

Title	日本語音声の振幅包絡に含まれる冗長性に関する検討
Author(s)	西野, 恭生
Citation	
Issue Date	2014-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12054
Rights	
Description	Supervisor: 鷓木祐史, 情報科学研究科, 修士

修 士 論 文

日本語音声の振幅包絡に含まれる冗長性に関する
検討

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

西野 恭生

2014年3月

修士論文

日本語音声の振幅包絡に含まれる冗長性に関する 検討

指導教員 鵜木 祐史 准教授

審査委員主査 鵜木 祐史 准教授
審査委員 赤木 正人 教授
審査委員 田中 宏和 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

1110047 西野 恭生

提出年月: 2014年2月

概要

音声は知覚的に頑健である。音声の言語情報（何と発話しているのか）の取得に関する重要な特徴は、音声の至るところに分散して冗長に含まれている。そのため、その一部が変形・削除されたとしても我々は容易に音声の言語情報を取得することができる。その一例として、雑音駆動音声の知覚が知られている。雑音駆動音声とは、音声の時間包絡線（振幅包絡）を取り出して、その振幅包絡で雑音を振幅変調することにより、時間微細構造を持たず、音声の振幅包絡情報のみを持った信号である。この雑音駆動音声の先行研究より、音声の了解性に関する情報は、音声の振幅包絡に多く含まれていることが報告されている。しかし、これらの先行研究では、振幅包絡に冗長性があるのかどうかまでは深く踏み込んで検討されていない。そのため、音声の振幅包絡に含まれる冗長性を解明する。

音声の冗長性とは、音声知覚に関係した特徴が音声信号に分散して冗長に含まれており、その中のいくつかが失われたとしても何と言っているのか正確に聴き取ることができる性質である。この性質を利用した技術として、音声圧縮や音声符号化技術がある。これらは、ヒトの錯覚を利用し聴き取りに影響を与えない特徴を、オリジナルの音声信号から削除・変形させている。そこで、振幅包絡の特徴を削除や変形といった制御をしても言語情報の取得に影響のない特徴に着目しなければならない。その特徴として、変調成分に着目した。振幅包絡の変調成分の周波数が $4 \sim 16$ Hz は音声了解度に重要であり、変調成分の周波数が $2 \sim 8$ Hz に特有のピークを持つと報告されている。しかしながら、この特徴を変化させたときに言語情報の取得にどのような影響を与えるのかは検討していない。そのため、振幅包絡に含まれる変調成分を系統的に制御し、言語情報の取得が可能であるかどうか調査する。

振幅包絡に含まれる変調成分を系統的に制御するにあたって、音声のモーラの時間構造とモーラの時間構造よりも短い時間長の子音の時間構造に着目した。モーラの時間構造に着目した聴取実験では、1 モーラの時間長を1周期としたときに得られる変調成分の上限周波数が言語情報の取得が可能である境界であることが分かった。子音の時間構造に着目した聴取実験では、子音を表現する振幅包絡を再現するだけの変調成分が必要と考え、変調成分の上限周波数は高いと予測した。しかし、聴き取り可能だった変調成分の上限周波数は低かった。これは、子音部の振幅包絡を再現しなくとも言語情報の取得が可能であることを示している。子音の言語情報の取得が可能であったのは、子音部の振幅包絡の立ち上がりにより言語情報の取得を可能としていたと考えられる。

言語情報の取得に必要な変調成分は、モーラの時間構造の観点からは、モーラの時間構造を再現できるだけの変調成分が保存されていればよく、子音（摩擦音）の観点からは、振幅包絡の立ち上がりの違いによって、言語情報の取得が可能であることが分かった。

これらの結果から、低域通過フィルタにより削除された変調成分が日本語音声の振幅包絡に含まれる冗長性だと解釈できる。

目次

第1章	序論	2
1.1	はじめに	2
1.2	音声の冗長性の定義	2
1.3	音声の振幅包絡	3
1.4	研究の目的	3
1.5	本論文の構成	6
第2章	日本語音声におけるモーラと子音	7
2.1	日本語音声のモーラ	7
2.2	日本語音声の母音・子音	7
2.2.1	母音	7
2.2.2	子音	7
2.3	振幅包絡の制御方法	11
2.3.1	雑音駆動音声	11
2.3.2	周波数帯域分割	11
2.3.3	振幅包絡の抽出方法	12
2.3.4	振幅包絡の制御方法	13
2.4	本研究の着眼点	13
2.4.1	振幅包絡に含まれる変調成分の制御によるモーラへの影響	13
2.4.2	子音の振幅包絡に含まれる変調成分の制御による子音知覚への影響	13
第3章	日本語音声のモーラの振幅包絡における時間的構造	15
3.1	日本語音声の振幅包絡の変調成分が言語情報の取得に与える影響の検討	15
3.1.1	実験の目的	15
3.1.2	実験参加者	15
3.1.3	刺激音	15
3.1.4	実験手続き	16
3.1.5	実験結果	16
3.1.6	考察	18
3.2	日本語音声の各周波数帯域の振幅包絡の変調成分が言語情報の取得に与える影響の検討	20
3.2.1	実験の目的	20

3.2.2	実験参加者	20
3.2.3	刺激音	20
3.2.4	実験手続き	20
3.2.5	実験結果	21
3.2.6	考察	21
3.3	まとめ	28
第4章	日本語音声の子音の振幅包絡における時間的構造	29
4.1	日本語音声の子音の振幅包絡の変調成分が言語情報の取得に与える影響の検討	29
4.1.1	実験の目的	29
4.1.2	実験参加者	29
4.1.3	刺激音	29
4.1.4	実験手続き	30
4.1.5	実験結果	33
4.1.6	考察	43
4.2	まとめ	43
第5章	総合検討	45
5.1	日本語音声の振幅包絡の時間的構造に対する検討	45
5.2	日本語音声の振幅包絡に含まれる冗長性に対する検討	45
第6章	結論	46
6.1	本研究で明らかになったこと	46
6.2	今後の展望	47
	謝辞	48
	参考文献	48
	付録	51
	学会発表リスト	176

目 次

1.1	本研究のアプローチの概念図	5
2.1	日本語音声の子音の分類	8
2.2	音声波形のタイプ分類	9
2.3	雑音駆動音声の作成のブロックダイアグラム	11
2.4	刺激音作成のブロックダイアグラム	13
2.5	振幅包絡の変調成分の制御による変化の例： $/za/$ の第 10 周波数帯域	14
3.1	実験環境	17
3.2	低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化	19
3.3	第 1 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化	22
3.4	第 5 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化	23
3.5	第 10 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化	24
3.6	第 1 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化	25
3.7	第 5 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化	26
3.8	第 10 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化	27
4.1	$/sa/$ の音声波形と振幅包絡	31
4.2	$/za/$ の音声波形と振幅包絡	32
4.3	変調成分を変化させたときの $/a/$ の正答率の変化	34
4.4	変調成分を変化させたときの $/ha/$ の正答率の変化	35
4.5	変調成分を変化させたときの $/sa/$ の正答率の変化	36
4.6	変調成分を変化させたときの $/za/$ の正答率の変化	37
4.7	$/sa/$ と $/za/$ のフィッティング曲線	38
4.8	原信号 $/a/$ に対する回答の割合の変化	39
4.9	原信号 $/ha/$ に対する回答の割合の変化	40

4.10	原信号/sa/に対する回答の割合の変化	41
4.11	原信号/za/に対する回答の割合の変化	42
4.12	変調成分の上限周波数が (a) 2 Hz と (b) 5 Hz のときの /sa/ と /za/ の振幅包絡の形状	44
6.1	第1周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	51
6.2	第2周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	52
6.3	第3周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	53
6.4	第4周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	54
6.5	第5周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	55
6.6	第6周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	56
6.7	第7周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	57
6.8	第8周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	58
6.9	第9周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	59
6.10	第10周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化	60
6.11	第1周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	61
6.12	第2周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	62
6.13	第3周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	63
6.14	第4周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	64
6.15	第5周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	65
6.16	第6周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	66
6.17	第7周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化	67

6.18	第 8 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均 正答モーラ数の変化	68
6.19	第 9 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均 正答モーラ数の変化	69
6.20	第 10 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平 均正答モーラ数の変化	70
6.21	/a/ の音声波形	71
6.22	/ka/ の音声波形	72
6.23	/sa/ の音声波形	73
6.24	/ta/ の音声波形	74
6.25	/na/ の音声波形	75
6.26	/ha/ の音声波形	76
6.27	/ma/ の音声波形	77
6.28	/ya/ の音声波形	78
6.29	/ra/ の音声波形	79
6.30	/wa/ の音声波形	80
6.31	/ga/ の音声波形	81
6.32	/za/ の音声波形	82
6.33	/da/ の音声波形	83
6.34	/ba/ の音声波形	84
6.35	/pa/ の音声波形	85
6.36	/kya/ の音声波形	86
6.37	/sya/ の音声波形	87
6.38	/tya/ の音声波形	88
6.39	/nya/ の音声波形	89
6.40	/hya/ の音声波形	90
6.41	/mya/ の音声波形	91
6.42	/rya/ の音声波形	92
6.43	/gya/ の音声波形	93
6.44	/zya/ の音声波形	94
6.45	/bya/ の音声波形	95
6.46	/pya/ の音声波形	96
6.47	/a/ を 10 個の周波数帯域分割した際の各振幅包絡	97
6.48	/ka/ を 10 個の周波数帯域分割した際の各振幅包絡	98
6.49	/sa/ を 10 個の周波数帯域分割した際の各振幅包絡	99
6.50	/ta/ を 10 個の周波数帯域分割した際の各振幅包絡	100
6.51	/na/ を 10 個の周波数帯域分割した際の各振幅包絡	101
6.52	/ha/ を 10 個の周波数帯域分割した際の各振幅包絡	102

6.53	/ma/ を 10 個の周波数帯域分割した際の各振幅包絡	103
6.54	/ya/ を 10 個の周波数帯域分割した際の各振幅包絡	104
6.55	/ra/ を 10 個の周波数帯域分割した際の各振幅包絡	105
6.56	/wa/ を 10 個の周波数帯域分割した際の各振幅包絡	106
6.57	/ga/ を 10 個の周波数帯域分割した際の各振幅包絡	107
6.58	/za/ を 10 個の周波数帯域分割した際の各振幅包絡	108
6.59	/da/ を 10 個の周波数帯域分割した際の各振幅包絡	109
6.60	/ba/ を 10 個の周波数帯域分割した際の各振幅包絡	110
6.61	/pa/ を 10 個の周波数帯域分割した際の各振幅包絡	111
6.62	/kya/ を 10 個の周波数帯域分割した際の各振幅包絡	112
6.63	/sya/ を 10 個の周波数帯域分割した際の各振幅包絡	113
6.64	/tya/ を 10 個の周波数帯域分割した際の各振幅包絡	114
6.65	/nya/ を 10 個の周波数帯域分割した際の各振幅包絡	115
6.66	/hya/ を 10 個の周波数帯域分割した際の各振幅包絡	116
6.67	/mya/ を 10 個の周波数帯域分割した際の各振幅包絡	117
6.68	/rya/ を 10 個の周波数帯域分割した際の各振幅包絡	118
6.69	/gya/ を 10 個の周波数帯域分割した際の各振幅包絡	119
6.70	/zya/ を 10 個の周波数帯域分割した際の各振幅包絡	120
6.71	/bya/ を 10 個の周波数帯域分割した際の各振幅包絡	121
6.72	/pya/ を 10 個の周波数帯域分割した際の各振幅包絡	122
6.73	/sa/ の第 1 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	123
6.74	/za/ の第 1 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	124
6.75	/sa/ の第 2 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	125
6.76	/za/ の第 2 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	126
6.77	/sa/ の第 3 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	127
6.78	/za/ の第 3 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	128
6.79	/sa/ の第 4 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	129
6.80	/za/ の第 4 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	130
6.81	/sa/ の第 5 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数 で制御した場合の変化	131

6.82	/za/ の第 5 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	132
6.83	/sa/ の第 6 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	133
6.84	/za/ の第 6 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	134
6.85	/sa/ の第 7 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	135
6.86	/za/ の第 7 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	136
6.87	/sa/ の第 8 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	137
6.88	/za/ の第 8 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	138
6.89	/sa/ の第 9 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	139
6.90	/za/ の第 9 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	140
6.91	/sa/ の第 10 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	141
6.92	/za/ の第 10 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化	142
6.93	被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)	143
6.94	被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)	144
6.95	被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)	145
6.96	被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)	146
6.97	被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)	147
6.98	被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)	148
6.99	被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)	149
6.100	被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)	150

6.101被験者 C：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）	151
6.102被験者 C：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /sa/）	152
6.103被験者 C：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /ha/）	153
6.104被験者 C：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /za/）	154
6.105被験者 D：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）	155
6.106被験者 D：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /sa/）	156
6.107被験者 D：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /ha/）	157
6.108被験者 D：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /za/）	158
6.109被験者 E：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）	159
6.110被験者 E：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /sa/）	160
6.111被験者 E：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /ha/）	161
6.112被験者 E：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /za/）	162
6.113被験者 F：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）	163
6.114被験者 F：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /sa/）	164
6.115被験者 F：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /ha/）	165
6.116被験者 F：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /za/）	166
6.117被験者 G：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）	167
6.118被験者 G：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /sa/）	168
6.119被験者 G：低域通過フィルタのカットオフ周波数に対する正答率（原信号 /ha/）	169

6.120被験者 G : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)	170
6.121被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)	171
6.122被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)	172
6.123被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)	173
6.124被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)	174

表 目 次

2.1 周波数帯域分割の境界周波数	12
6.1 ヘッドホンの周波数特性	175

第1章 序論

1.1 はじめに

音声は、知覚的に頑健である。たとえば、電話はオリジナルの音声から 300 ~ 3400 Hz の周波数帯域の情報（帯域制限された音声）のみを伝送し、会話を成立させている。帯域制限された音声、つまり、オリジナルの音声のある程度壊しているにも関わらず、何と言っているのかを我々は正確に聴き取ることができる（言語情報を正確に取得することができる）。また、音声信号の一部が削除され、明らかに了解度が低下する場合でも、削除された部分に別の音（たとえば白色雑音）が挿入されると、あたかも音声が無傷であるかのように知覚される。この現象を音韻修復という [1]。これらは、言語情報に関する重要な特徴が音声信号の至るところに分散して冗長に含まれているため、その一部を削除・変形しても我々は容易に言語情報を取得することができることを意味している。このように、削除・変形しても言語情報の取得に影響を与えない特徴のことを、音声の冗長性という。音声の冗長性を生かした技術として、音声圧縮や音声符号化技術がある [2]。これらの技術は、音声の冗長性を削除することで実現している。そのため、音声の冗長性の解明は、音声情報処理技術の発展につながる重要な課題である。

これまでに、音声の了解性に関する情報は、音声の時間包絡線（振幅包絡）に多く含まれていることが報告されている。その根拠の一つとして、雑音駆動音声のように無声音であっても言語情報を正確に聴き取ることができることが知られている。これらの研究では、音声の振幅包絡の情報のみで、言語情報が取得できることを示している。しかしながら、その振幅包絡の情報に冗長性があるのかどうかまで踏み込んで検討されていない。

1.2 音声の冗長性の定義

ヒトは、音声知覚に関係した特徴は音声信号に分散して冗長に含まれており、その中のいくつかが失われたとしても何と話しているのか正確に聴き取ることができる [3]。このように、オリジナルの音声信号から失われた特徴が音声知覚に影響を与えないことを音声の冗長性と定義する。

1.3 音声の振幅包絡

音声の振幅包絡には，音声の了解性や個人性に関係する情報が多く含まれていると報告されている．

まず，音声の了解性が振幅包絡に含まれていることについて述べる．Drullman は，100 ～ 6400 Hz の範囲を 1/4 オクターブ帯域で分割し各帯域から振幅包絡を抽出し，同じ帯域幅の帯域雑音を変調した音声は了解性があると報告している [4]．Shannon ら [5] は，振幅包絡の情報のみを保存している雑音駆動音声を用いて，振幅包絡の時間変動が英語音声知覚にどのように影響するのかを調査した．彼らは，雑音駆動音声を作成する際の周波数帯域分割数（分割数に合わせて周波数帯域幅をあるルールで変調）と振幅包絡を抽出する際の低域通過フィルタのカットオフ周波数とを系統的に変化させ，それらの組み合わせが言語情報の取得に与える影響を調査した．その結果，四つの周波数帯域の振幅包絡が保存されていれば，言語情報の聴き取りが十分可能であることを示した．この結果は，時間微細構造がない音声でも，振幅包絡さえ保存されていれば言語情報が取得可能であることを示しており，振幅包絡にも言語情報が含まれていると解釈することができる．また，同様の研究が日本語についても行われており，日本語でも 四つの周波数帯域の振幅包絡情報のみで言語情報の聴き取りが可能であることが示されている [6]．さらに，上田 & 中島 [7] は，20 個の周波数帯域に分割した音声の各周波数帯域の振幅包絡変動を变量として因子分析を行った．その結果，得られた各周波数帯域の因子得点係数のグラフの交差点が，雑音駆動音声の研究で明らかになった 四つの周波数帯域の境界周波数とほぼ同じであることを示した．この結果は，日本語だけでなく，英語やドイツ語といった他の言語を用いても同様に得られている [8]．これは，四つの周波数帯域の振幅包絡に含まれる変動要因になんらかの意味があることを示唆している．

次に，音声の個人性が振幅包絡に含まれていることについて述べる．風間らは，狭帯域音声波形包絡線の帯域間相互相関行列を用いた発話独立型の話者識別が可能であることを報告している [9]．

最後に，Drullman らによって，変調成分が 4 ～ 16 Hz は音声了解度に需要であると報告している [10, 11]．Arai ら [12] は，変調成分の周波数が 2 ～ 8 Hz に特有のピークを持つと報告している．しかしながら，この特徴を変化させたときに音声知覚がどのように変化するのかについては検討していないため，振幅包絡に含まれる変調成分に着目し，研究を進める．

1.4 研究の目的

上述したように，音声の振幅包絡には音声の了解性が含まれていることが報告されている．しかし，音声の言語情報を取得するにあたって，この振幅包絡に冗長性があるのかどうかまでは検討されていない．また，野口ら [13] は，雑音駆動音声の四つの周波数帯域の一つを除去した際の言語情報の聴き取りを調査した．その結果，第 1 帯域 (50 ～ 570

Hz) の除去は子音の伝達量の減少を招き，第 2 帯域 (570 ~ 1850 Hz) の除去は母音の情報量を減少させ，第 3 帯域 (1850 ~ 4000 Hz) の除去はいくつかの母音の知覚に対して若干影響を与え (大きな影響ではない)，第 4 帯域 (4000 ~ 7000 Hz) の除去は全体的に情報量を減少させることを明らかにした．同様の周波数帯域分割で，複数の周波数帯域の除去が雑音駆動音声の知覚に与える影響を伊佐次ら [14] によって，報告されている．この結果は，(1) 子音の有声性は第 1 帯域 (50 ~ 570 Hz) とその他の周波数帯域との時間的な関係が重要であること，(2) 子音の調音方法，調音位置は第 1 周波数帯域の役割が重要であり，帯域除去によって伝達情報量が大きく低下し，残存する周波数帯域が 2 つ以下になると，ほとんど知覚が困難になること，(3) 母音については，特定の帯域を除去することで様々な異聴が引き起こされること，を明らかにした．これらの研究では，各周波数帯域の情報が言語情報の取得に与える影響を明らかにしている．

しかし，周波数帯域を除去するという制御では，全体のスペクトル包絡を変形させてしまう (全周波数帯域のスペクトル包絡の形状を眺めたときに，該当箇所にくぼみを作ってしまう，見かけ上そのまわりにスペクトルピークがあるような構成をつくってしまう)，そのため，得られた結果が，純粹に振幅包絡の変動の違いによる影響なのか，スペクトル包絡形状にみられる急激な変動の違いによる影響なのか，切り分けて議論することができない．また，周波数帯域の除去では振幅包絡のいかなる特徴が言語情報の取得に影響を与えているのかも詳細に調査することができない．

そこで本研究は，上述した問題意識をもって，音声の振幅包絡に含まれる冗長性に関する検討することが目的である．そのために，音声の振幅包絡に含まれる特徴を系統的に変化させに言語情報の取得に与える影響を調査し，言語情報の取得に重要な特徴を明らかにする．そして，言語情報の取得に影響を与えない特徴が，音声の冗長性であると推定する．図 1.1 に本研究のアプローチの概念図を示す．

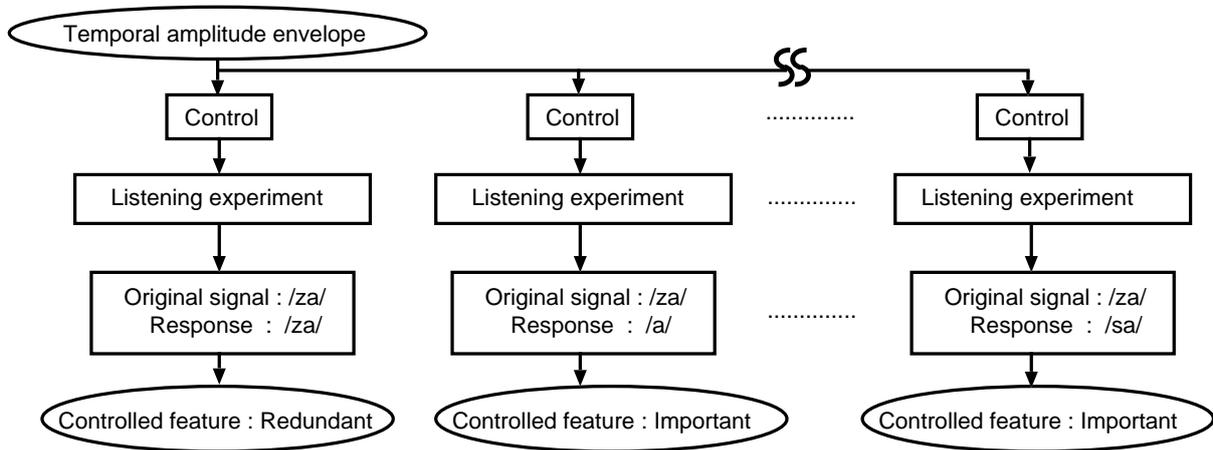


図 1.1: 本研究のアプローチの概念図

1.5 本論文の構成

本論文は、6章で構成される。

第1章

この章では、研究の背景、音声の冗長性の定義、音声の振幅包絡、研究の目的を述べた。これらにより、本論文で音声の振幅包絡に含まれる冗長性に関する検討を行うことを示した。

第2章

日本語音声のモーラ、母音・子音について述べる。そして、振幅包絡の制御方法について述べ、この制御方法によって日本語音声のモーラ、子音の知覚に与える影響を予想する。

第3章

日本語音声の振幅包絡の変調成分が言語情報の取得に与える影響を検討する。そのために第2章で説明した振幅包絡の制御方法に則って、刺激音の作成と実験手続きを設定し、二つの聴取実験を行う。そして、日本語音声のモーラにおける時間的構造に着目した際の言語情報の取得に与える影響を考察する。

第4章

日本語音声の子音の振幅包絡の変調成分が言語情報の取得に与える影響を検討する。そのために第2章で説明した時間的構造の制御方法に則って、刺激音の作成と実験手続きを設定し、聴取実験を行う。このとき、子音には摩擦音を使用する由を述べる。そして、日本語音声の子音における時間的構造に着目した際の言語情報の取得に与える影響を考察する。

第5章

第3章と第4章で得られた結果より、日本語音声の振幅包絡の変調成分が言語情報の取得に与える影響を総合検討する。そして、言語情報の取得に与える影響から、音声の振幅包絡に含まれる冗長性について検討する。

第6章

本研究で得られた結果を要約し、今後の展望を述べる。

第2章 日本語音声におけるモーラと子音

2.1 日本語音声のモーラ

日本語音声のリズムを記述する一つとして「モーラ」という単位がある。モーラとは、日本語音声学で「拍」と呼ばれており、同時的なリズムの単位である [15]。たとえば、日本語の伝統的詩歌（俳句・川柳・短歌）の五七五，五七五七七というリズムは、モーラを単位として数えられた長さである。長音（ー：長い持続時間で発音される音），撥音（ん），促音（っ）のように語頭に立たない音でも1モーラとして数える [16]。

次の節では、モーラを形成している「母音・子音」について述べる。

2.2 日本語音声の母音・子音

2.2.1 母音

日本語音声の母音は、 $/a/$ 、 $/i/$ 、 $/u/$ 、 $/e/$ 、 $/o/$ の5個である [17]。声帯の振動による周期的なパルスで、固定された声道を共振させて、母音を生成する。また、この声道の形状の断面積によって、フォルマント周波数が決められる。主に、舌の調音位置で生成される母音が決めるが、顎や唇の位置も、生成された母音に影響を与える。また、軟口蓋の位置もわずかながら生成された母音に影響を与える。

母音は子音に比べて、通常長い継続時間を持ち、スペクトルも比較的明確である。このように、母音は通常容易にかつ確実に認識されることができるので、人間による音声認識でも、機械による音声認識でも、重要な役割を果たしている。母音の調音形状は、舌の調音位置（前舌，中舌，後舌）と舌の調音位置の高さ（狭母音，中母音，広母音）によって分類される。この分類に従うと、 $/i/$ と $/e/$ は前舌母音， $/a/$ は中舌母音， $/u/$ と $/o/$ は後舌母音となる。前舌母音は高い周波数で共振しており，中舌母音は広い周波数領域にわたってエネルギーが均衡しており，後舌母音は低い周波数情報が優勢である [16]。

2.2.2 子音

日本語音声の子音は、 $/k/$ 、 $/s/$ 、 $/t/$ 、 $/n/$ 、 $/h/$ 、 $/m/$ 、 $/y/$ 、 $/r/$ 、 $/w/$ 、 $/g/$ 、 $/z/$ 、 $/d/$ 、 $/b/$ 、 $/p/$ 、 $/ky/$ 、 $/sy/$ 、 $/ty/$ 、 $/ny/$ 、 $/hy/$ 、 $/my/$ 、 $/ry/$ 、 $/gy/$ 、 $/zy/$ 、 $/by/$ 、 $/py/$ の25個である。これらは、音声生成の観点から、共鳴音と阻害音に分類される。さらに、共鳴

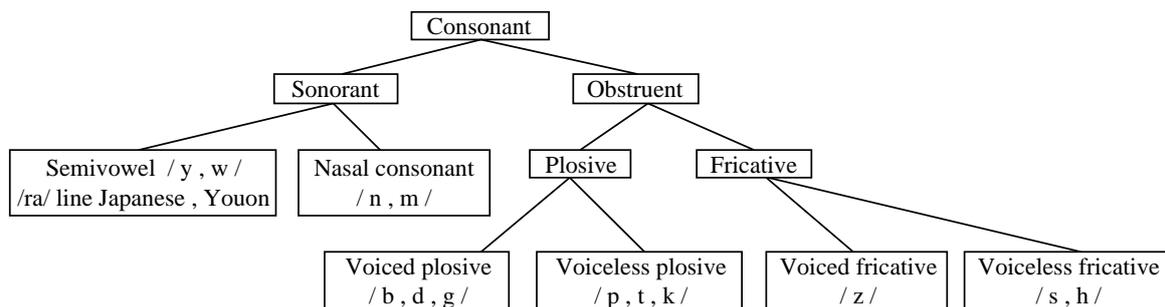


図 2.1: 日本語音声の子音の分類

音は、半母音・鼻子音・ラ行・拗音に、阻害音は、摩擦音・破裂音に分類される [18, 19] . この分類の系統図を図 2.1 に示す .

本研究では、音声の振幅包絡の形状に着目するため、これらの子音を音声波形の形状の違いにより、拗音タイプ、半母音・鼻子音・ラ行タイプ、破裂音タイプ、摩擦音タイプの四つに分けた . それぞれのタイプの代表的な波形を図 2.2 に示す .

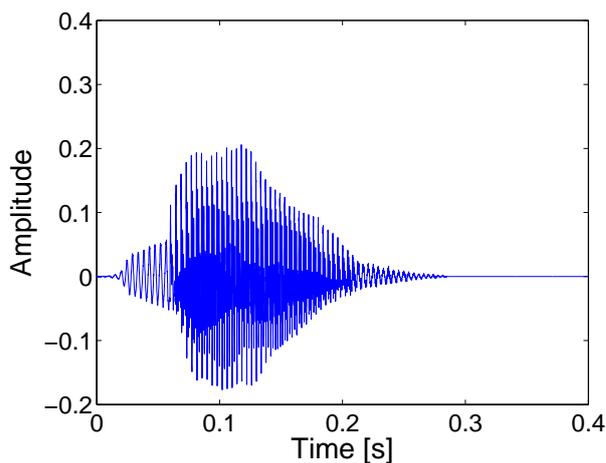
半母音・鼻子音・ラ行

半母音には、ヤ行音 /y/ とワ行音 /w/ があり、渡りのある母音のような音韻であるため、半母音と呼ばれている . /y/ と /w/ は、調音の構えとしては母音の /i/ と /u/ に類似しているが、調音位置での声道の狭めが母音よりやや強いのが特徴である . どのように発話されているか /y/ を例に説明する . 例えば「や」であれば、/ya/ である . この場合、/y/ は母音 /e/ (/i/) 付近から母音 /a/ へと緩やかに変化し、発話される . また、日本語の音声には、拗音と呼ばれる子音と半母音 /y/ の結合で作られる子音がある . /ky/, /ny/, /hy/, /my/, /ry/, /gy/, /by/, /py/ の音声である . 特徴は、前半の子音に続いて、/y/ の特徴である /e/ (/i/) から後続母音 (/a/, /u/, /o/) への緩やかな変化である . そのため、連続母音 (例えば、/ia/, /iu/, /ea/) との区別は困難である .

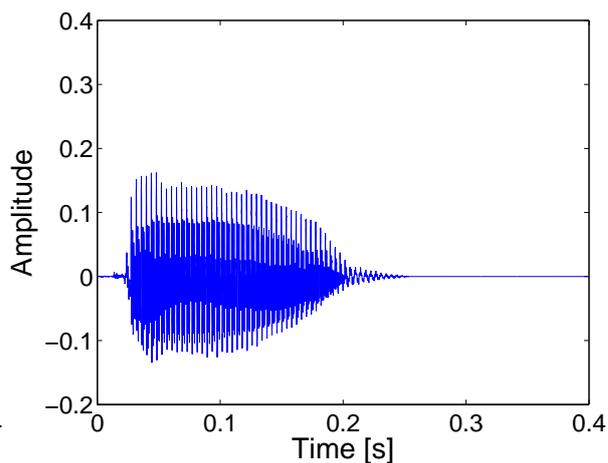
鼻子音には、マ行音 /m/, ナ行音 /n/ がある . 声帯の振動を持ち、声道が途中で閉鎖され、軟口蓋が下げられるので、声道から音波が放出される . 口腔は前方へは閉鎖されているが、咽頭とは音響的に結合されている . このようにして、口腔は共鳴箱として働き、共振周波数を持つ音響エネルギーがトラップされる . 放射される音に関しては、口腔の共振周波数は、反共振として現れる . つまり、音の伝達関数における零点に対応する .

日本語のラ行音 /r/ は、音素として生起頻度は /k/ などと共に高いが、その音響的特徴の研究は比較的少ない . 音声学的には、弾き音、側面音、ふるえ音などに分類される . 音韻環境や人によって特徴が大きく変わり、連続音声中では音響的变化が小さく、検出が難しい場合もある . 調音点の類似から音響的特徴が /n/, /d/ と近く、それらとの類別はかなり困難である .

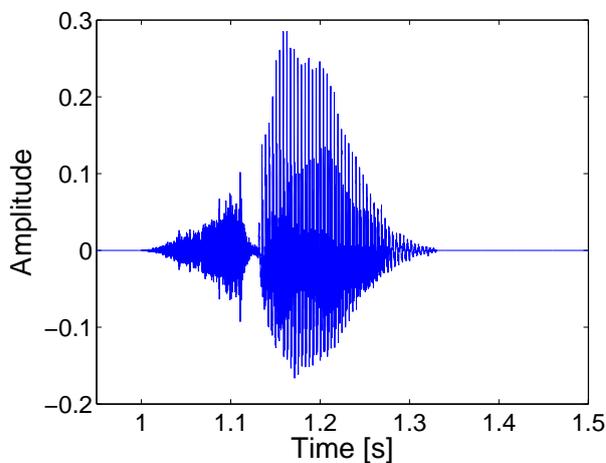
これらの音声は、共鳴音に分類される . 図 2.2(a) に示すように、半母音・鼻子音・ラ行は、後続母音 /a/ が発声される前にしっかりと子音部分が発声されている . しかし、子



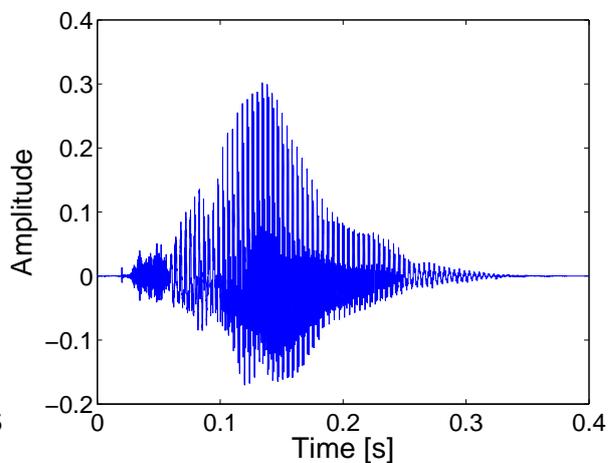
(a) 半母音・鼻子音・ラ行タイプの音声波形：/ma/



(b) 破裂音タイプの音声波形：/pa/



(c) 摩擦音タイプの音声波形：/sa/



(d) 拗音タイプの音声波形：/zya/

図 2.2: 音声波形のタイプ分類

音部分の時間長や立ち上がり部分の形状は、子音の種類によらずおおよそ同じであり、振幅包絡の形状に大きな違いがみられない。

破裂音

破裂音には、無声破裂音 パ行音 /p/、タ行音 /t/、カ行音 /k/ と有声破裂音 バ行音 /b/、ダ行音 /d/、ガ行音 /g/ がある。無声破裂音を時間軸に沿って観察すると、無音部、破裂部、気音部、母音と続く。これに対して有声破裂音は、破裂の時点で声帯が振動している音であり、声帯のバズ音、破裂部、母音と続く。無声破裂音と有声破裂音の差異は、基本的には破裂の瞬間から声帯が振動を始まるまでの時間 voice onset time (VOT) があり、欧米語では、VOT が 20 ms を境にして、これより長い場合が無声破裂音、短い場合が有声破裂音とされる。しかし、日本語の通常の発声では必ずしもこの数値が当てはまらず、VOT が 10 ms 前後の無声破裂音も /p/ などに多い。

図 2.2(b) に示すように、破裂音は、後続母音 /a/ が発声される前に、子音を表現する破裂部が発声される。この破裂部の時間間隔は非常に短く、急峻な立ち上がり立ち下がりを持っている。

摩擦音

摩擦音には、無声摩擦音であるサ行音 /s/ とハ行音 /h/、有声摩擦音のザ行音 /z/ がある。摩擦音は、声帯からの定常的な気流を、狭めの持った声道に吹きこんで生成される。/h/ の生成は狭めは唇の近く、/s/ の生成は口腔の中央付近である。このように、無声摩擦音の生成系は、狭めでの雑音源とこの狭めによって分けられた二つの空洞を持つことになる。音声は唇から放出されるので、前の空洞から放出されるので、後ろの空洞は、鼻子音の場合と同様にエネルギーをトラップする働きをし、音声中に半共振を生じさせる。/z/ は、/s/ に対応している。ただし、/z/ は声帯が振動しているので、低域のエネルギーが観測できる。また、声門と唇の間に存在する声道の狭めで、気流は周期性をもつ乱流となる。よって、有声摩擦音のスペクトルには二つの明確なスペクトル要素を持つことが予想される。

図 2.2(c) に示すように、摩擦音は、摩擦部がある程度の長さを持って発声されている。また、摩擦音には /ha/、/sa/、/za/ があり、それぞれの子音部分の時間長と摩擦部の立ち上がりの鋭さが異なる。

拗音

拗音には、/ky/、/sy/、/ty/、/ny/、/hy/、/my/、/ry/、/gy/、/zy/、/by/、/py/ がある。図 2.2(d) に示すように、拗音は、子音 + 半母音 + 後続母音の構成をとる。そのため、他の子音より複雑な構成をしている。

2.3 振幅包絡の制御方法

2.3.1 雑音駆動音声

本研究では，音声の振幅包絡に着目するために，振幅包絡の情報のみを保存している雑音駆動音声の作成方法を基にし，刺激音を作成する [20, 21] .

雑音駆動音声とは，時間微細構造を持たず，雑音を時間的な振幅包絡で変調した音声のことである．この音声は，(1) 振幅変調モデルを仮定し，(2) 原音声を複数の周波数帯域に分割し，(3) 各周波数帯域の振幅を抽出し，(4) 時間微細構造（キャリア信号に相当）を帯域制限されたガウス性白色雑音に置き換え，(5) 両者を掛け合わせて再合成することで得られる．図 2.3 に雑音駆動音声の作成のブロックダイアグラムを示す．

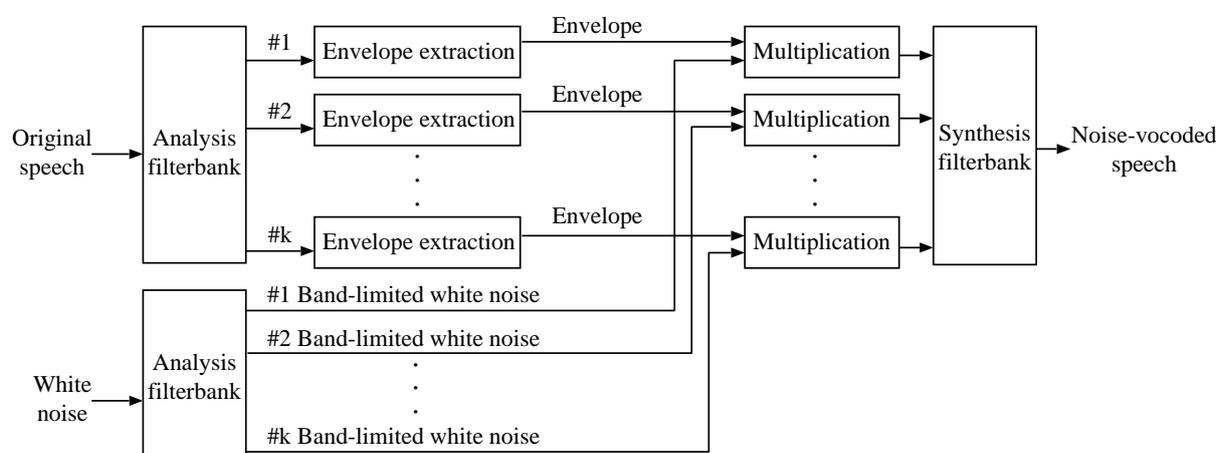


図 2.3: 雑音駆動音声の作成のブロックダイアグラム

2.3.2 周波数帯域分割

雑音駆動音声の先行研究では，周波数帯域分割方法として様々な分割の仕方がある [5, 6] . 例えば，野口ら [13] による分割方法は，50 ~ 570 Hz , 570 ~ 1850 Hz , 1850 ~ 4000 Hz , 4000 ~ 7000 Hz の四つの周波数帯域である．しかし，この分割方法では，周波数帯域幅が広すぎるため，どの周波数帯域の振幅包絡に，言語情報が含まれているのか詳細に調査することができない．そこで，本研究では，振幅包絡の変動をより細かく制御できるように周波数帯域分割数を 10 とした．また，人間の聴覚機構をできるだけ忠実に模擬するために，聴覚フィルタモデルに基づいて周波数帯域分割を行った．この聴覚フィルタの周波数帯域幅は，等価矩形帯域幅（ ERB_N : equivalent rectangular bandwidth）で表現される [22] . この ERB_N と中心周波数 f Hz との関係式を下記に示す．

$$ERB_N = 24.7 \left(\frac{4.37f}{1000} + 1 \right) \quad [\text{Hz}] \quad (2.1)$$

表 2.1: 周波数帯域分割の境界周波数

Band number	ERB _N -number	Frequency [Hz]
第 1 周波数帯域	2 ~ 5	54.94 ~ 163.06
第 2 周波数帯域	5 ~ 8	163.06 ~ 312.26
第 3 周波数帯域	8 ~ 11	312.26 ~ 518.54
第 4 周波数帯域	11 ~ 14	518.54 ~ 803.27
第 5 周波数帯域	14 ~ 17	803.27 ~ 1196.48
第 6 周波数帯域	17 ~ 20	1196.48 ~ 1739.50
第 7 周波数帯域	20 ~ 23	1739.50 ~ 2489.39
第 8 周波数帯域	23 ~ 26	2489.39 ~ 3524.97
第 9 周波数帯域	26 ~ 29	3524.97 ~ 4955.09
第 10 周波数帯域	29 ~ 32	4955.09 ~ 6930.06

また，ERB_N の幅を 1 として周波数軸を変形した ERB_N-number は下記の式で示される．

$$\text{ERB}_N\text{-number} = 21.4 \log_{10} \left(\frac{4.37f}{1000} + 1 \right) \quad (2.2)$$

この ERB_N-number を用いて，周波数帯域分割を行う．人間の聴覚機構を忠実に模擬するためには，各 ERB_N-number の周波数帯域幅をもつ聴覚フィルタで分割した音声を個別に操作しなければならない．しかし，予備検討を行ったところ，音声の周波数構造が忠実に再現されすぎており，そこに含まれる言語情報量が多いため，振幅包絡の変化による影響を効果的に示すことができないことが分かった．これは，音声の冗長性を示す良い例であるが，振幅包絡に含まれる情報の変化を観察する本研究の目的にそぐわない．そこで，音声の周波数スペクトルに含まれる言語情報のある程度取り除くために，ERB_N-number が，2 から 32 まで，3 ずつひとまとめにして分割し，総周波数帯域分割数は 10 個とした．ERB_N-number と周波数帯域幅の対応を，表 2.1 に示す．また，各周波数帯域を，低い周波数帯域から第 1 周波数帯域，第 2 周波数帯域，…，第 10 周波数帯域と呼ぶ．

2.3.3 振幅包絡の抽出方法

周波数帯域分割した 10 個の周波数帯域から，次式を用いて振幅包絡 $e_k(t)$ を抽出した [23] ．

$$e_k(t) = \text{LPF}[|x_k(t) + j\text{Hilbert}(x_k(t))|] \quad (2.3)$$

ただし， $x_k(t)$ は 周波数帯域分割したときの各信号， $\text{Hilbert}(\cdot)$ は Hilbert 変換， $\text{LPF}[\cdot]$ は低域通過フィルタ， k は分割番号 ($k = 1, 2, \dots, 10$) である．

2.3.4 振幅包絡の制御方法

低域通過フィルタを用いて式 (2.3) によって抽出した振幅包絡に含まれる変調成分を制御した。ただし、カットオフ周波数が低い低域通過フィルタのフィルタ特性を向上させるために、原信号のサンプリング周波数から 250 Hz にダウンサンプリングした上で、低域通過フィルタに通した。制御した後は、250 Hz から原信号のサンプリング周波数にアップサンプリングした。この制御した振幅包絡を用いて、同周波数帯域幅の帯域雑音を変調し再合成することで、刺激音を作成した。図 2.3 に刺激音作成のブロックダイアグラムを示す。

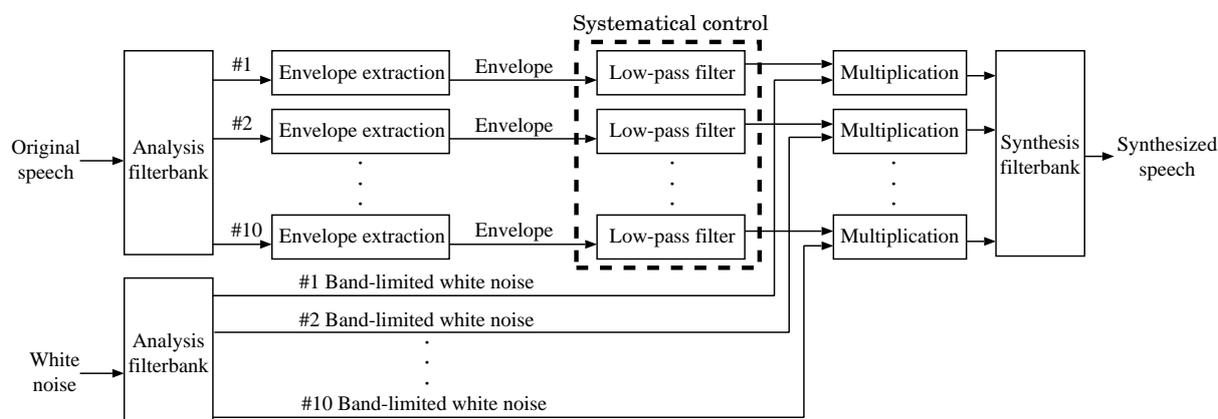


図 2.4: 刺激音作成のブロックダイアグラム

2.4 本研究の着眼点

2.4.1 振幅包絡に含まれる変調成分の制御によるモーラへの影響

モーラに着目するため、4 モーラ単語の振幅包絡に含まれる変調成分を制御し、言語情報の取得に与える影響を調査する。この単語の振幅包絡の変調成分の上限周波数を低くしていくと、オリジナルの音声の振幅包絡に含まれる変調成分を徐々に減少させているため、言語情報の取得は段階的にできなくなると予測する。

2.4.2 子音の振幅包絡に含まれる変調成分の制御による子音知覚への影響

モーラよりも短い時間構造に着目するため、単音節音声 (1 モーラ) の振幅包絡に含まれる変調成分を制御し、言語情報の取得に与える影響を調査する。上記と同様に、この音声の振幅包絡の変調成分の上限周波数を低くしていくと、オリジナルの音声の振幅包絡に含まれる変調成分を徐々に減少させているため、言語情報の取得は段階的にできなくなると予測する。

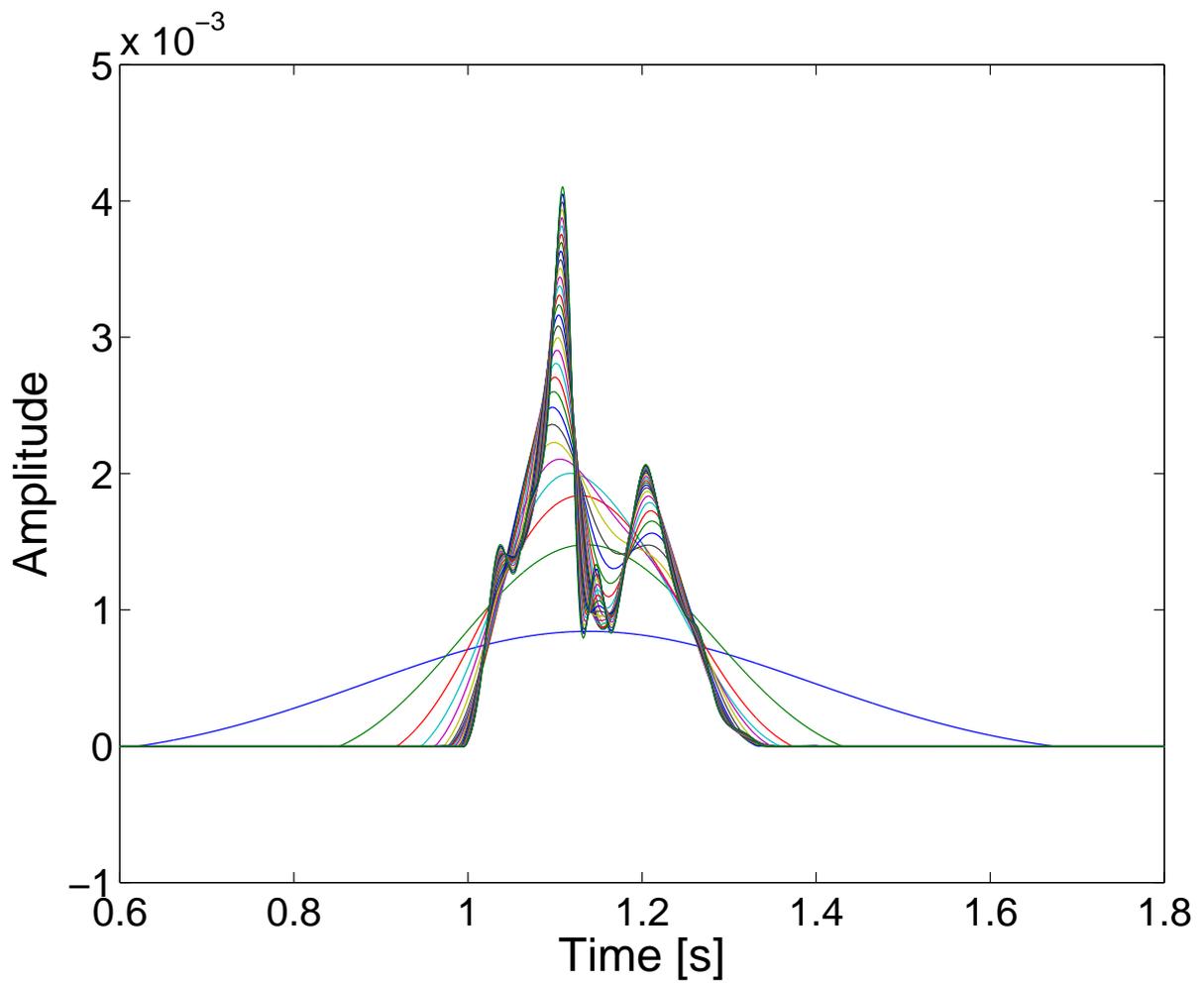


図 2.5: 振幅包絡の変調成分の制御による変化の例 : /za/ の第 10 周波数帯域

第3章 日本語音声のモーラの振幅包絡における時間的構造

3.1 日本語音声の振幅包絡の変調成分が言語情報の取得に与える影響の検討

3.1.1 実験の目的

本実験の目的は、日本語音声（4 モーラ単語）の振幅包絡に含まれる変調成分を制御することで、言語情報の取得がどのように変化するのかを聴取実験により確かめることである。

