader/
> sudo cp hdsploader/multiface_firmware_rev11.dat /lib/firmware/hdsploader/
```

以上で， hdsp コマンドがインストールされて Multiface の操作と設定が可能になった．

hdsp コマンドによる設定

必要な hdsp コマンドは hdsploader， hdspconf， hdspmixer の３つである．これらのコマンドを順に説明する．

- **hdsploader**
  hdsploader は，multiface の FPGA を初期化する firmware プログラム（multiface_firmware_rev11.dat）をアップロードするコマンドである．以下に実行例を示す．エラーメッセージ (Hwdep ioctl error on card hw:0 : Device or resource busy.) が表示されるが，ここでは問題ないので無視してよい．同一システム上で 2 度以上実行すると，エラーメッセージが表示される．OS 起動中に実行されるシステムの場合には，この作業で必ずこのエラーメッセージが表示される．

```
# hdsploader
   hdsploader - firmware loader for RME Hammerfall DSP cards
   Looking for HDSP + Multiface or Digiface cards :
   Card 0 : RME Hammerfall DSP + Digiface at 0xf9df0000, irq 16
   Upload firmware for card hw:0
   Hwdep ioctl error on card hw:0 : Device or resource busy.
   Card 1 : HDA Intel at 0xfdffc000 irq 30
```

- **hdspconf**
  hdspconf を実行すると，Multiface の設定用のウィンドウが開く．サンプリング周波数を設定できる．他の項目は Multiface のドキュメントを参考にする．図 8.4 に設定用のウィンドウ例を示す．設定可能なサンプリングレートは，32kHz， 44.1kHz， 48kH， 64kHz， 88.2KkHz， 96kHz である．ここでは，32kHz が選択されている．

- **hdspmixer**
  hdspmixer を実行すると，ミキサーの GUI が表示される．入力レベル，出力レベルの調節と確認が可能である．図 8.5 にミキサーの表示例を示す．

8.2.2 Multiface を用いた HARK での録音テスト

録音に先立ち機材を確認する．使用機材を表 8.3 に示す．

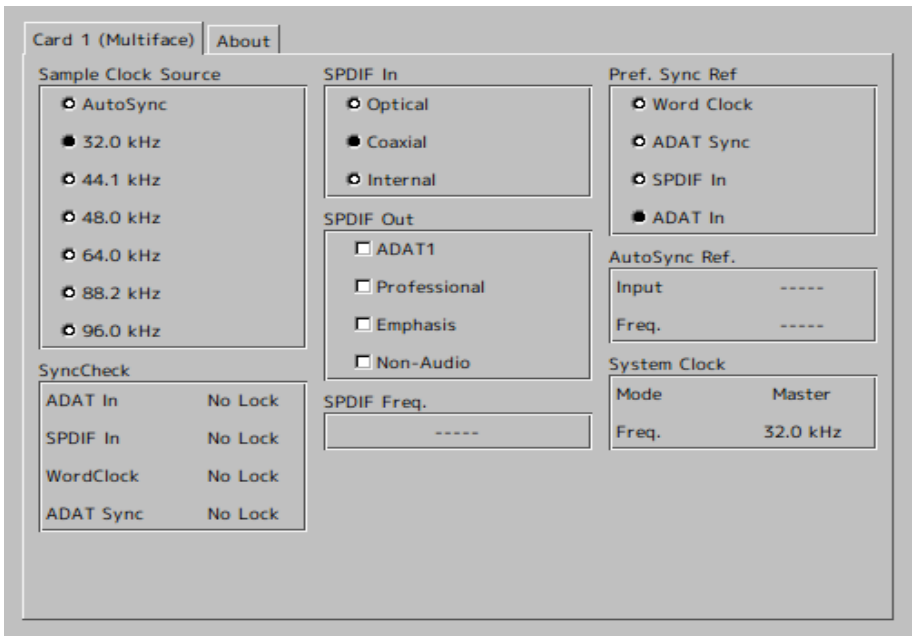