3.1.2 実験参加者

男性 4 名の実験参加者に対して、聴取実験を行った。実験参加者全員に対して標準的な聴力検査（RION AA-72B オーディオメータを使用）を両耳に行い、125 ~ 8000 Hz 間のオクターブ周波数に対して正常な聴力（聴力レベルで 12 dB 以下）を有することを確認した。

3.1.3 刺激音

原音声信号を 2.3.2 で示した周波数分割の方法で 10 個の周波数帯域に分割し、2.3.3 で示した振幅包絡の抽出方法で振幅包絡を抽出した。得られた各振幅包絡を制御するために 2.3.4 で示した方法で、各振幅包絡に含まれる変調成分を制御した。このとき、低域通過フィルタのカットオフ周波数は、1 Hz から 30 Hz まで、1 Hz 刻みとし、各振幅包絡に含まれる変調成分を系統的に変化させた。そして、この制御された各振幅包絡を用いて、同周波数帯域幅の帯域雑音を変調し再合成することで、刺激音を作成した。

原信号として、親密度別単語了解度試験用データセット（FW03）[24] から親密度が 1.0 ~ 2.5 の 4 モーラ単語を用いた。文章や親密度が高い単語であると、前後の言語情報によって、類推や補完が生じる可能性があるため、親密度の低い単語を用いた。このデータセットは、サンプリング周波数が 48000 Hz、量子化精度が 16 bit である。また、音声の前後に無音区間を設け、刺激音の長さが 4000 ms になるようにした。無音区間は、原信

号の前に 1000 ms, 原信号の後に 4000 ms になるまでとした。なお, このデータセットを使って了解度を計算するためには, データセットをすべて使う必要がある。ここでは, 実験の目的, および実験にかかる時間を勘案し, その一部のみを使用した。そのため, 単語了解度を算出することはできない点に注意されたい [25]。

3.1.4 実験手続き

実験方法

実験参加者には, 全 180 個 (30 (振幅包絡の制御数) × 6 (振幅包絡を制御する低域通過フィルタの各カットオフ周波数に用意された音声の数)) をランダムに呈示し, 刺激音の言語情報がどのような内容であったかを回答させた。回答は, PC のキーボードで行った。刺激音 180 個を 1 セット 45 個に分割し, 各実験参加者は 5 セット行った。それぞれのセットの間には休憩時間を設けた。また, 実験の始めには, 安定した実験結果を得るために 1 セットのウォーミングアップ試行を行った。

分析方法

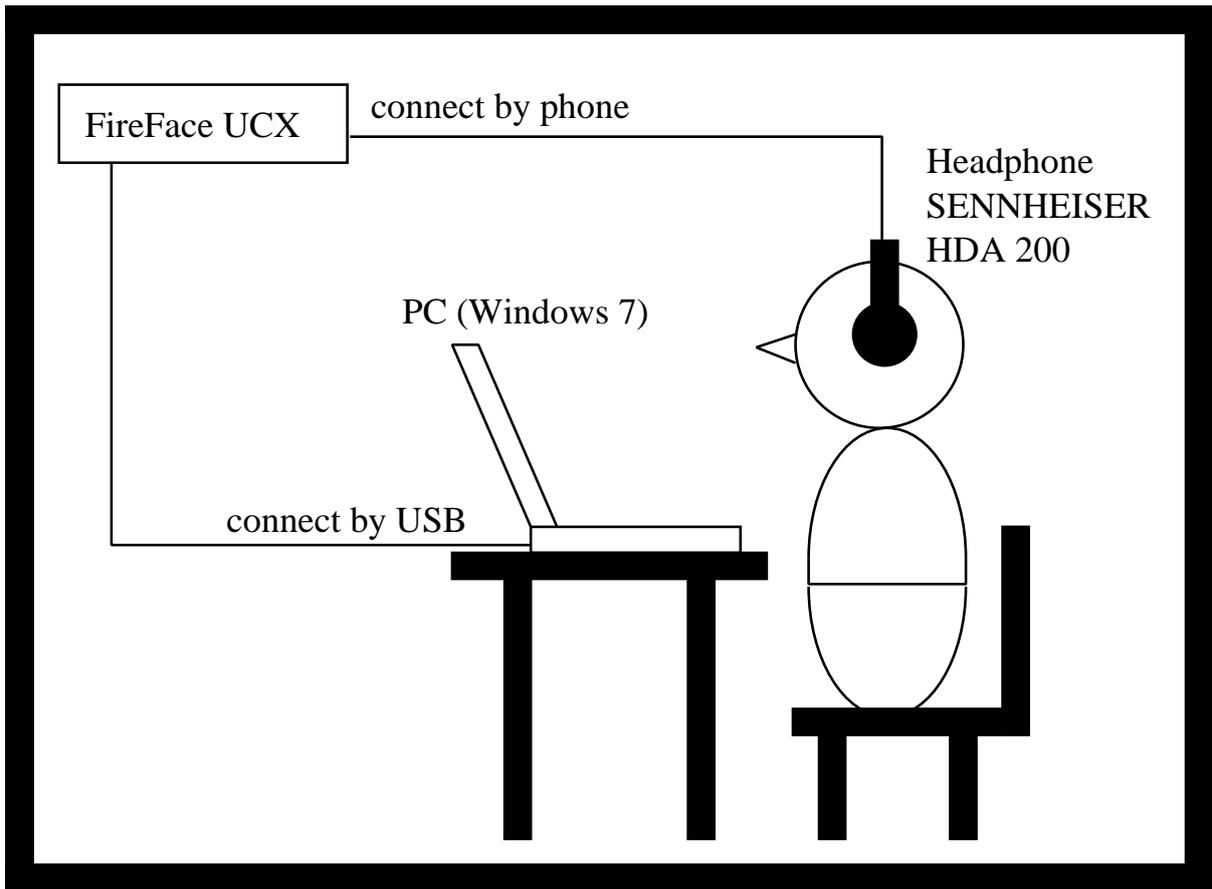
原信号の言語情報と聴取者の回答をモーラ単位で比較した。たとえば, 原信号の言語情報が “やきにく”, 聴取者の回答が “あきにう” であった場合, 第 2, 第 3 モーラ目は正答, 第 1, 第 4 モーラ目は不正答となるため, 正答モーラ数は 2 となる。一つのカットオフ周波数で 6 個の音声に対して回答を得たので, 各カットオフ周波数で正答モーラ数を全て加算し音声の個数である 6 で除算したものを平均正答モーラ数と呼ぶ。

実験機器

実験は, 防音室 (暗騒音の A 特性音圧レベル: 25.8 dB) にて行った。刺激音は, PC (Windows 7), オーディオインターフェイス (Fireface UCX), ヘッドホン (SENNHEISER HDA 200) を用いて, 実験参加者の両耳に呈示した。このときの実験環境を図 3.1 に示す。ヘッドホンからの出力レベルは, B&K HATS, B&K NEXUS, B&K type 2231 モジュール型精密騒音計を利用して, 実験前に毎回校正を行った。

3.1.5 実験結果

図 3.2 に, 低域通過フィルタの各カットオフ周波数に対する実験参加者 4 名分の平均正答モーラ数を示す。低域通過フィルタのカットオフ周波数が約 5 Hz 未満になると平均正答モーラ数が徐々に下がっており, 低域通過フィルタのカットオフ周波数が 1 Hz になると平均正答モーラ数は, ほぼ 0 になった。低域通過フィルタのカットオフ周波数が 5 Hz



Sound proofroom

图 3.1: 实验环境

以上で平均正答モーラ数が一定となるため，低域通過フィルタのカットオフ周波数が 6 Hz 以上で付加される変調成分は正答モーラ数に大きく影響しないことが分かった．

3.1.6 考察

今回の実験で使用した 4 モーラ単語の持続時間は約 1000 ms であるため，1 モーラの長さは約 250 ms となる．これを 1 周期と考えると，振幅包絡に含まれる変調成分の上限周波数が 4 Hz であれば，今回使用した音声のモーラの時間構造を再現できるといえる．実験結果より，約 5 Hz 未満の変調成分を削除した際に平均正答モーラ数が減少するのは，このモーラの時間構造が壊されたためであると考えられる．そのため，日本語音声のモーラの時間構造を再現する変調成分を保存さえしていれば，言語情報の取得が可能であることが分かった．

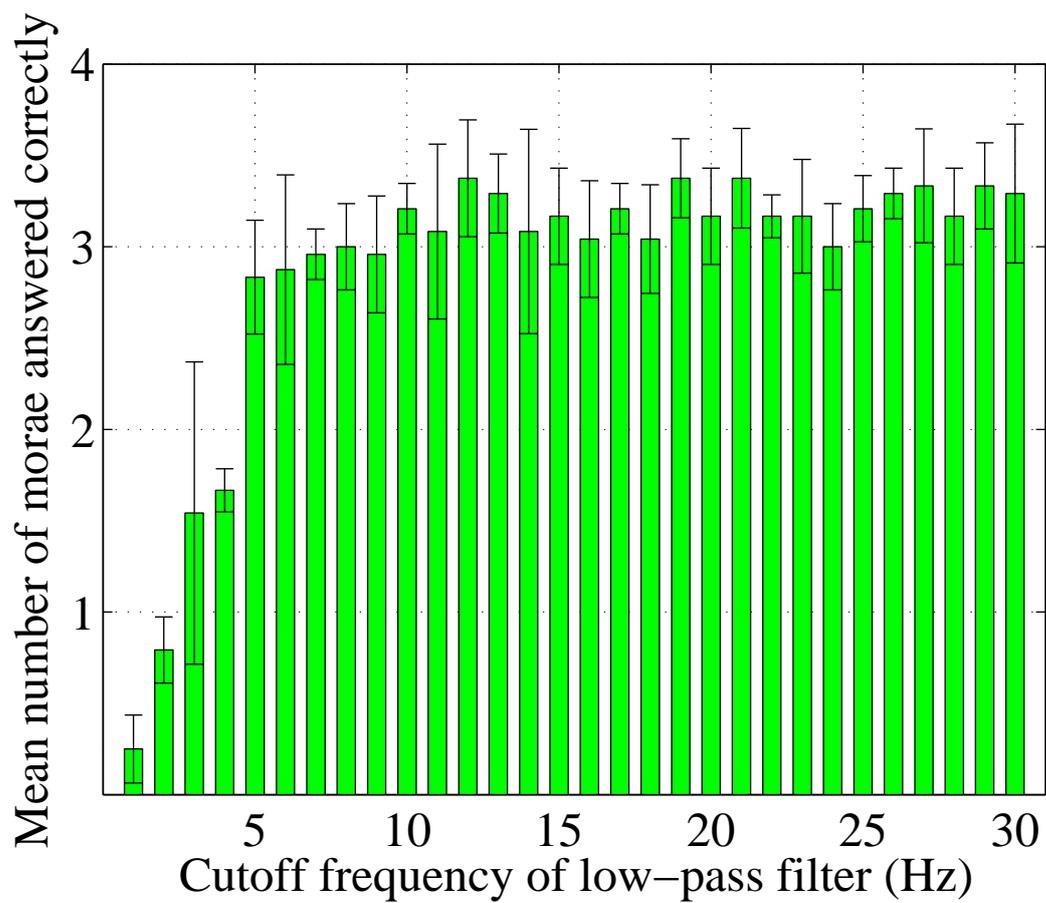


図 3.2: 低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化

3.2 日本語音声の各周波数帯域の振幅包絡の変調成分が言語情報の取得に与える影響の検討

3.2.1 実験の目的

本実験では、各周波数帯域の振幅包絡に含まれる変調成分を一つずつ制御し、周波数帯域によって変調成分が聴き取りやすさに与える影響が変化するかを聴取実験により確かめた。

3.2.2 実験参加者

実験参加者は、3.1.2と同様である。

3.2.3 刺激音

得られた振幅包絡の変調成分を、低域通過フィルタを用いて系統的に制御した。10個の周波数帯域中、一つの周波数帯域の振幅包絡は、低域通過フィルタのカットオフ周波数を5～19 Hzまで、1 Hz刻みで変化させた。残りの九つの周波数帯域の振幅包絡は、低域通過フィルタのカットオフ周波数を5 Hzと固定した。原音源は、3.1.3と同様である。

3.2.4 実験手続き

実験方法

実験参加者には、全900個の刺激（10（制御する周波数帯域の総数）×15（振幅包絡の制御数）×6（各カットオフ周波数に用意された音声の数））をランダムに呈示し、刺激音の言語情報がどのような内容であったかを回答させた。回答は、PCのキーボードで行った。900個の刺激音を1セット75個に分割し、各実験参加者は14セット行った。1日に7セットとし、2日に分けて実験を行った。それぞれのセットの間には休憩時間を設けた。また、1日の実験の始めには、安定した実験結果を得るために1セットのウォーミングアップ試行を行った。

分析方法・実験機器

分析方法と実験機器は、3.1.4と同様である。

3.2.5 実験結果

周波数帯域の低域通過フィルタのカットオフ周波数に対する実験参加者 4 名分の平均正答モーラ数の結果の一部を，図 3.3, 3.4, 3.5 に示す（残りは付録を参照）．図 3.3 は第 1 周波数帯域，図 3.4 は第 5 周波数帯域，図 3.5 は第 10 周波数帯域の結果である．周波数帯域の低域通過フィルタのカットオフ周波数によって，言語情報の取得に差はみられなかった．他の周波数帯域の結果も同様であった．また，周波数帯域別の低域通過フィルタのカットオフ周波数に対する実験参加者 4 名の母音の平均正答モーラ数の一部を，図 3.6, 3.7, 3.8 に示す（残りは付録を参照）．図 3.6 は第 1 周波数帯域，図 3.7 は第 5 周波数帯域，図 3.8 は第 10 周波数帯域の結果である．母音の平均正答モーラ数は約 4 モーラであることから，母音の知覚はほとんどできていることがわかった．他の周波数帯域の結果も同様であった．

今回の実験では，一つの周波数帯域の振幅包絡を原音声から抽出した振幅包絡の情報に近づける処理を行ったが，言語情報の聴き取りやすさが顕著に変化する結果は得られなかった．母音の平均正答モーラ数の結果より，母音の知覚はほとんどできていることから，子音の知覚が母音より劣っていることが分かった．

3.2.6 考察

今回の実験で，振幅包絡に含まれる変調成分の上限周波数の変化による影響が見られなかった原因の一つとして，言語情報の取得に関わる周波数帯域が，複数に渡っている可能性がある．例えば，上田 & 中島 [7] の研究では，二つの離れた周波数帯域で，共通の因子得点が高くなることが示されており，いくつかの周波数帯域の変動が共変調している必要があるのかもしれない．そのため，一つの周波数帯域の振幅包絡のみを変化させても言語情報の取得に影響はなかったと考えられる．また，母音は，時間的変動が一定時間の間安定しており，変動成分の変化に頑健である可能性が高い．そのため，モーラ構造さえ保存されていれば，細かい変動成分を取り除いても頑健に知覚できたと考えられる．一方，半母音や拗音といった子音は，時間的な変動（わたり）がその知覚に重要だといわれている．変調成分の上限周波数が高い場合，細かな時間的変動の情報を持っている．よって，変調成分の上限周波数が 5 Hz 以上には子音の情報が含まれていると予想される．今回行った実験では，実験に用いる子音の種類について考慮していなかった．そのため，子音の種類別の正答率を求めることができなかった．

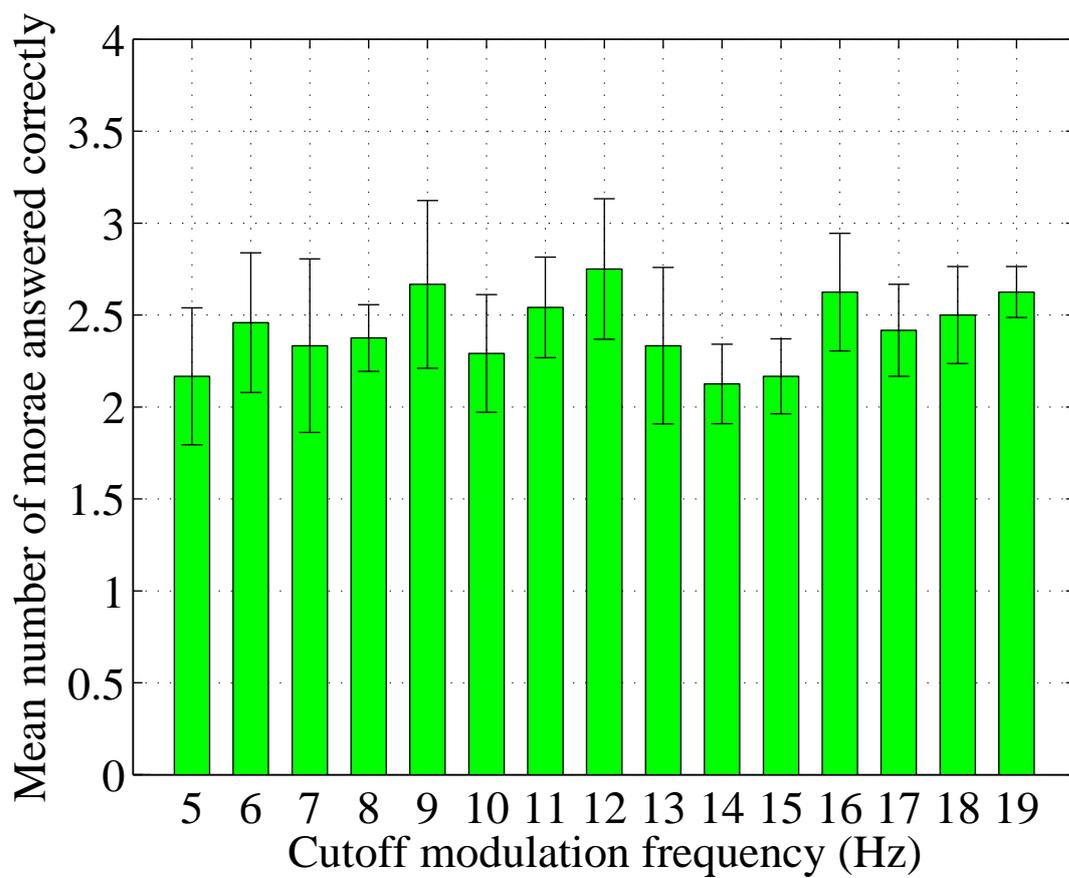


図 3.3: 第 1 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化

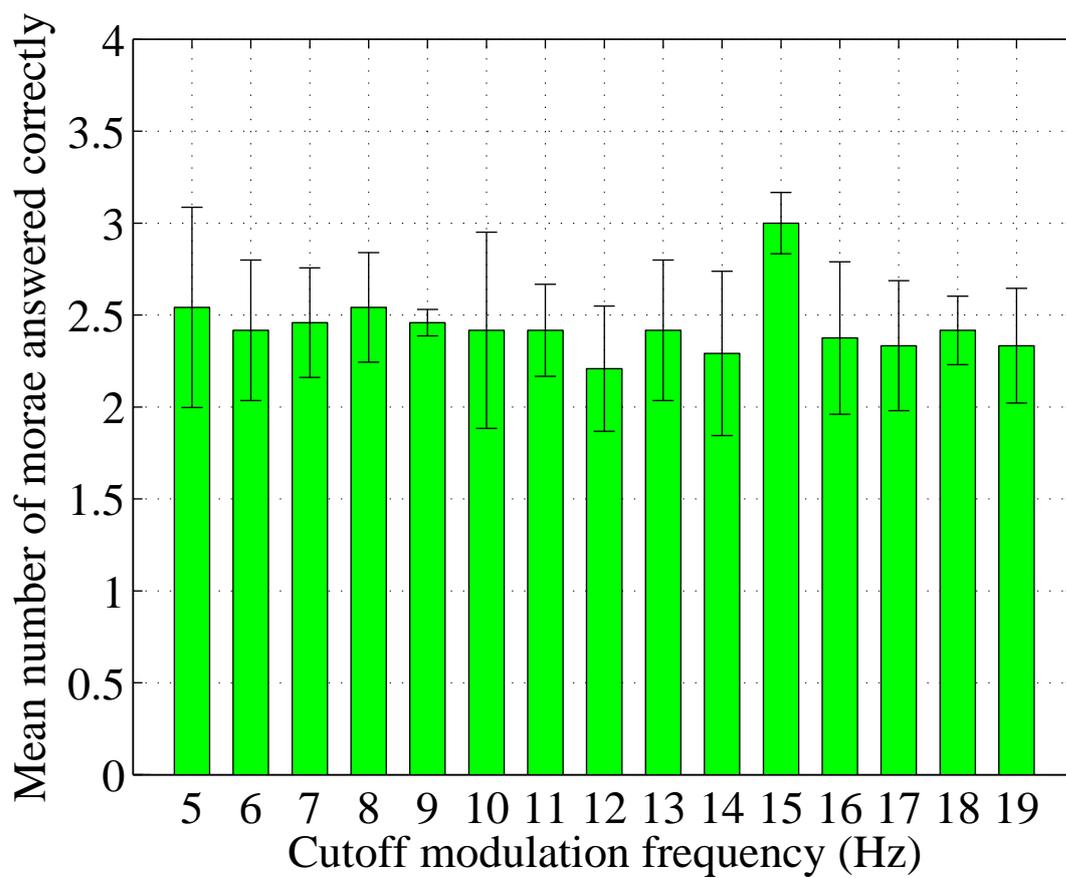


図 3.4: 第 5 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化

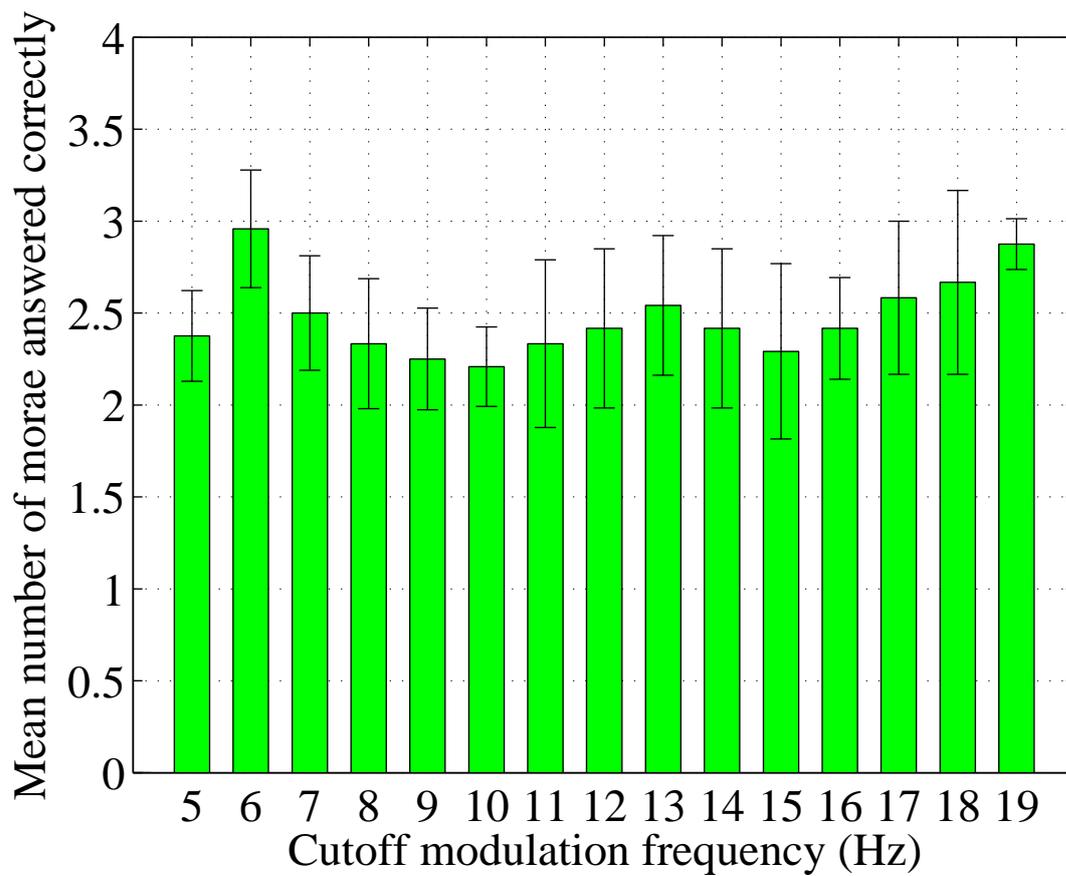


図 3.5: 第 10 周波数帯域の低域通過フィルタのカットオフ周波数に対する平均正答モーラ数の変化

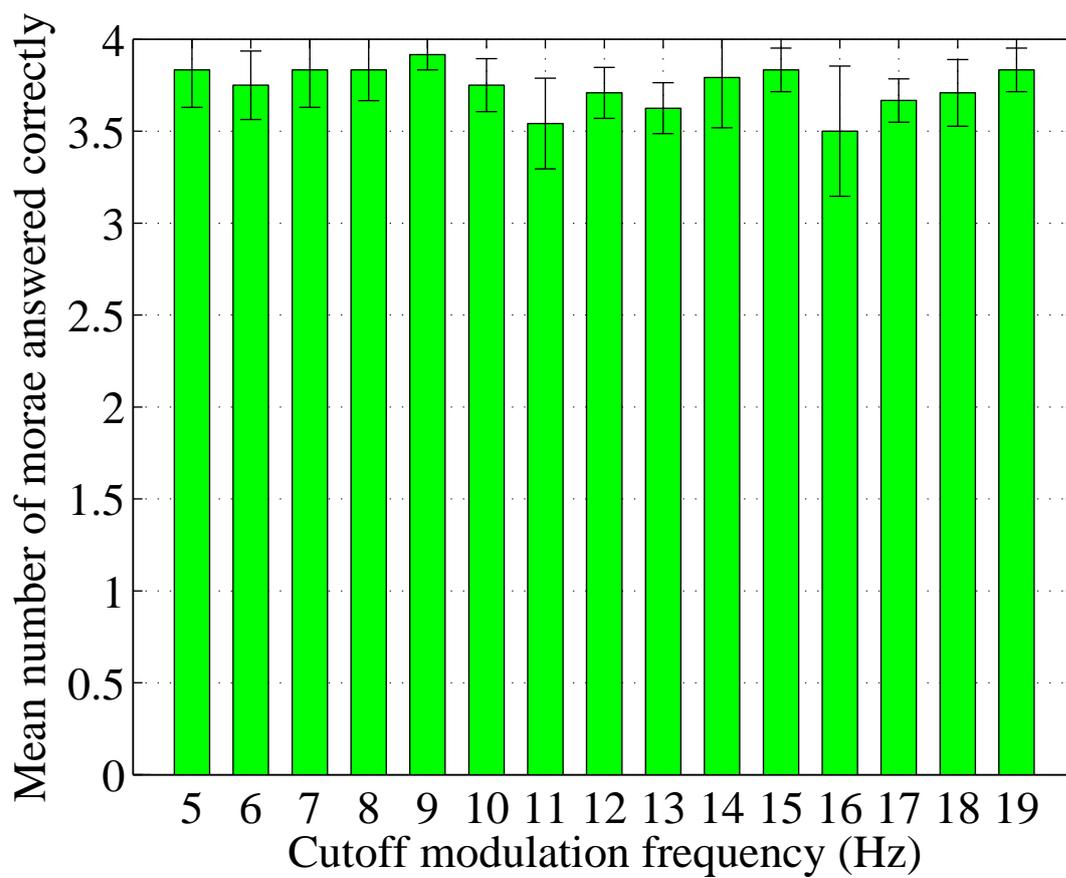


図 3.6: 第 1 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化

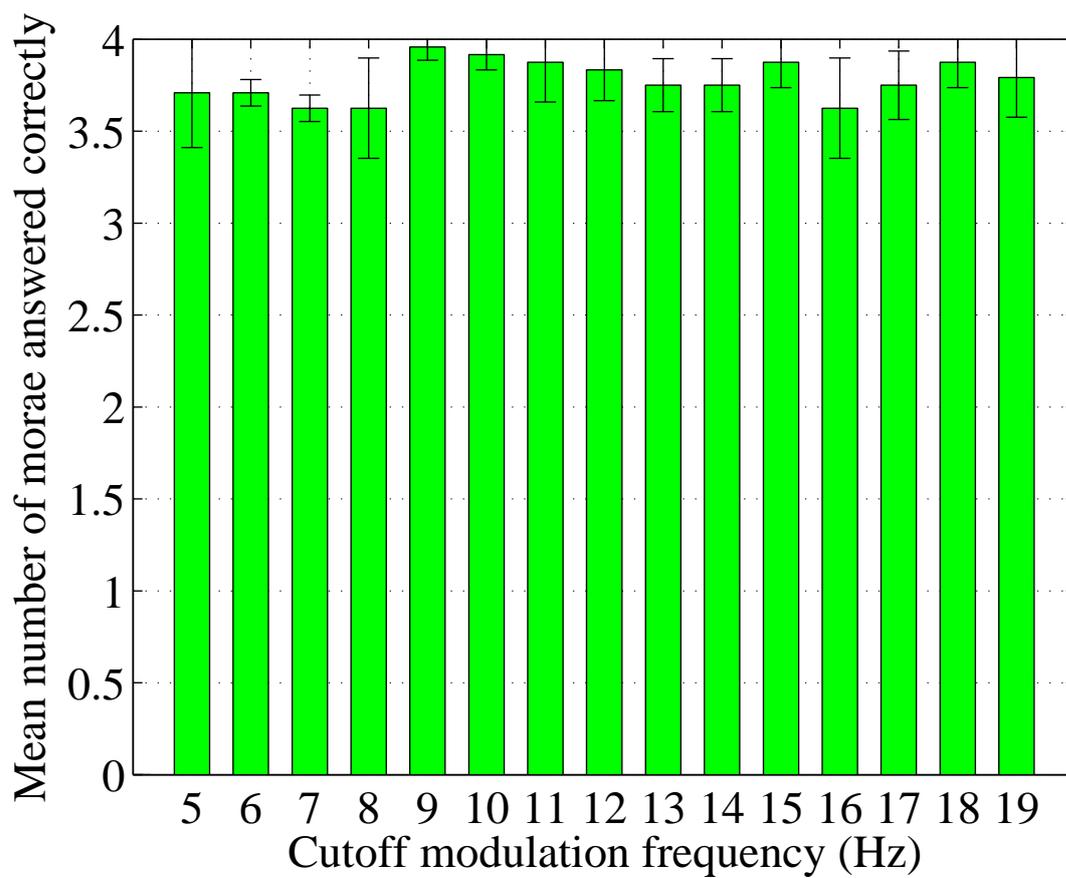


図 3.7: 第 5 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化

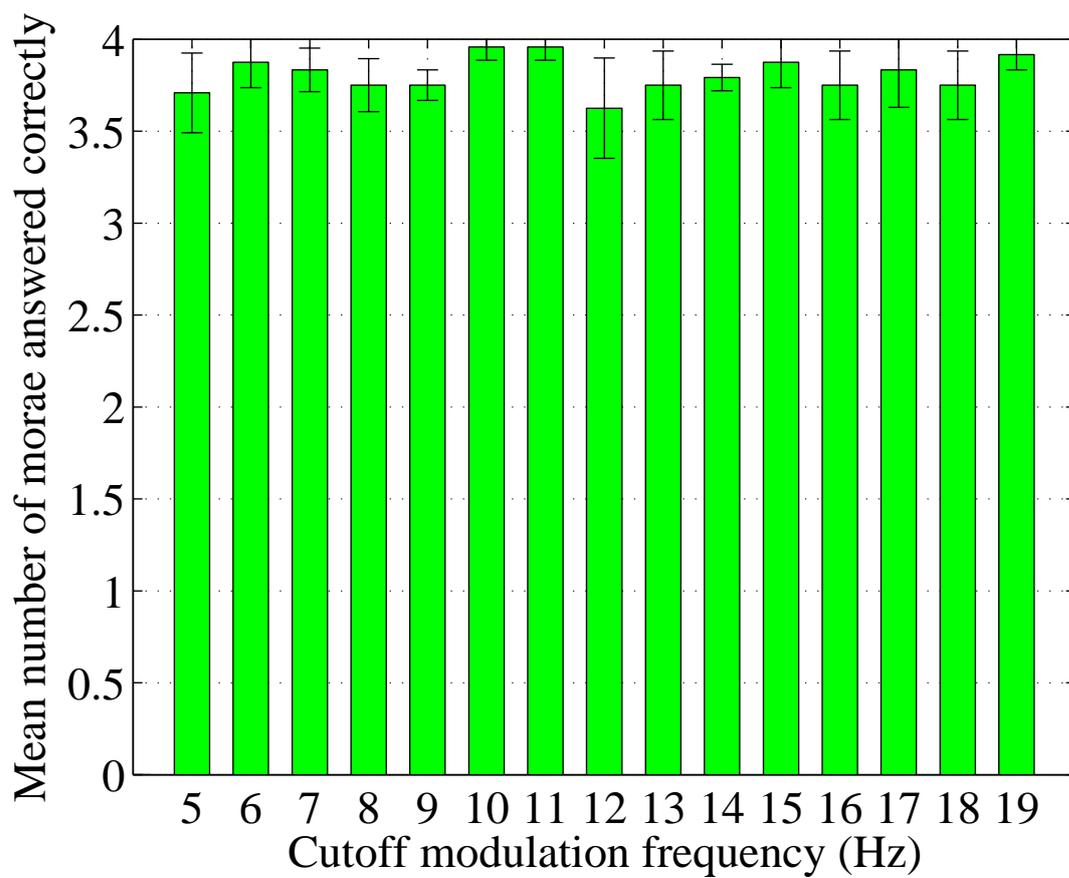


図 3.8: 第 10 周波数帯域の低域通過フィルタのカットオフ周波数に対する母音の平均正答モーラ数の変化

3.3 まとめ

まず 3.1 での実験では、音声の振幅包絡の変調成分という特徴に着目し、この特徴が言語情報の取得にどのような影響を与えているのか調査をした。音声の振幅包絡に含まれる変調成分の上限周波数が 5 Hz で平均正答モーラ数がおおよそ 3 となり 5 Hz 以上は平均正答モーラ数が一定になった。この結果から、振幅包絡にモーラの時間構造の情報が含まれていることが分かった。

次に 3.2 での実験では、音声の各周波数帯域の振幅包絡の変調成分に着目し、どの周波数帯域の振幅包絡に言語情報の取得に関わる情報がどのくらい含まれているのか調査した。その結果、一つの周波数帯域の振幅包絡を変化させただけでは平均正答モーラ数に顕著な変化はみられなかった。この結果から、一つの周波数帯域の振幅包絡に単独で含まれている言語情報を見つけることができなかった。しかし、他の周波数帯域は低域通過フィルタのカットオフ周波数が 5 Hz で制御した振幅包絡であるため、モーラの時間構造は保存されている。つまり、モーラの時間構造を表現する変調成分を保存してさえいれば言語情報の取得が可能であることが頑健に分かった。

最後に、振幅包絡にモーラという日本語の基本的な時間構造の情報が含まれていることは明らかとなったが、各周波数帯域の振幅包絡に含まれる言語情報については明らかにすることができなかった。その原因として、実験に用いた音声の子音や母音について考慮していなかった点が考えられる。変調成分の上限周波数が 4 ~ 5 Hz でモーラの基本的な時間変動に対応しているとすると、子音に関する情報は高い変調成分に含まれている可能性がある。

次に行う聴取実験では、子音の知覚が変調成分の上限周波数の変化によって子音知覚に影響を与えているのかどうかを、調査する。

第4章 日本語音声の子音の振幅包絡における時間的構造

4.1 日本語音声の子音の振幅包絡の変調成分が言語情報の取得に与える影響の検討

4.1.1 実験の目的

本実験の目的は、子音（摩擦音）の振幅包絡の変調成分を系統的に制御したとき、子音（摩擦音）の知覚がどのように変化するかを明らかにすることである。

4.1.2 実験参加者

8名の実験参加者（男性6名、女性2名）に対して聴取実験を行った。実験参加者全員に対して標準的な聴力検査（RION AA-72B オーディオメータを利用）を両耳に行い、125～8000 Hz 間のオクターブ周波数に対して正常な聴力（聴力レベルで 12 dB 以下）を有することを確認した。

4.1.3 刺激音

子音の振幅包絡を制御する低域通過フィルタのカットオフ周波数は、1 Hz から 30 Hz まで、1 Hz 刻みとした。

ここで、用いた子音について説明する。2.2.2 で説明した子音の振幅包絡の形状より、本研究で着目する子音について検討する。まず、拗音は他の複数の子音を組み合わせたような振幅包絡の形状をしており複雑である。そのため、今回の実験では拗音を使用しない。

次に、半母音・鼻子音・ラ行は、子音部分の時間長がどれもおよそ同じ長さであるため、振幅包絡の変調成分を制御しても、同じような振幅包絡の形状になってしまい振幅包絡の違いによる影響をうまく表現できない可能性が高い。そのため、今回の実験では半母音・鼻子音・ラ行を使用しない。

次に、破裂音について検討する。後述する刺激音の作成方法に沿って破裂音を使用した刺激音を作成し予備実験を行ったが、破裂音の言語情報を取得することができなかった。今回は、振幅包絡の制御において、30 Hz 以上の変調成分をカットしている。そのため、

破裂音の子音部分の非常に短く急峻な振幅包絡の変化が低い変調成分のみでは表現できないためと考えられる。破裂部の急峻な立ち上がりを表現するためには、高い周波数帯域において低域通過フィルタのカットオフ周波数を高くする必要がある。しかし、周波数帯域ごとにカットオフ周波数を変えると実験条件が煩雑になってしまう。そのため、今回の実験では破裂音を使用しない。

最後に、摩擦音は、子音ごとに摩擦部の時間長が段階的に変化している。また、特に立ち上がり部分の振幅包絡の形状に大きな違いがみられる。例として、/sa/ の振幅包絡の形状を図 4.1 に、/za/ の振幅包絡の形状を図 4.2 に示す。/sa/ と /za/ の子音部分の長さは、それぞれおおよそ 0.13 秒 と 0.05 秒である。先行研究のモーラ構造と同様に、これらを 1 周期と考えると子音の時間構造を再現するためには、おおよそ 8 Hz と 20 Hz の変調成分が必要であると考えられる。振幅包絡の変調成分を制御し、摩擦部の振幅包絡の形状を段階的に変化させることで、この子音部分の長さの違いが子音の知覚に影響すると予測される。そのため、今回の実験では、摩擦音の知覚が、振幅包絡の制御により変化するかどうかに着目する。

親密度別単語理解度試験用データセット (FW03) [24] にある女性話者 fhi の四つの単音節音声 (/a/ , /sa/ , /ha/ , /za/) を原信号とした。また、原信号の前後に無音区間を設けた。原信号の前に 1000 ms の、原信号の後に 2000 ms になるまでの無音区間を設けた。

4.1.4 実験手続き

実験方法

実験参加者には、全 120 (30 (振幅包絡の制御数) × 4 (四つの単音節音声, /a/ , /ha/ , /sa/ , /za/)) 個の刺激音をランダムに呈示し、刺激音が /a/ , /ha/ , /sa/ , /za/ のどれに聴こえたかを回答させた (四肢強制選択)。回答は、PC のキーボードで行った。刺激音 120 個を 1 セットとし、各実験参加者は 5 セット分の刺激に対して回答を行った。それぞれのセットの間には、休憩時間を設けた。また、実験の始めには、安定した実験結果を得るために、1 セットのウォーミングアップ試行を行った。

実験機器

実験は、防音室 (暗騒音は A 特性音圧レベル 25.8 dB) にて行った。刺激の呈示には、PC (Windows 7)、オーディオインターフェイス (Fireface UCX)、ヘッドホン (SENNHEISER HDA 200) を使用した。ヘッドホンからの出力レベルは、B&K HATS, B&K NEXUS, B&K type 2231 モジュール型精密騒音計を利用して、実験前に毎回校正された。

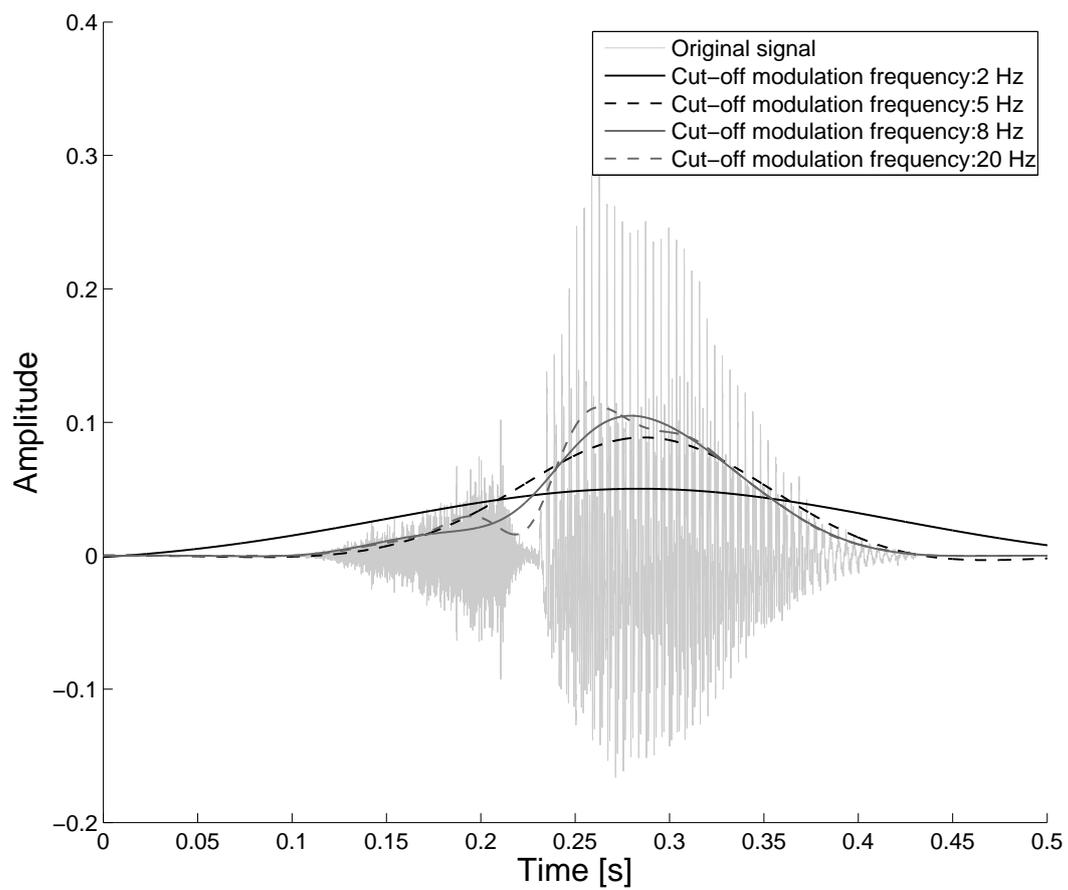


図 4.1: /sa/ の音声波形と振幅包絡

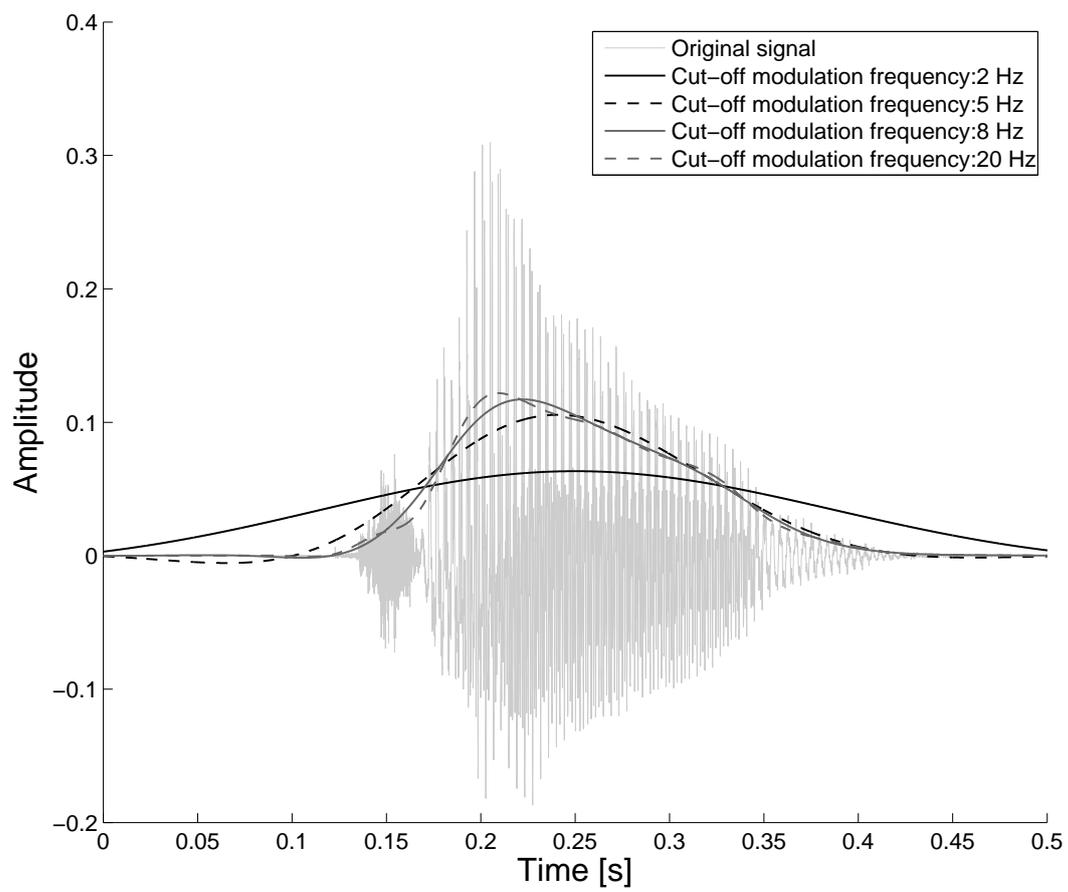


図 4.2: /za/ の音声波形と振幅包絡

4.1.5 実験結果

図 4.3, 図 4.4, 図 4.5, 図 4.6 に, それぞれ /a/, /ha/, /sa/, /za/ のに対する実験参加者 8 名分の変調成分の変化に対する正答率を示す. 縦軸が正答率, 横軸が低域通過フィルタのカットオフ周波数, エラーバーが標準偏差, 青線が S 字フィッティングを表す. 図 4.3 と図 4.4 より, /a/ と /ha/ の正答率は低く, エラーバーが大きいことが示された. 図 4.5 と図 4.6 より, /sa/ と /za/ の正答率は, 振幅包絡に含まれる変調成分の上限周波数が高くなると 100 % に近づいており, 正確に聴き取れていたといえる. 振幅包絡に含まれる変調成分の上限周波数が低くなると, /sa/ は /za/ に比べて, 正答率が上昇するカットオフ周波数は低くなっており, 変調成分の上限周波数が低い状態で聴き取ることが可能であることが示された. ここで, 図 4.7 に /sa/ と /za/ のフィッティング曲線のみを示す. 図 4.7 より, /sa/ は 50 % を変調成分がおよそ 2 Hz で, /za/ は 50 % を変調成分およそ 3 Hz で超えていることが分かった. また, /za/ のフィッティング曲線と図 4.11 より, /za/ は変調周波数の上限周波数が 5 Hz よりも高いときは, 頑健に知覚していることが分かった. それに比べて, /sa/ は変調成分の上限周波数が低い時でもある程度知覚しているが, 上限周波数が高い値になっても, 僅かながら異聴を起こしていることが, 図 4.10 から分かった.

図 4.8, 図 4.9, 図 4.10, 図 4.11 に, /a/, /ha/, /sa/, /za/ が呈示されたときの実験参加者 8 名分の回答の割合を示す. 縦軸が低域通過フィルタのカットオフ周波数, 横軸が四つの回答の割合である. 図 4.8 と図 4.9 より, /a/ と /ha/ は, 互いに聴き間違えていることが示された. 図 4.10 より, 原信号が /sa/ の場合の回答の割合をみると, /za/ と異聴されることは少ない. 図 4.11 より, 原信号が /za/ の場合の回答の割合をみると, 変調成分の上限周波数が低い場合では, /sa/ と知覚されることが分かった. 変調成分の上限周波数を高くすると /za/ と知覚されることが示された.

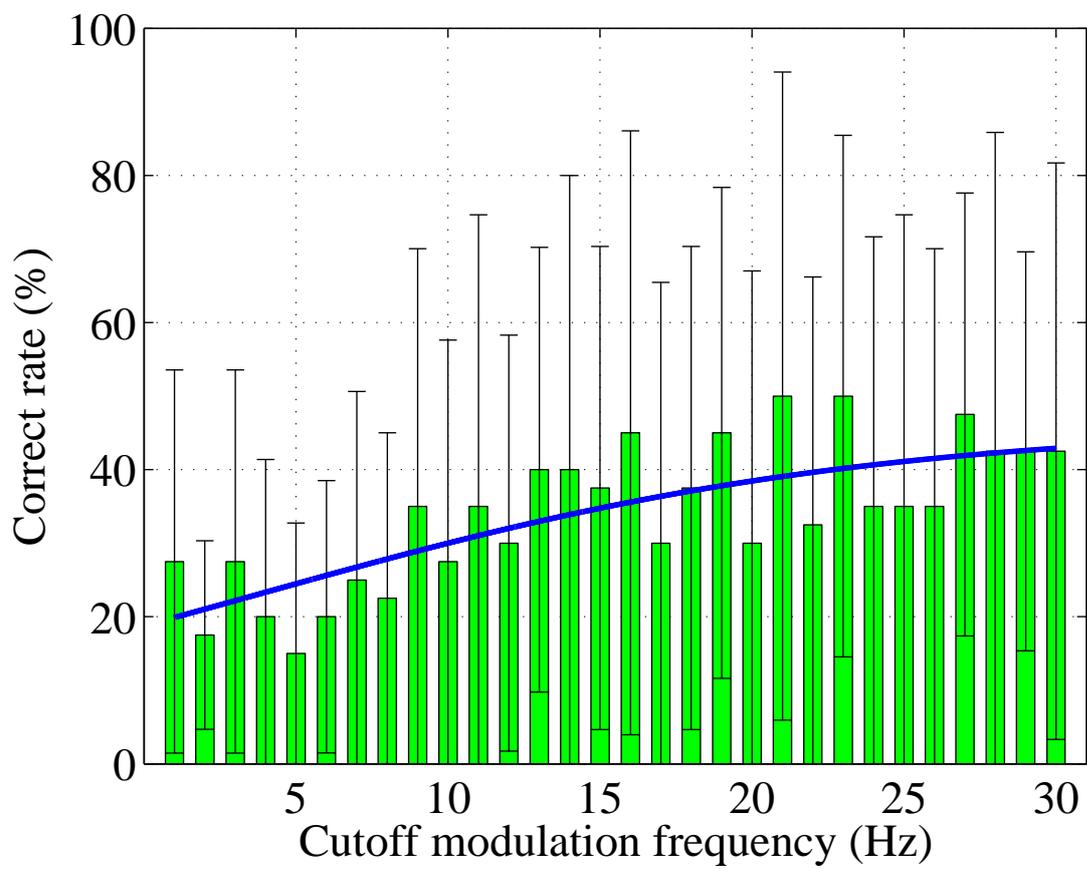


図 4.3: 変調成分を変化させたときの/a/の正答率の変化

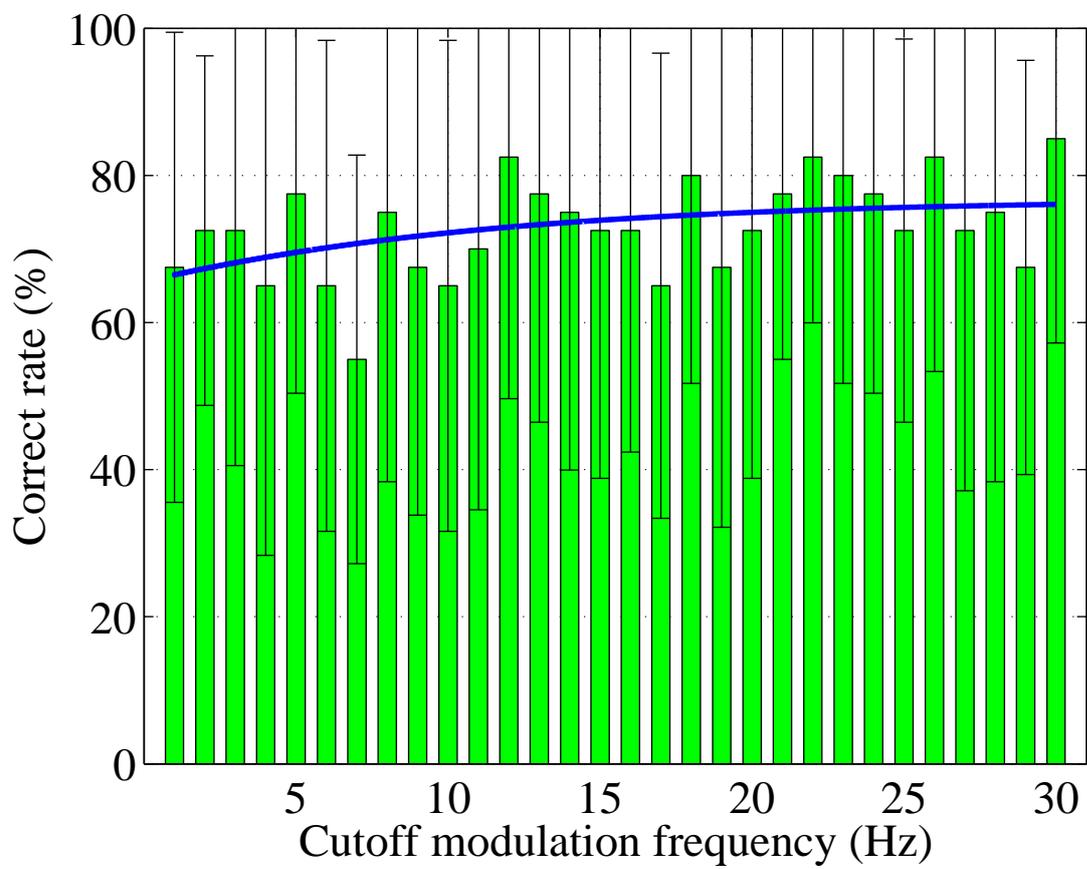


図 4.4: 変調成分を変化させたときの/ha/の正答率の変化

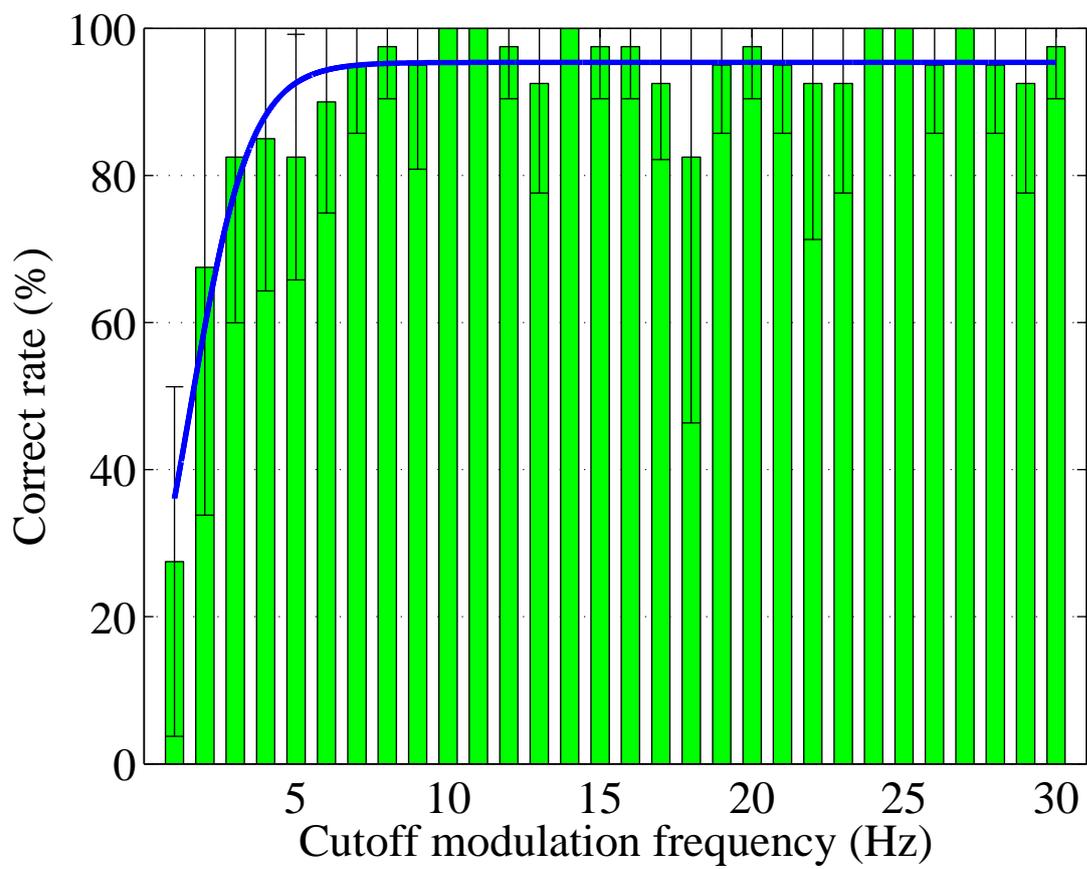


図 4.5: 変調成分を変化させたときの/sa/の正答率の変化

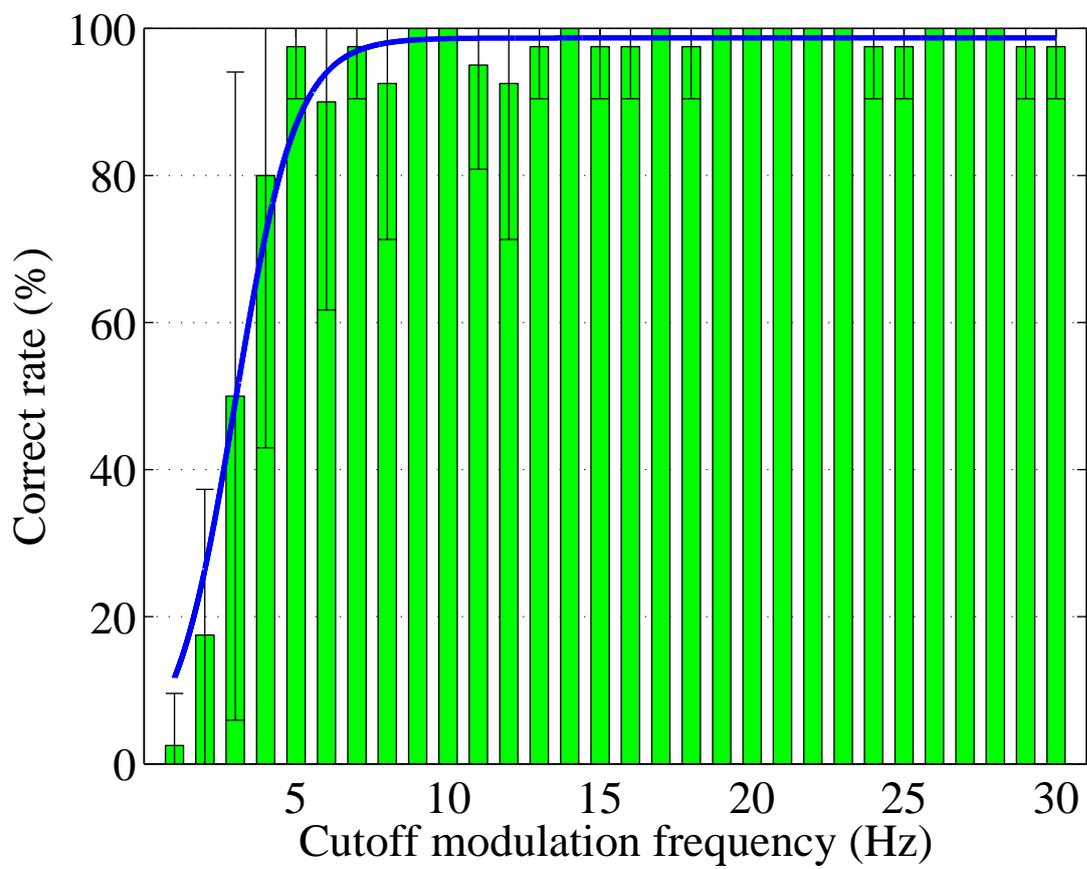


図 4.6: 変調成分を変化させたときの/za/の正答率の変化

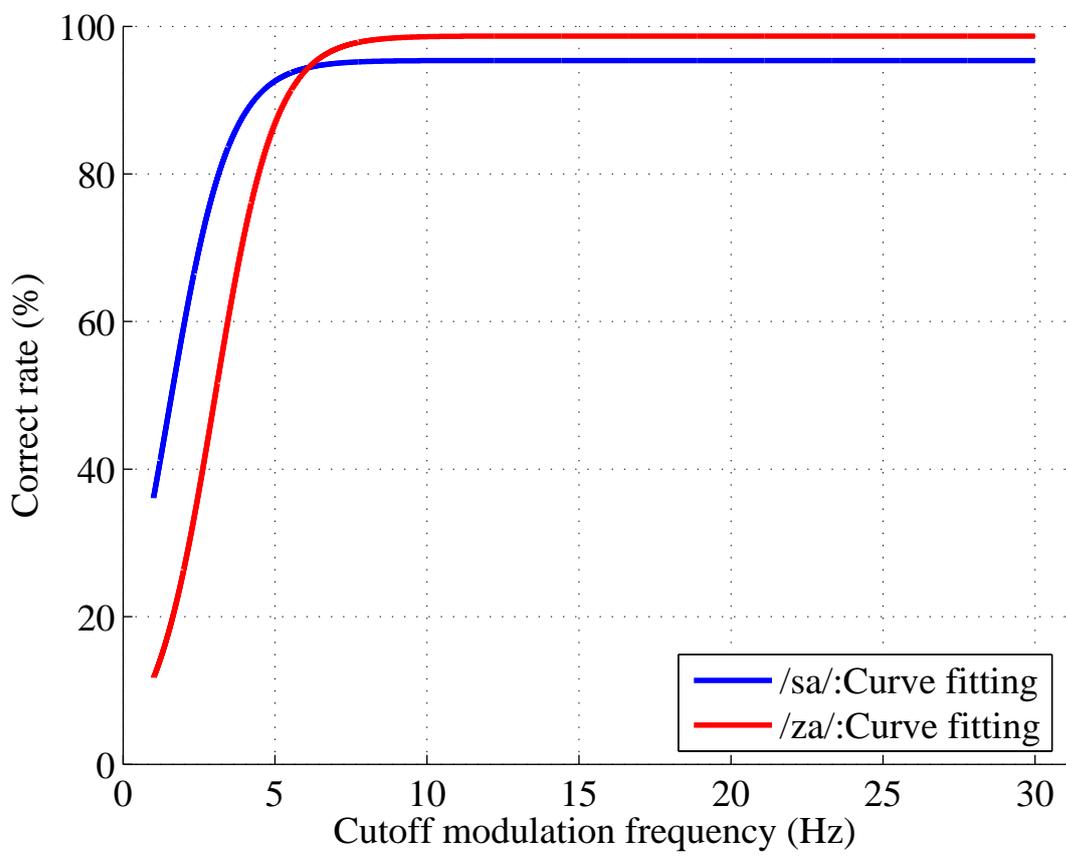


図 4.7: /sa/ と /za/ のフィッティング曲線

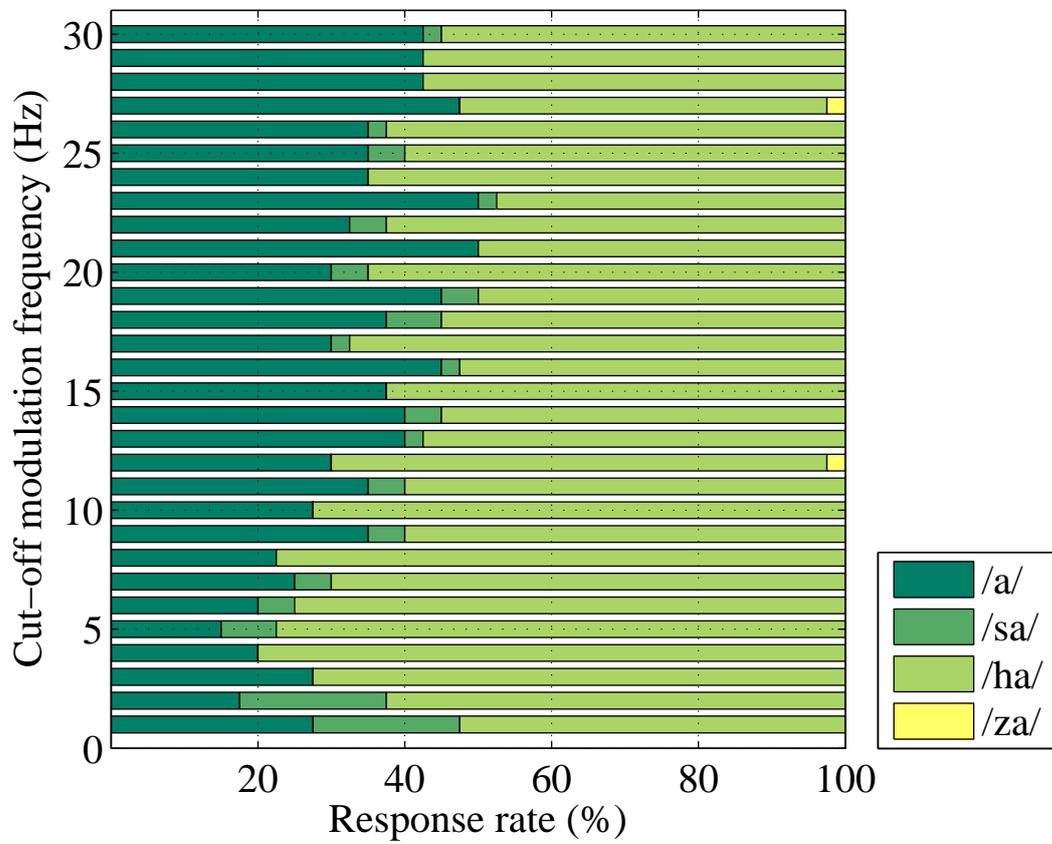


図 4.8: 原信号 /a/ に対する回答の割合の変化

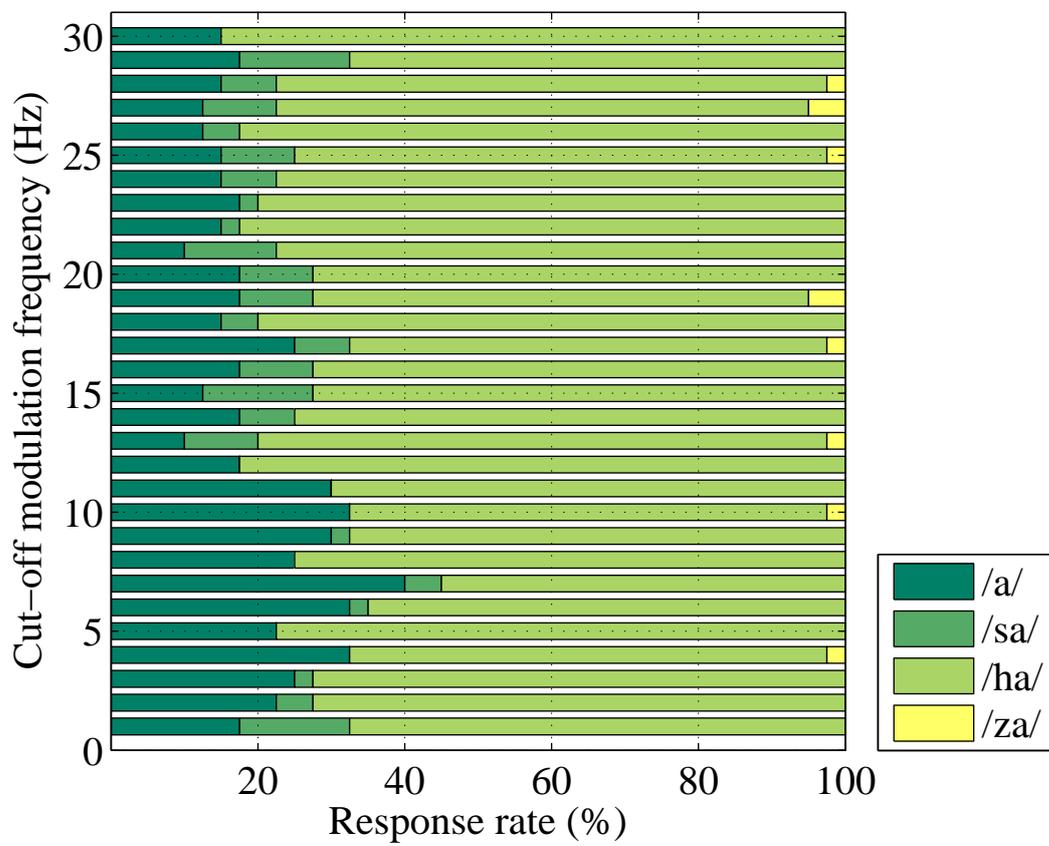


図 4.9: 原信号 /ha/ に対する回答の割合の変化

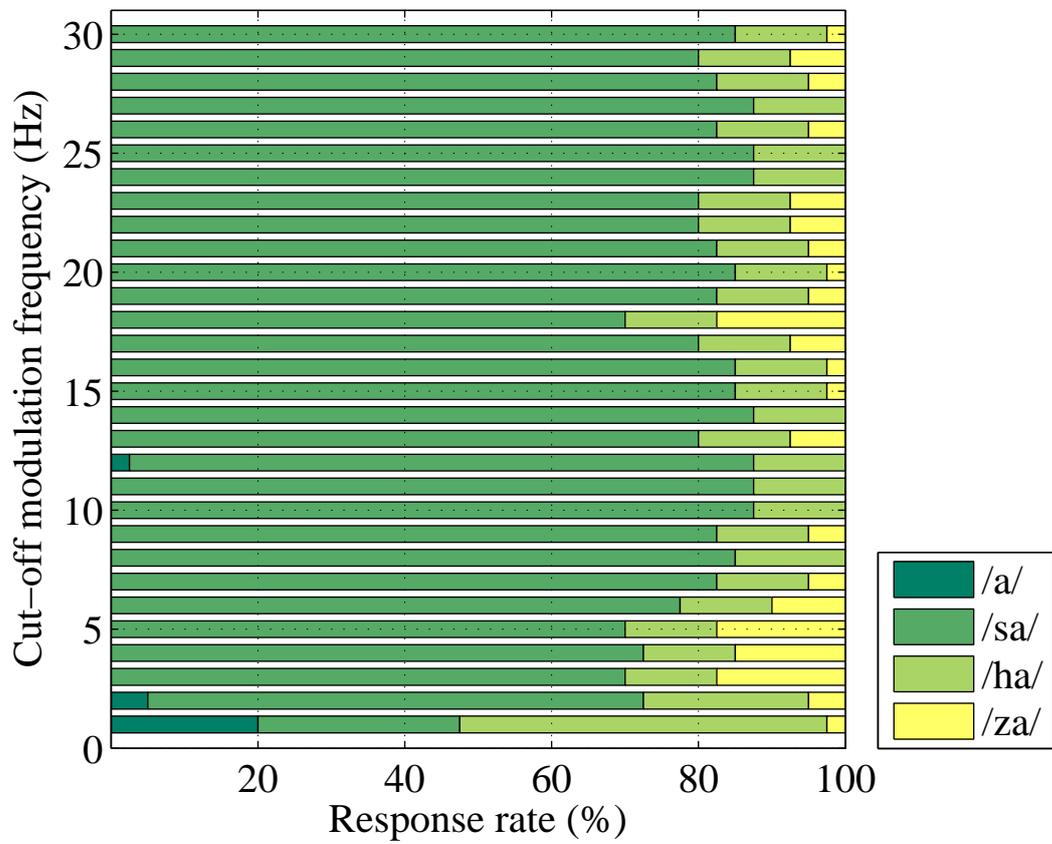


図 4.10: 原信号/sa/に対する回答の割合の変化

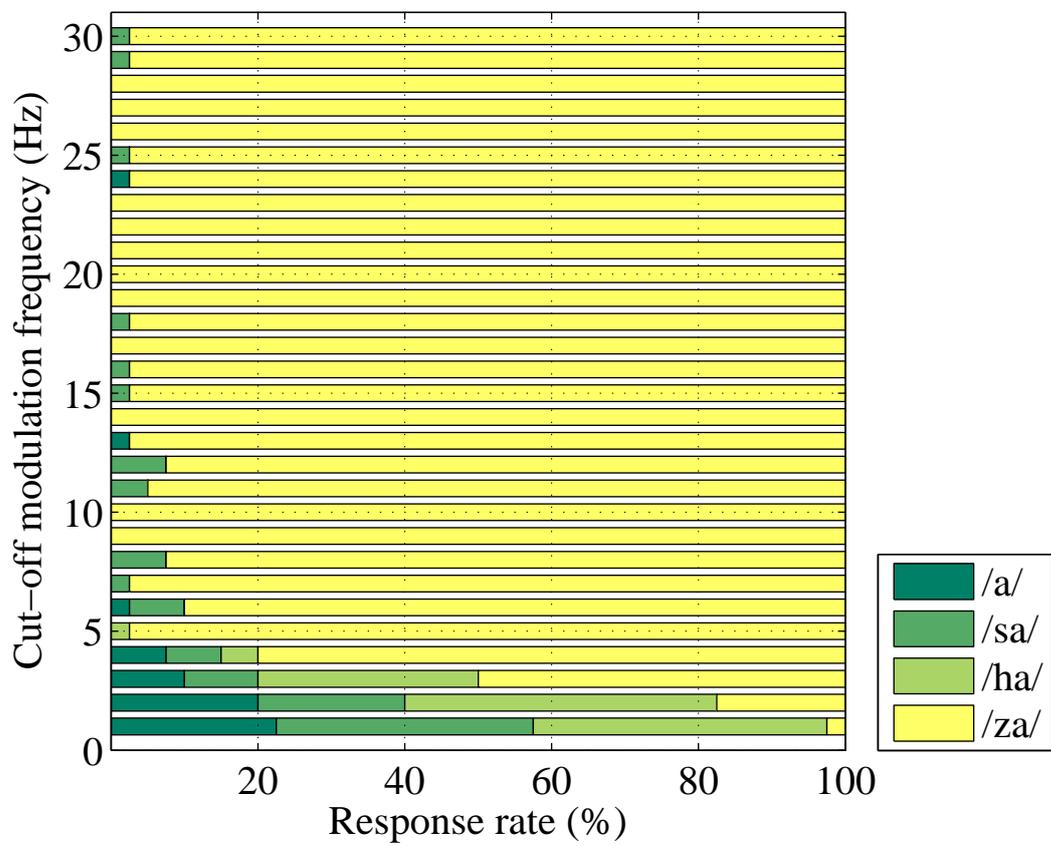


図 4.11: 原信号/za/に対する回答の割合の変化

4.1.6 考察

/sa/ と */za/* の子音部分の振幅包絡構造を再現するためには、それぞれの変調成分の上限周波数が 8 Hz と 20 Hz であると予測した。しかし、*/sa/* も */za/* も予測していた変調成分の上限周波数よりも低かったが、知覚が可能であることが分かった。これは、子音部分の立ち上がりの僅かな振幅包絡の形状の違いさえ再現できれば、子音の違いを聴き取ることができると考えられる。図 4.12 より、音声波形をカットオフ周波数が 2 Hz の低域通過フィルタで通したときの振幅包絡を見比べると、ほとんど違いがない。そのため、*/za/* という音声を聴いても、*/sa/* という音声に知覚してしまう。しかし、カットオフ周波数が 5 Hz の低域通過フィルタで通したときの振幅包絡を見比べると、*/sa/* の子音部分に比べて、*/za/* の子音部分は急峻な振幅包絡の形状である。つまり、振幅包絡の後続母音へのわたりの傾きによって、*/sa/* と */za/* を知覚しており、この僅かな傾きの違いにより、*/sa/* と */za/* を聴き分けていたと考えられる。

図 4.7 より、*/sa/* と */za/* のフィッティング曲線が変調成分の上限周波数が 1 ~ 5 Hz 付近までの経過を見ると、言語情報の取得が可能になるまでの差がはっきりと分かる。これにより、*/sa/* と */za/* に必要な変調成分がそれぞれ違うことが考えられる。

4.2 まとめ

本実験では、子音（摩擦音）の振幅包絡の変調成分を制御したとき、子音（摩擦音）の知覚が変化するかどうかを聴取実験を行い調査した。その結果、変調成分が少ないとき、*/za/* を呈示しているにもかかわらず、*/sa/* と知覚していることが分かった。しかし、変調成分を増やすと、*/za/* と聴き取ることができるようになった。これは、子音の振幅包絡の変調成分を制御したことで、子音知覚が変化したことを示している。*/sa/* と */za/* の聴き分けに重要な情報は、子音から後続母音へのわたりの傾きの違いであり、振幅包絡の変調成分の制御によって、この傾きが変化すると考えられる。

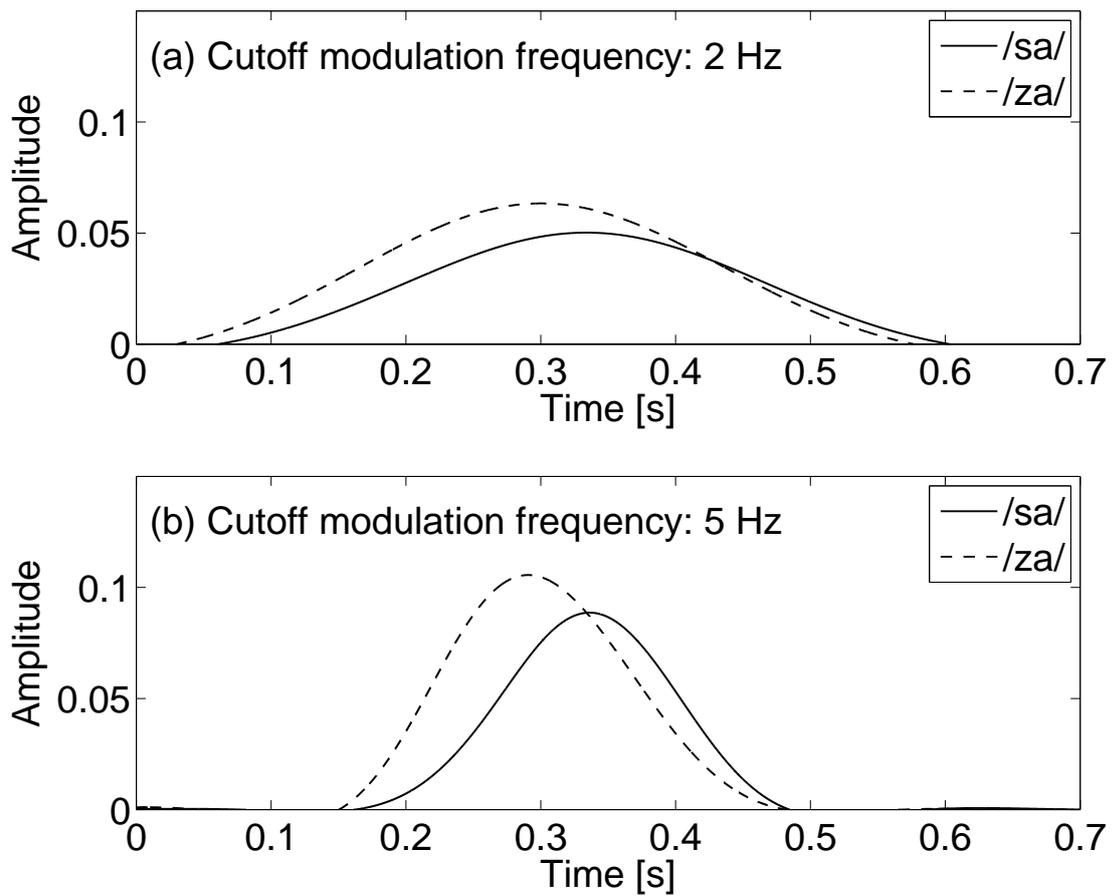


図 4.12: 変調成分の上限周波数が (a) 2 Hz と (b) 5 Hz のときの /sa/ と /za/ の振幅包絡の形状

第5章 総合検討

5.1 日本語音声の振幅包絡の時間的構造に対する検討

本研究で行った聴取実験は、日本語音声の振幅包絡に含まれる変調成分を制御し、言語情報の取得にどのような影響を与えるのかを明らかにした。モーラ構造を再現できる変調成分を保存してさえいれば、言語情報の取得が可能であることが分かった。また、モーラ構造を再現するという観点をモーラ構造の時間長よりも短い子音・母音に当てはめた場合、子音部（摩擦部）・母音部の振幅包絡の形状を再現する変調成分を保存すれば、子音の聴き取りがより正確になるのではないかと予測を立てた。しかし、摩擦部・母音部の振幅包絡の形状をそれぞれ詳細に保存しなくても、大まかな振幅包絡でも摩擦音を聴き取ることができた。これは、摩擦部から母音部への振幅包絡の傾きの違いにより、言語情報の取得ができることが分かった。

これらによって、日本語音声の振幅包絡の時間的構造は、オリジナルの音声の振幅包絡の時間構造を再現しなくても、言語情報の取得が可能であることが分かった。

5.2 日本語音声の振幅包絡に含まれる冗長性に対する検討

本研究の聴取実験では、振幅包絡に含まれる変調成分が言語情報の取得にどのような影響を与えるのかを明らかにした。つまり、言語情報の取得に最低限必要とされる振幅包絡に含まれる変調成分の上限周波数を明らかにした。これを音声の冗長性の観点からみると、モーラ構造を再現するだけ変調成分よりも多い変調成分、すなわち、変調成分の上限周波数がモーラ構造を再現する最低の値よりも高い周波数で保存される変調成分は冗長であると示唆される。また、摩擦音の振幅包絡の立ち上がりの違いを再現できる変調成分を保存してさえいれば言語情報の取得が可能であるので、その変調成分の上限周波数よりも高い変調成分は冗長であると示唆される。

言語情報の取得が可能となる振幅包絡に含まれる変調成分よりも多い変調成分は、日本語音声の冗長性であると示唆される。

第6章 結論

6.1 本研究で明らかになったこと

本研究では、音声の振幅包絡に含まれる冗長性の解明のために、振幅包絡の変調成分を系統的に制御し、言語情報の取得に与える影響を調査した。そのために、モーラの時間構造と子音の時間構造、2点に着目して聴取実験を行った。一つ目に、振幅包絡のモーラの時間構造に着目した。4モーラ単語を用いて、その単語の振幅包絡に含まれる変調成分を系統的に制御し、言語情報の取得ができるのかどうかを聴取実験により検討した。その結果、言語情報の取得には、モーラの時間構造を再現できる変調成分を保存していれば可能であることを明らかにした。

二つ目に、モーラの時間構造よりも短い時間構造を持つ子音の時間構造に着目した。単音節音声（摩擦音： $/sa/$ 、 $/ha/$ 、 $/za/$ ）を用いて、その単音節音声の振幅包絡に含まれる変調成分を系統的に制御し、言語情報の取得ができるかどうかを聴取実験により検討した。上記の実験より、言語情報の取得にはモーラの時間構造を再現できる変調成分を保存していれば可能であることより、子音・母音の表現している振幅包絡に含まれる変調成分は、モーラの時間構造よりも短い時間構造であるため、変調成分の上限周波数は高いと予測した。聴取実験を行った結果、予測していた変調成分の上限周波数よりも低い上限周波数で言語情報の取得が可能であることが分かった。これは、摩擦音の振幅包絡の立ち上がりの違いにより、摩擦音の言語情報を聞き取っていたと示唆される。つまり、摩擦音は振幅包絡の立ち上がりによって、我々は言語情報の取得をしていると示唆される。今回、単音節音声は摩擦音を使用したがる、他の子音についても、振幅包絡の立ち上がりが異なる子音に関しては、摩擦音と同様に振幅包絡の立ち上がりの僅かな違いにより聞き取りが可能だと考えられる。

これらの聴取実験より得られた結果から、音声の振幅包絡に含まれる冗長性に関して考察すると、各単音節音声の振幅包絡の立ち上がりの違いを再現できる変調成分を保存してさえいれば言語情報の取得が可能であるので、その変調成分の上限周波数よりも高い変調成分は冗長であると示唆される。そのため、音声の振幅包絡に含まれる冗長性は、変調成分に含まれることが明らかになった。

6.2 今後の展望

本研究の今後の展望として、摩擦音以外の子音の振幅包絡に含まれる変調成分を制御したときに、言語情報の取得にどのような影響を与えるのかである。本研究では、摩擦音である /sa/ , /ha/ , /za/ それぞれの子音部の振幅包絡の時間長や立ち上がりの違いが、他の子音に比べて顕著に現れていたため、変調成分の上限周波数を変化させることで聞き分けを容易に可能としていたと考えられる。しかし、子音部の時間長や立ち上がりが、ほぼ同じであるような子音では、振幅包絡の立ち上がりを再現できたとしても、同じような振幅包絡の形状をしてしまい、異聴を起こしてしまう恐れがある。たとえば、破裂音のように、破裂部は極めて短い時間長であり、振幅包絡の立ち上がりは、どの破裂音も急峻であるため、違いが得られない可能性がある。そのため、他の子音でも同様の聴取実験を行うことによって、言語情報の取得に用いる特徴として振幅包絡の立ち上がりとは別の特徴が明らかになると考えられる。

本論文で明らかにした音声の振幅包絡に含まれる冗長性が、音声情報処理技術やヒトが音声を聴くメカニズムの双方に利益をもたらし、我々の生活をより豊かに、今後の音響学の発展することを願って本論文の結びとする。

謝辞

本研究を進めるにあたり，終始御指導ならびに御助言頂いた 北陸先端科学技術大学院大学情報科学研究科 鷓木祐史 准教授 に心からお礼申し上げます。

本研究と実験を進めるにあたり，終始御指導，御助言，並びに日頃から熱心に御討論して頂きました 北陸先端科学技術大学院大学情報科学研究科 宮内良太 助教 に心からお礼申し上げます。

また，折に触れてご指導いただきました 北陸先端科学技術大学院大学情報科学研究科 赤木正人 教授，森川大輔 助教 に心より感謝致します。加えて，研究を遂行していく上で，熱心な議論と多面にわたる協力を賜った 北陸先端科学技術大学院大学情報科学研究科 党建武 教授，田中宏和 准教授，末光厚夫助教，川本真一助教に厚く御礼申し上げます。

研究を行うにあたり，多面に渡りご協力いただいた鷓木研究室ならびに，赤木研究室の諸先輩方および皆様に感謝致します。

最後に，大学院在学中に自由な研究の場を与えていただき，暖かく見守ってくれた両親，祖父母，弟に心から感謝致します。

参考文献

- [1] 柏野牧夫, “音韻修復：消えた音声を修復する脳 (小特集 音と映像で体験できる聴覚の不思議な世界,” 日本音響学会誌, **61**(5), 263–268, 2005.
- [2] 青木直史, “冗長性から見た情報技術 やさしく理解する原理と仕組み,” 講談社, 2011.
- [3] 柏野牧夫, “雑音と音声の知覚,” 日本音響学会研究発表会講演論文集, 1998(2), 229–232, 1998.
- [4] Drullman, R., “Temporal envelope and fine structure cues for speech intelligibility,” *J. Acoust. Soc. Am.*, **1**, 585–592, 1997.
- [5] Shannon, R. V., Zeng, F. -G., Kamath, V., Wygonski, J., and Ekelid., M., “Speech recognition with primarily temporal cues,” *Science*, **270**, 303–304, 1995.
- [6] Riquimaroux, H., “Perception of noise-vocoded speech sounds: Sentences, words, accents and melodies,” *Acoust. Sci. & Tech.*, **27**, 325–331, 2006.
- [7] 上田和夫, 中島祥好, “臨界帯域フィルターを用いた日本語音声の因子分析,” 日本音響学会聴覚研究会資料, **38**(3), 203–208, 2008.
- [8] Ueda, K., Nakajima, Y., “A consistent clustering of power fluctuations in british english, french, german, and japanese,” *Trans. Tech. Comm. Psychol. Physiol. Acoust.*, **38**(8), 771–776, 2008.
- [9] 風間道子, 東山三樹夫, 山崎芳男, “狭帯域音声波形包絡線の帯域間相関行列に現れる話者情報,” 電子情報通信学会論文誌. A, 基礎・境界, **J92-A**(4), 205–215, 2009.
- [10] Drullman, Rob., Festen, Joost M., and Plomp, Reinier., “Effect of temporal envelope smearing on speech reception,” *J. Acoust. Soc. Am.*, **95**(2), 1053–1064, 1994.
- [11] Drullman, Rob., Festen, Joost M., and plomp, Reinier., “Effect of reducing slow temporal modulation on speech reception,” *J. Acoust. Soc. Am.*, **95**(5), 2670–2680, 1994.
- [12] Arai Takayuki, Pavel Misha, Hermansky Hynek, and Avendano Carlos, “Intelligibility of speech with filtered time trajectories of spectral envelopes,” *Proc. ICSLP 1996*, 2490–2493, 1996.

- [13] 野口聖人, 佐津川裕一, 上田和夫, 中島祥好, “周波数帯域の除去が雑音駆動音声の知覚に与える影響: 複数話者の分析,” 日本音響学会聴覚研究会資料, 42(2), 231–236, 2012.
- [14] 伊佐次伸也, 上田和夫, 中島祥好, “複数の周波数帯域の除去が雑音駆動音声の知覚に与える影響,” 日本音響学会聴覚研究会資料, 44(1), 45–51, 2014.
- [15] 日本音響学会 編, “新版 音響用語辞典,” コロナ社, 2003.
- [16] Juang, Biing-hwang, Rabiner, Lawrence., (古井貞熙), “音声認識の基礎 (上),” NTT アドバンステクノロジー株式会社, 1995.
- [17] 国際音声学会編, (神山孝夫, 竹林滋訳), “国際音声記号ガイドブック: 国際音声学会案内,” 大修館書店, 2003.
- [18] 板橋秀一 編, “音声工学” 森北出版株式会社, 2005.
- [19] Kent, R.D., Read, C., (荒井隆行, 菅原勉 監訳), “音声の音響分析,” 海文堂出版株式会社, 1996.
- [20] 力丸裕, “高度マンマシンインターフェイスと情報技術の応用展開: 第3章 劣化雑音音声の作成と知覚,” CMC 出版, 2009.
- [21] 力丸裕, “劣化雑音音声の聞こえ,” 日本音響学会, 61(5), 273–278, 2005.
- [22] Moore, B. C. J., “An introduction to the psychology of hearing, fifth edition,” 72–75, 2003.
- [23] Furukawa, M., Sakata, K., Unoki, M. and Akagi, M., “An improved method based on the mtf concept for restoring the power envelope from a reverberation signal,” *Acoust. Sci. & Tech.*, 25, 232–242, 2004.
- [24] 天野成昭, 近藤公久, 坂本修一, 鈴木陽一, “親密度別単語了解度試験用音声データセット (FW03),” NII 音声資源コンソーシアム, 2006.
- [25] 坂本修一, 鈴木陽一, 天野成昭, 小澤賢司, 近藤公久, 曾根敏夫, “親密度と音韻バランスを考慮した単語了解度試験用リストの構築,” 日本音響学会誌, 54(12), 842–849, 1998.