図 8.4: パラメータ設定ウィンドウ

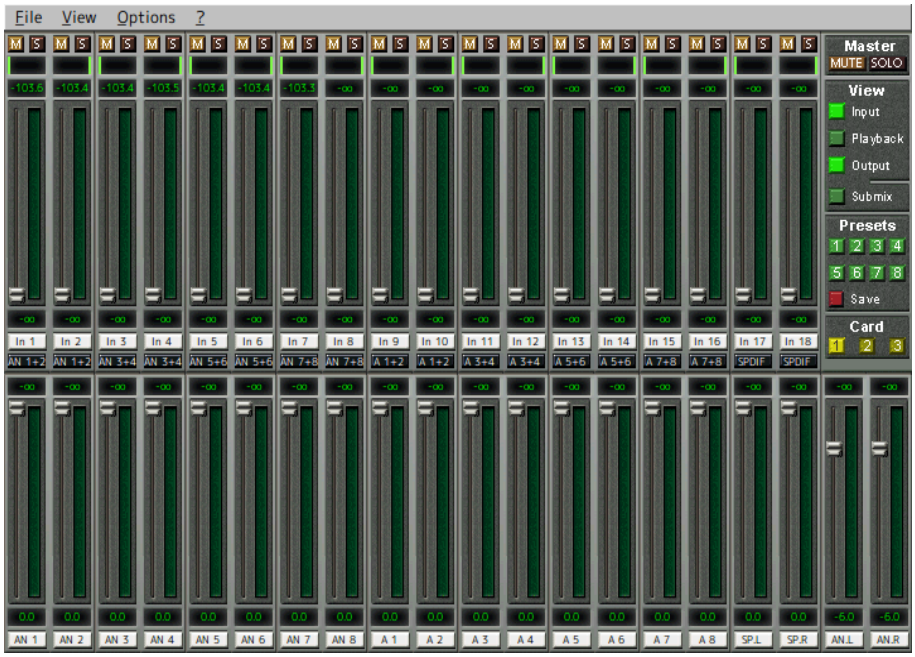


図 8.5: パラメータ設定ウィンドウ

HARK を用いた録音テストには，図 8.2 と同じネットワークファイル test.n を用い，各モジュールのプロパティのみを変更する．各プロパティの設定を表 8.4 に示す．ただし，DEVICE パラメータは，PC に接続されているデバイスに応じて番号が変化する可能性がある．ここでは，plughw:1，0 が Multiface を示しているものとする．

```
> ./test.n
```

とすることで，Multiface のアナログ入力 1 の音声が 3 秒間録音される．PCM 16 ビット量子化，32kHz サン

表 8.3: 使用機材一覧

機材名	説明
Hammerfall DSP Multiface	A/D，D/A 変換装置
Digital Audio CardBus Interface	Multiface 接続専用ポートを PCMCIA スロットに増設するカード
Octa Mic	マイクロホンプリアンプ

図 8.6: RME Hammerfall DSP Multiface (front)

図 8.7: RME Hammerfall DSP Multiface (Rear)

プリングの音声がリトルエンディアン RAW 形式で保存される．

図 8.8: PCMCIA CardBus Interface for Hammerfall DSP System

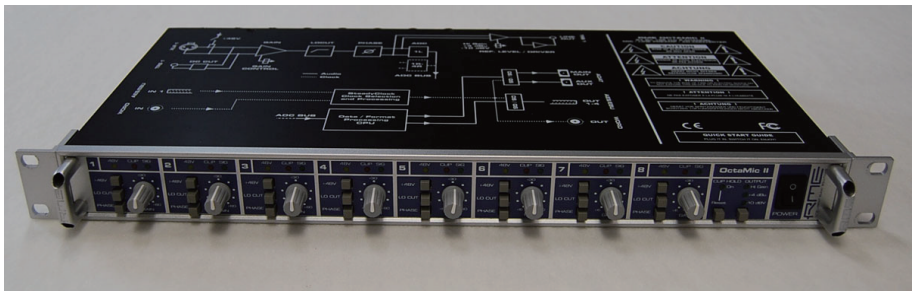


図 8.9: RME OctaMic (Front)

図 8.10: RME OctaMic (Rear)

表 8.4: 録音ネットワークのプロパティ

モジュール名	プロパティ名	値の型	設定値
AudioStreamFromMic	LENGTH	int	1024
	ADVANCE	int	320
	CHANNEL_COUNT	int	8
	SAMPLING_RATE	int	32000
	DEVICETYPE	string	ALSA
	DEVICE	string	plughw:1，0
ChannelSelector	SELECTOR	Object	<Vector <int > 0 >
SaveRawPCM	BASENAME	string	sep_
	ADVANCE	int	320
	BITS	int	16
Iterate	MAX_ITER	int	300

8.3 東京エレクトロンデバイス TD-BD-16ADUSB

本節では，HARK 開発チームが動作確認を行っているマルチチャネル A/D 装置の一つである東京エレクトロンデバイス の Inrevium シリーズの中の１つである TD-BD-16ADUSB （以下，東京エレクトロンデバイス TD-BD-16ADUSB を単に 16ADUSB と呼ぶ）を HARK で使用するための設定手順について説明する．