付録

日本語音声の各周波数帯域の振幅包絡の変調成分が言語情報の取得に与える影響の検討

本論文中に日本語音声の各周波数帯域の振幅包絡の変調成分が言語情報の取得に与える影響の検討にて、掲載していない実験結果の図と共に全て下記に示す。

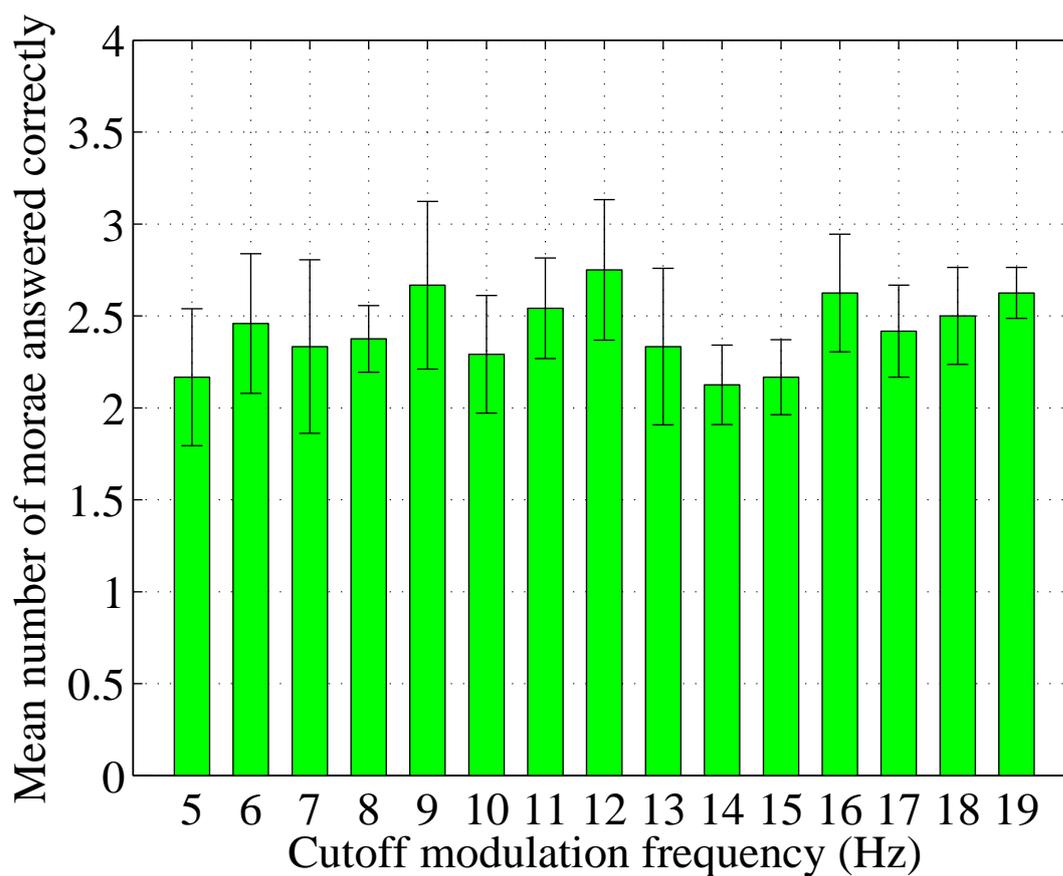


図 6.1: 第 1 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

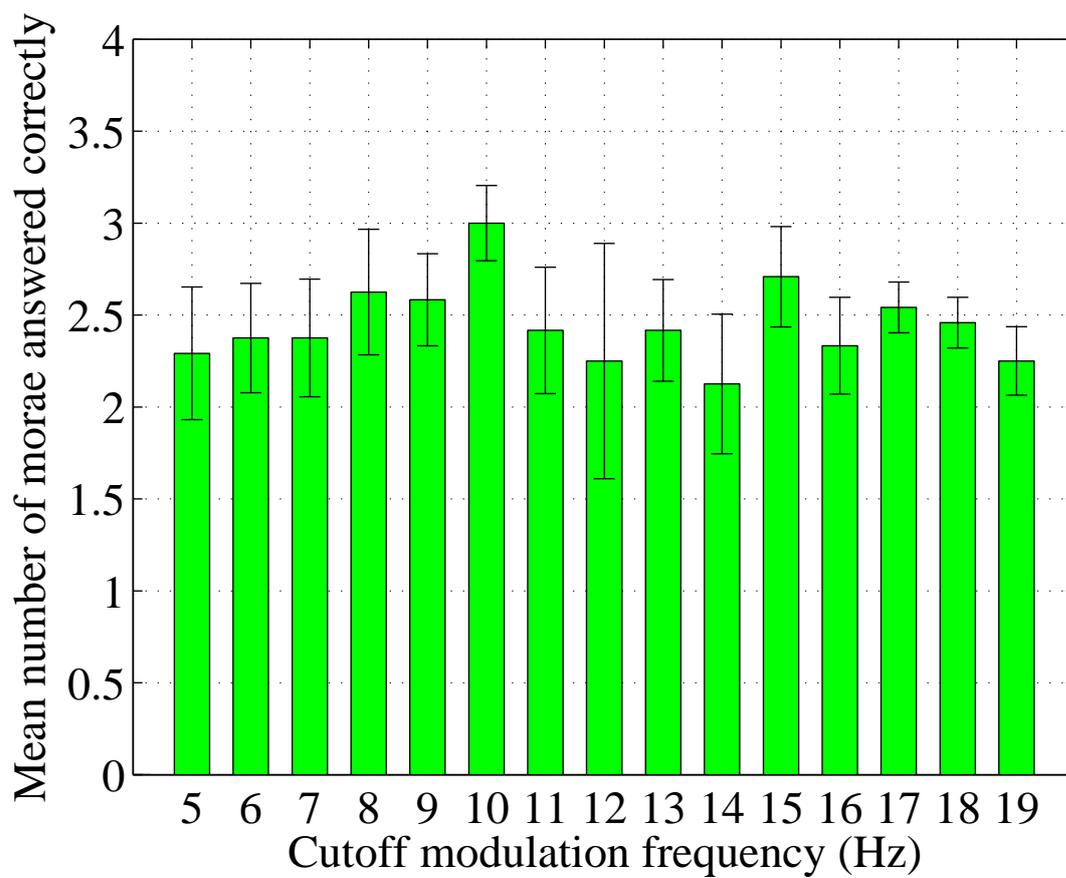


図 6.2: 第 2 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

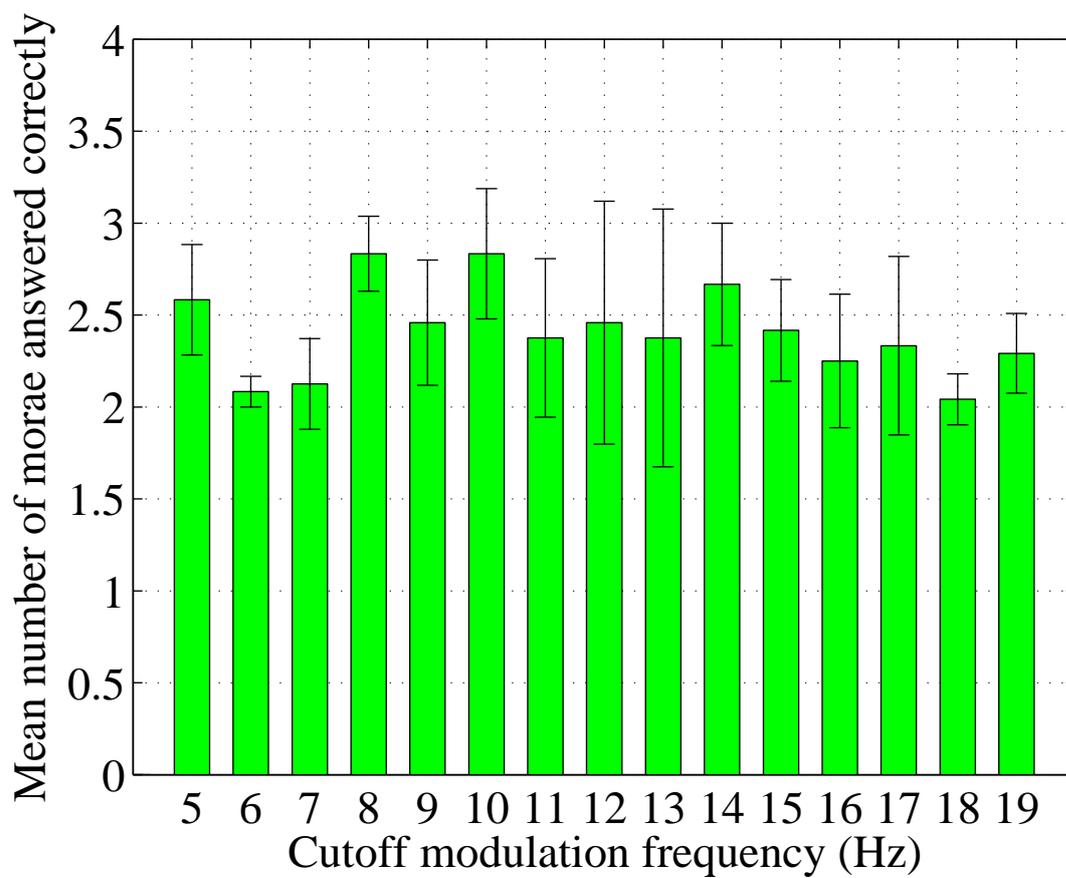


図 6.3: 第 3 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

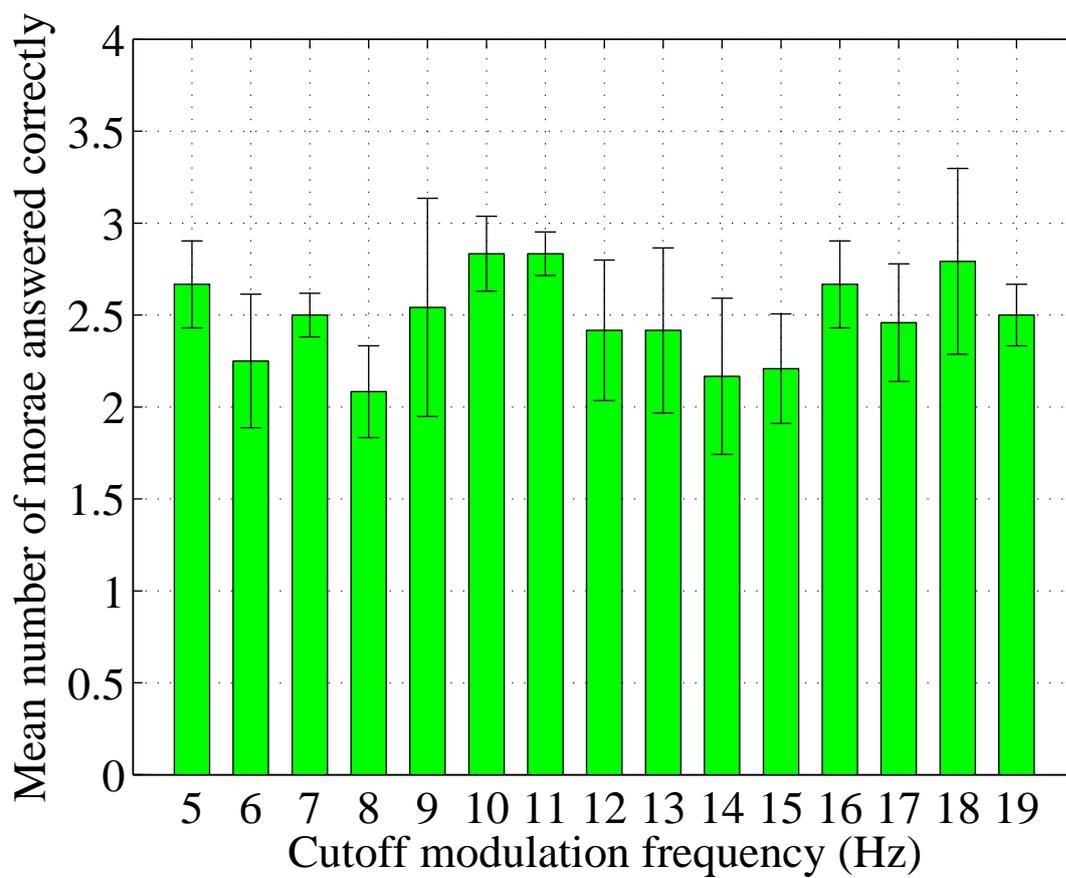


図 6.4: 第 4 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

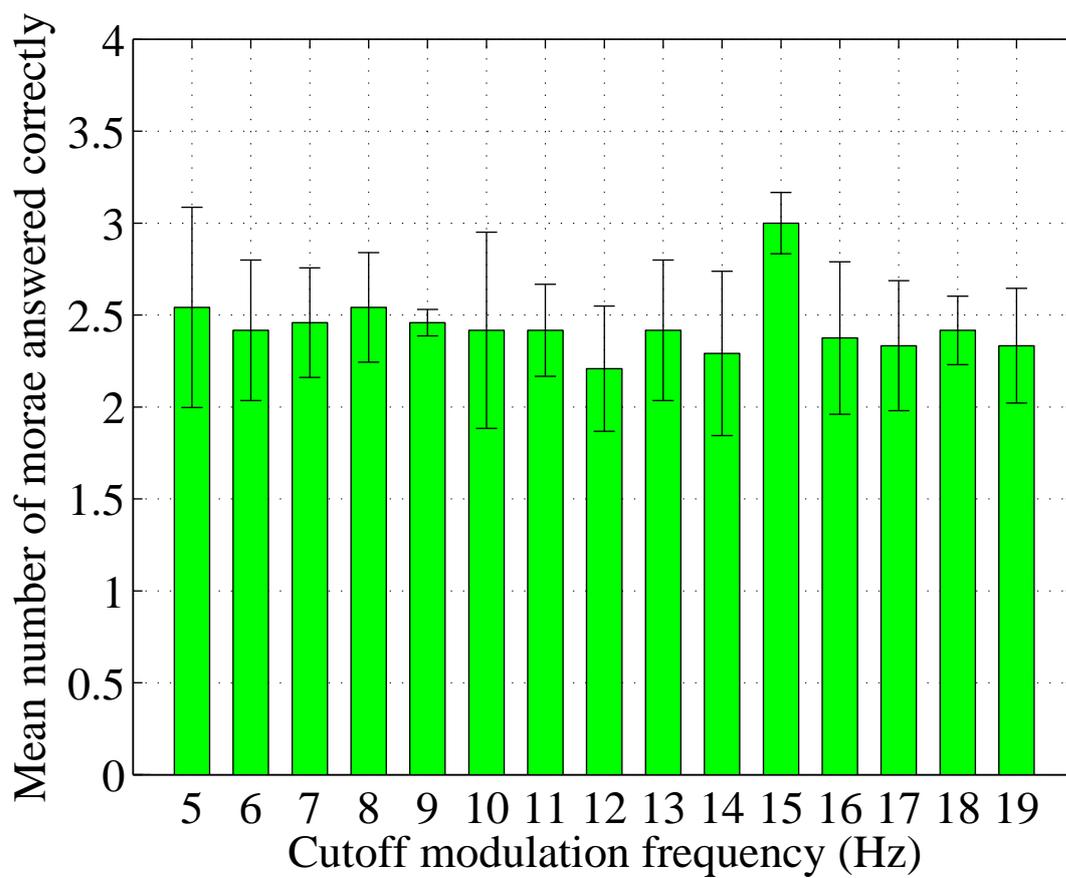


図 6.5: 第 5 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

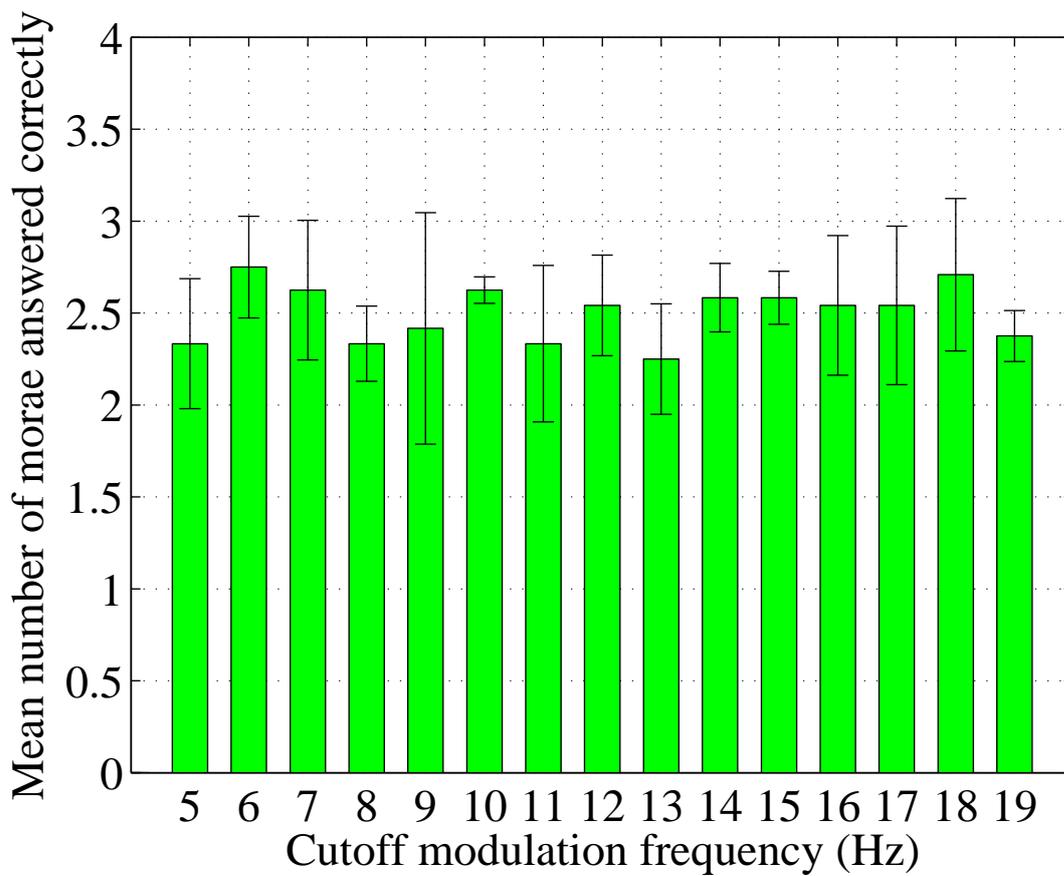


図 6.6: 第 6 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

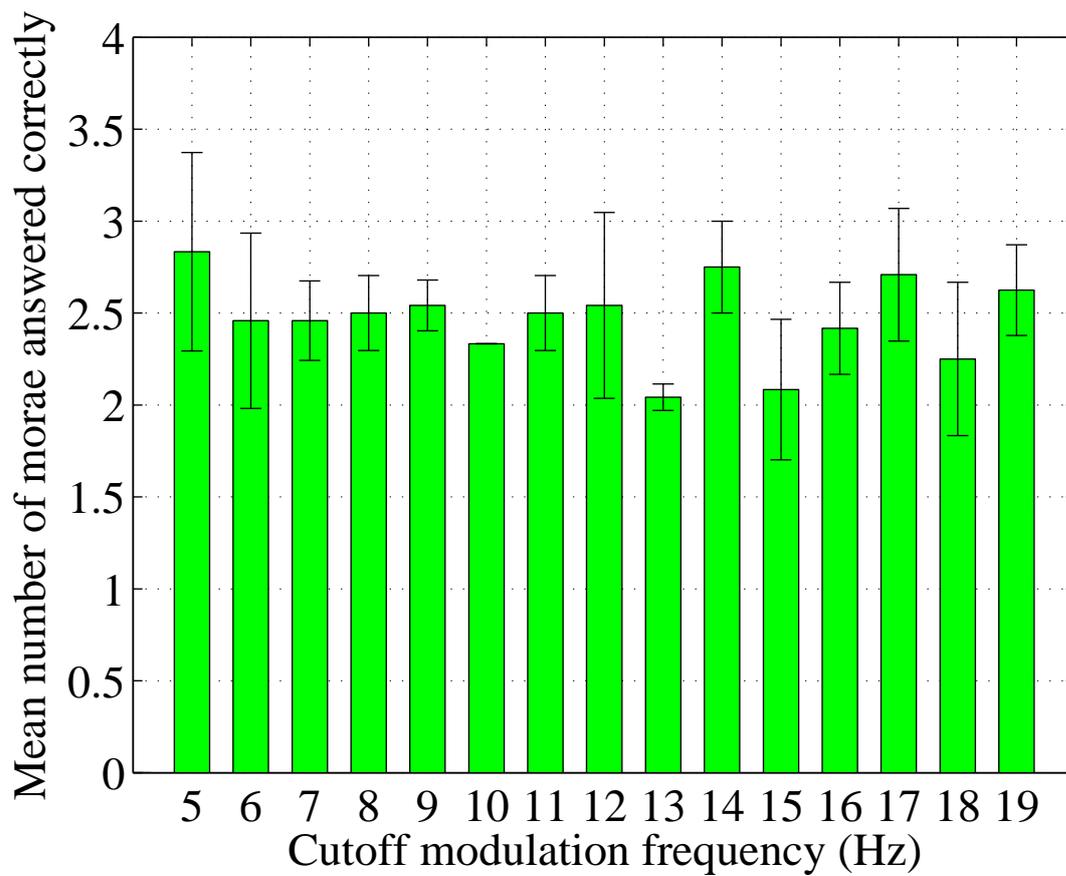


図 6.7: 第 7 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

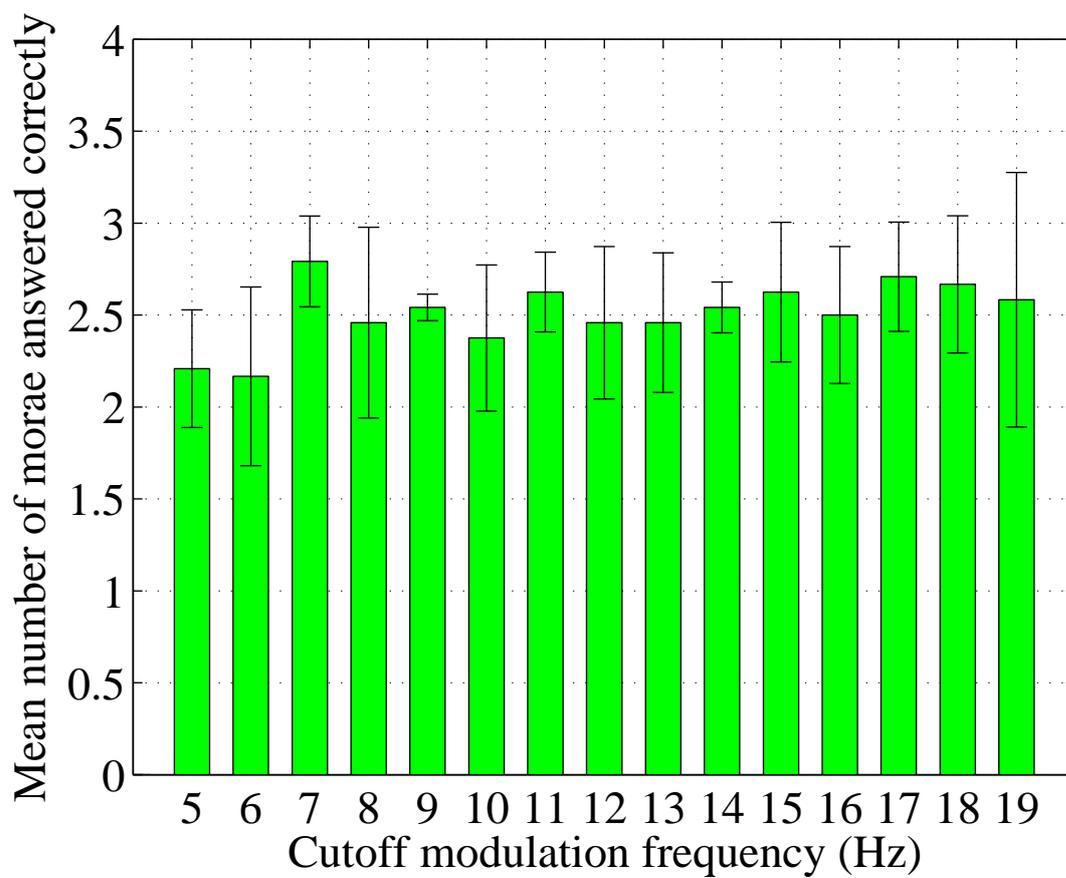


図 6.8: 第 8 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

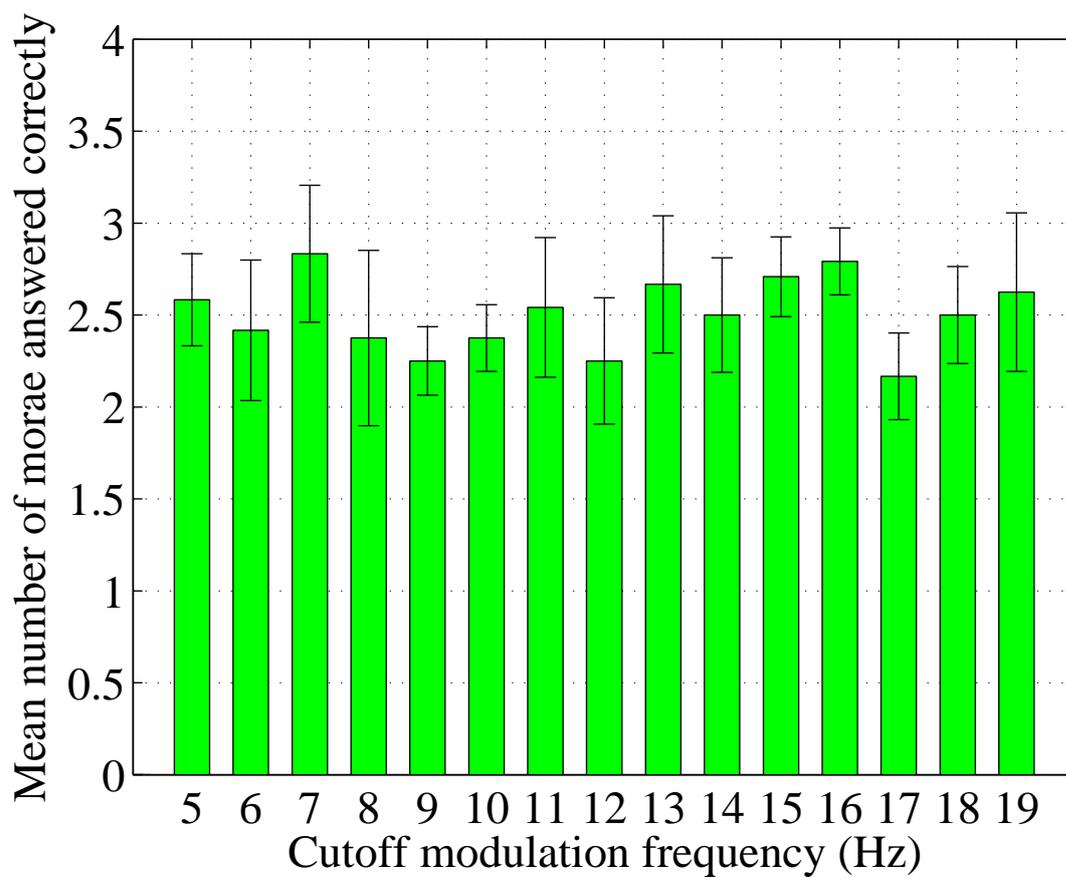


図 6.9: 第 9 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

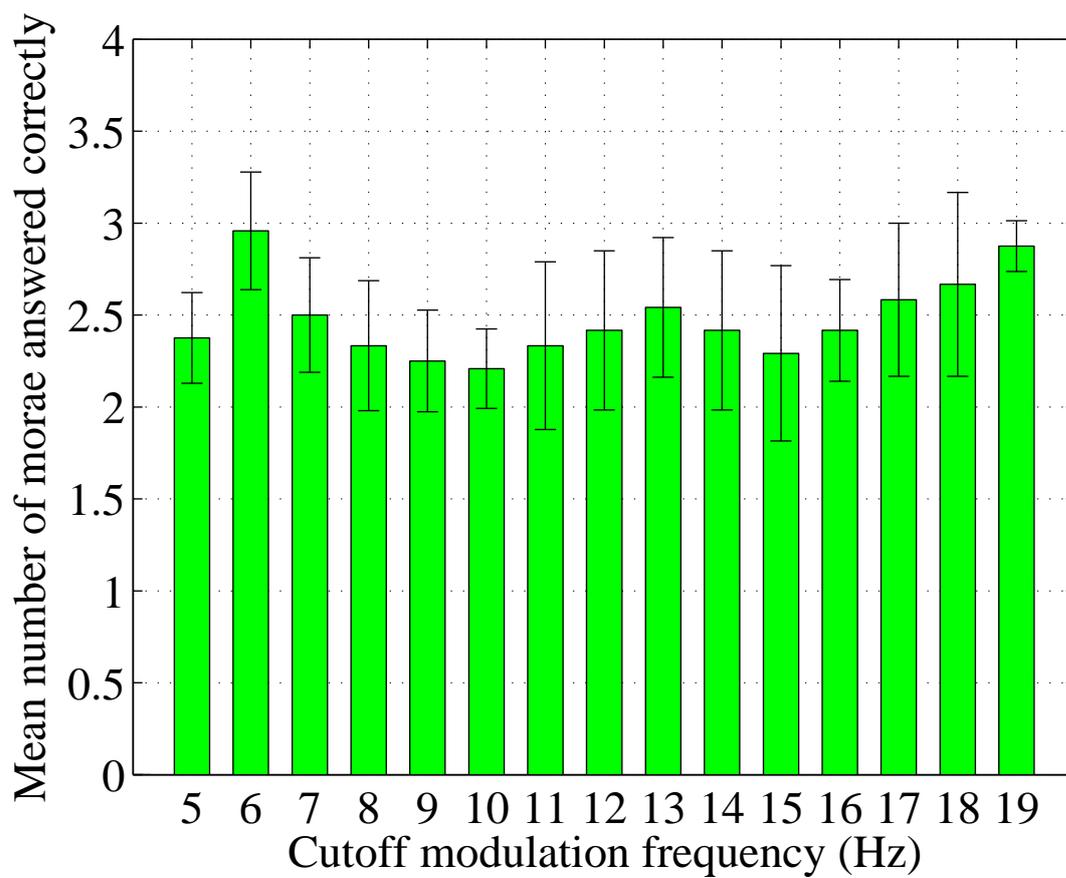


図 6.10: 第 10 周波数帯域の振幅包絡に含まれる変調成分の変化に対する平均正答モーラ数の変化

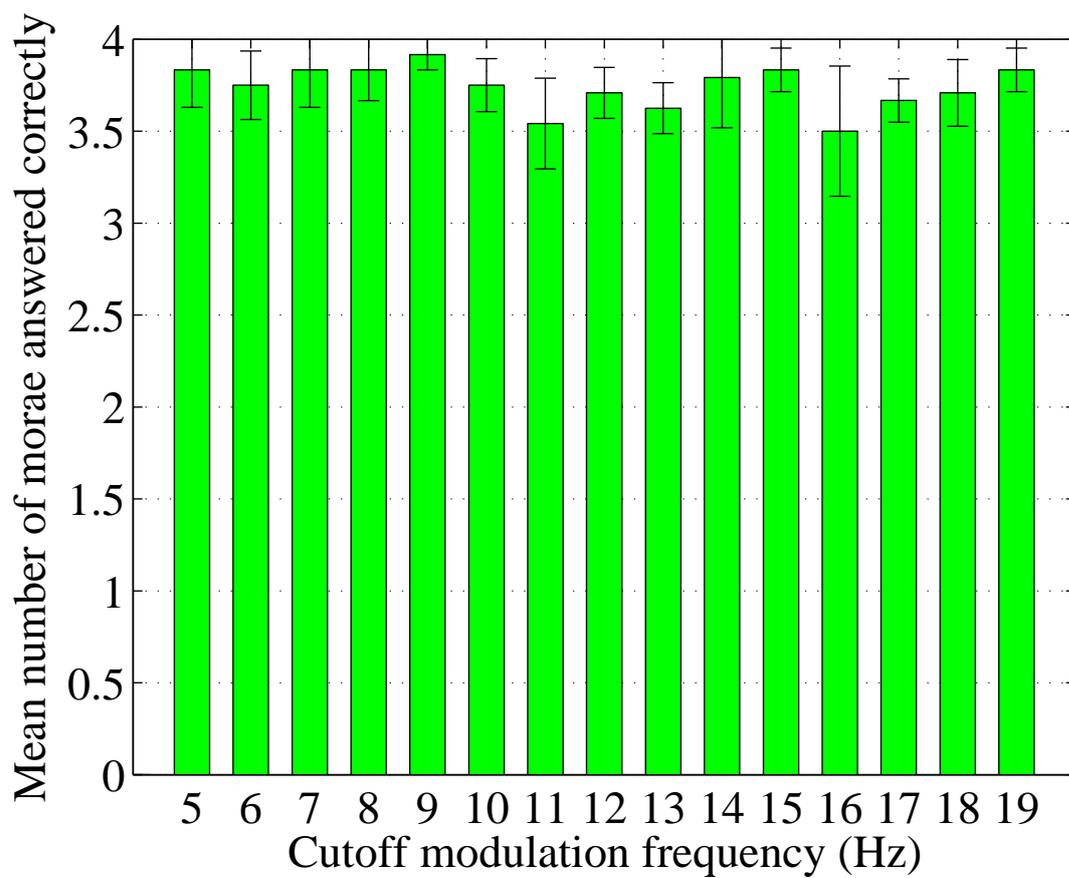


図 6.11: 第 1 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

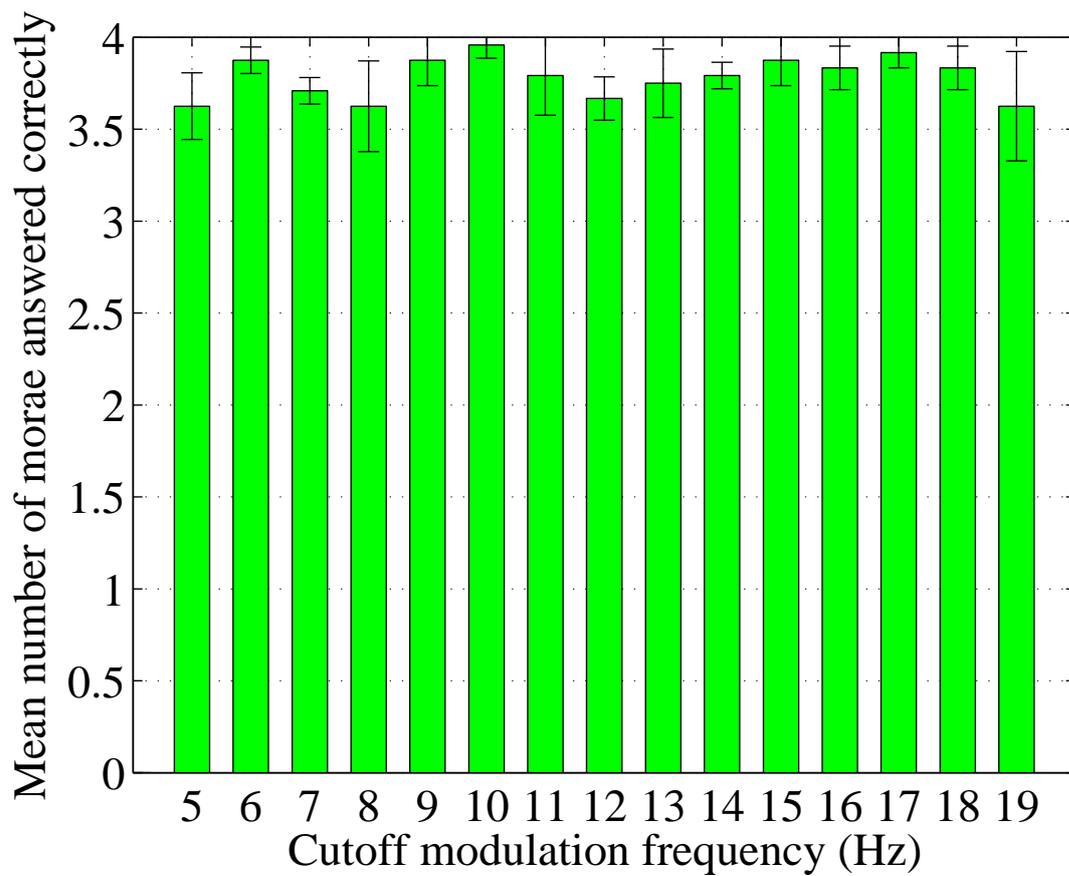


図 6.12: 第 2 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

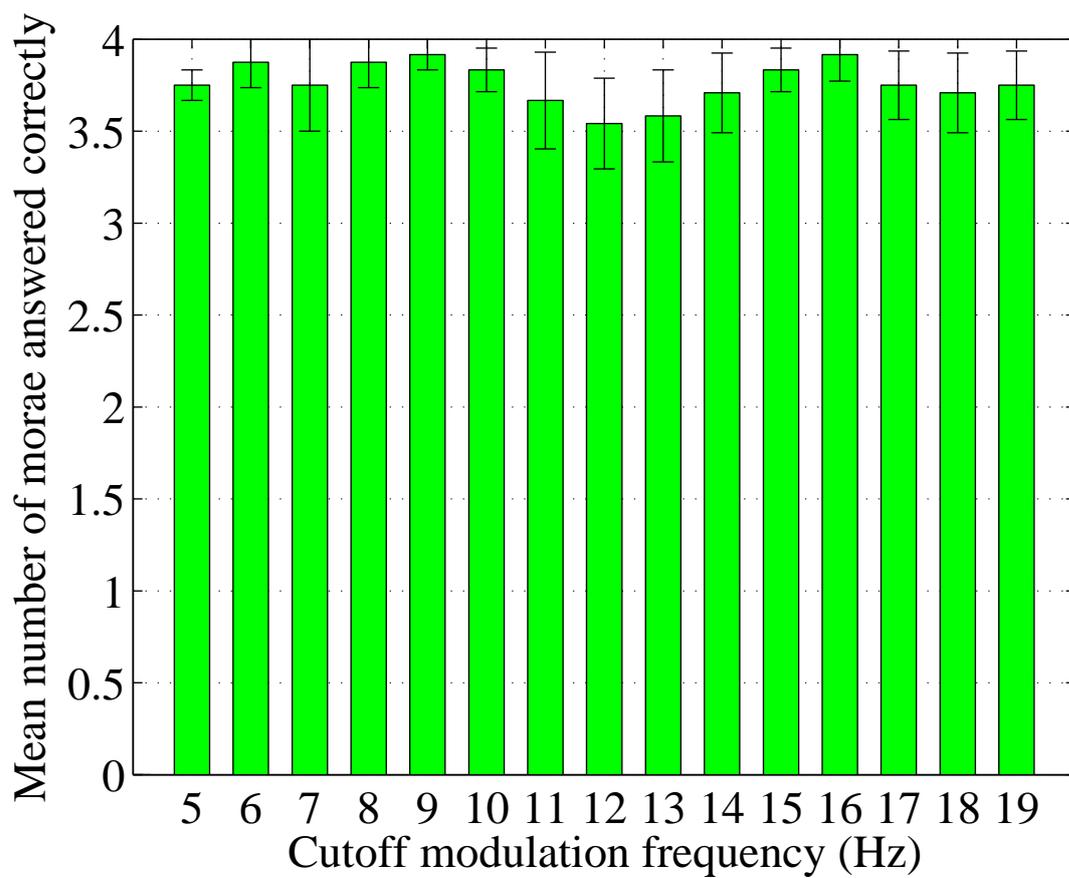


図 6.13: 第3周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

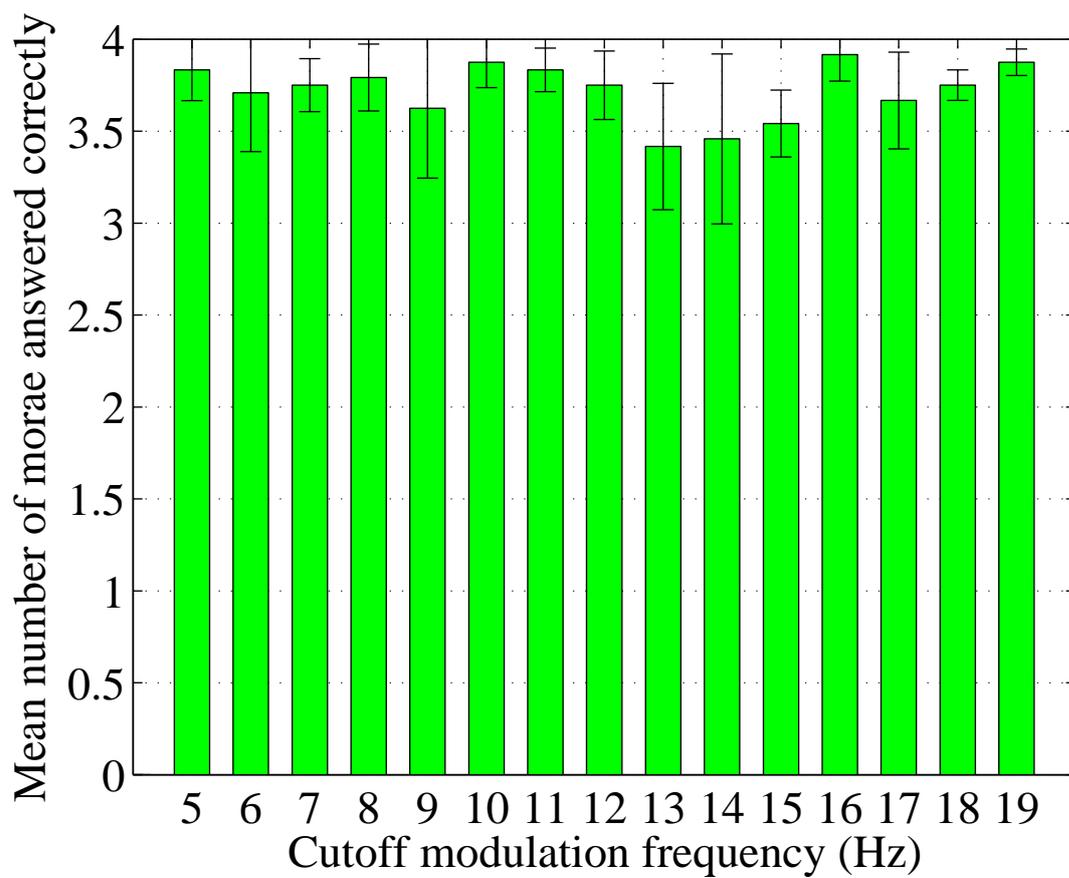


図 6.14: 第 4 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

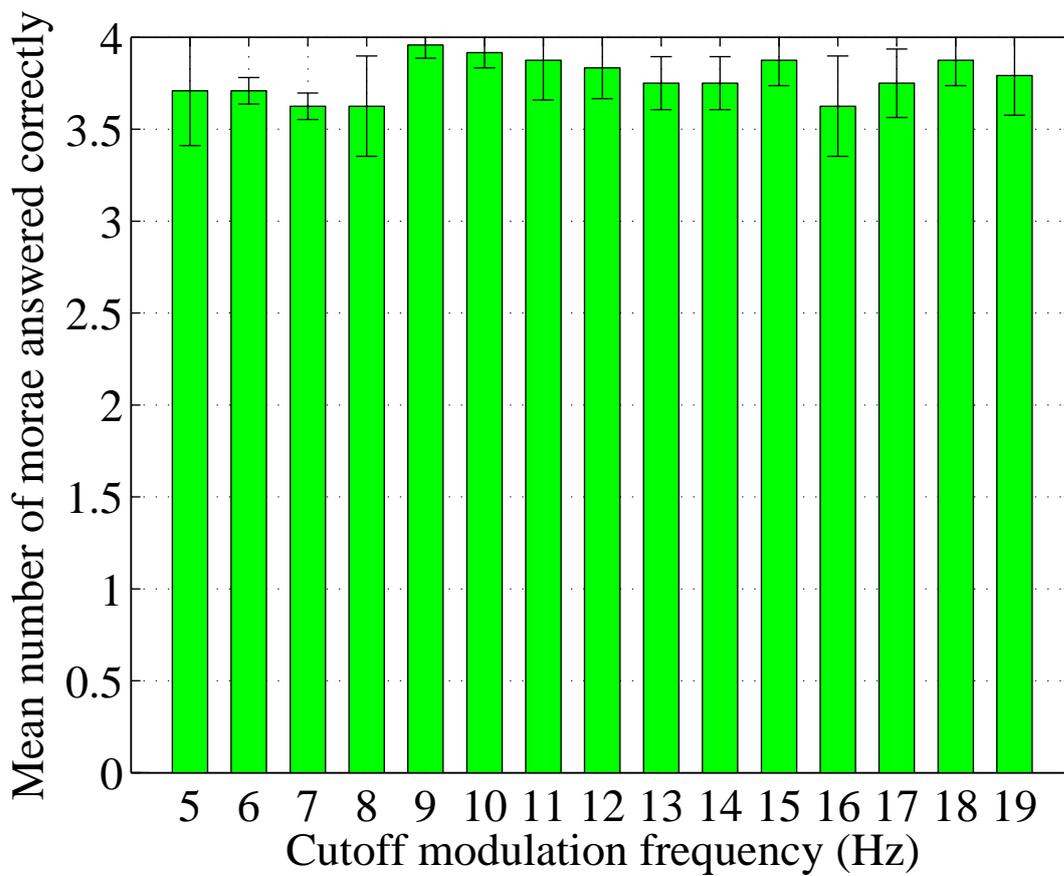


図 6.15: 第 5 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

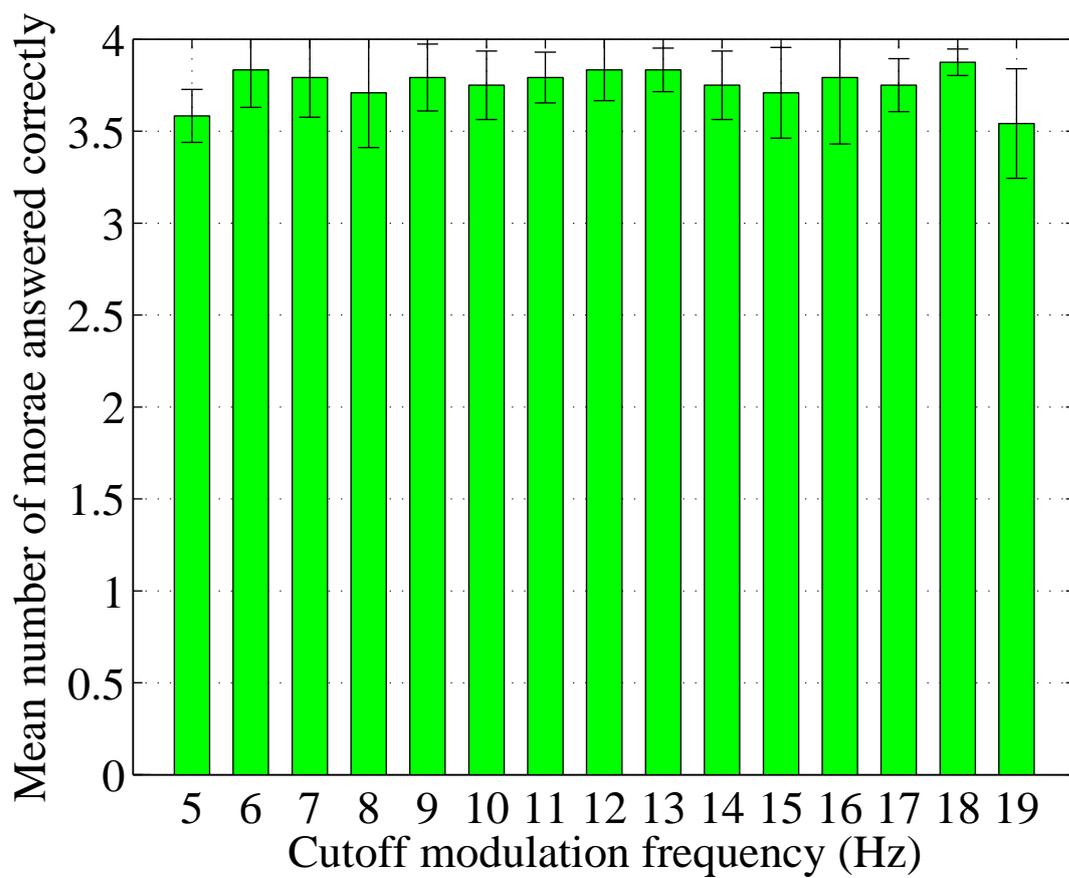


図 6.16: 第 6 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

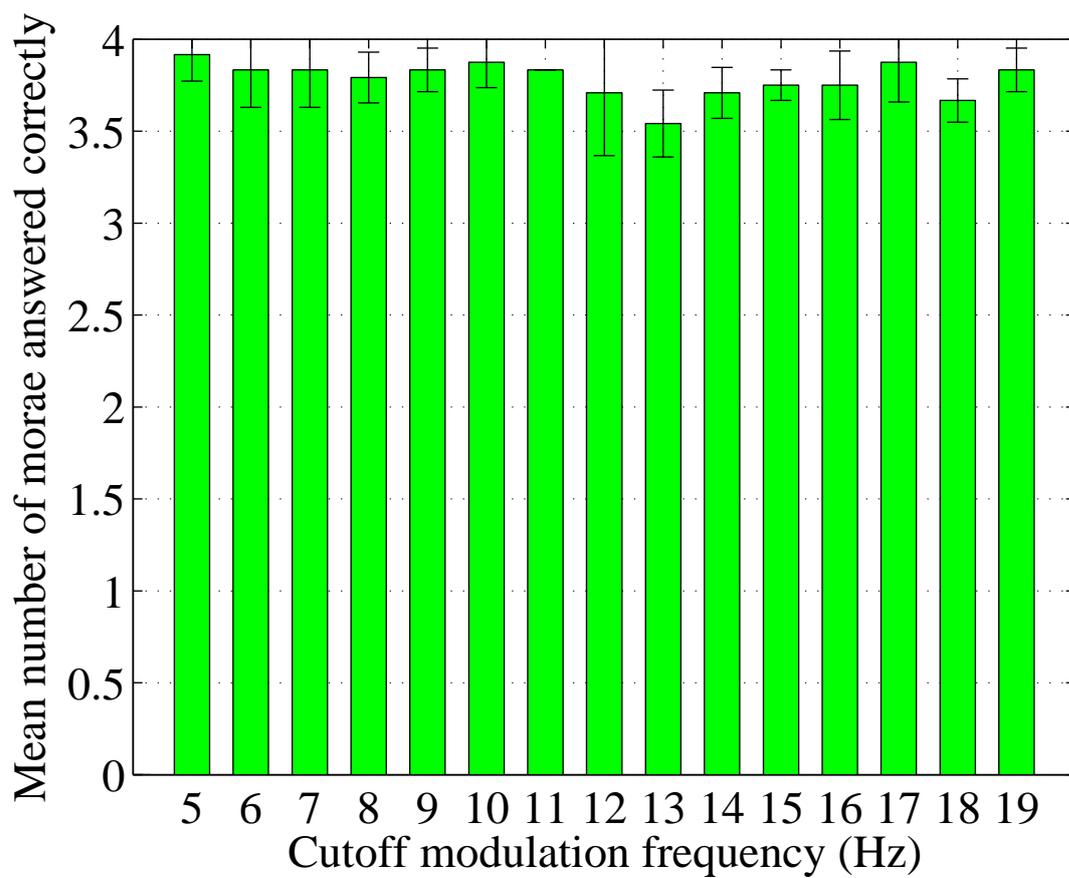


図 6.17: 第7周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

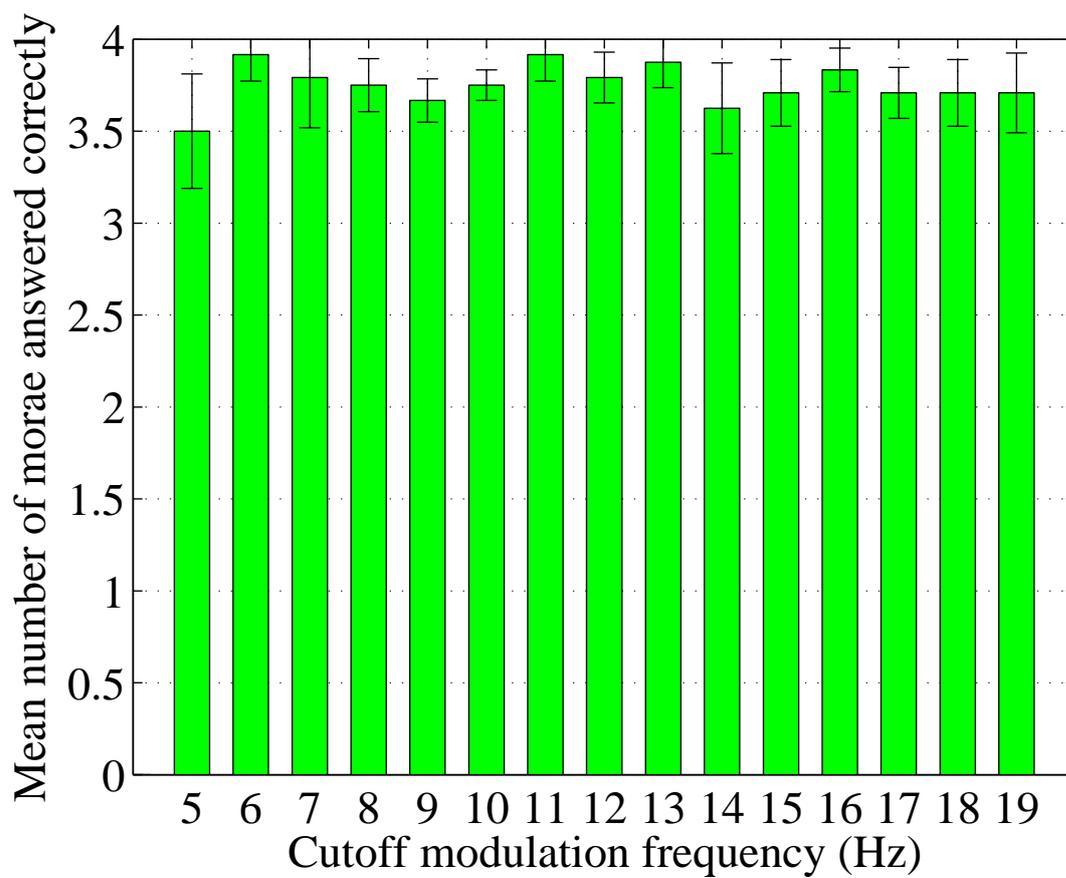


図 6.18: 第 8 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

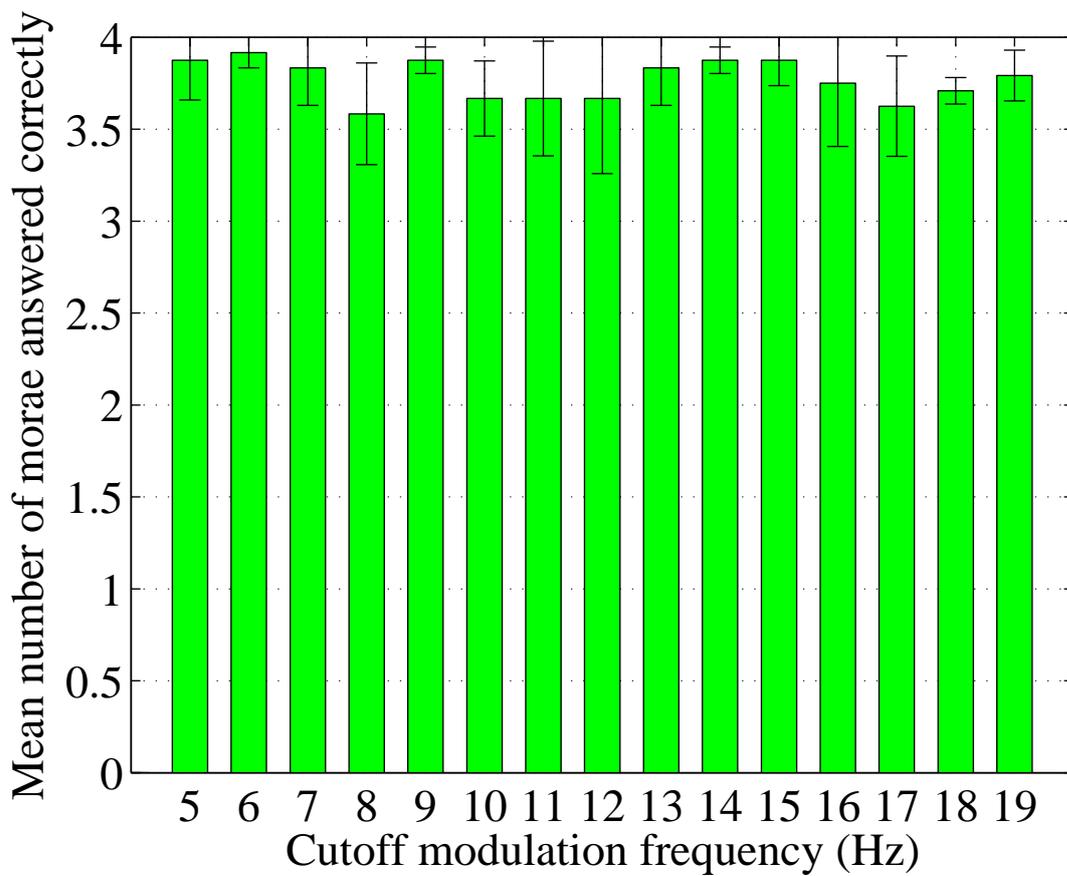


図 6.19: 第 9 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

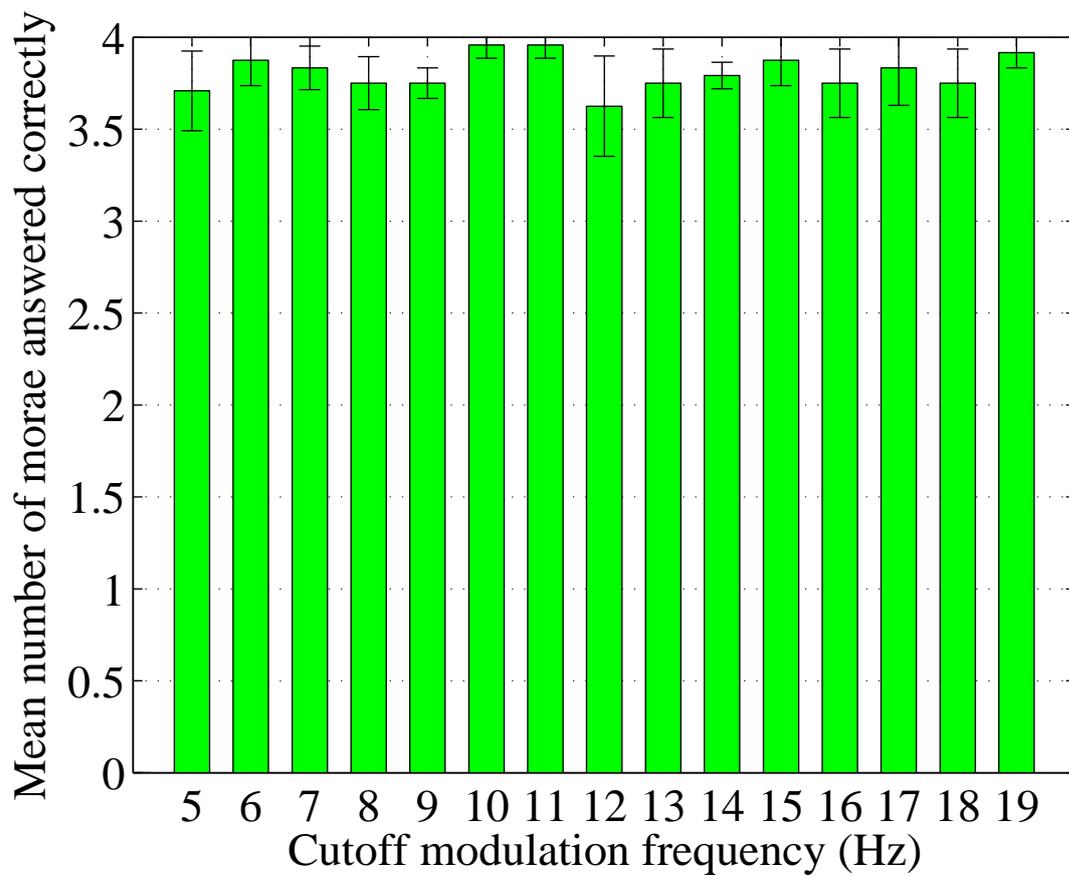


図 6.20: 第 10 周波数帯域の振幅包絡に含まれる変調成分の変化に対する母音の平均正答モーラ数の変化

音声波形

第4章にて、調査する子音を選定する際に使用した音声波形の図を下記に示す。

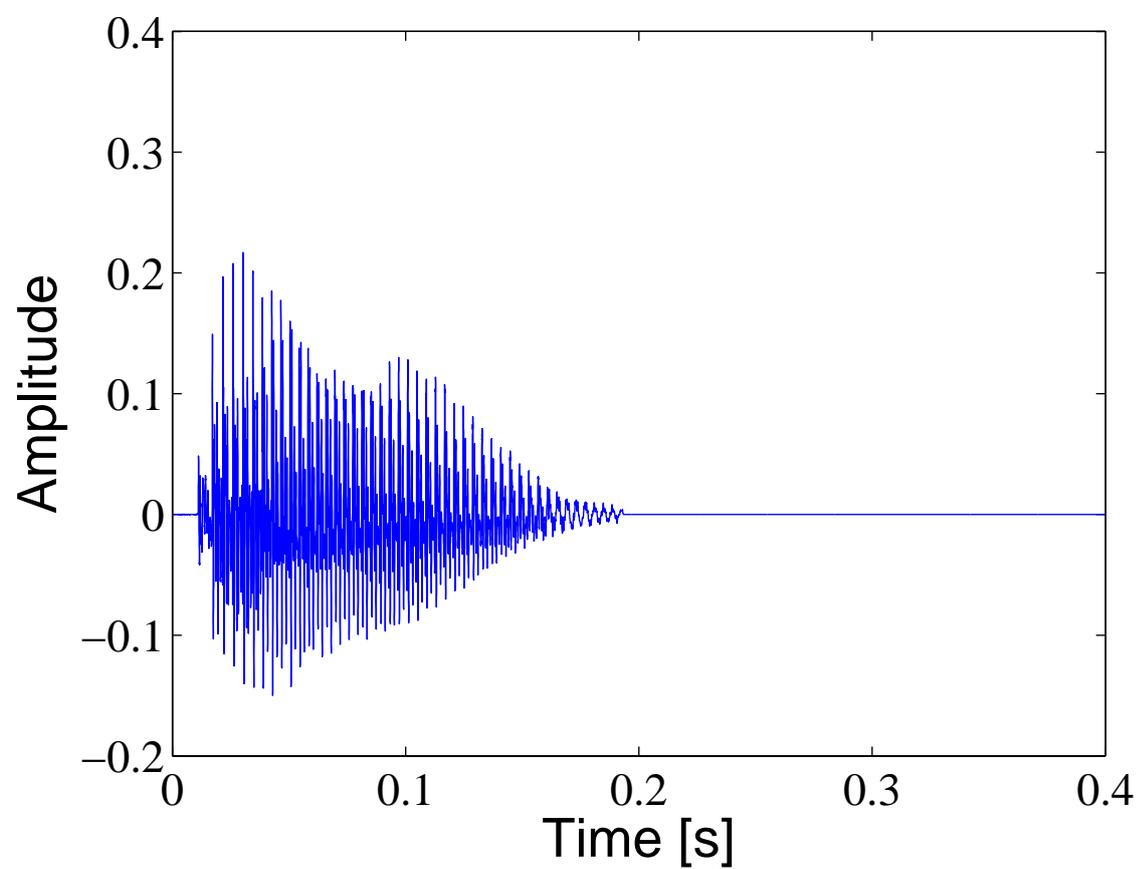


図 6.21: /a/ の音声波形

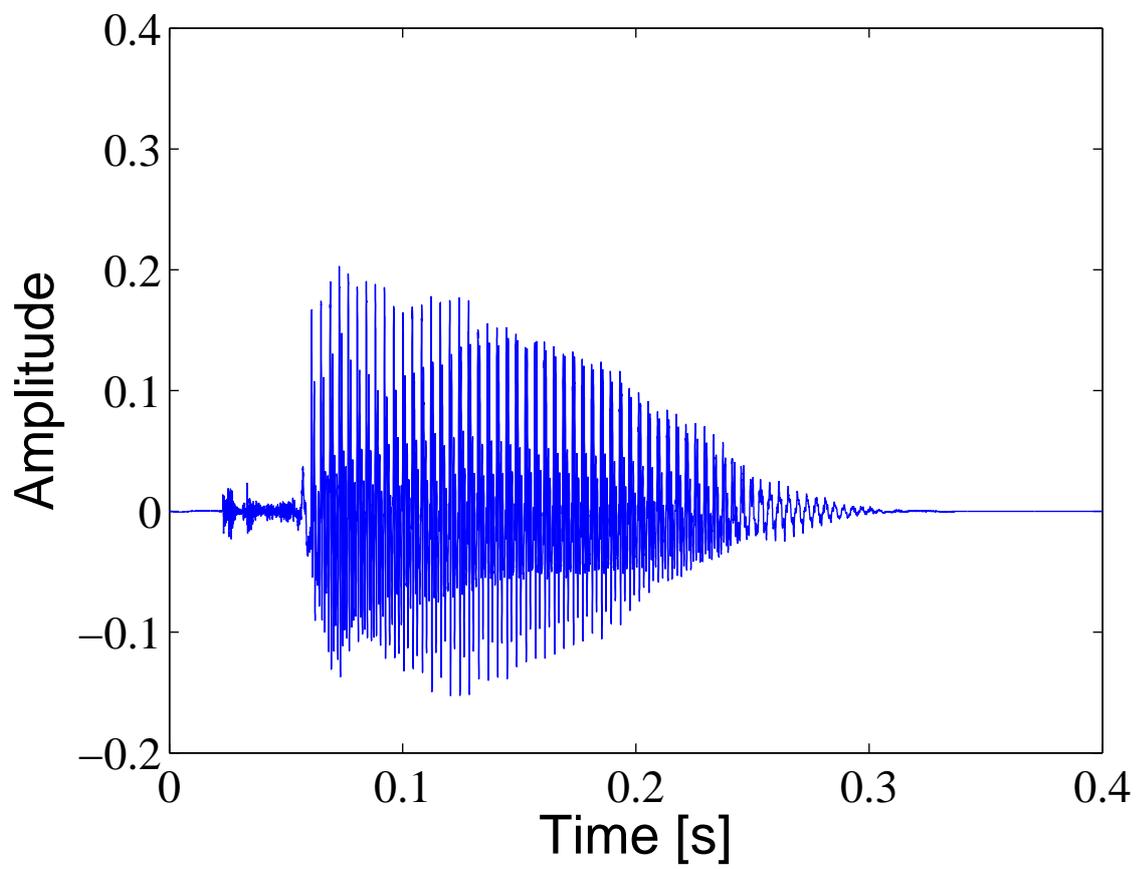


図 6.22: /ka/ の音声波形

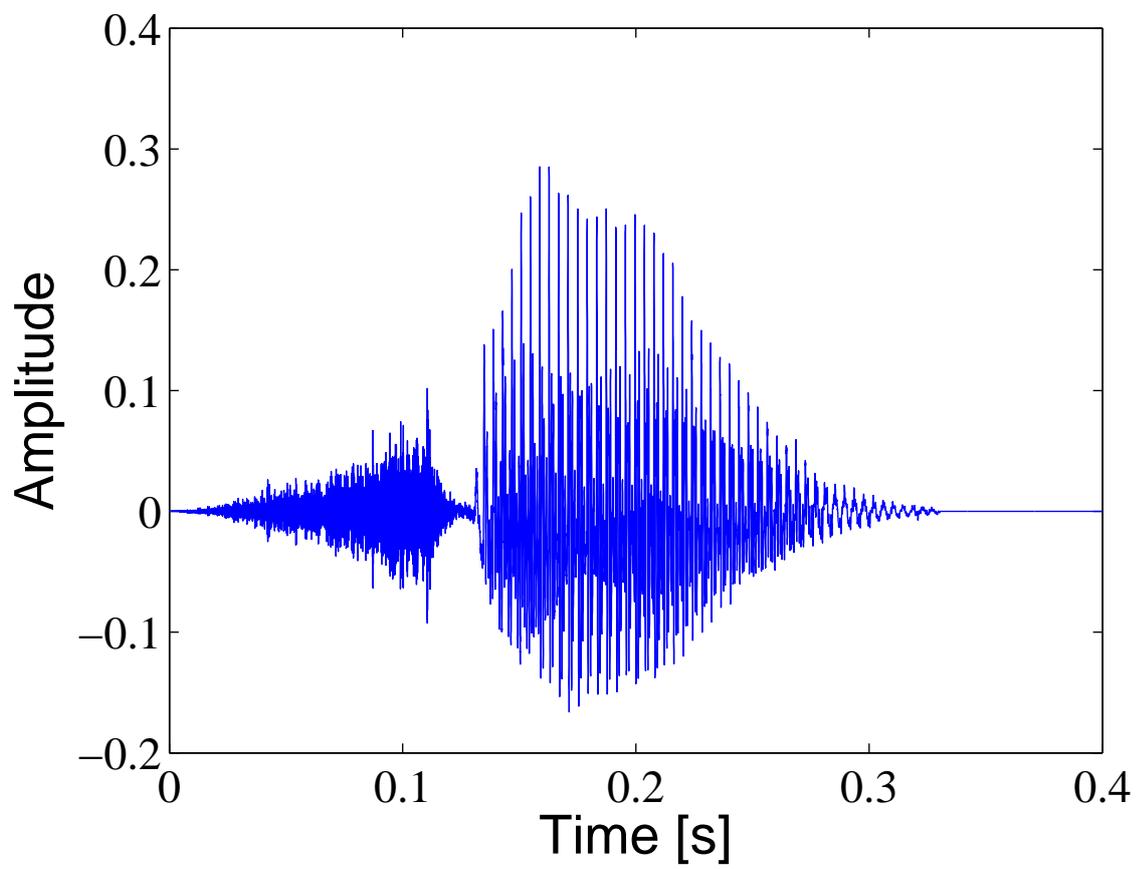


図 6.23: /sa/ の音声波形

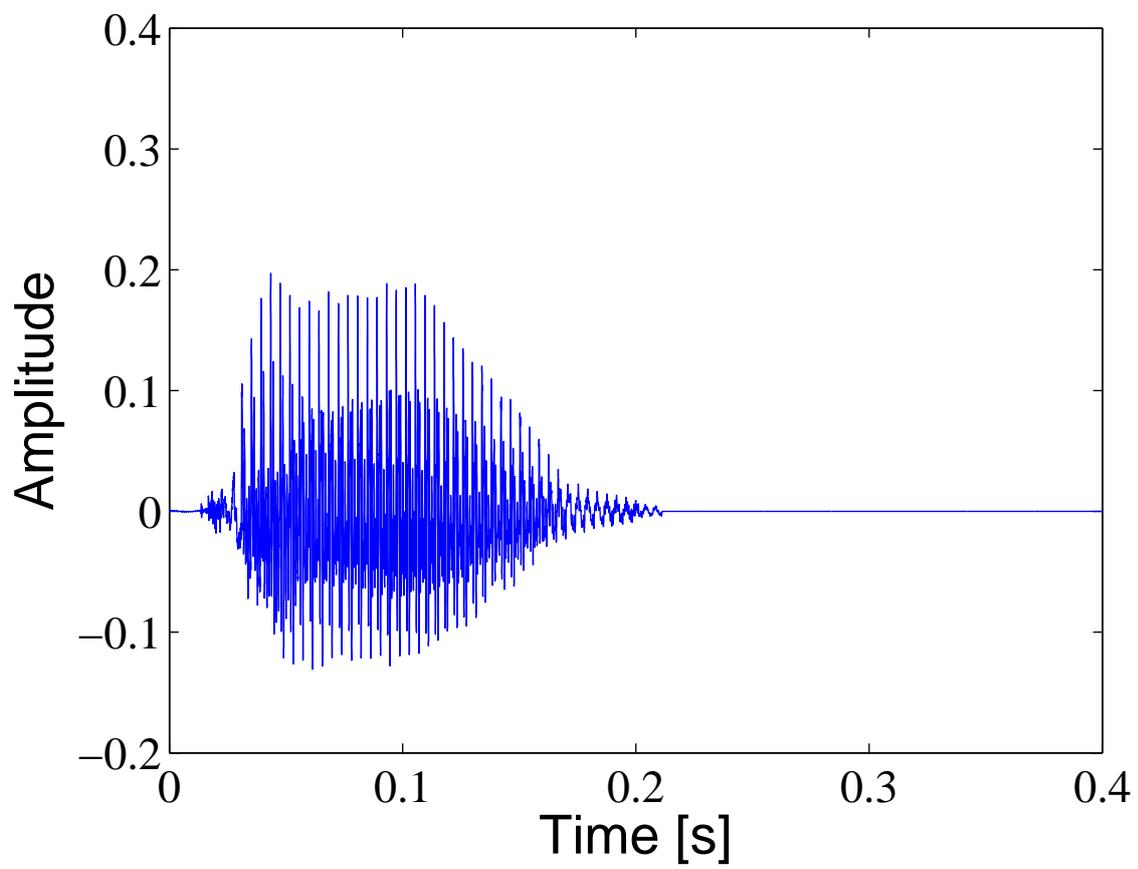


図 6.24: /ta/ の音声波形

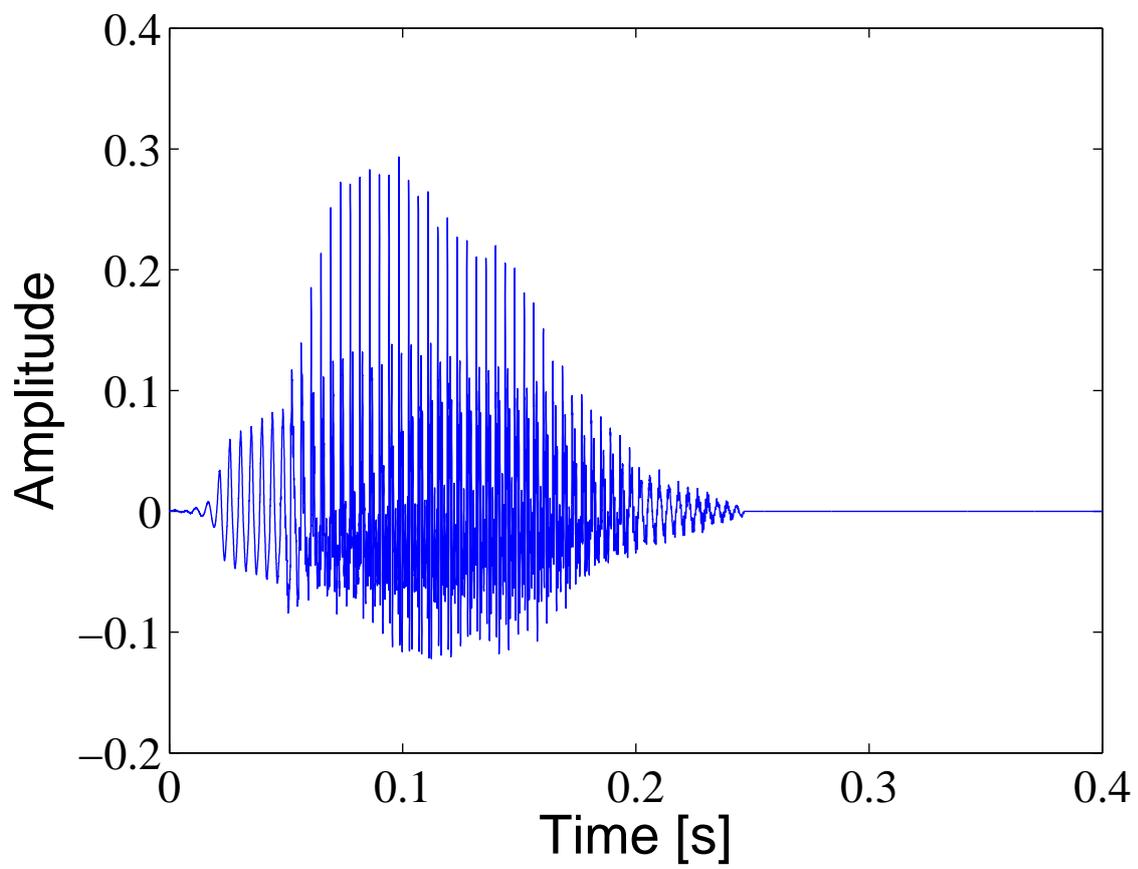


図 6.25: /na/ の音声波形

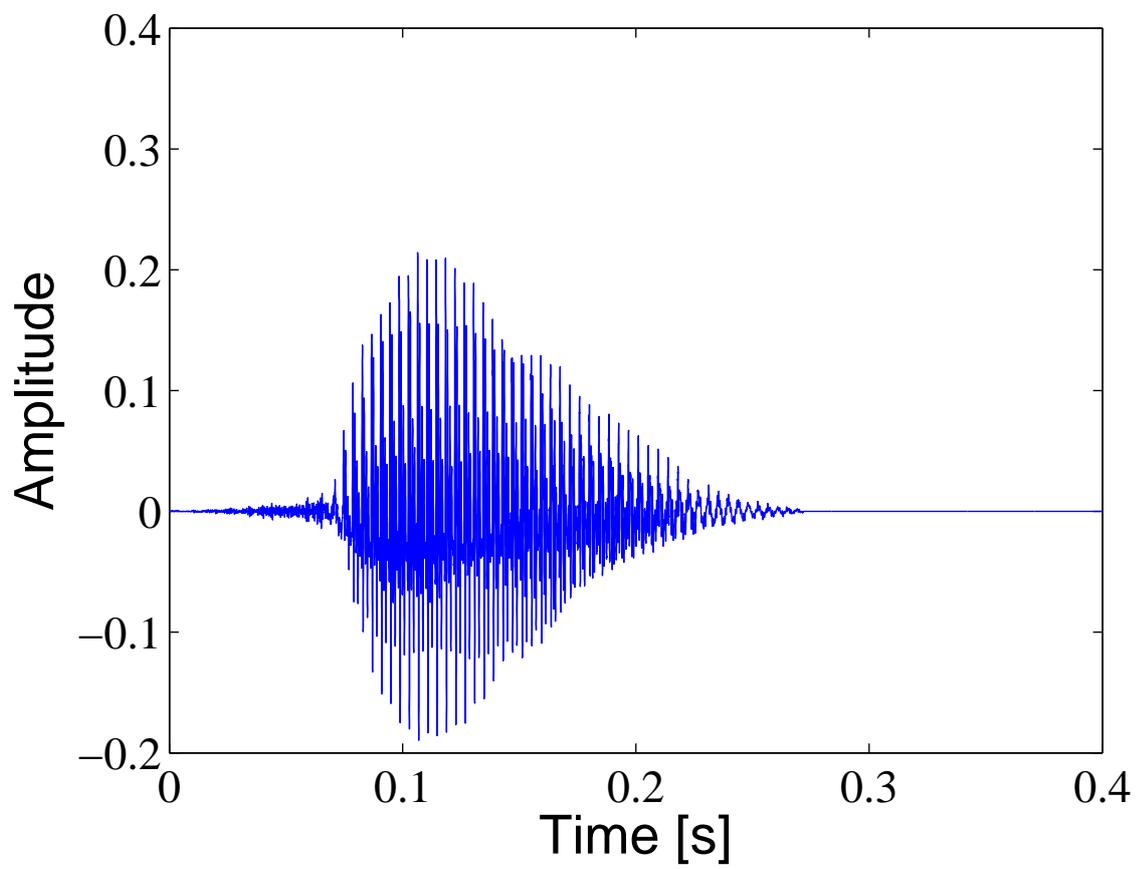


図 6.26: /ha/ の音声波形

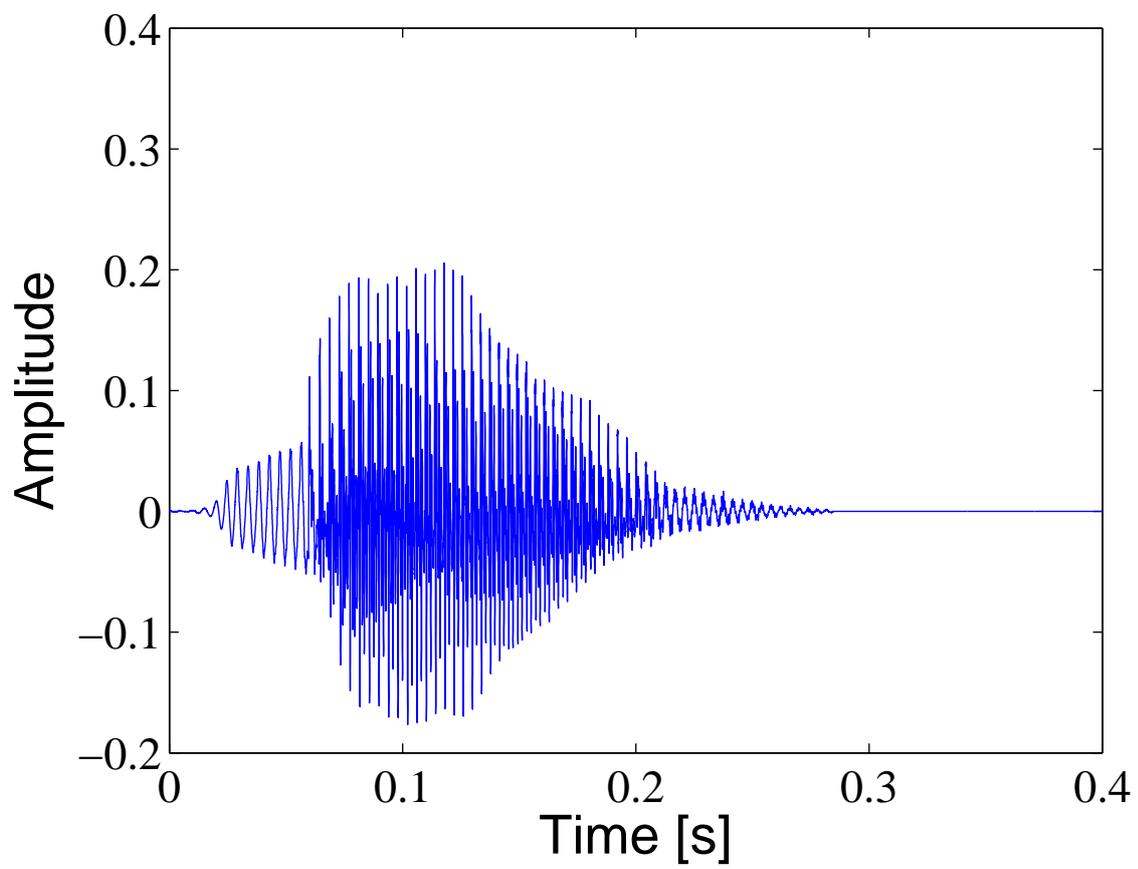


図 6.27: /ma/ の音声波形