16ADUSB は，単体で 16ch A/D 変換と 4 ch D/A 変換が可能な A/D，D/A 変換装置である．16ADUSB のドライバソフトウェアは，製品に附属している．Linux 用のカーネルモードドライバが附属しており使い勝手が良い．マイクロホンの接続では，プラグインパワー供給に対応しているため，プラグインパワー供給可能なコンデンサマイクロホンをそのまま接続できる．

16ADUSB のインストール方法，および，HARK 上での 16ADUSB を接続したマイクロホンからの音声録音方法を取り上げる．説明と動作確認は，Ubuntu 10.04 で行っている．

8.3.1 16ADUSB の PC への接続

16ADUSB は，名刺サイズの基盤に実装されている．ホスト PC との接続は，USB で行うため，接続が容易である．ハードウエアは，添付マニュアルを参照して，PC に接続する．

8.3.2 16ADUSB 用ソフトウェアのインストールと設定

16ADUSB を操作するためには，操作するホスト PC に，カーネルモードドライバをインストールする必要がある．デバイスに付属するマニュアルを参考にインストールを行う．サンプルプログラムもデバイスに付属しているので，そのサンプルプログラムを使用して録音ができることを確認しておく．

8.3.3 TD-BD-16ADUSB を用いた HARK での録音テスト

録音に先立ち機材を確認する．使用機材を表 8.5 に示す．

表 8.5: 使用機材一覧

機材名	説明
TD-BD-16ADUSB	A/D，D/A 変換装置
PC	TD-BD-16ADUSB 操作用 PC

HARK を用いた録音テストには，図 8.2 と同じネットワークファイル test.n を用い，各プロパティのみを変更する．各プロパティの設定を表 8.6 に示す．

```
> ./test.n
```

とすることで，16ADUSB のアナログ入力 1 の音声が 3 秒間録音される．PCM 16 ビット量子化，16kHz サンプリングの音声がリトルエンディアン RAW 形式で保存される．

表 8.6: 録音ネットワークのプロパティ

モジュール名	プロパティ名	値の型	設定値
AudioStreamFromMic	LENGTH	int	512
	ADVANCE	int	160
	CHANNEL_COUNT	int	16
	SAMPLING_RATE	int	16000
	DEVICETYPE	string	TDBD16ADUSB
	DEVICE	string	SINICH
ChannelSelector	SELECTOR	Object	<Vector <int > 0 >
SaveRawPCM	BASENAME	string	sep_
	ADVANCE	int	160
	BITS	int	16
Iterate	MAX_ITER	int	300