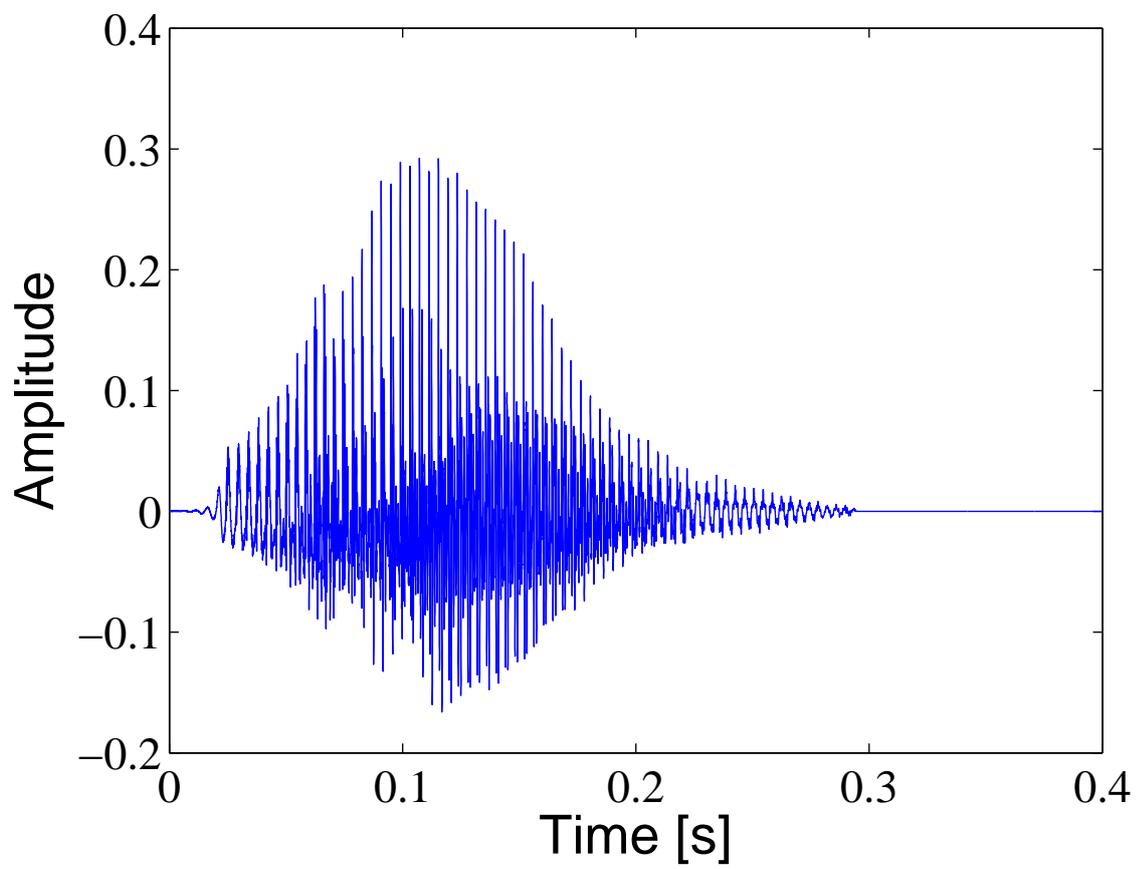


図 6.28: /ya/ の音声波形

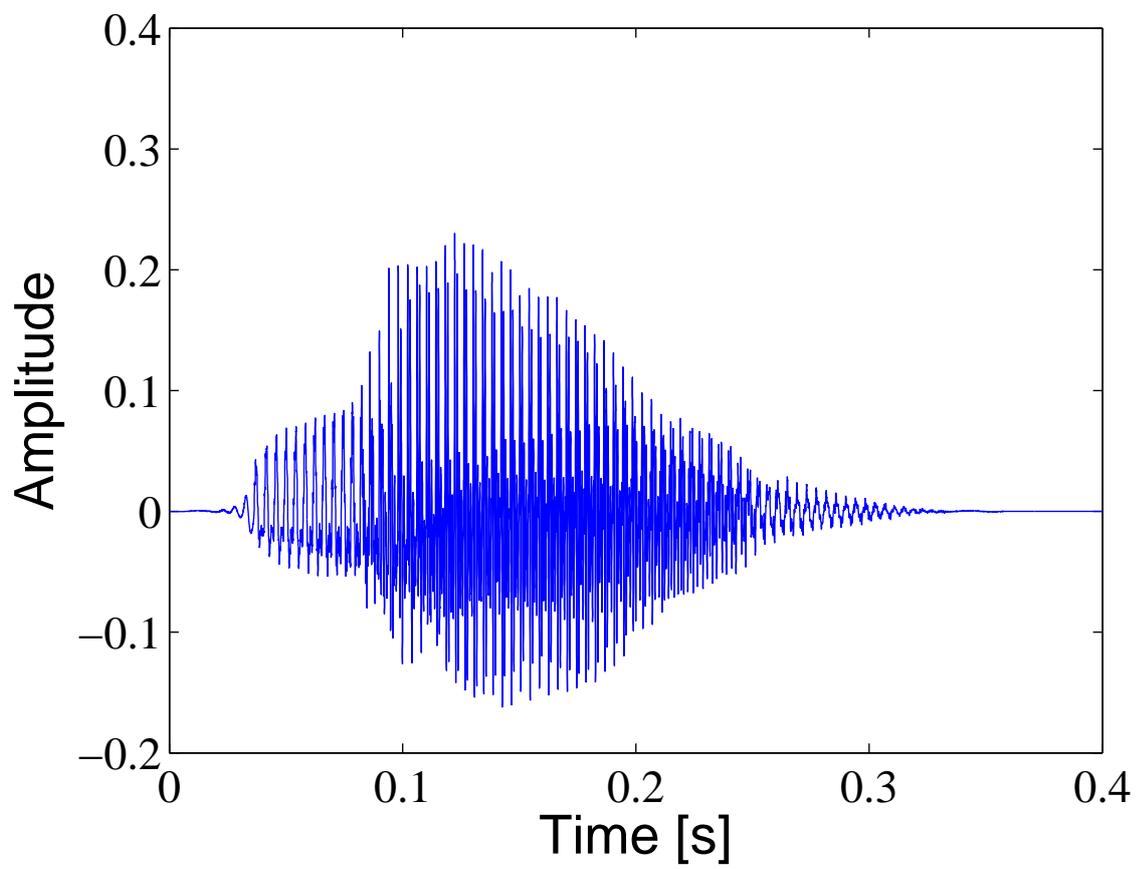


図 6.29: /ra/ の音声波形

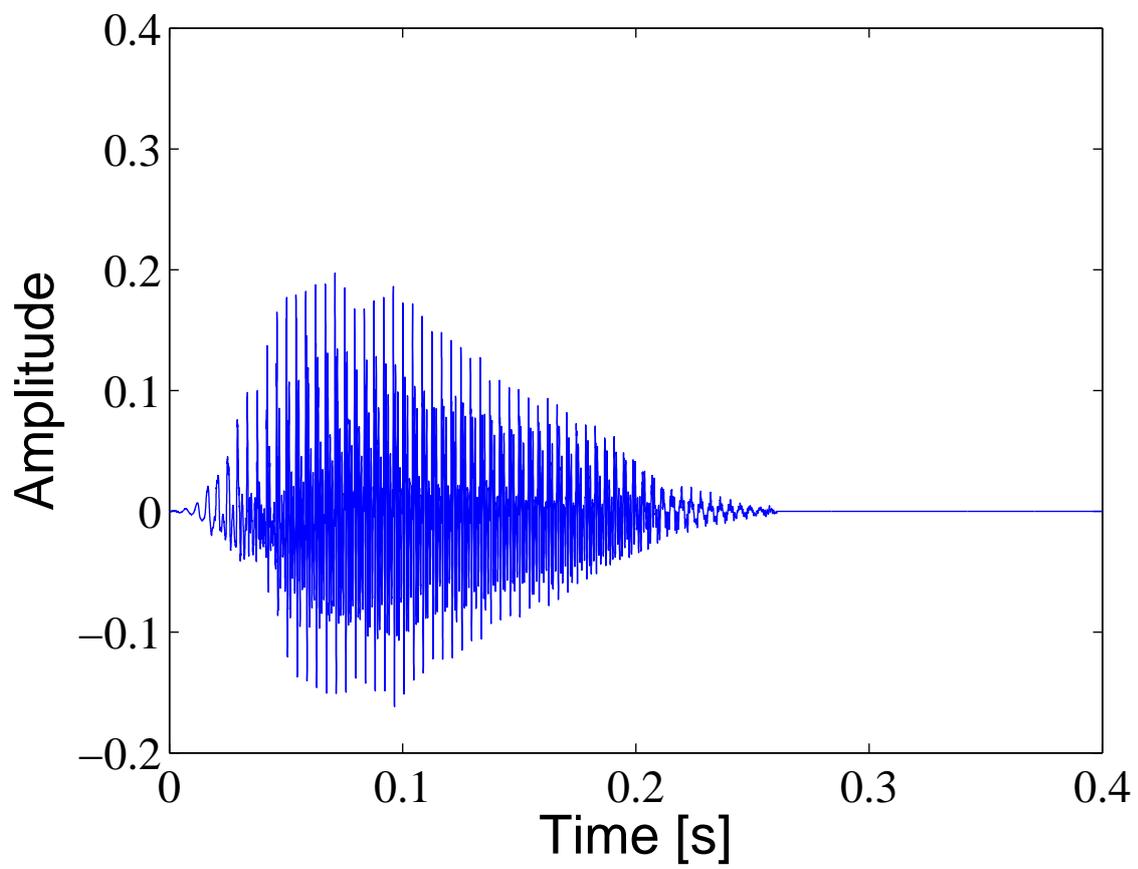


図 6.30: /wa/ の音声波形

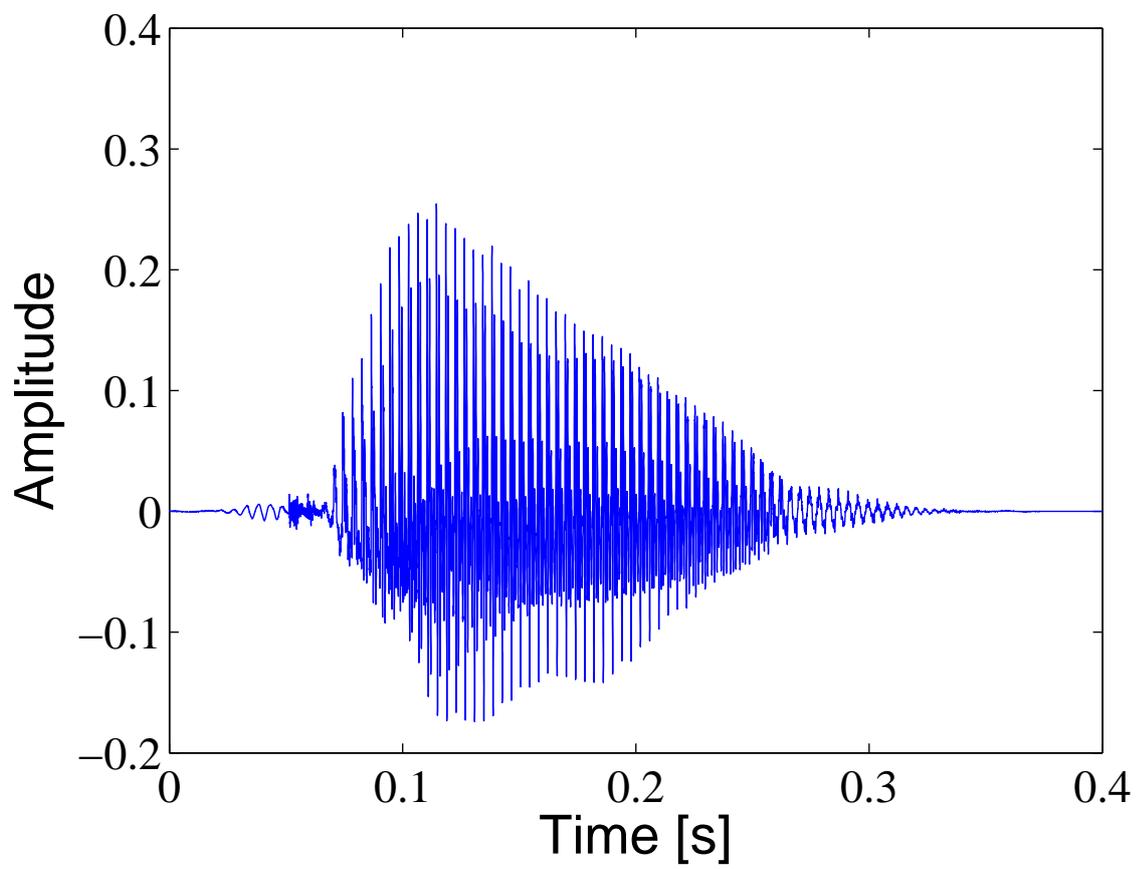


図 6.31: /ga/ の音声波形

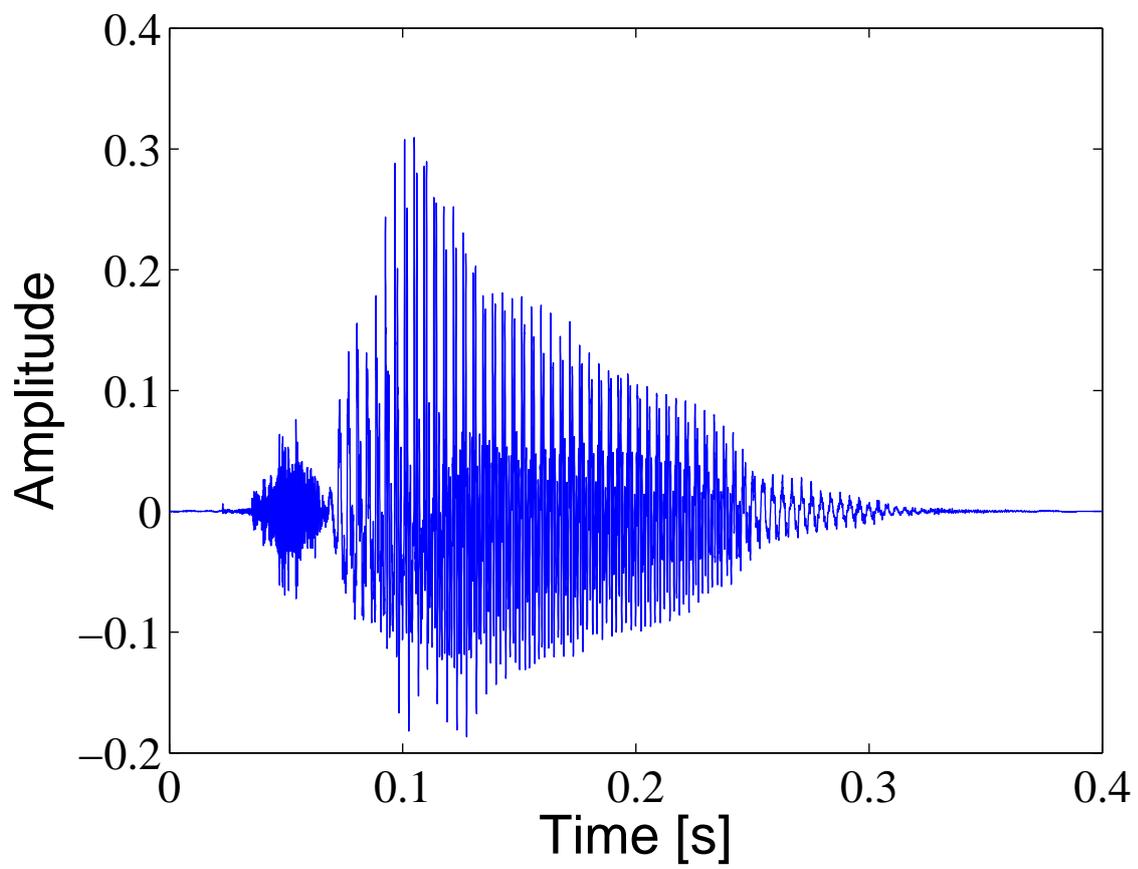


図 6.32: /za/ の音声波形

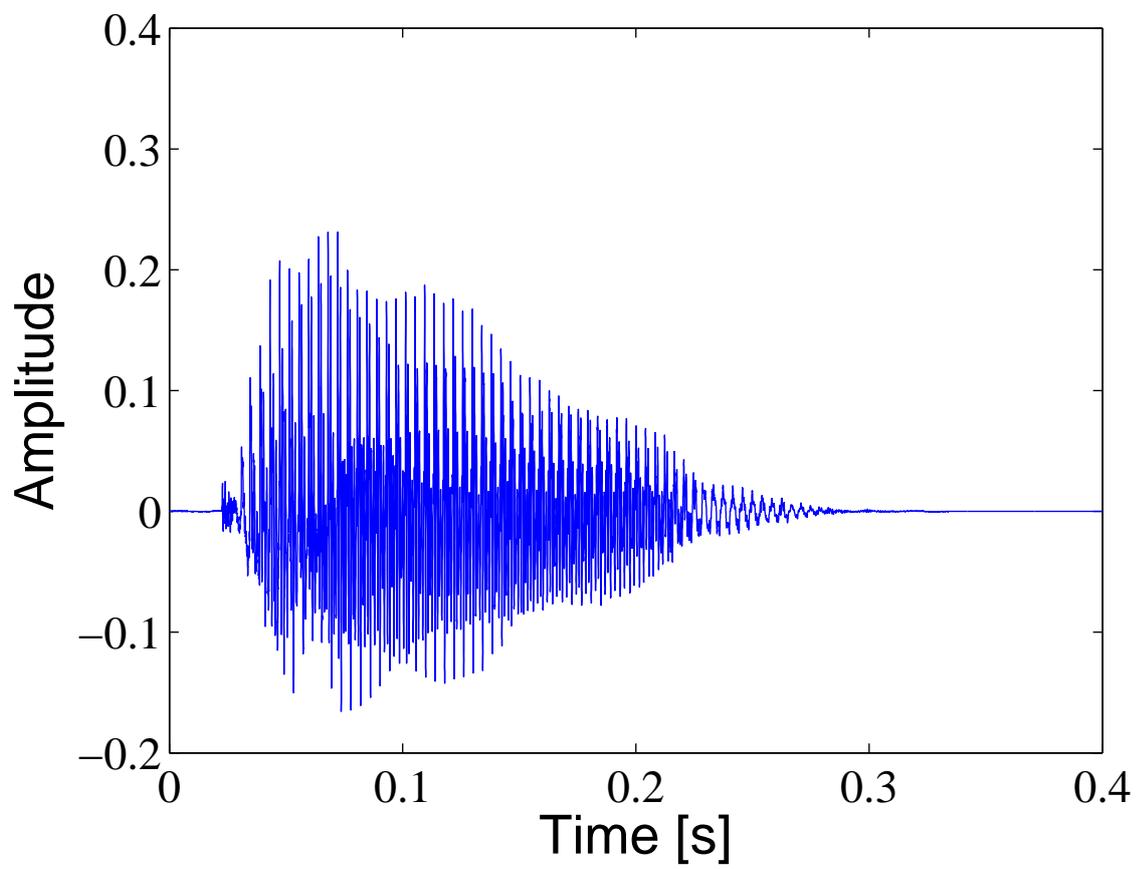


図 6.33: /da/ の音声波形

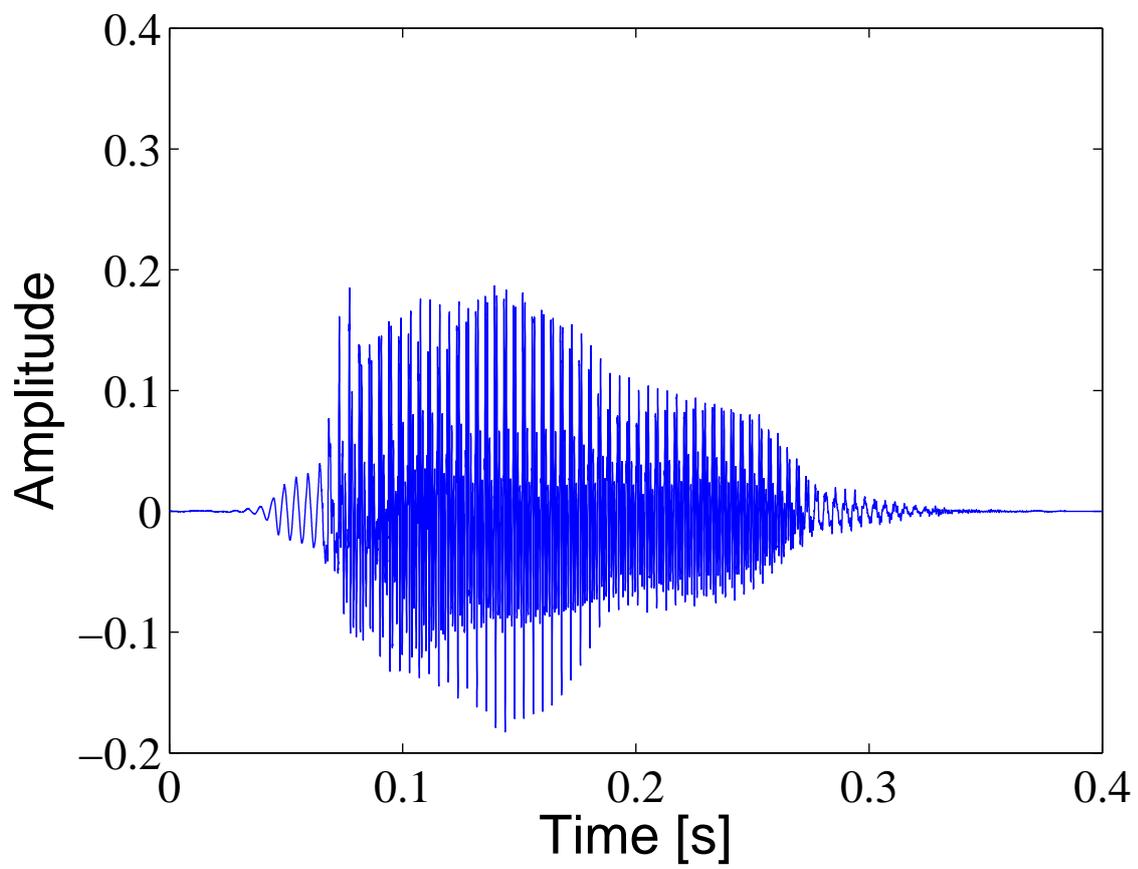


図 6.34: /ba/ の音声波形

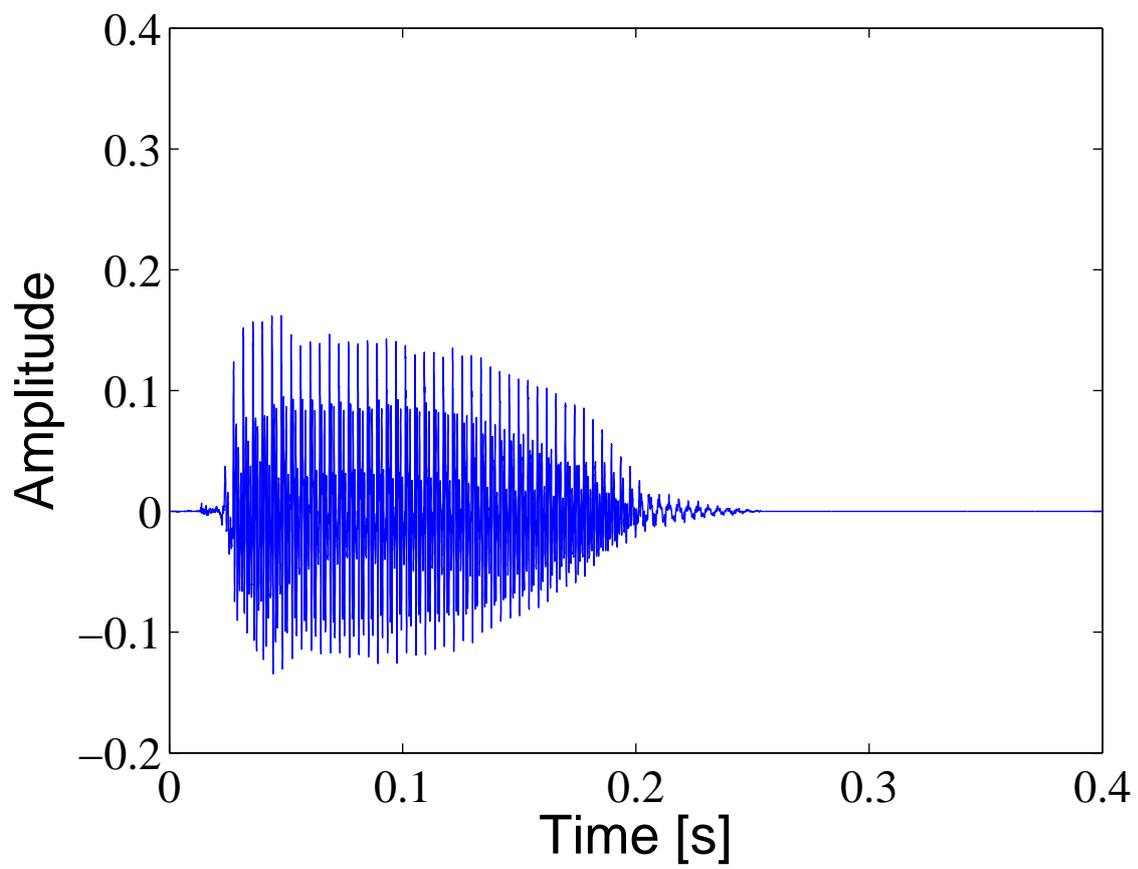


図 6.35: /pa/ の音声波形

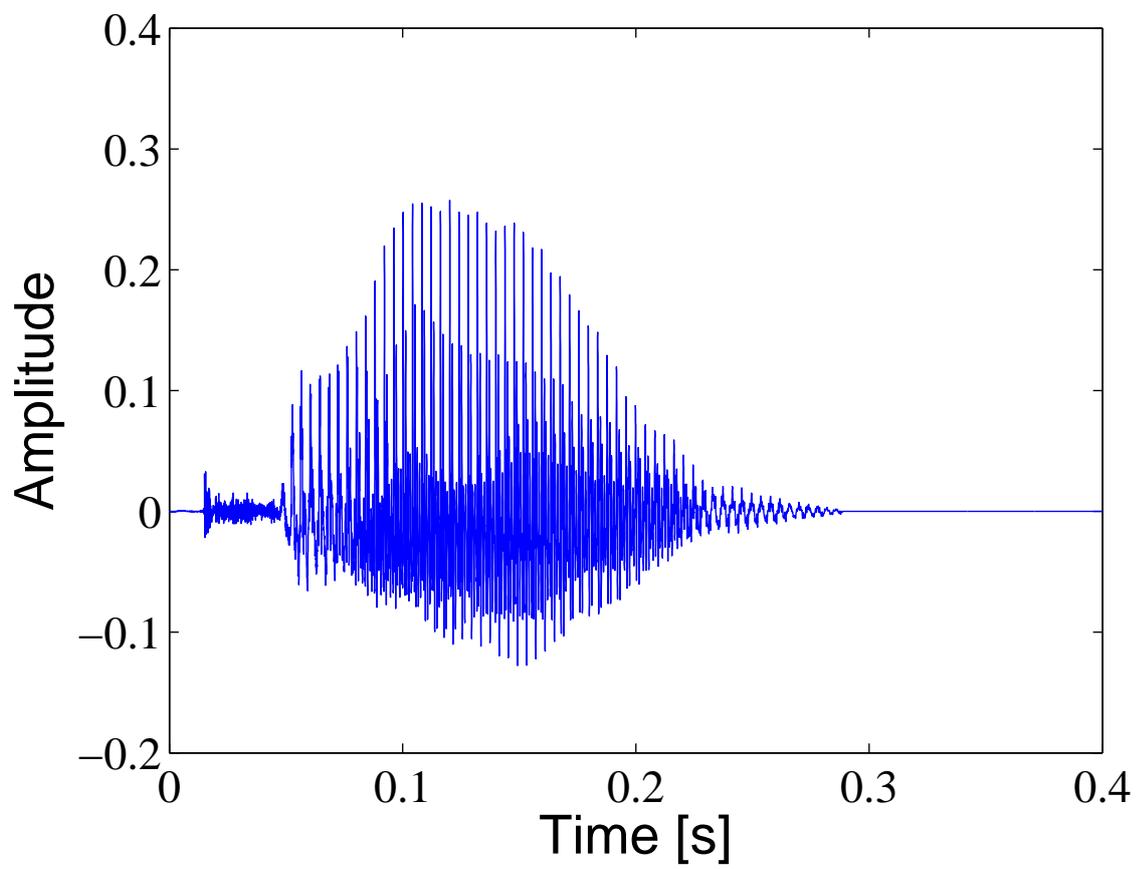


図 6.36: /kya/ の音声波形

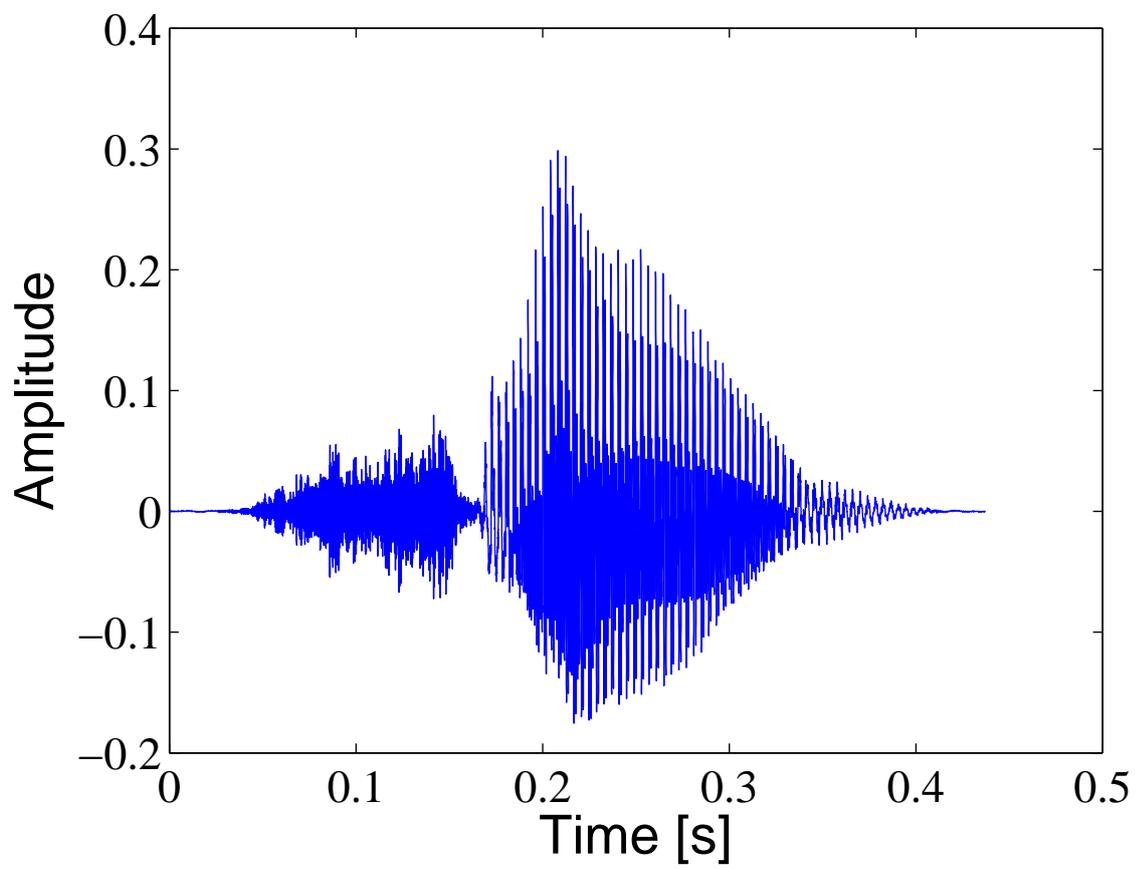


図 6.37: /sya/ の音声波形

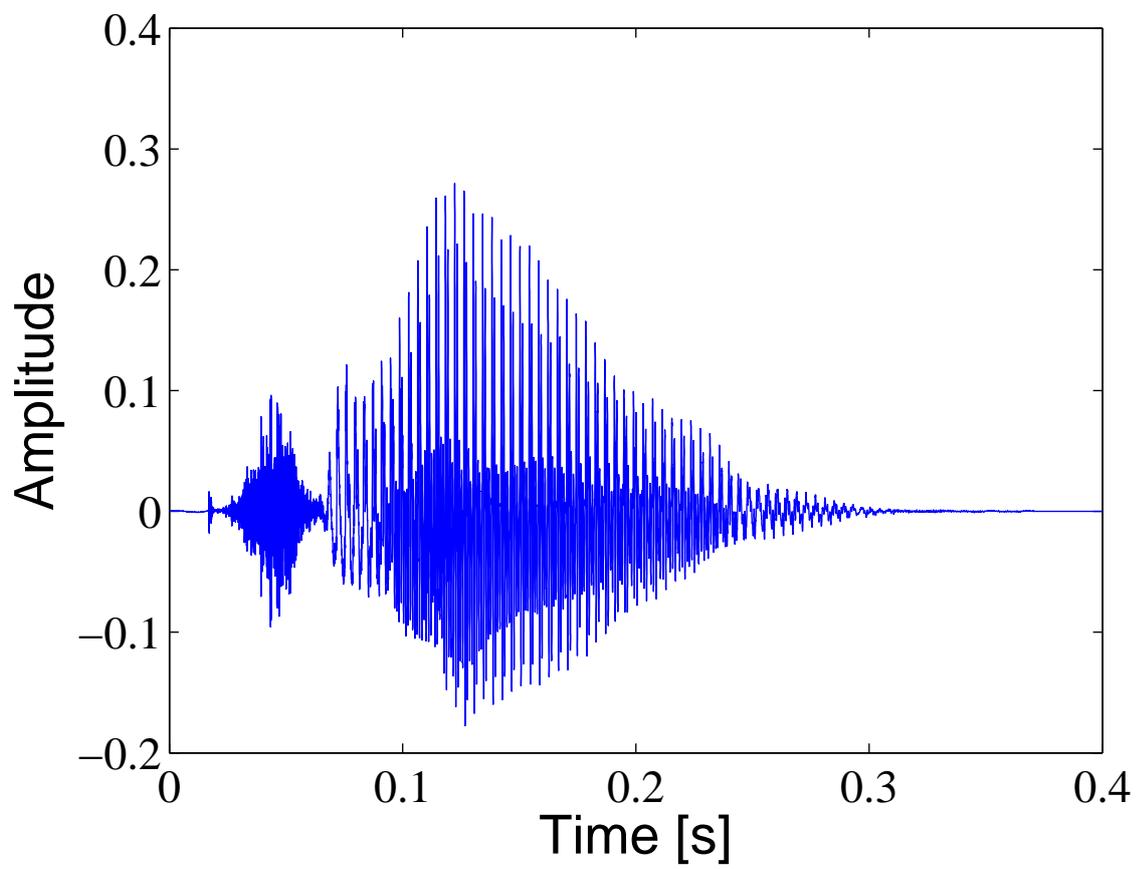


図 6.38: /tya/ の音声波形

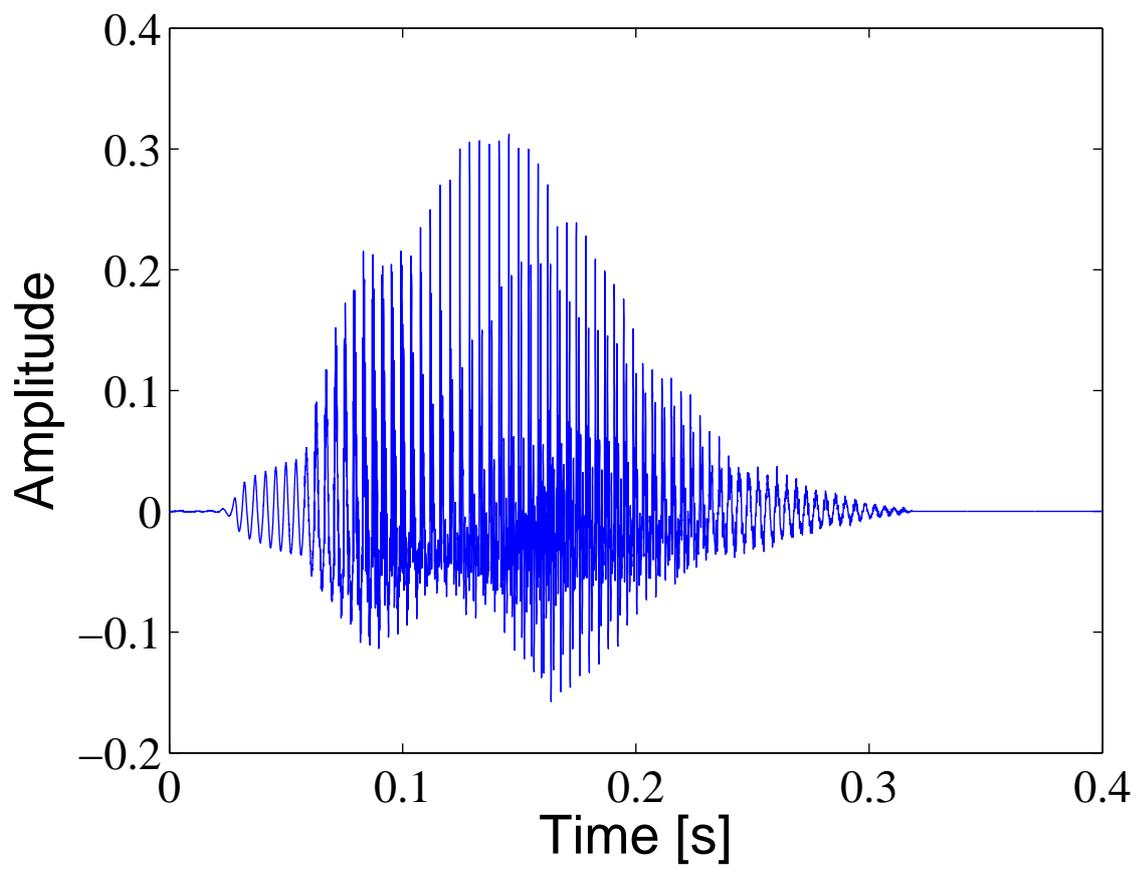


図 6.39: /nya/ の音声波形

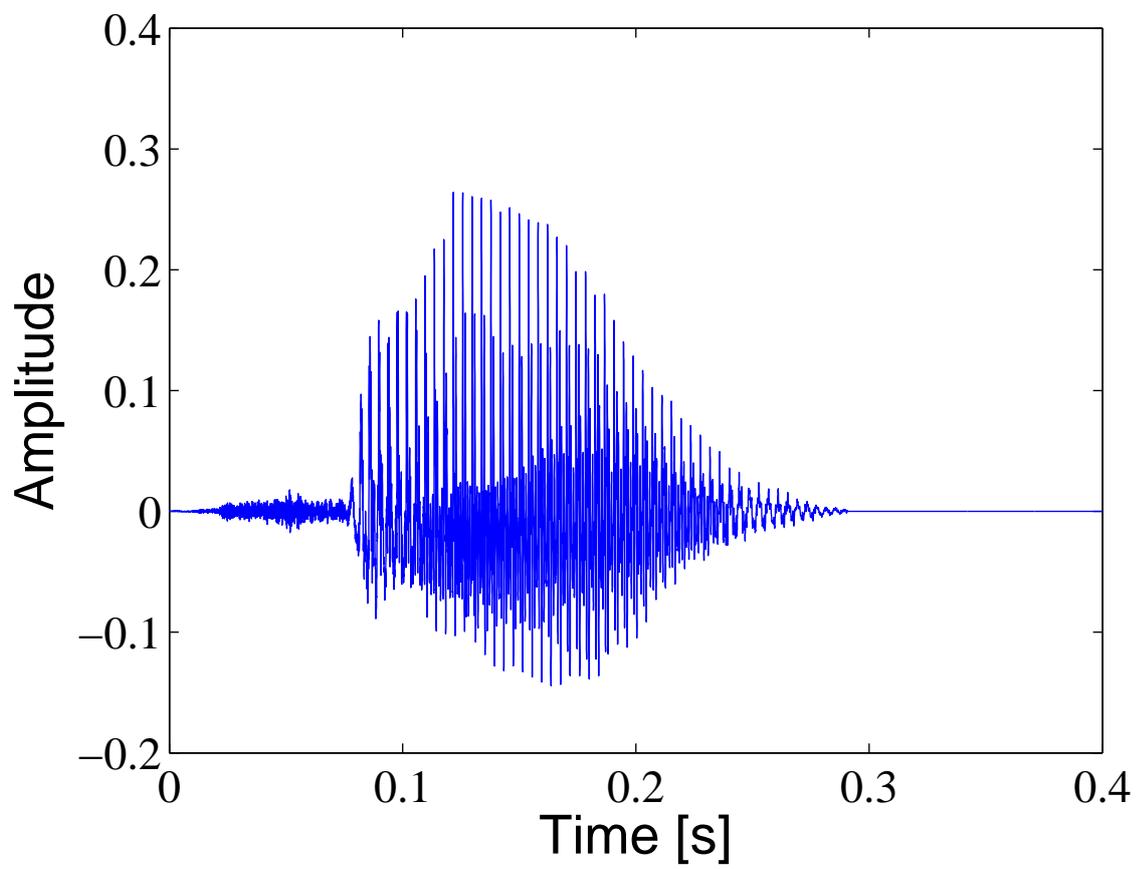


図 6.40: /hya/ の音声波形

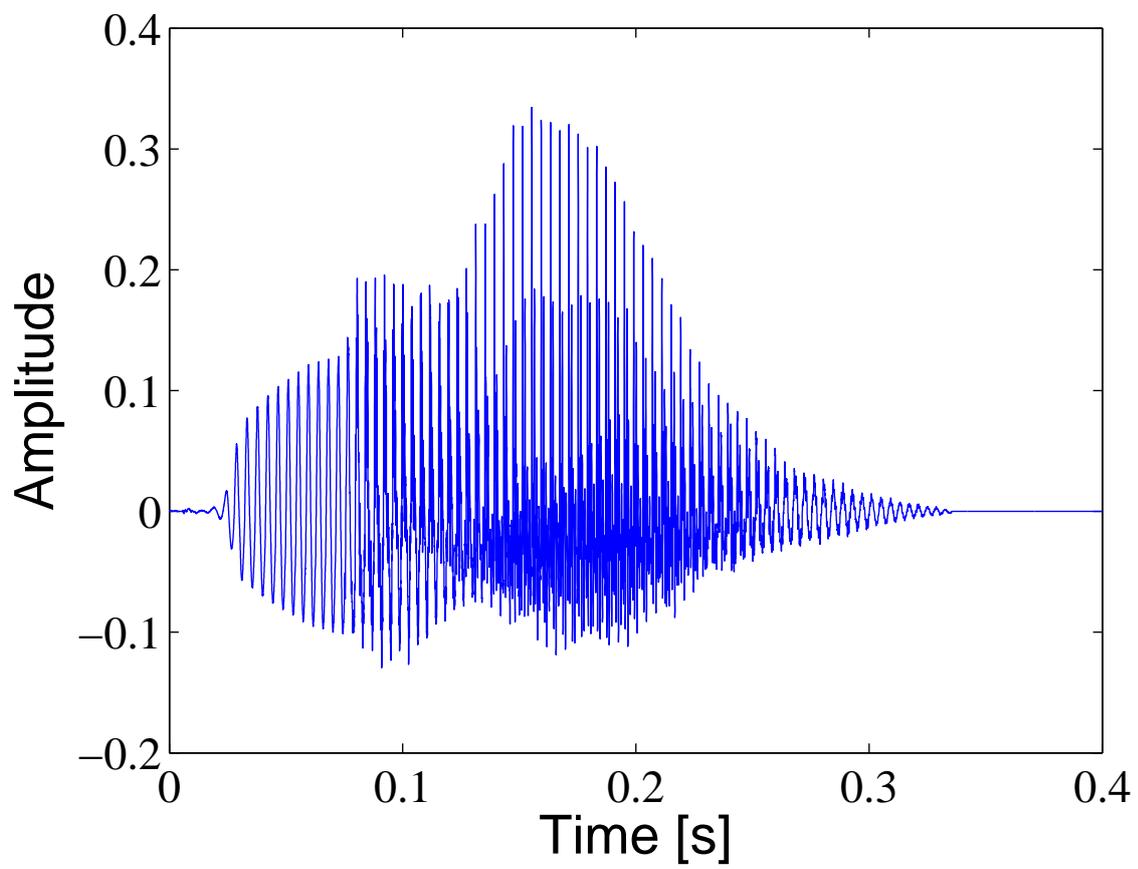


図 6.41: /mya/ の音声波形

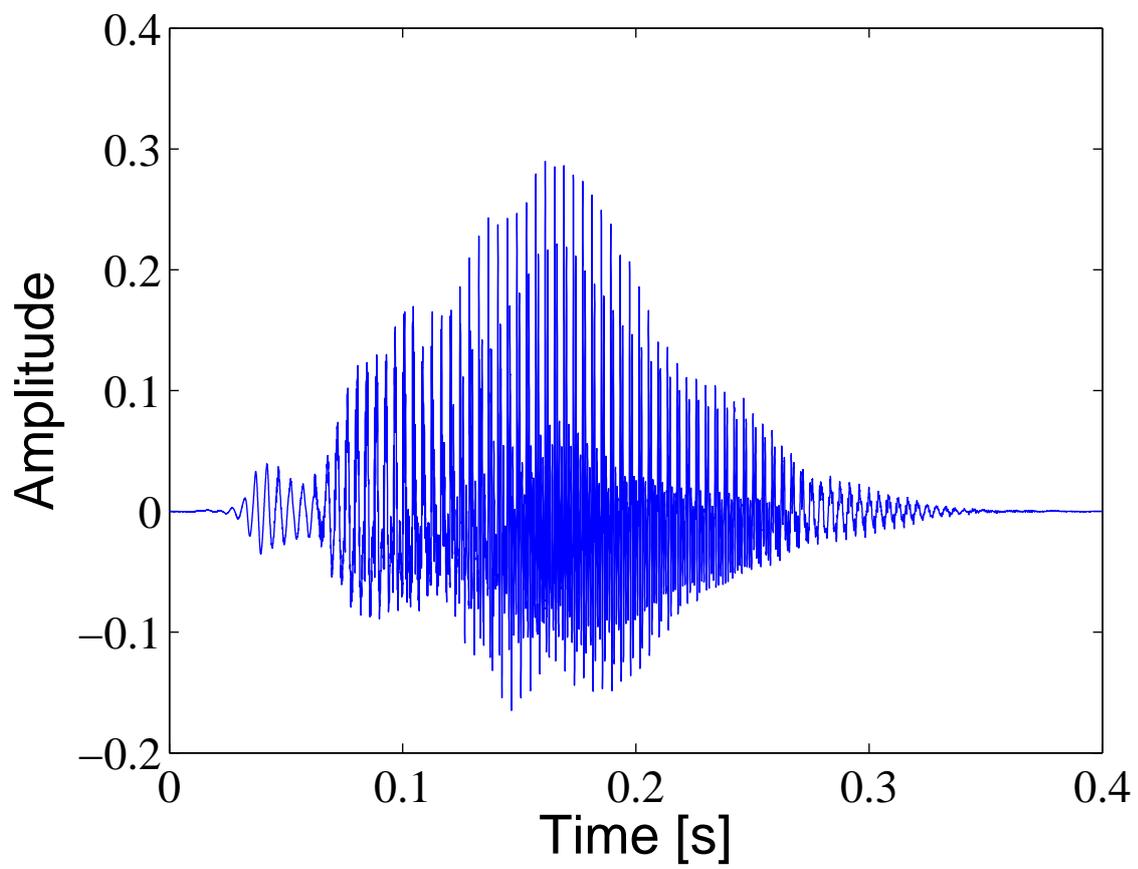


図 6.42: /rya/ の音声波形

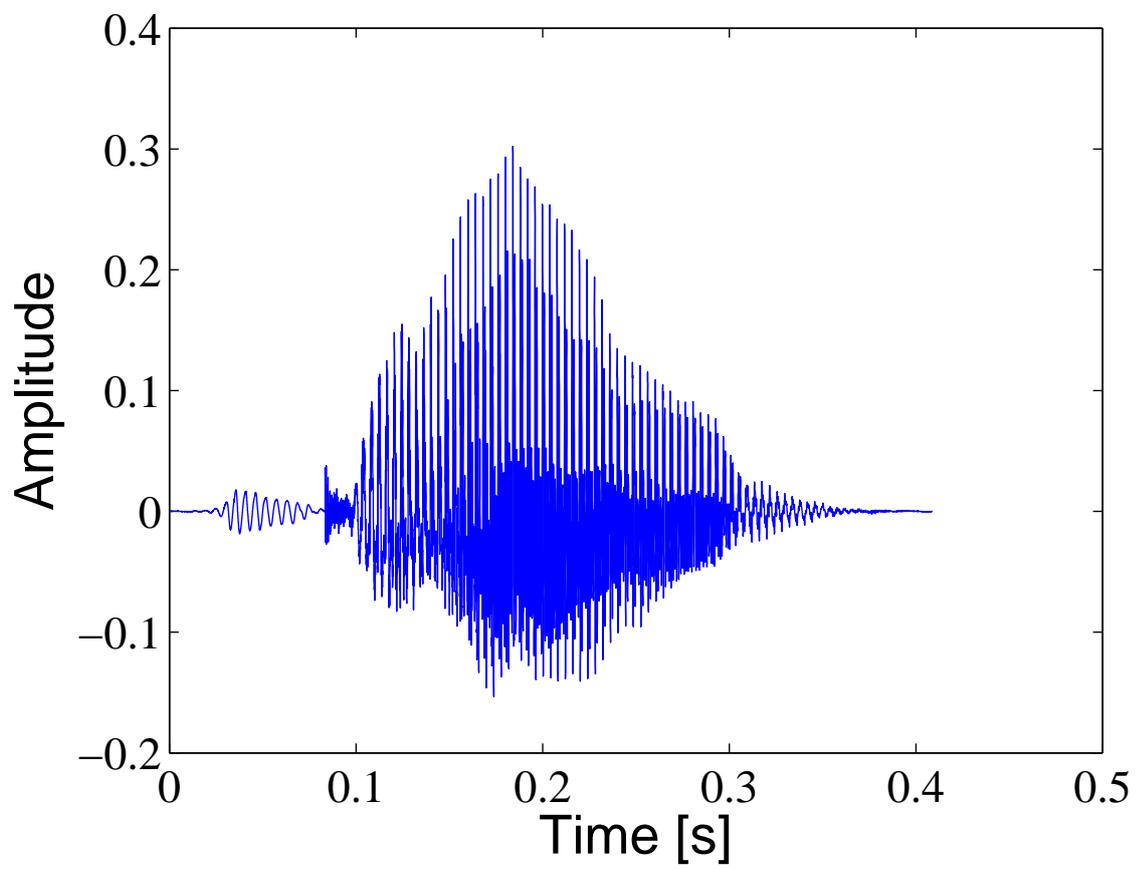


図 6.43: /gya/ の音声波形

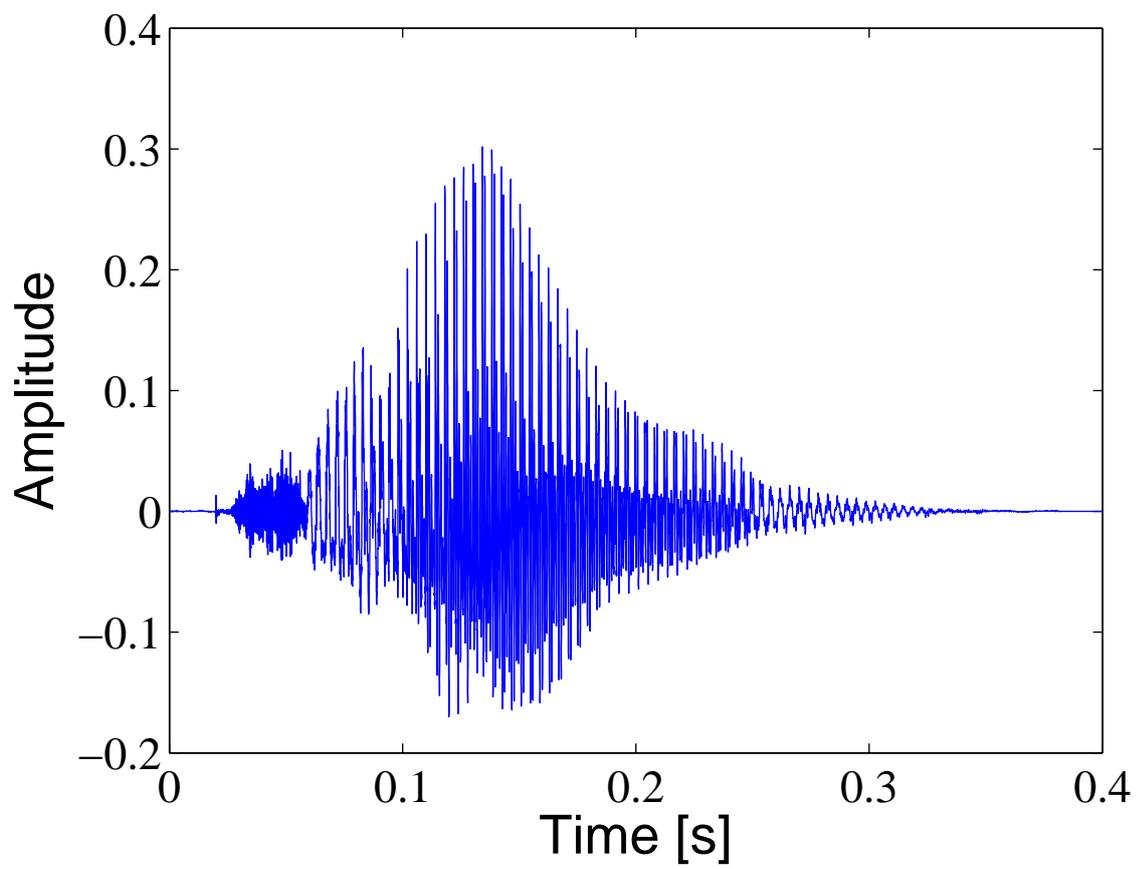


図 6.44: /zya/ の音声波形

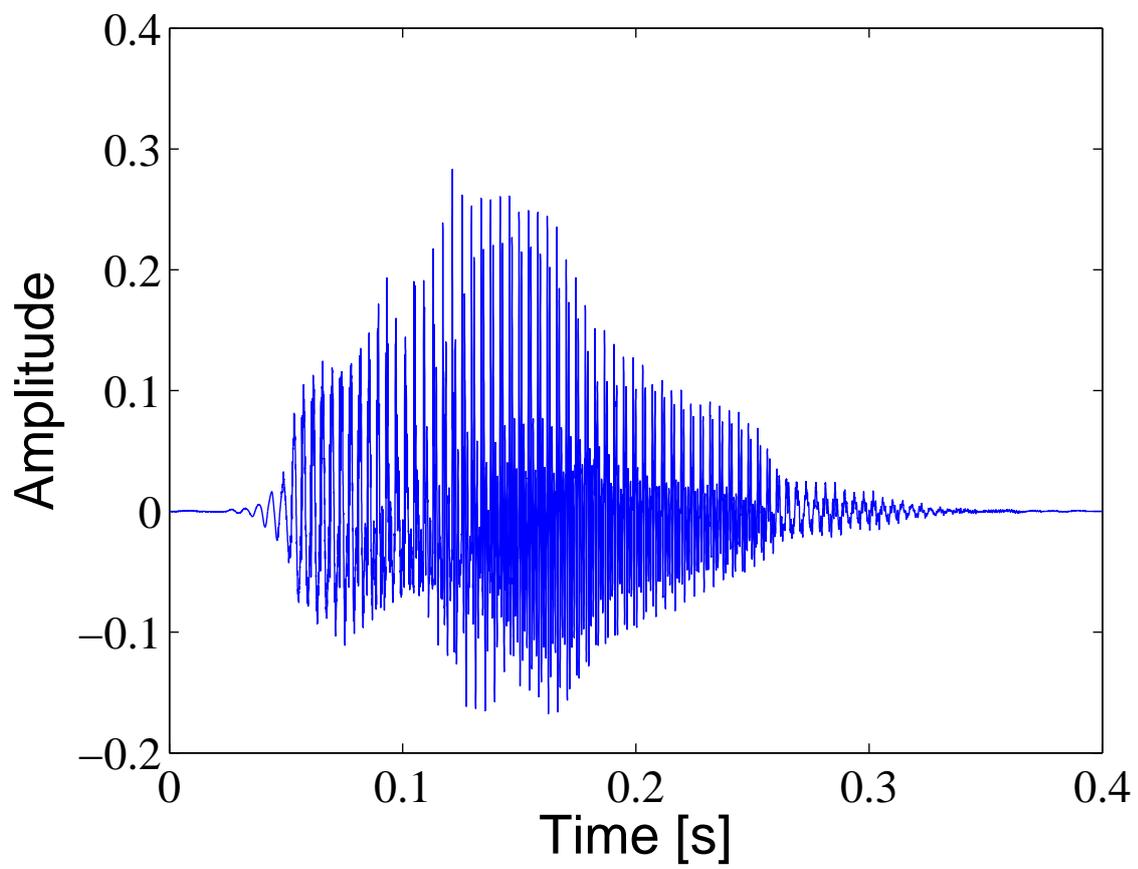


図 6.45: /bya/ の音声波形

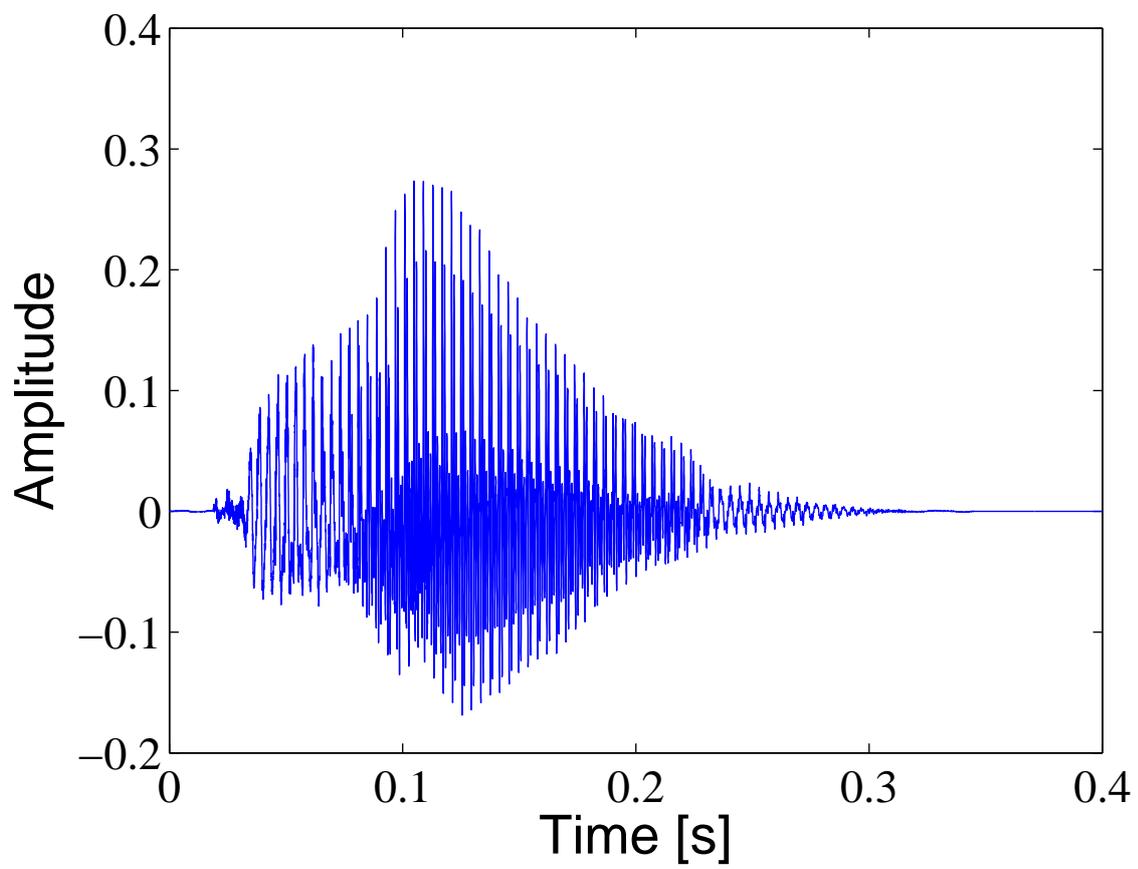


図 6.46: /pya/ の音声波形

単音節音声の各周波数帯域の振幅包絡

単音節音声の10個に周波数帯域分割した際、それぞれの周波数帯域の振幅包絡を図をそれぞれ下記に示す。

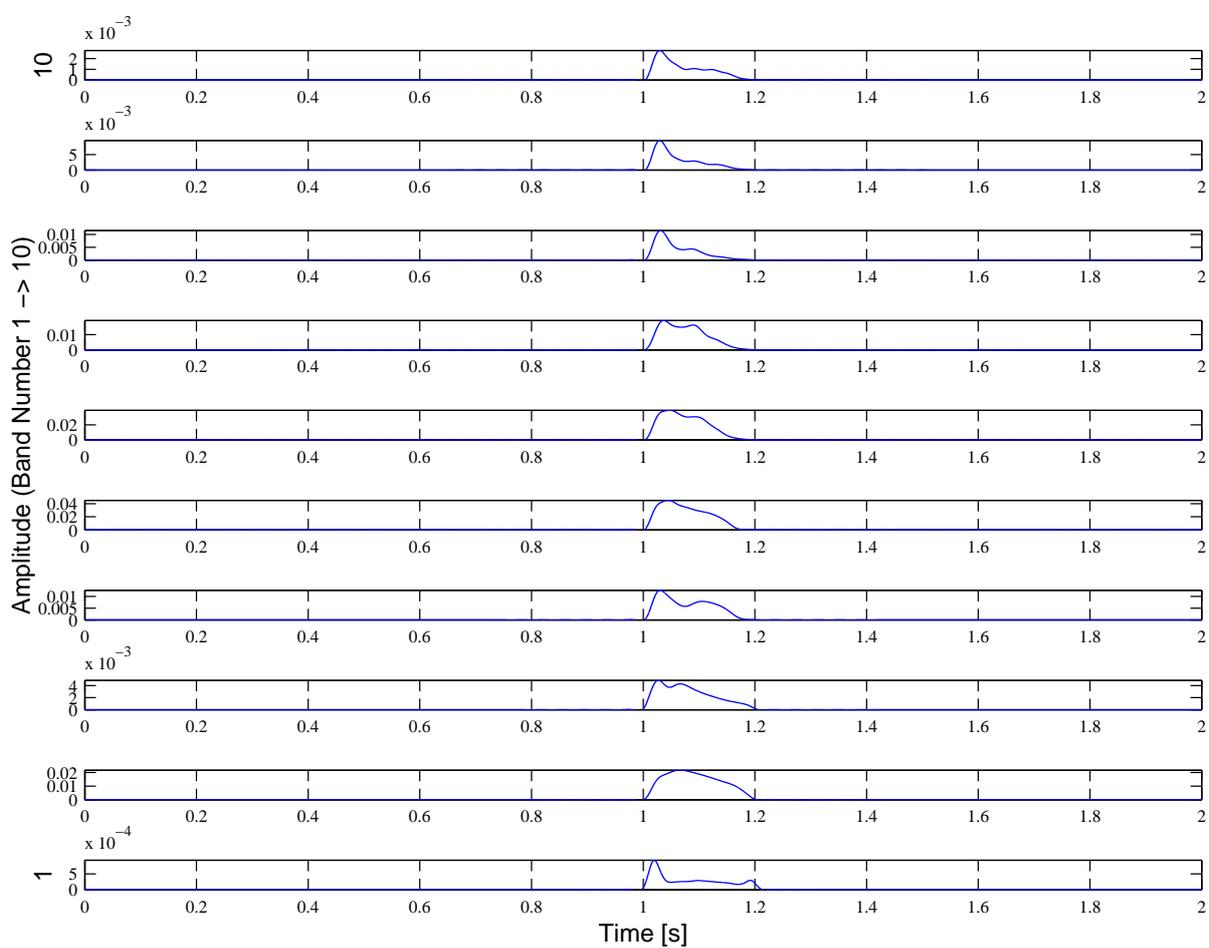


図 6.47: /a/ を 10 個の周波数帯域分割した際の各振幅包絡

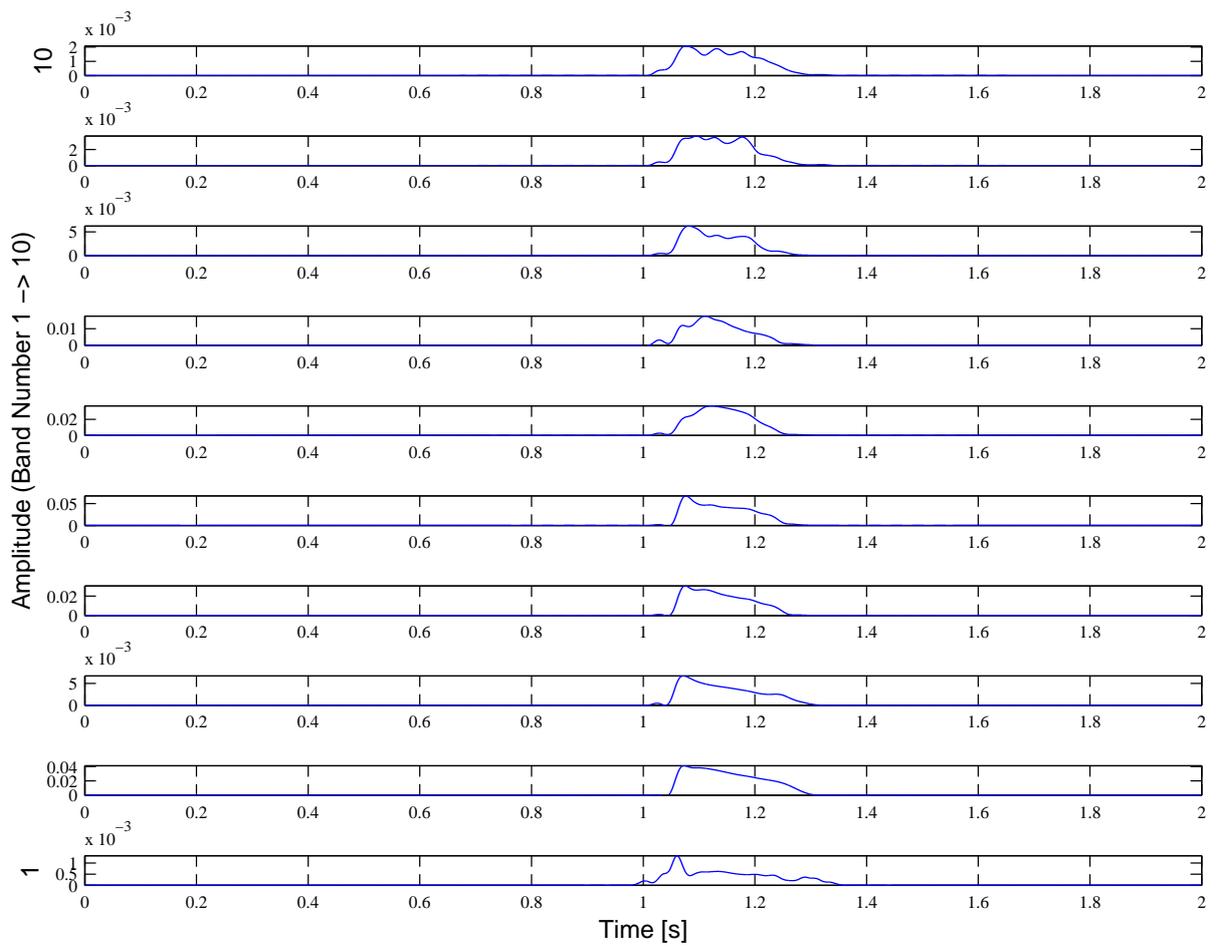


図 6.48: $/ka/$ を 10 個の周波数帯域分割した際の各振幅包絡

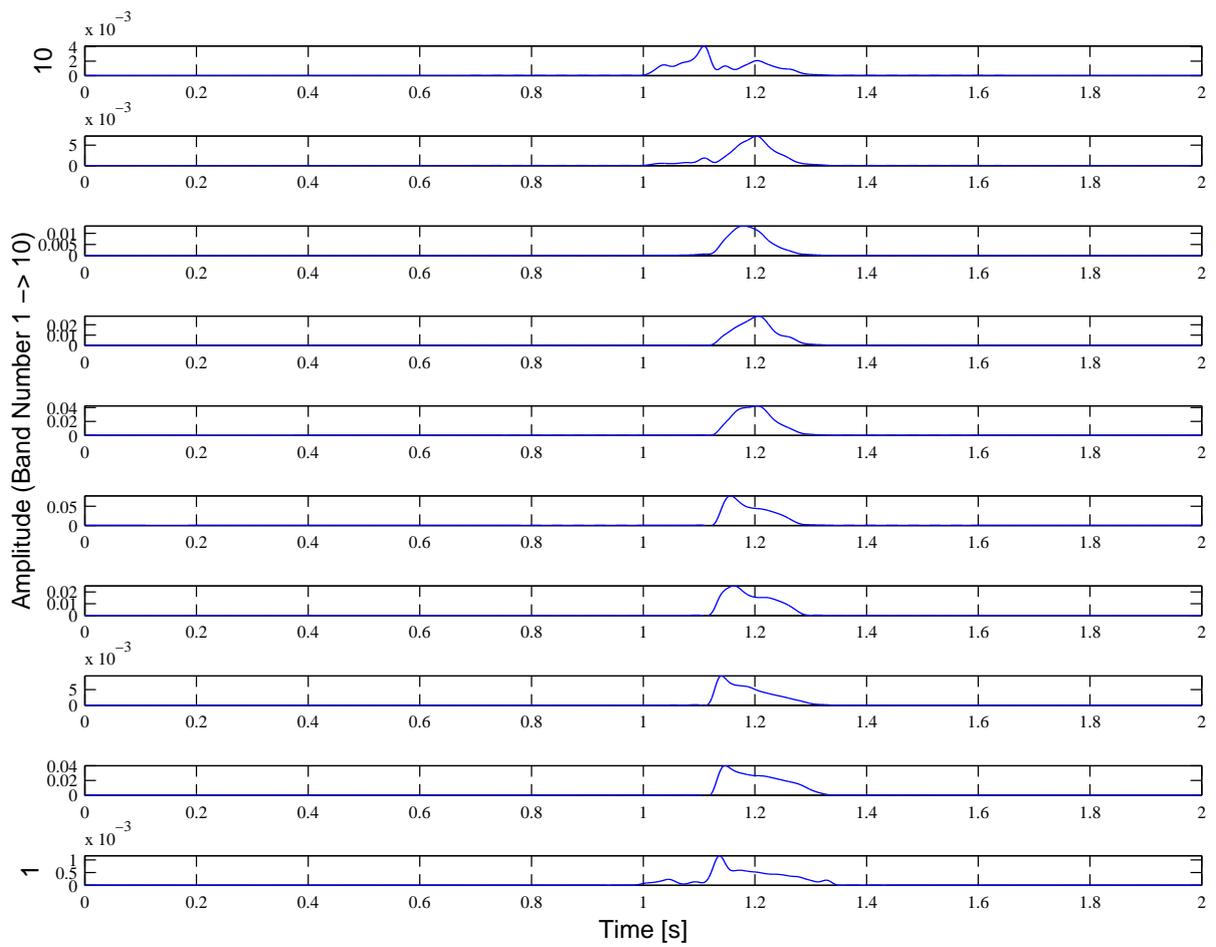


図 6.49: /sa/ を 10 個の周波数帯域分割した際の各振幅包絡

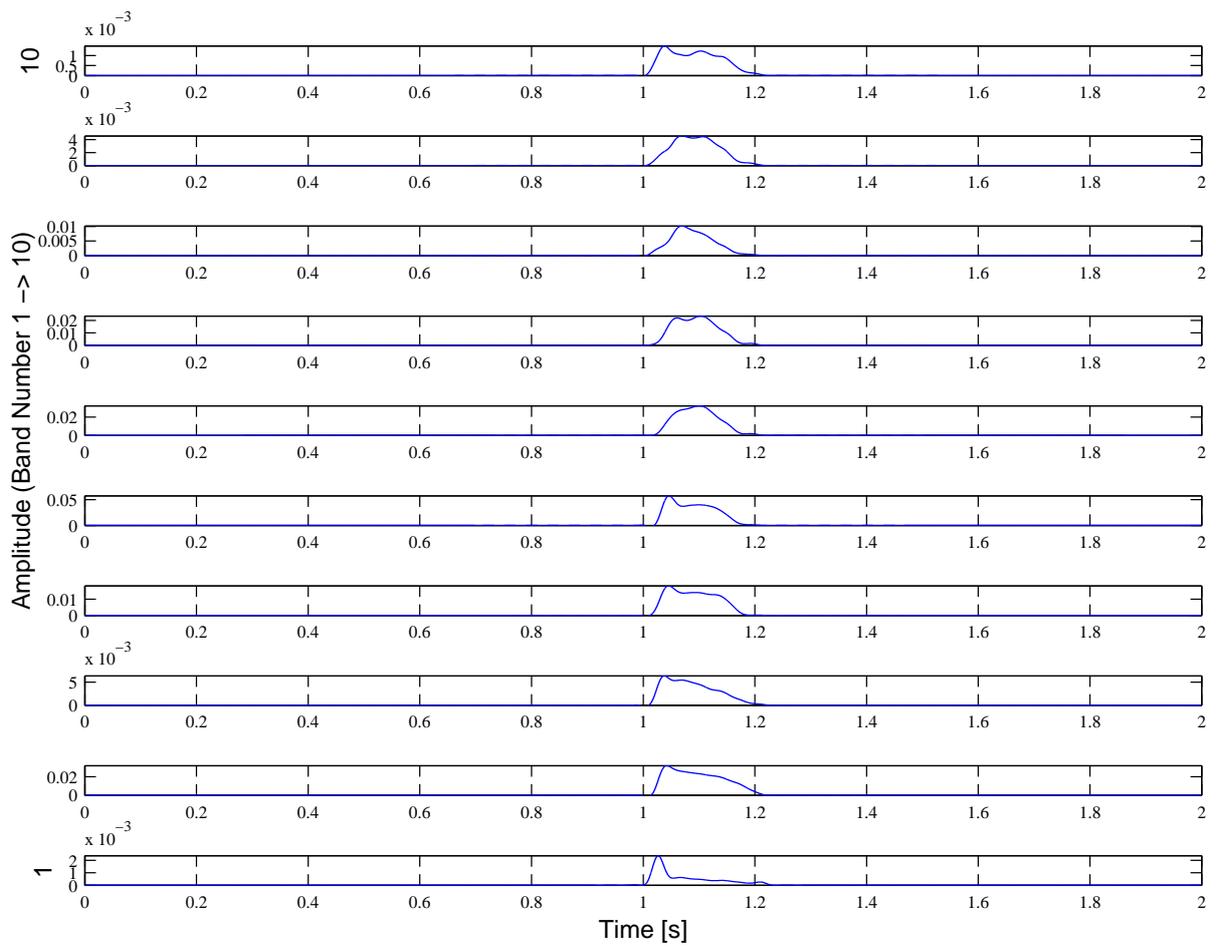


図 6.50: /ta/ を 10 個の周波数帯域分割した際の各振幅包絡

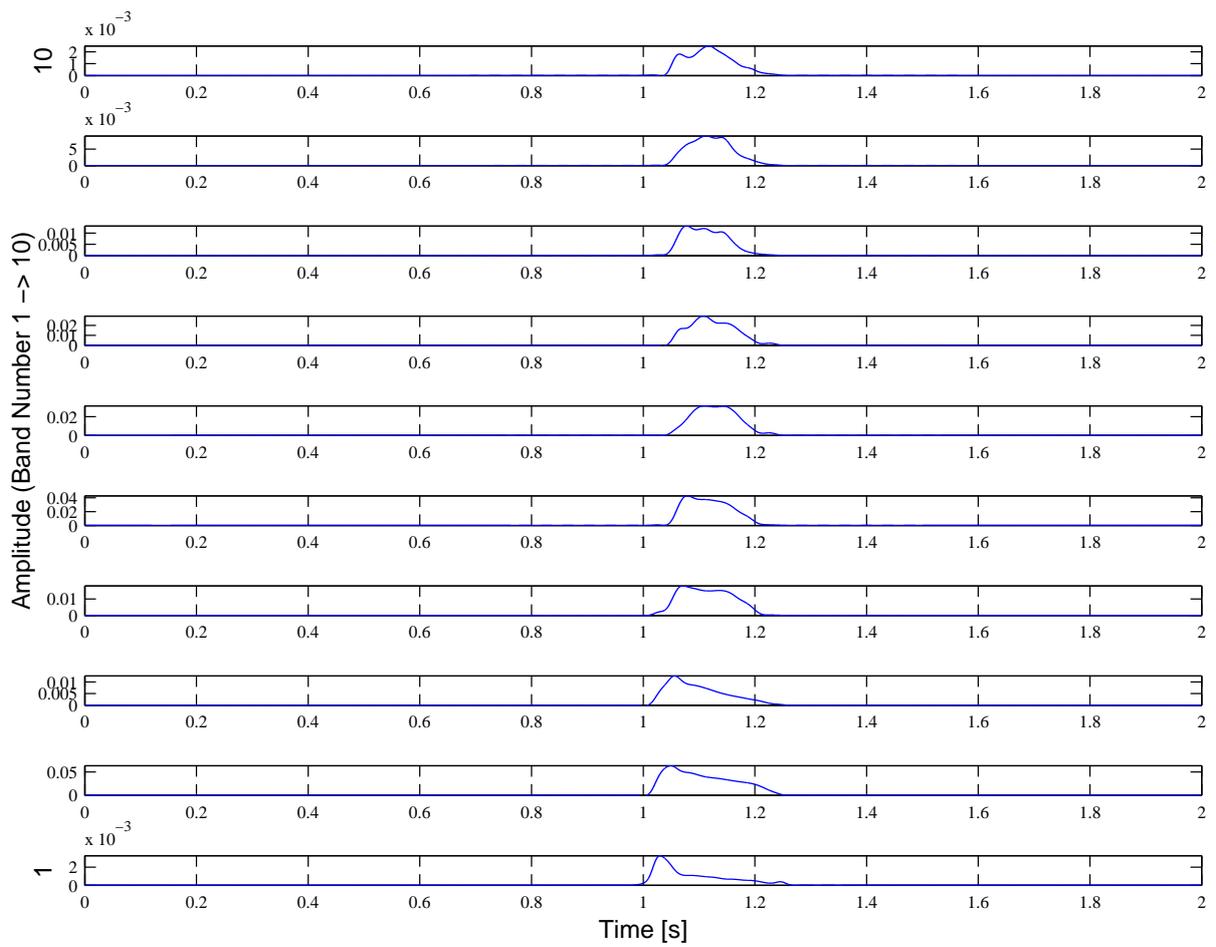


図 6.51: /na/ を 10 個の周波数帯域分割した際の各振幅包絡

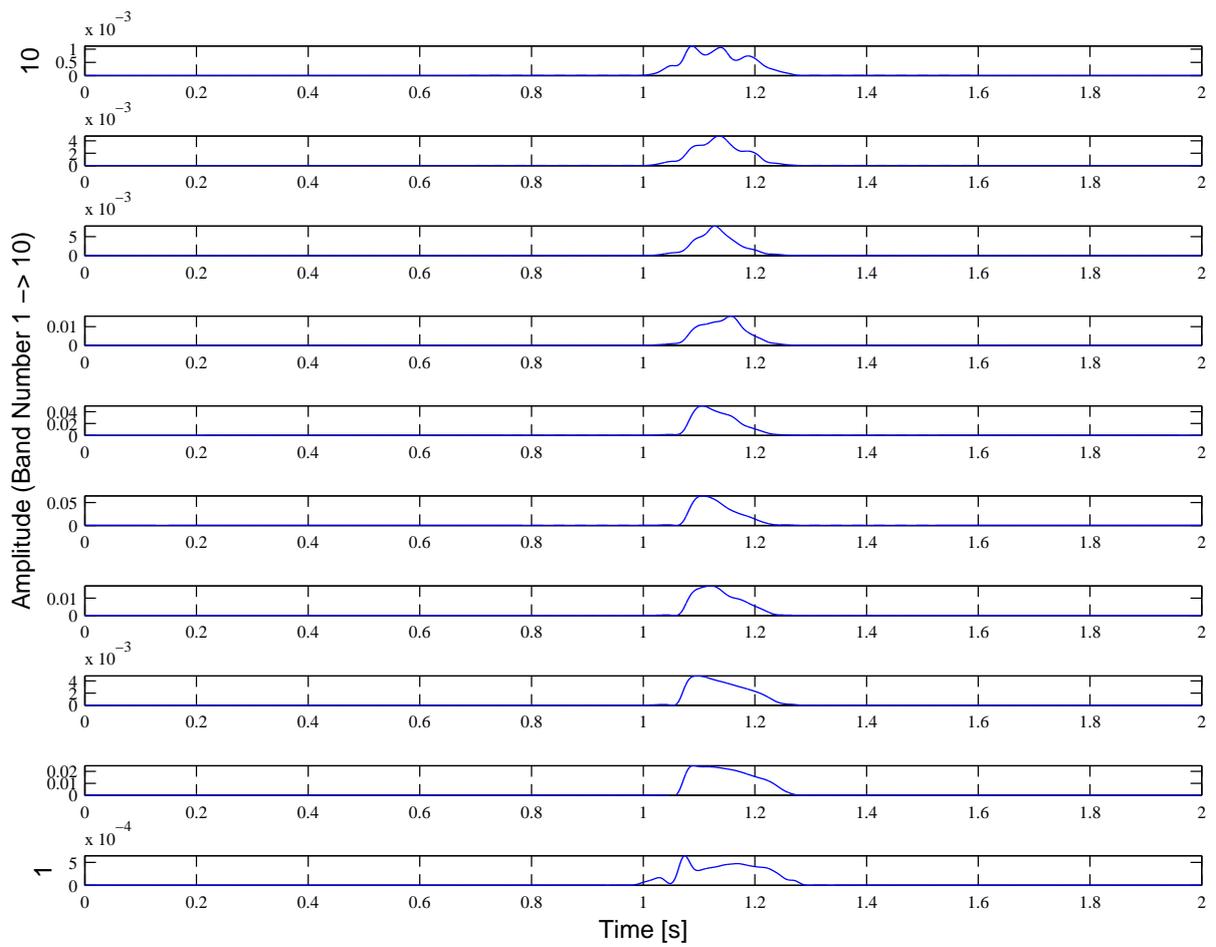


図 6.52: /ha/ を 10 個の周波数帯域分割した際の各振幅包絡

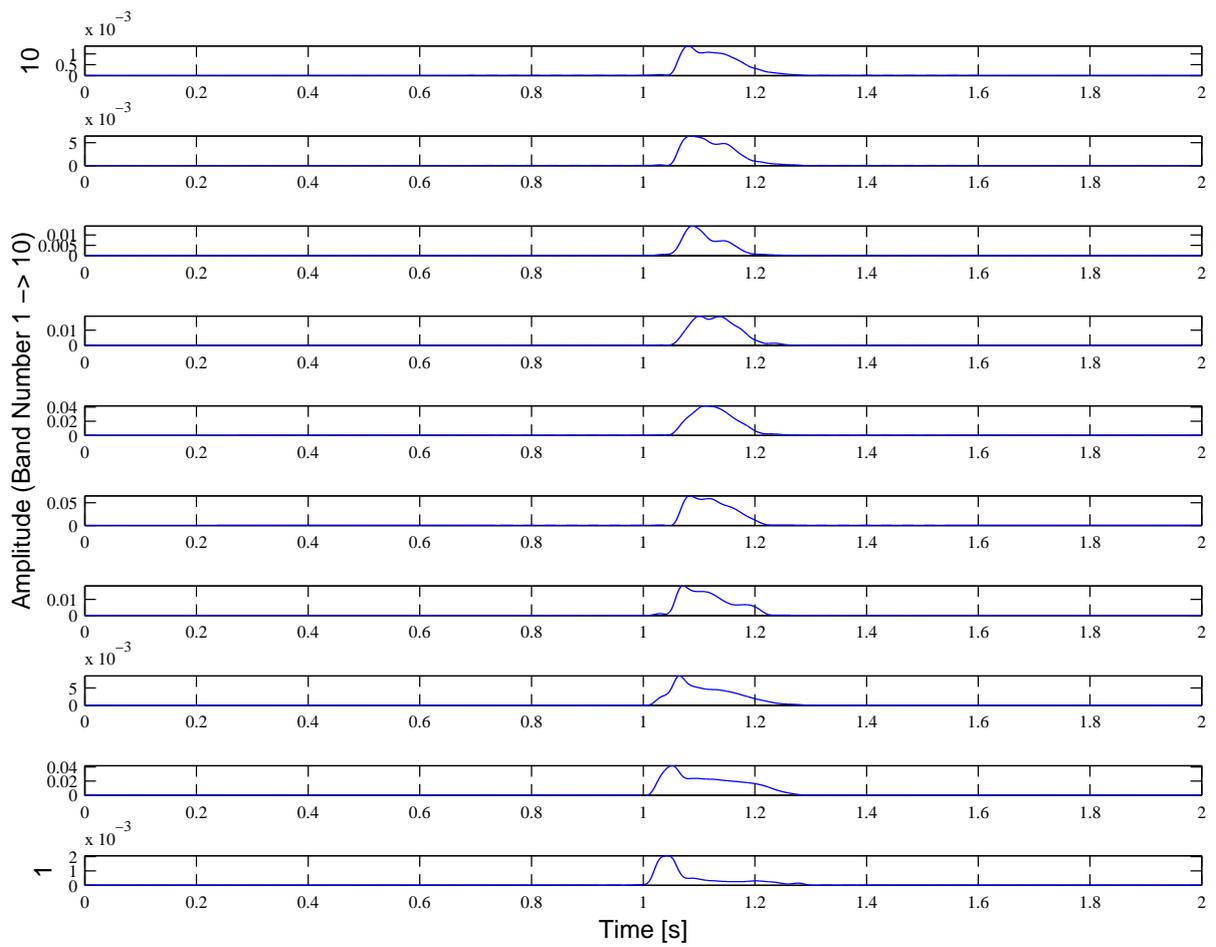


図 6.53: /ma/ を 10 個の周波数帯域分割した際の各振幅包絡

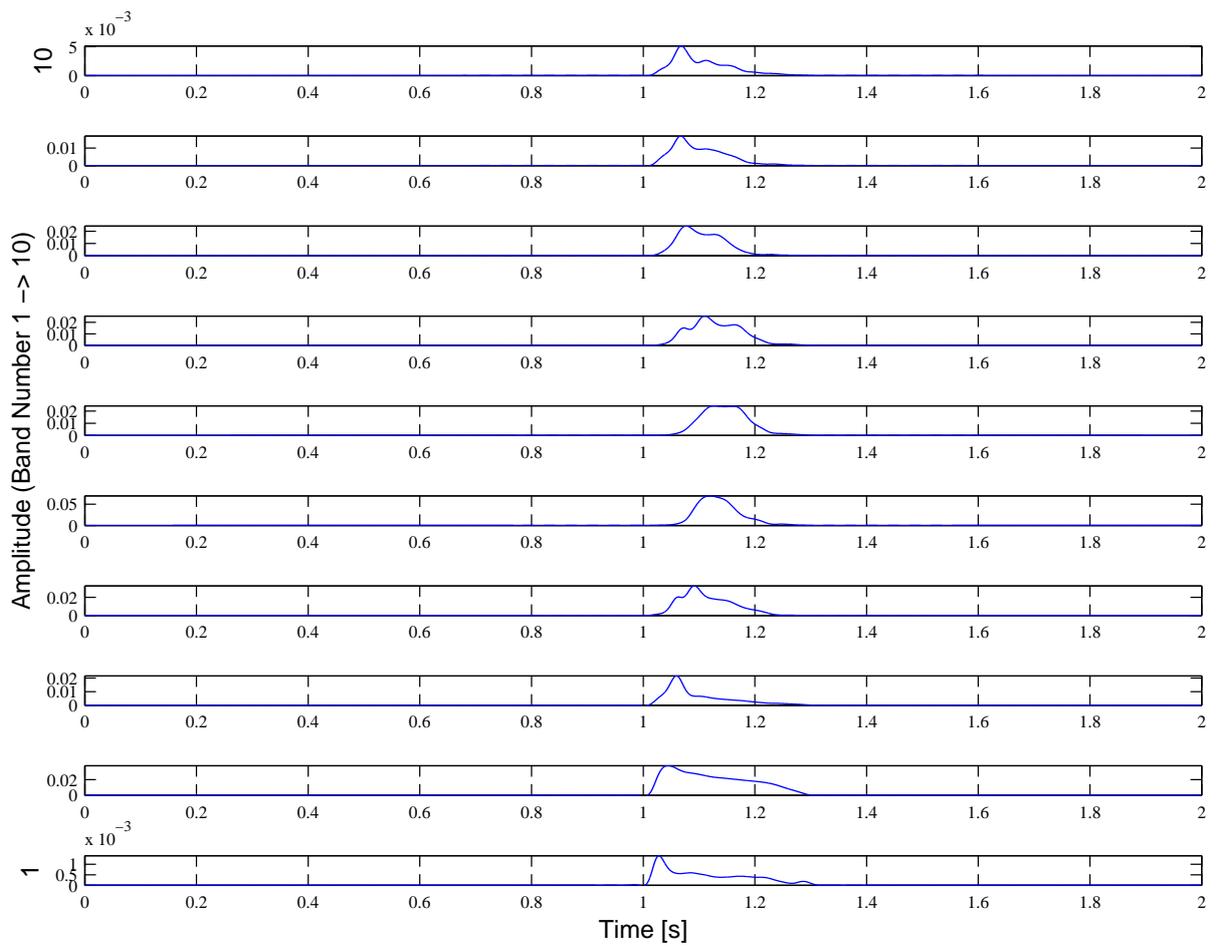


図 6.54: /ya/ を 10 個の周波数帯域分割した際の各振幅包絡

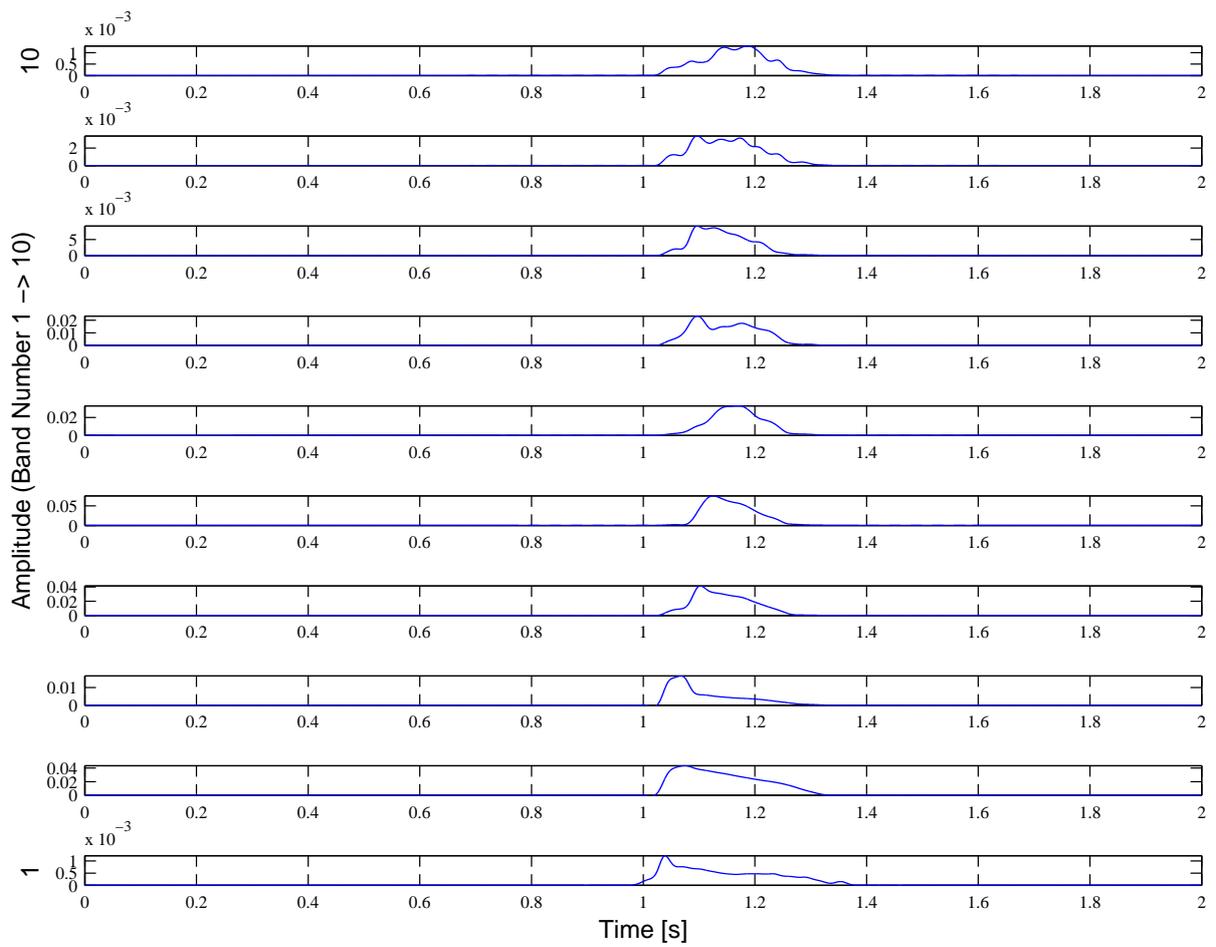


図 6.55: /ra/ を 10 個の周波数帯域分割した際の各振幅包絡

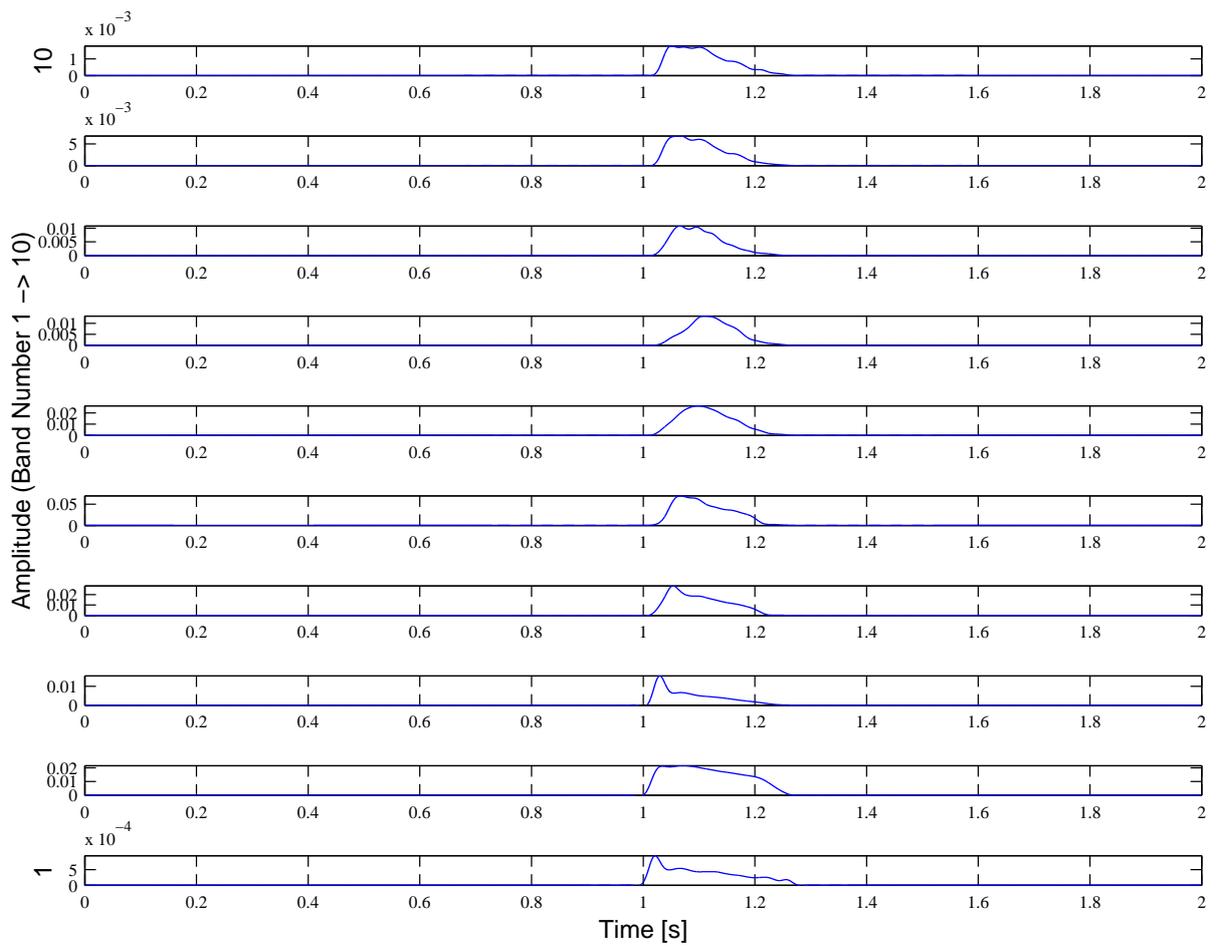


図 6.56: /wa/ を 10 個の周波数帯域分割した際の各振幅包絡

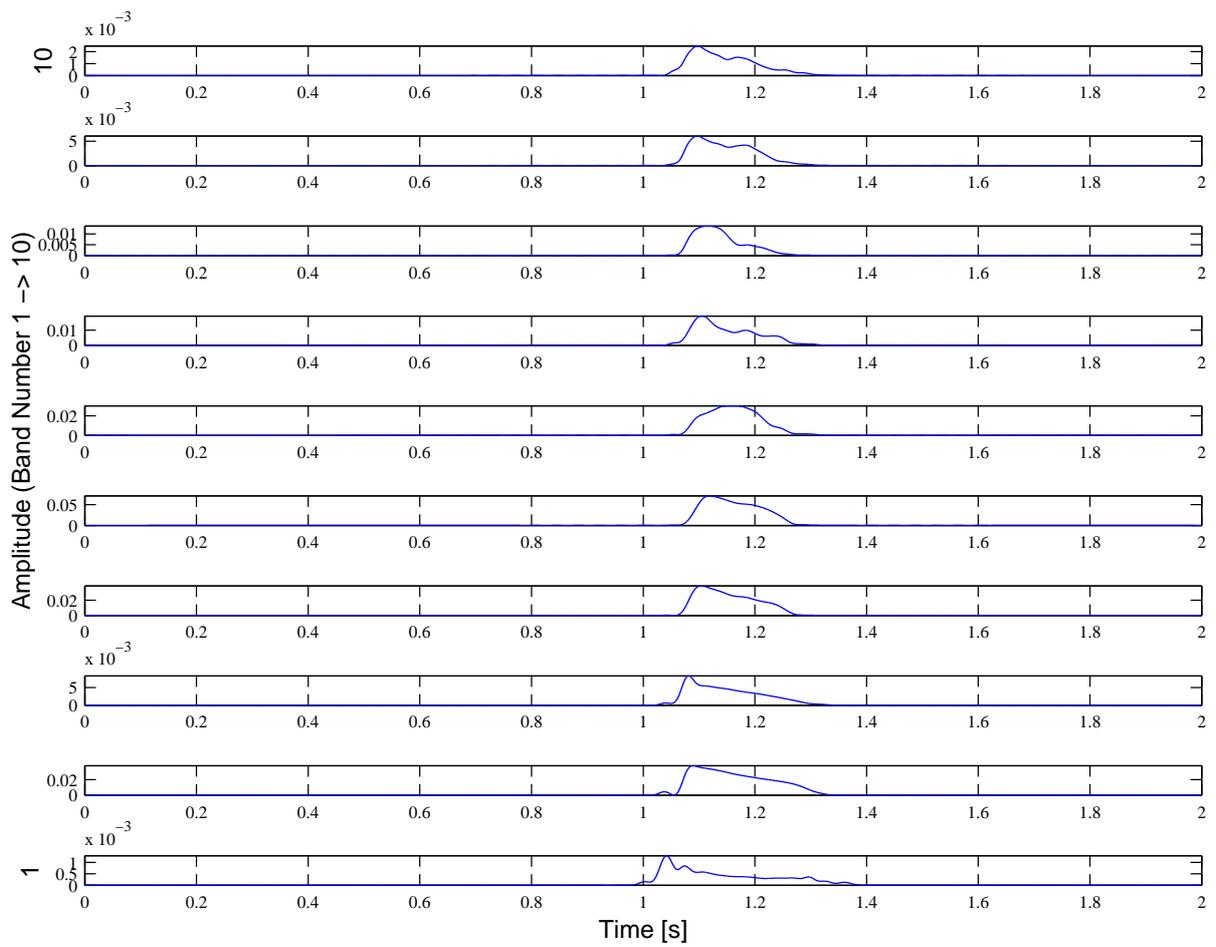


図 6.57: /ga/ を 10 個の周波数帯域分割した際の各振幅包絡

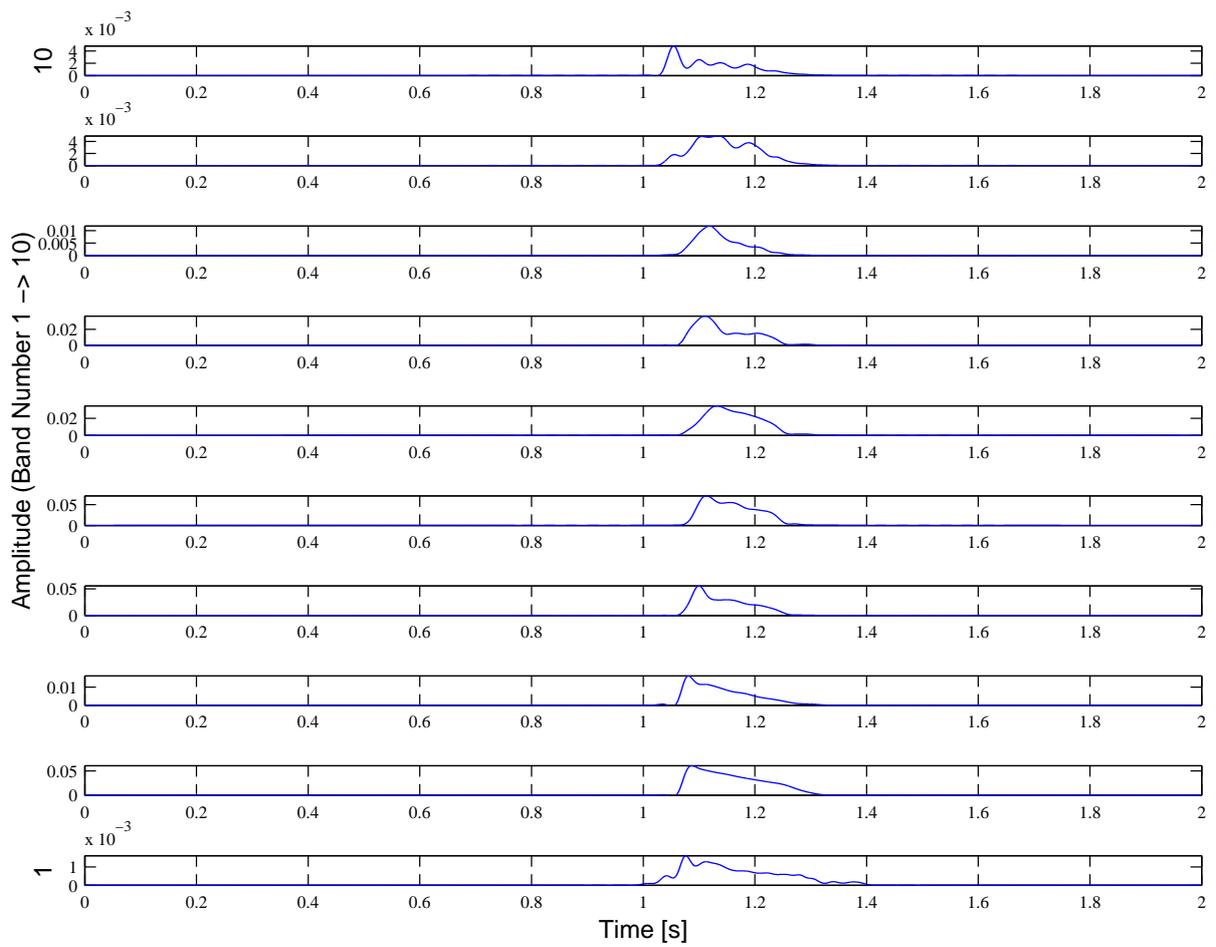


図 6.58: /za/ を 10 個の周波数帯域分割した際の各振幅包絡

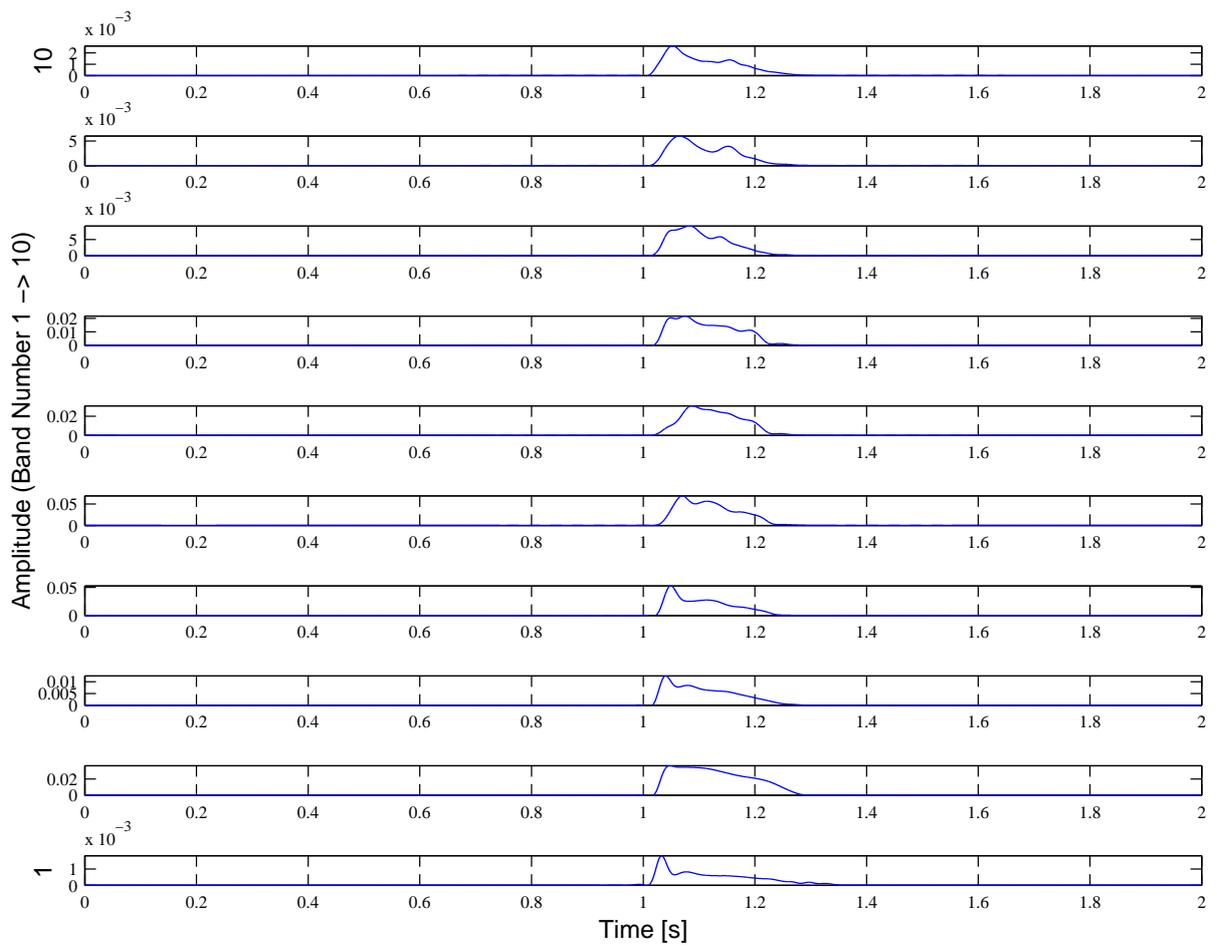


図 6.59: /da/ を 10 個の周波数帯域分割した際の各振幅包絡

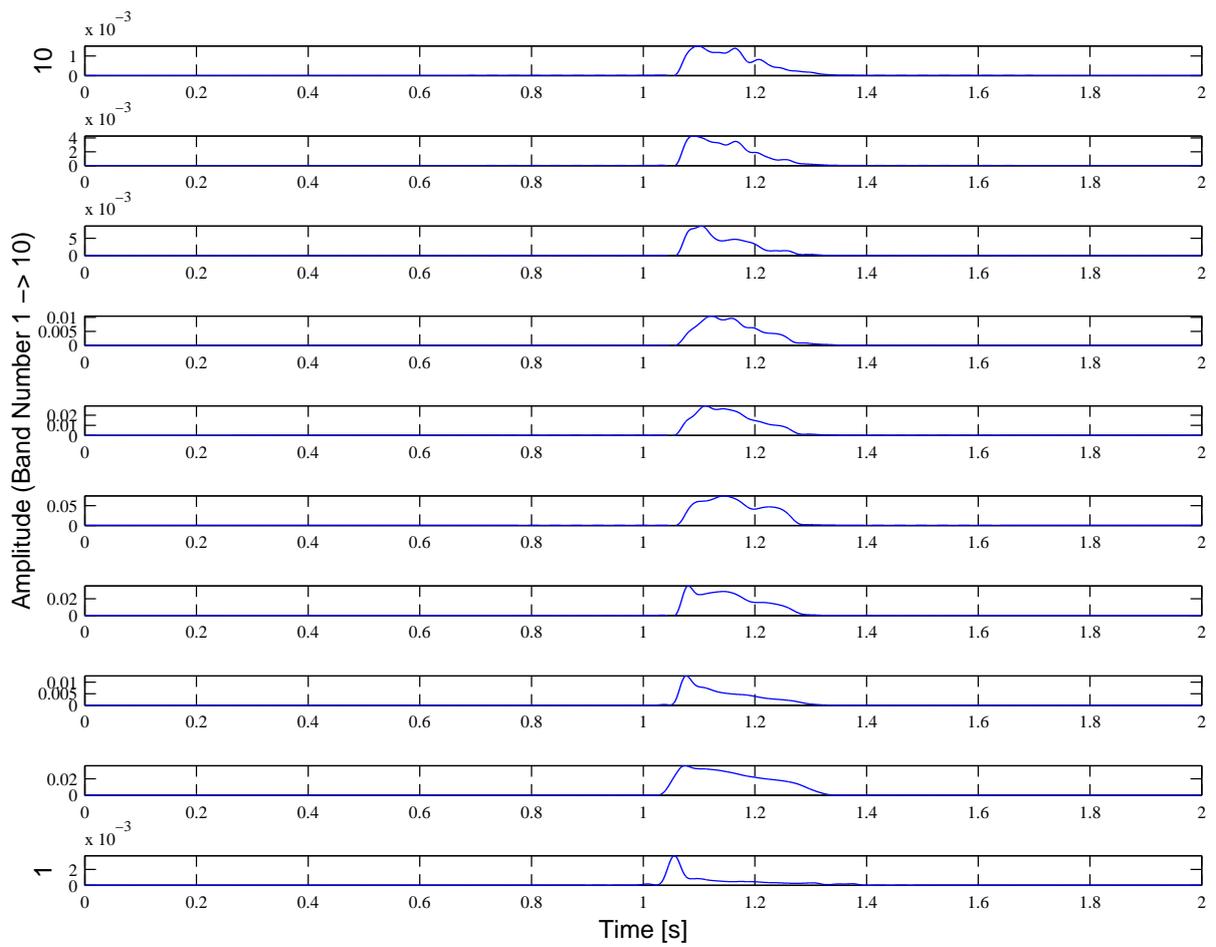


図 6.60: /ba/ を 10 個の周波数帯域分割した際の各振幅包絡

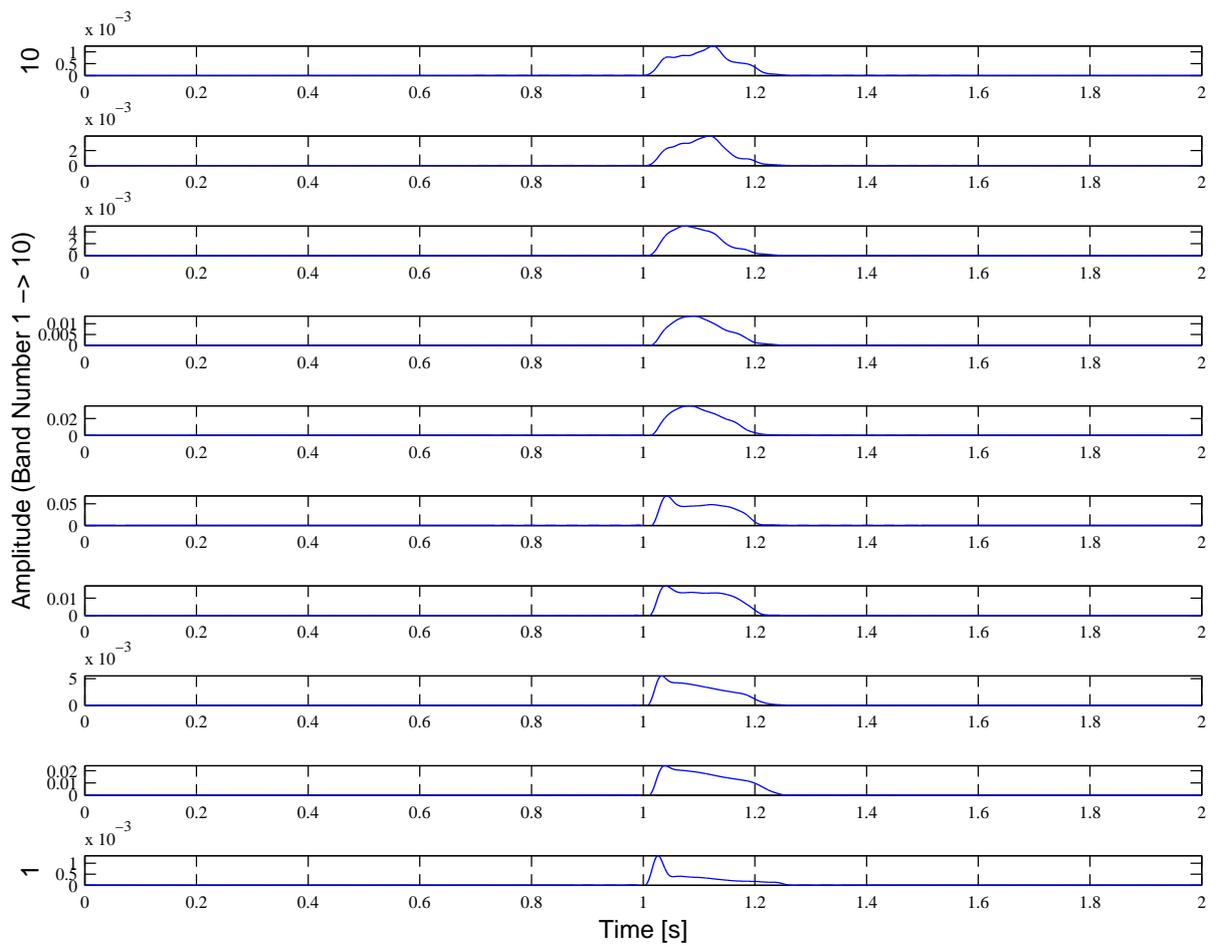


図 6.61: /pa/ を 10 個の周波数帯域分割した際の各振幅包絡

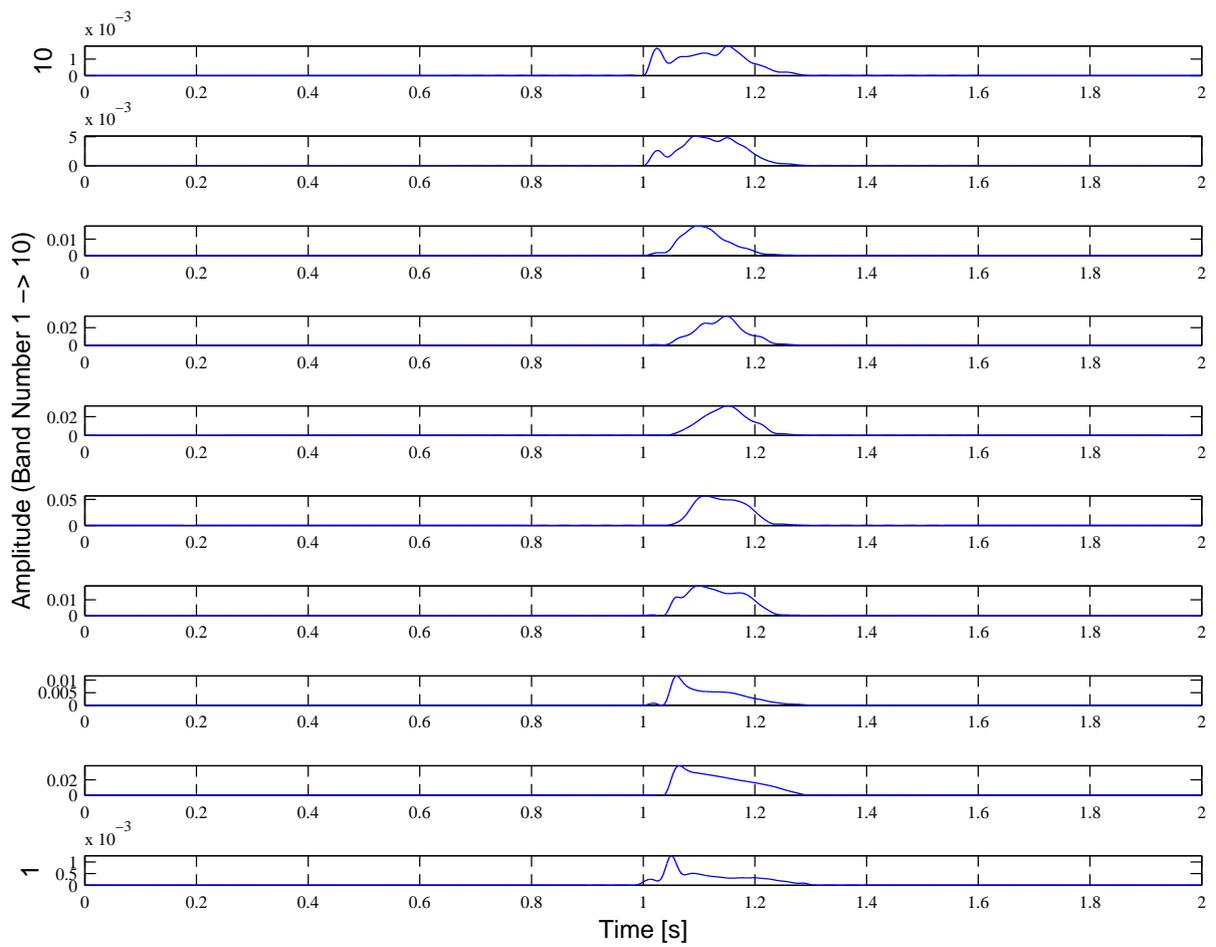


図 6.62: /kya/ を 10 個の周波数帯域分割した際の各振幅包絡

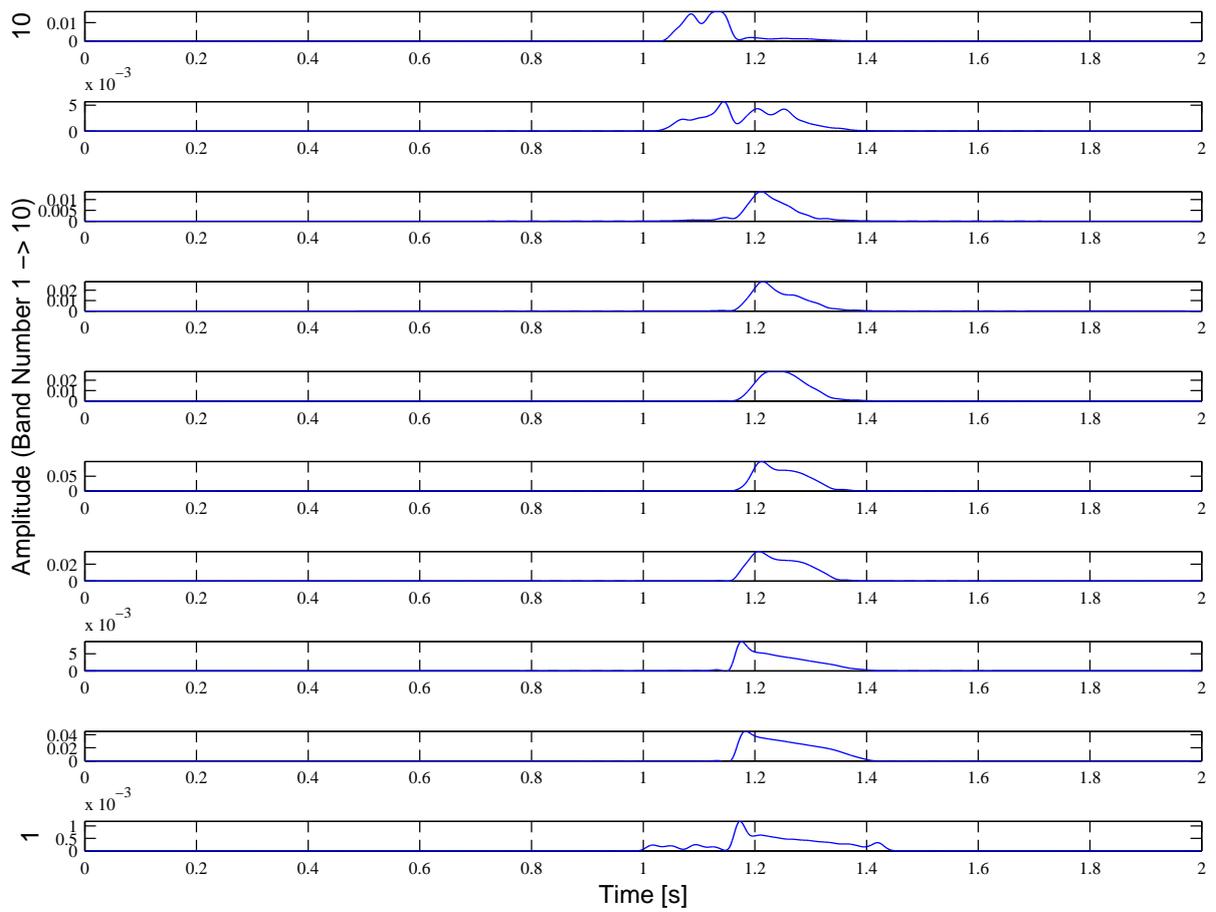


図 6.63: /sya/ を 10 個の周波数帯域分割した際の各振幅包絡

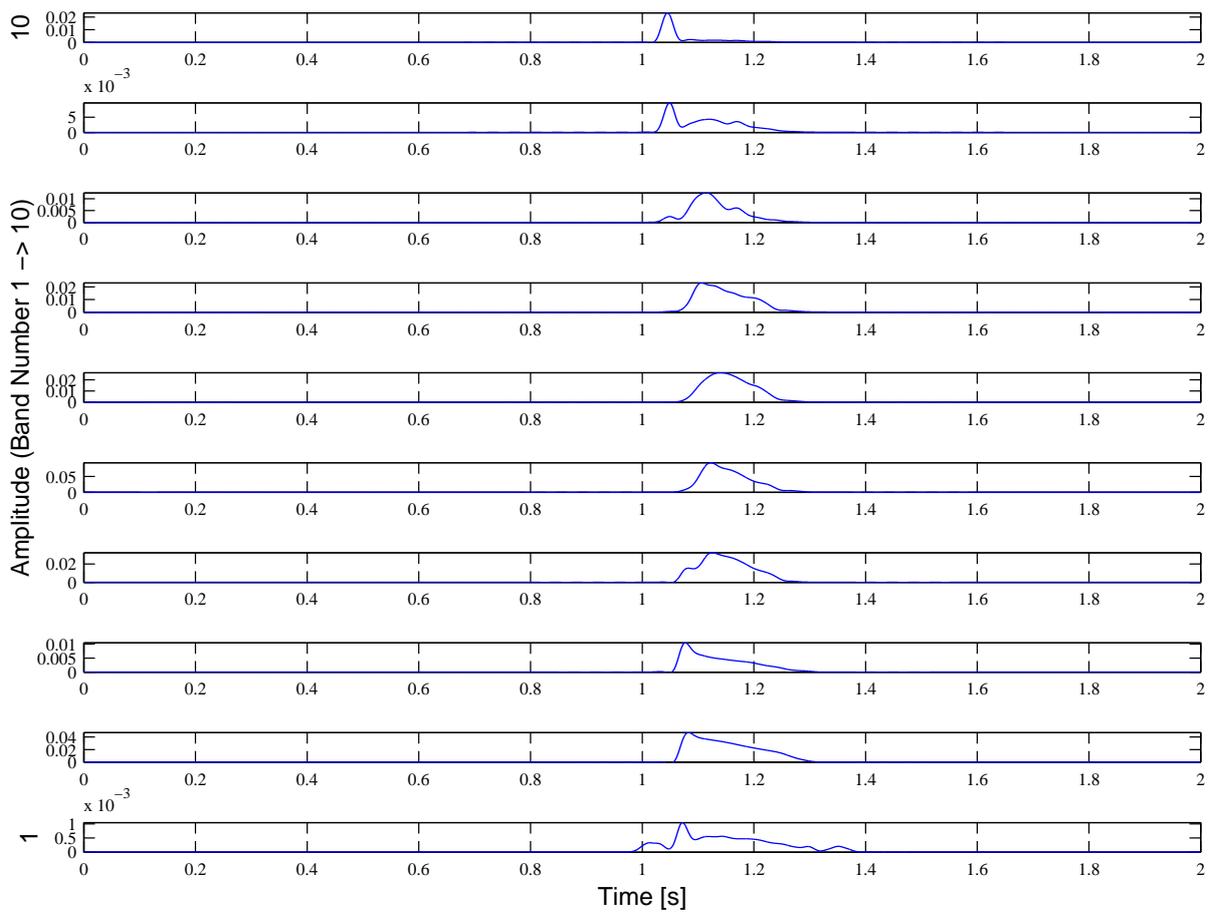


図 6.64: /tya/ を 10 個の周波数帯域分割した際の各振幅包絡

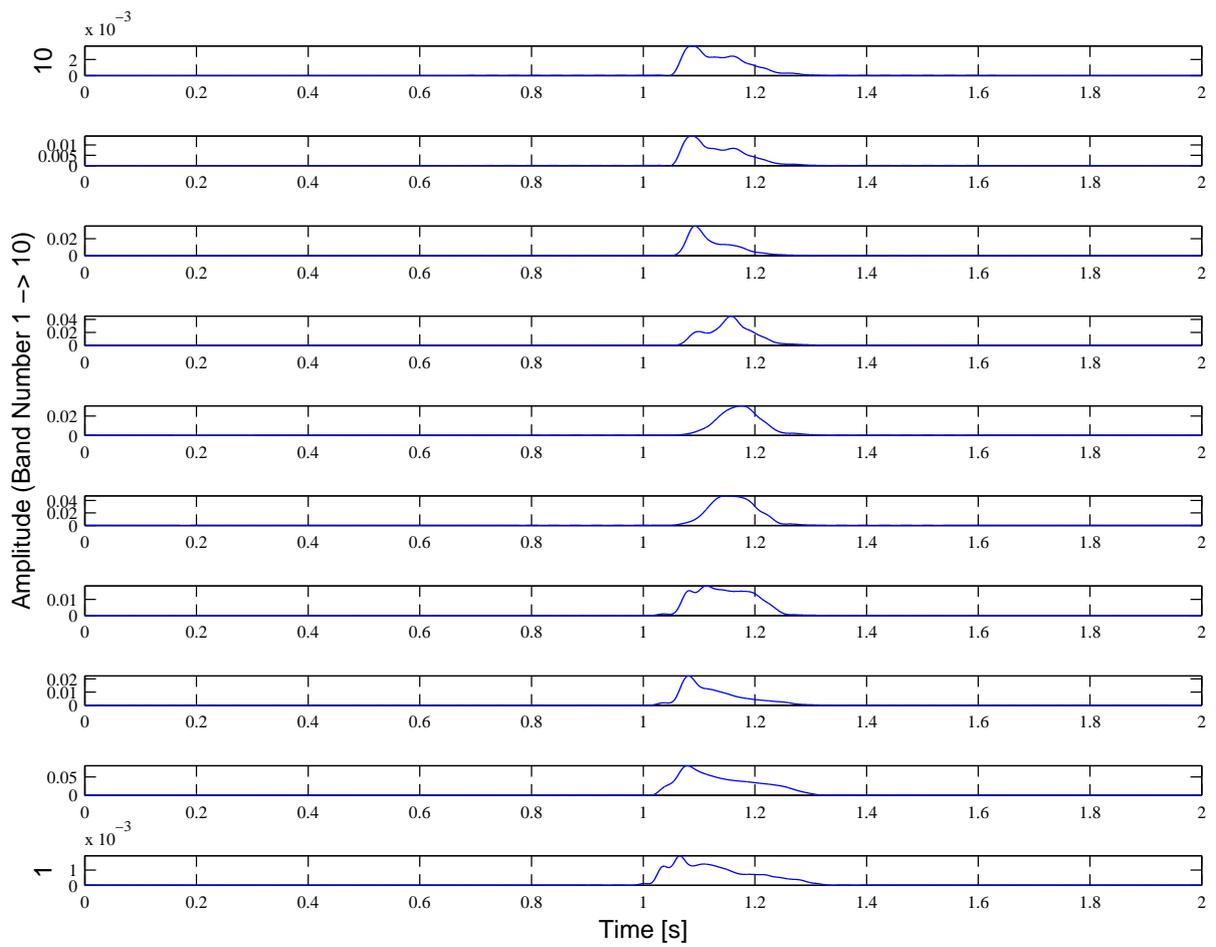


図 6.65: /nya/ を 10 個の周波数帯域分割した際の各振幅包絡

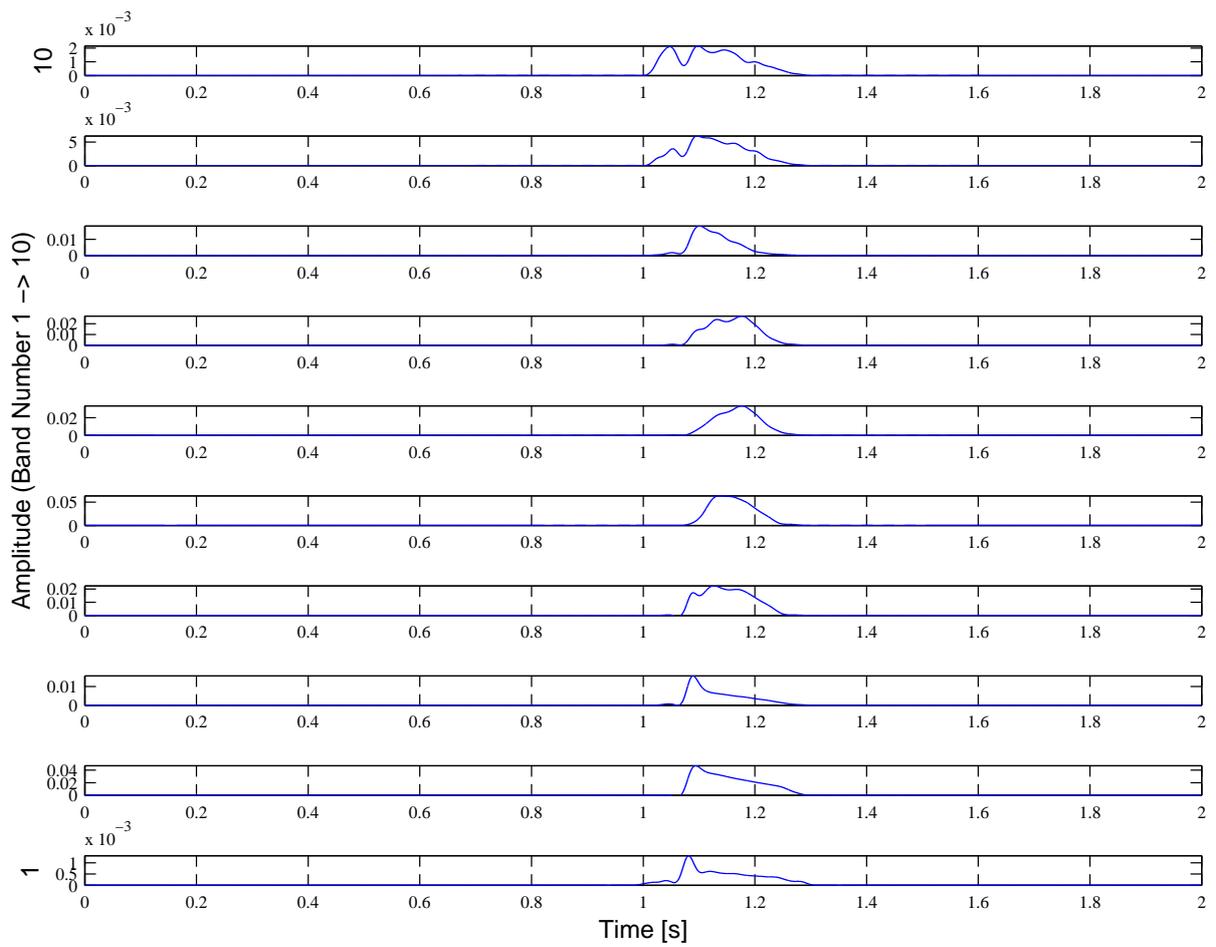


図 6.66: /hya/ を 10 個の周波数帯域分割した際の各振幅包絡

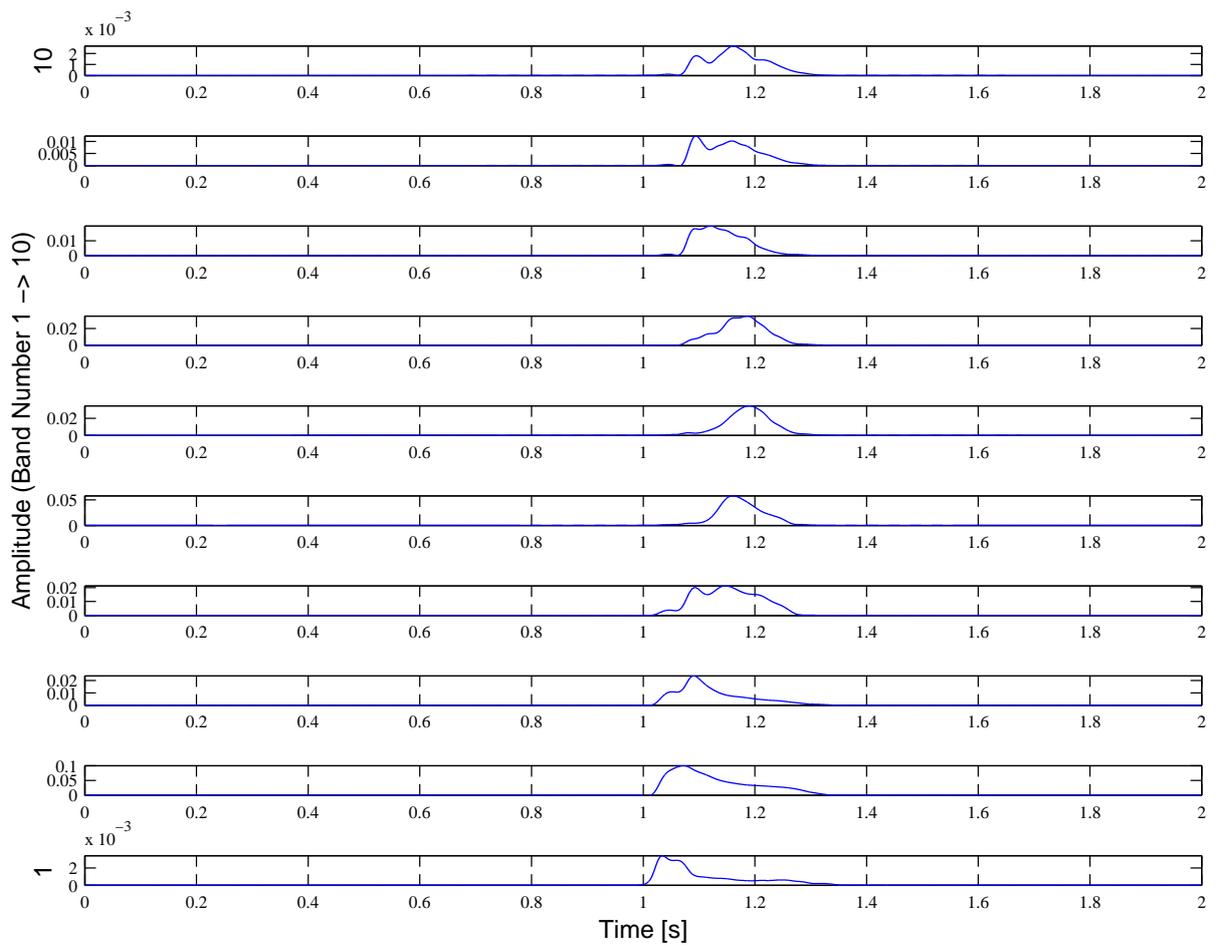


図 6.67: /mya/ を 10 個の周波数帯域分割した際の各振幅包絡

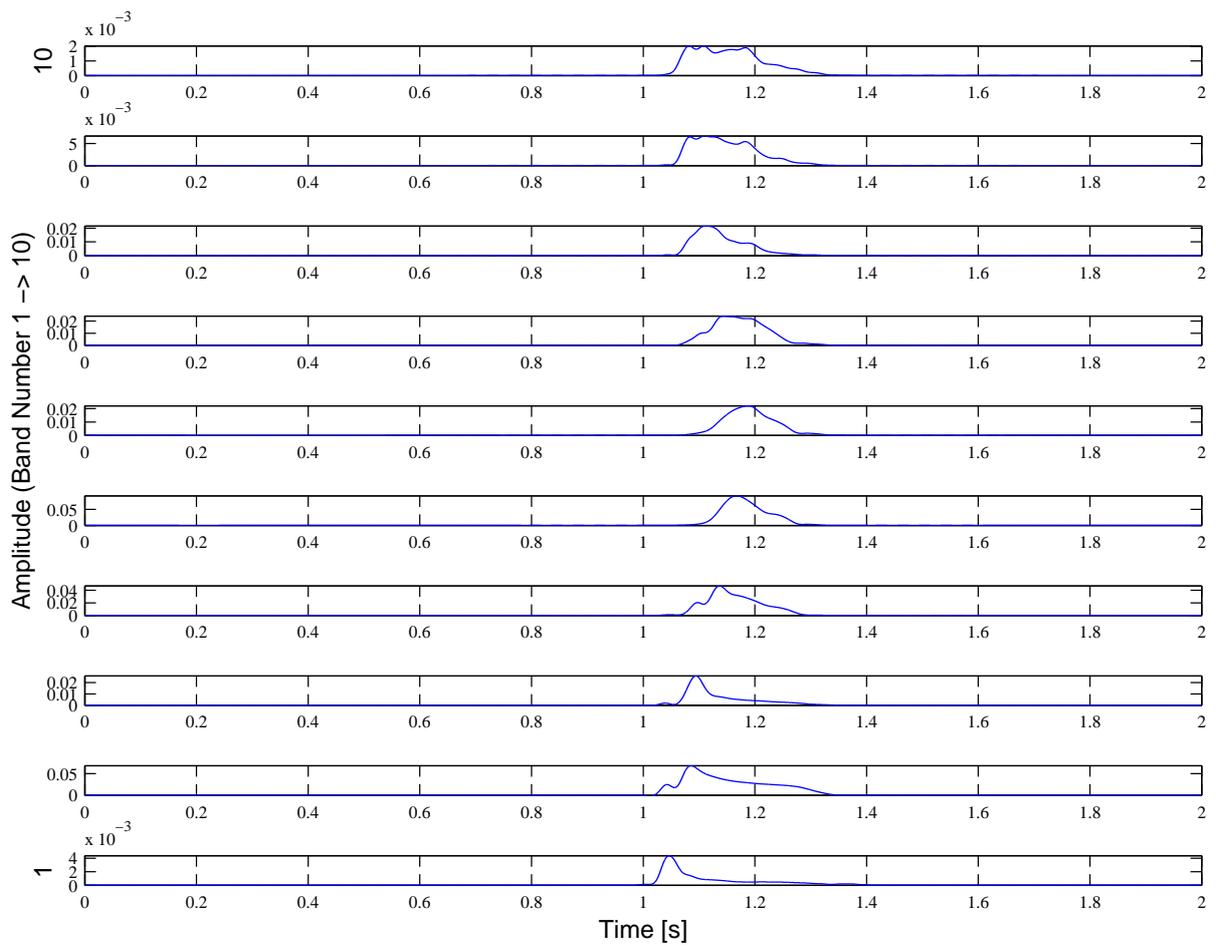


図 6.68: /rya/ を 10 個の周波数帯域分割した際の各振幅包絡

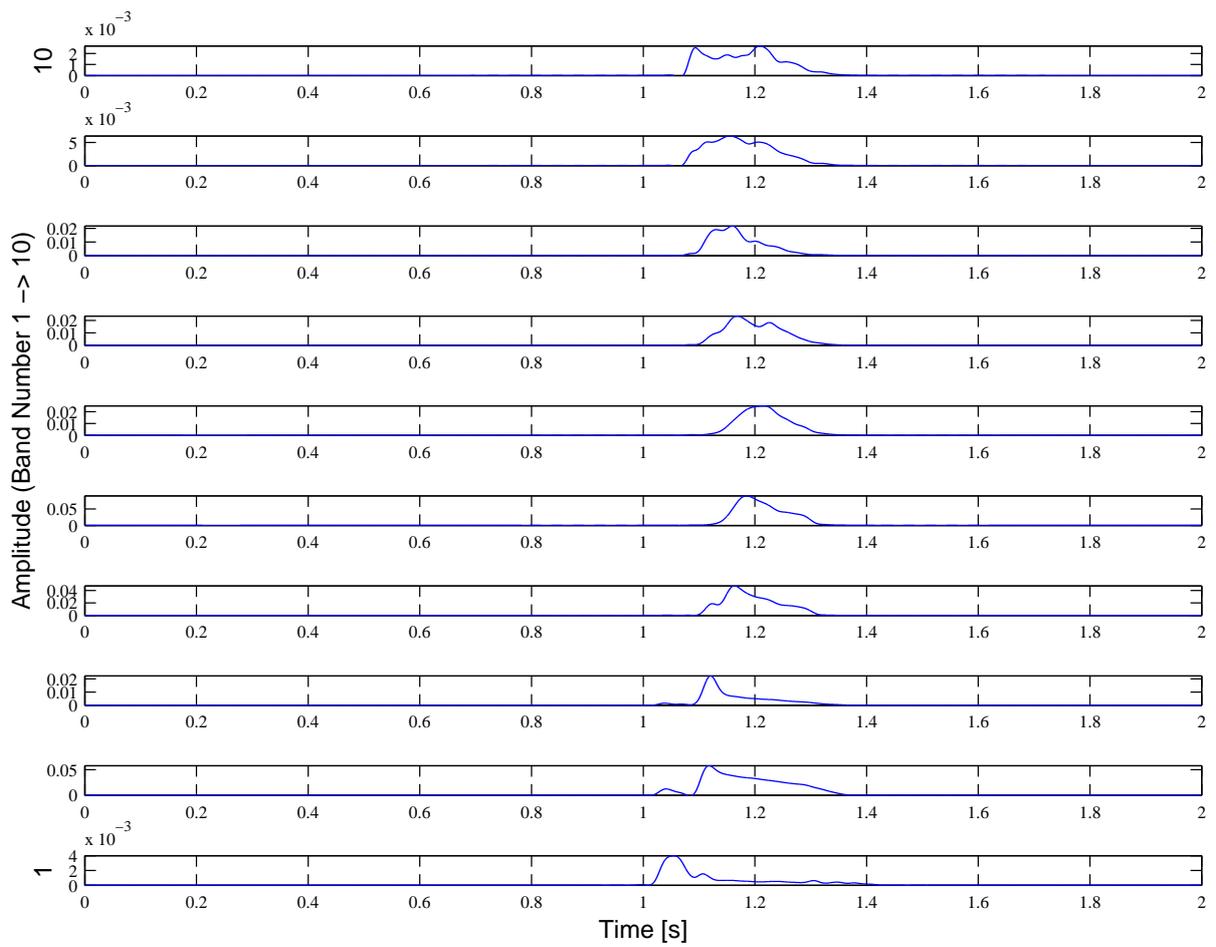


図 6.69: /gya/ を 10 個の周波数帯域分割した際の各振幅包絡

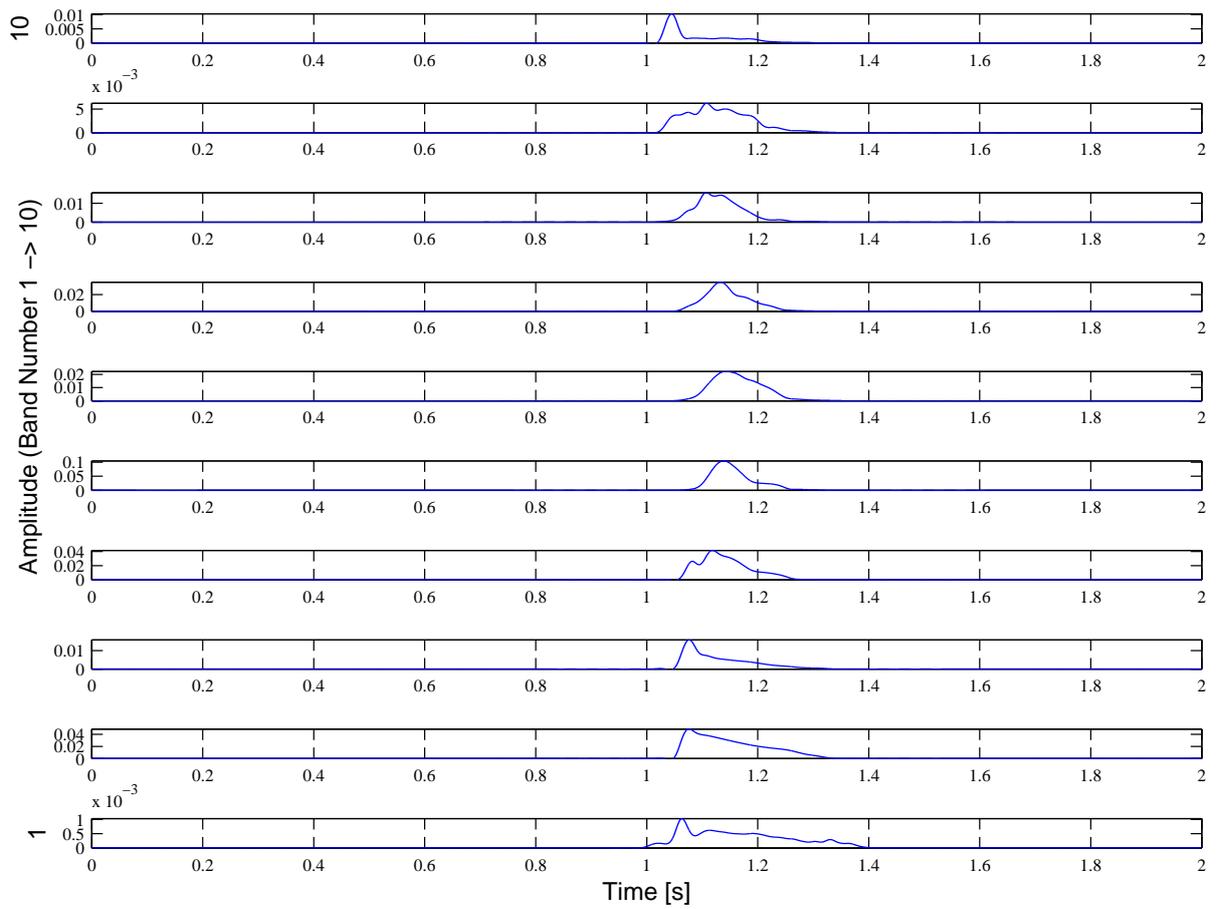


図 6.70: /zya/ を 10 個の周波数帯域分割した際の各振幅包絡

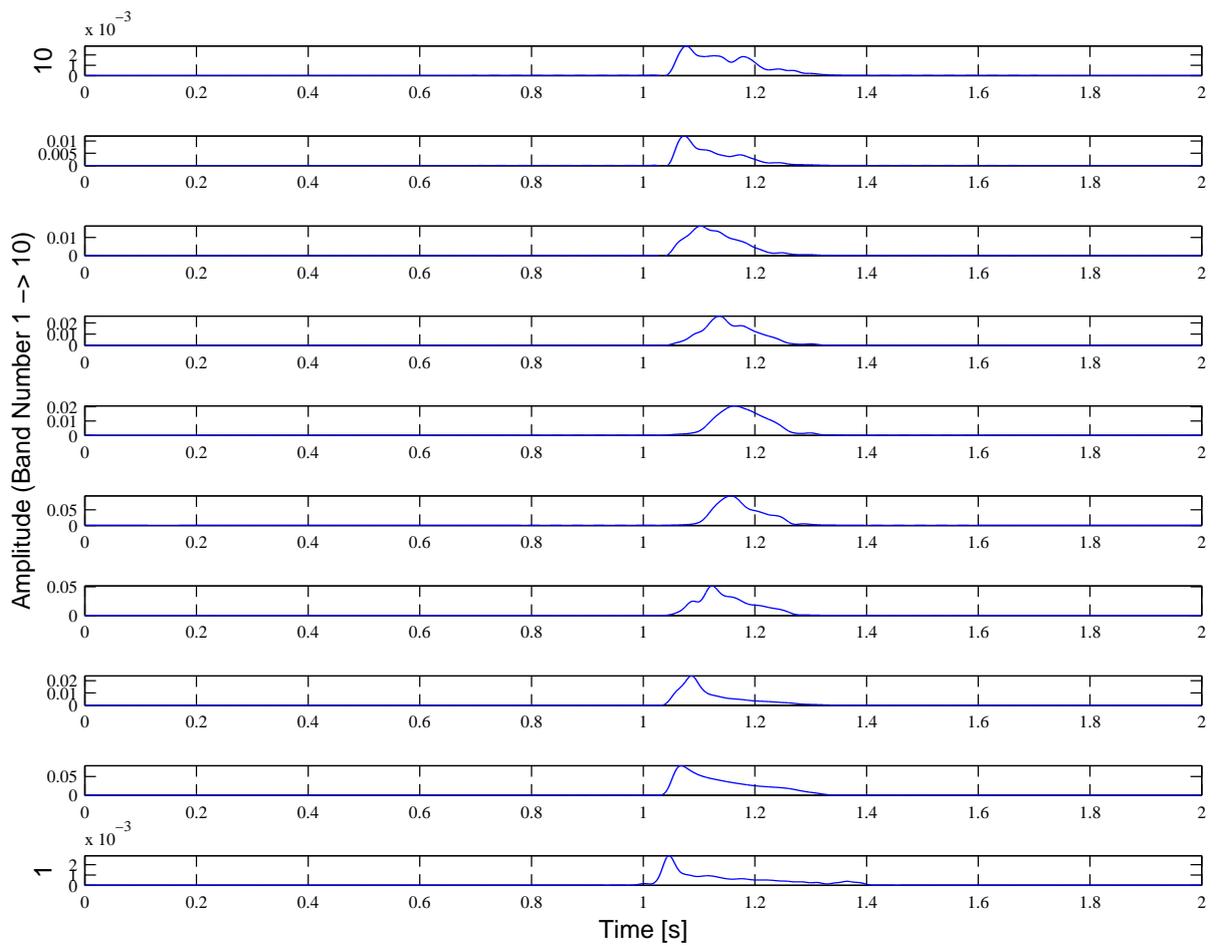


図 6.71: /bya/ を 10 個の周波数帯域分割した際の各振幅包絡

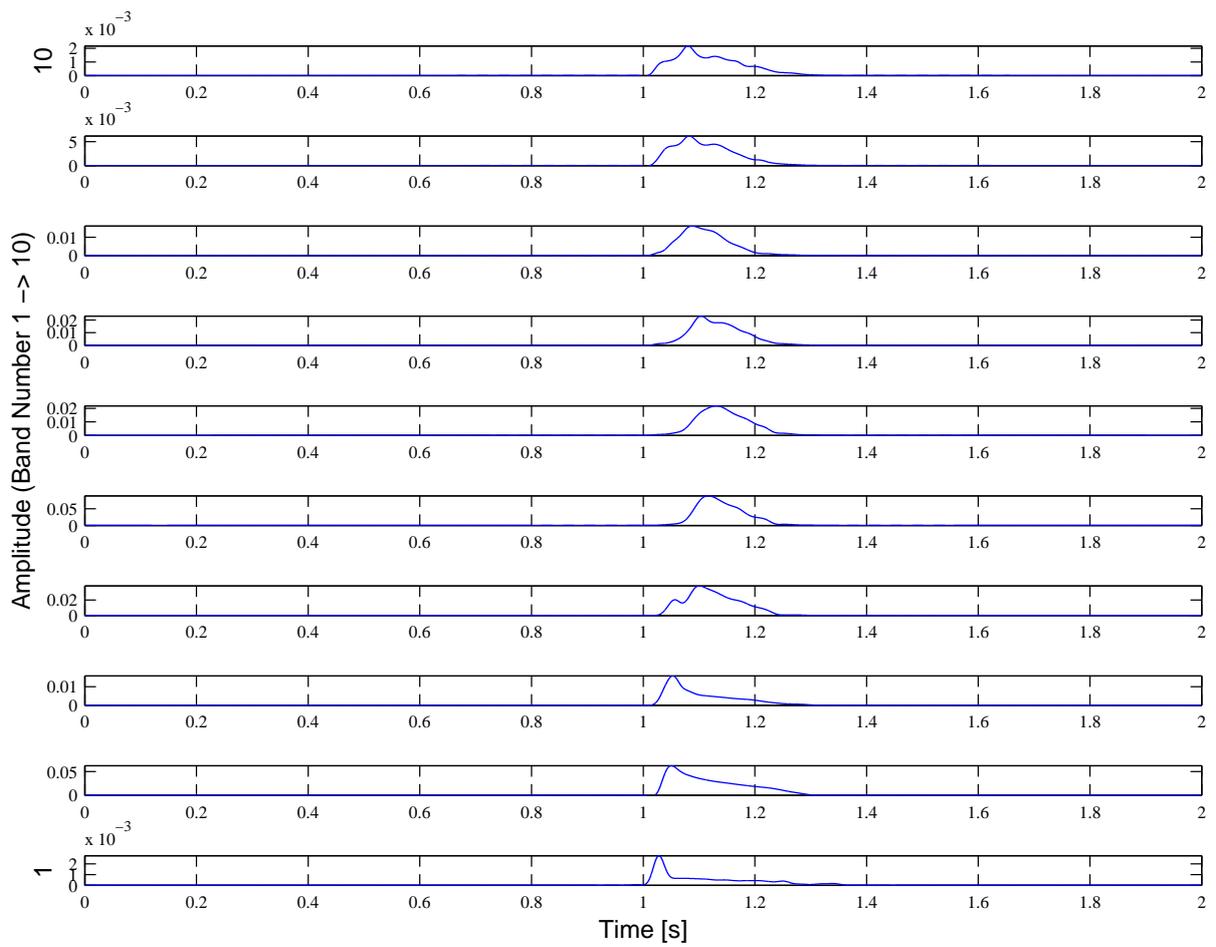


図 6.72: /pya/ を 10 個の周波数帯域分割した際の各振幅包絡

/sa/ と /za/ の各周波数帯域の振幅包絡を制御

本研究で用いた単音節音声 /sa/ と /za/ の各周波数帯域の振幅包絡を低域通過フィルタで制御した際、それぞれの振幅包絡の形状をそれぞれ下記に示す。ただし、示す図は低域通過フィルタのカットオフ周波数が、1、5、30 Hz で制御した場合である。

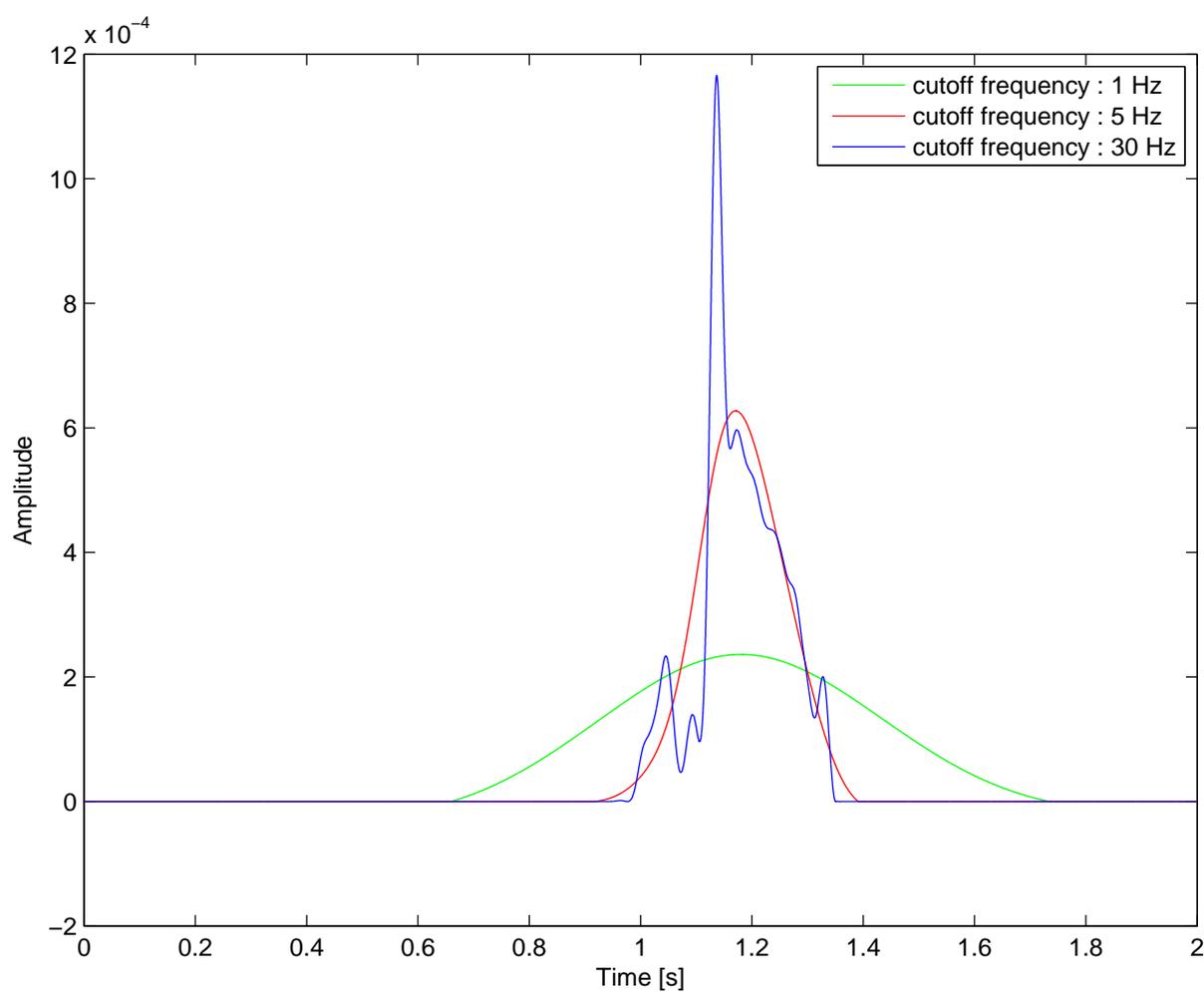


図 6.73: /sa/ の第 1 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

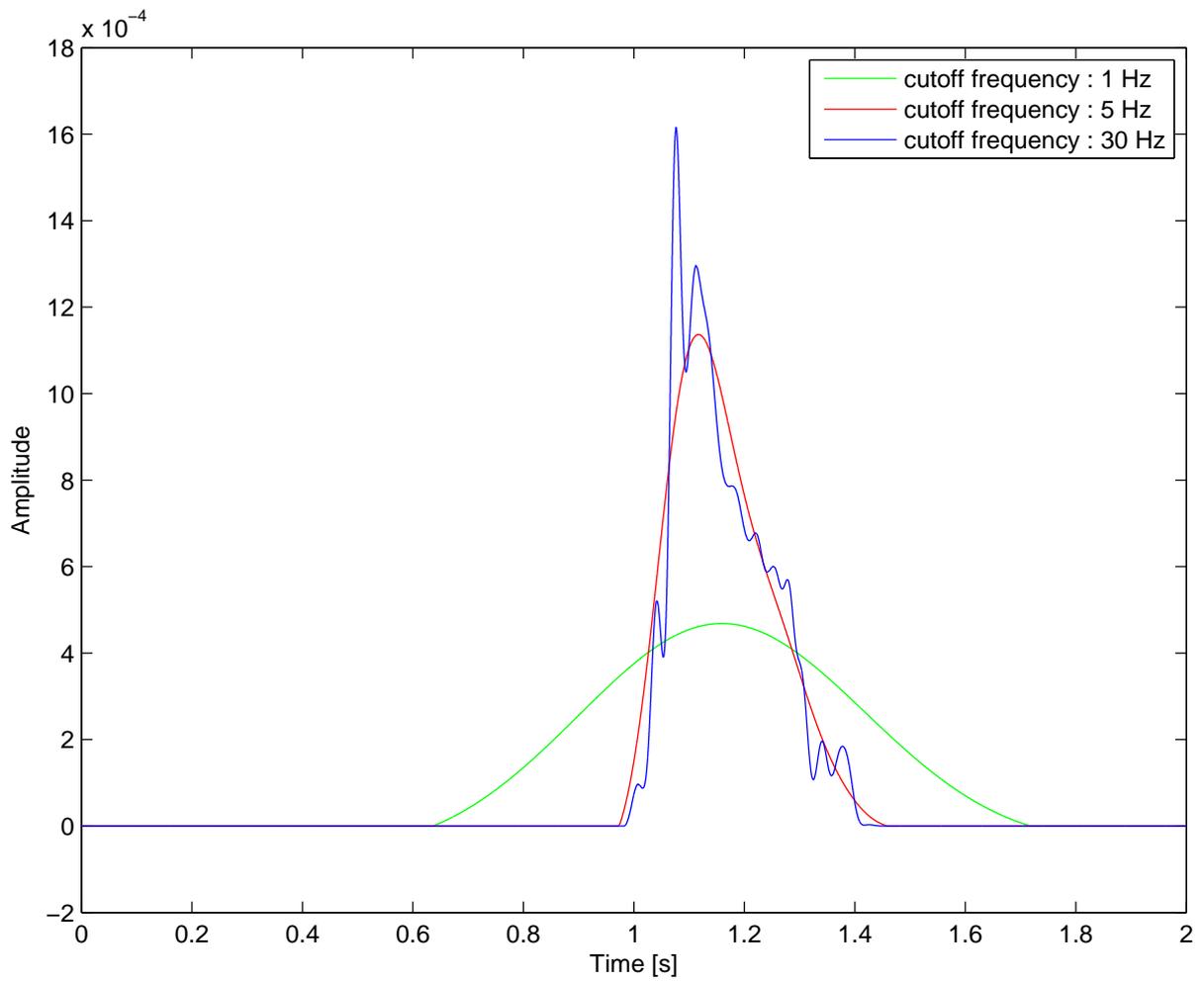


図 6.74: /za/ の第 1 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

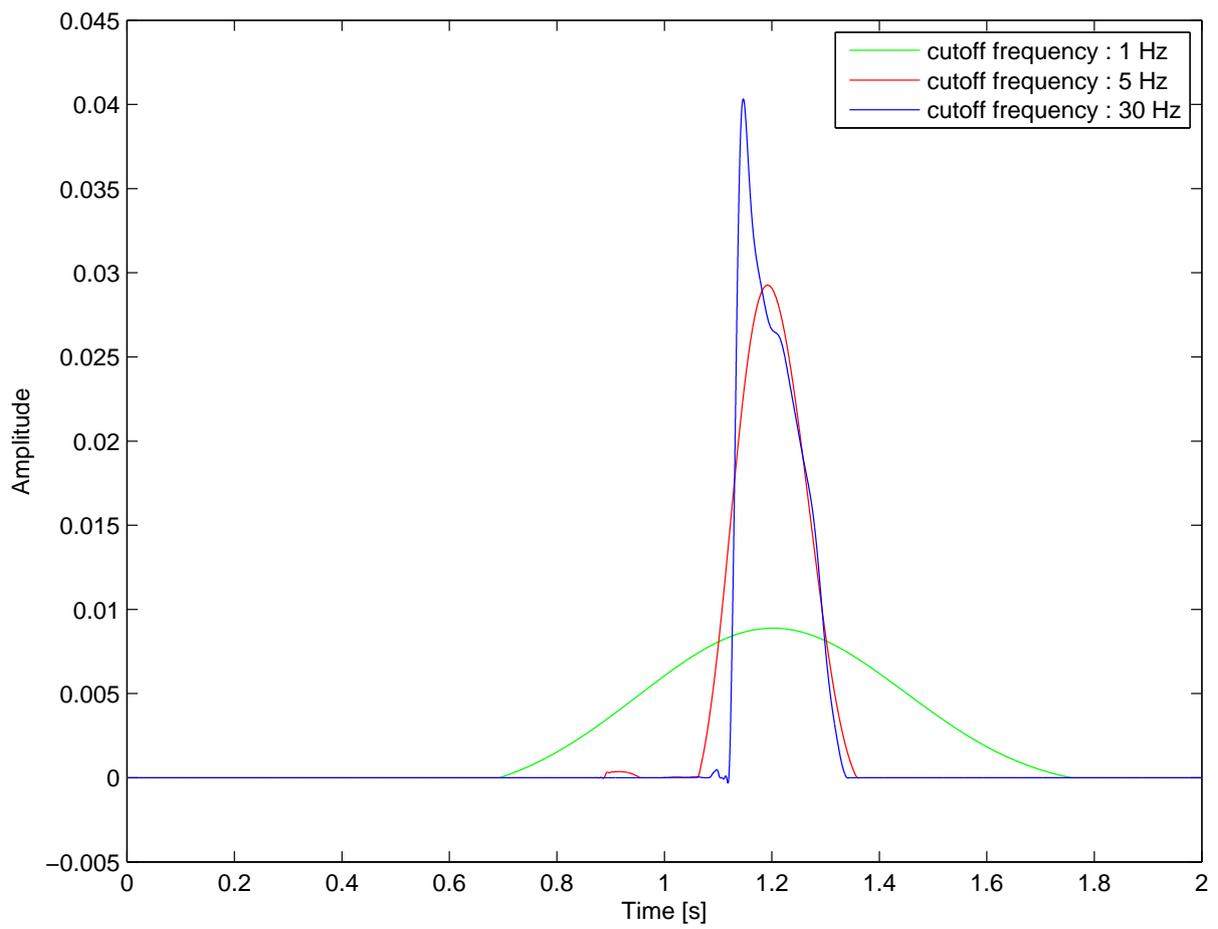


図 6.75: /sa/ の第 2 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

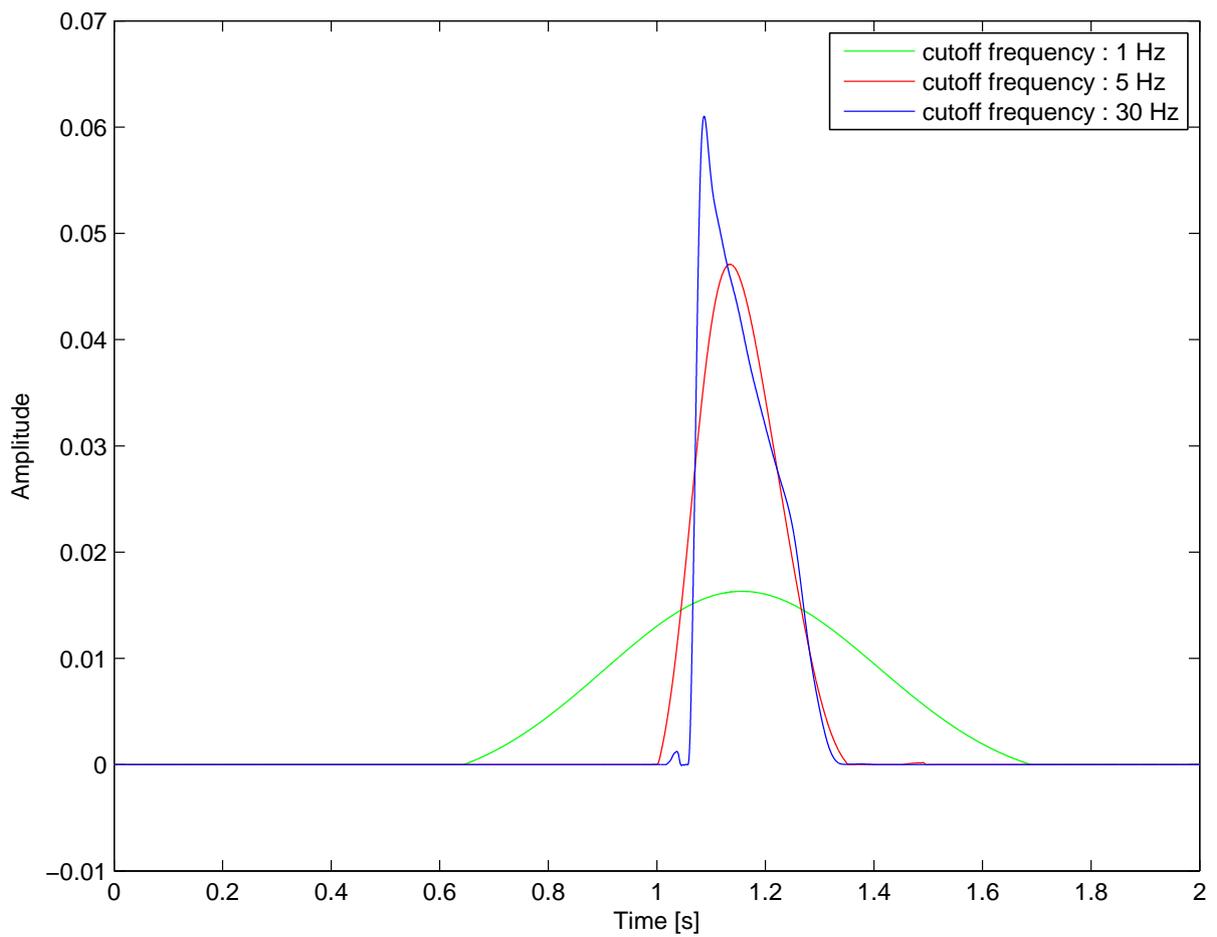


図 6.76: /za/ の第 2 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

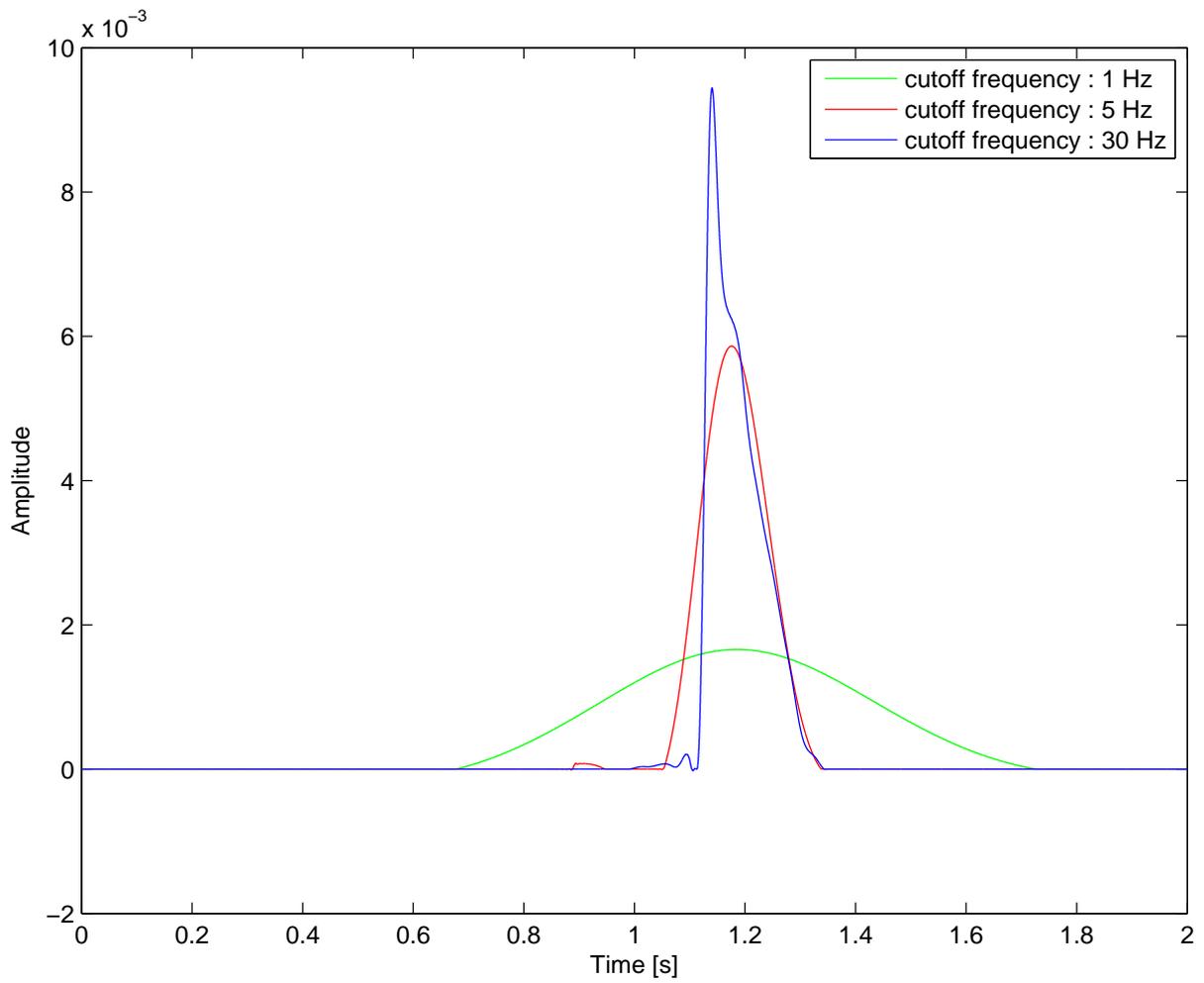


図 6.77: /sa/ の第 3 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

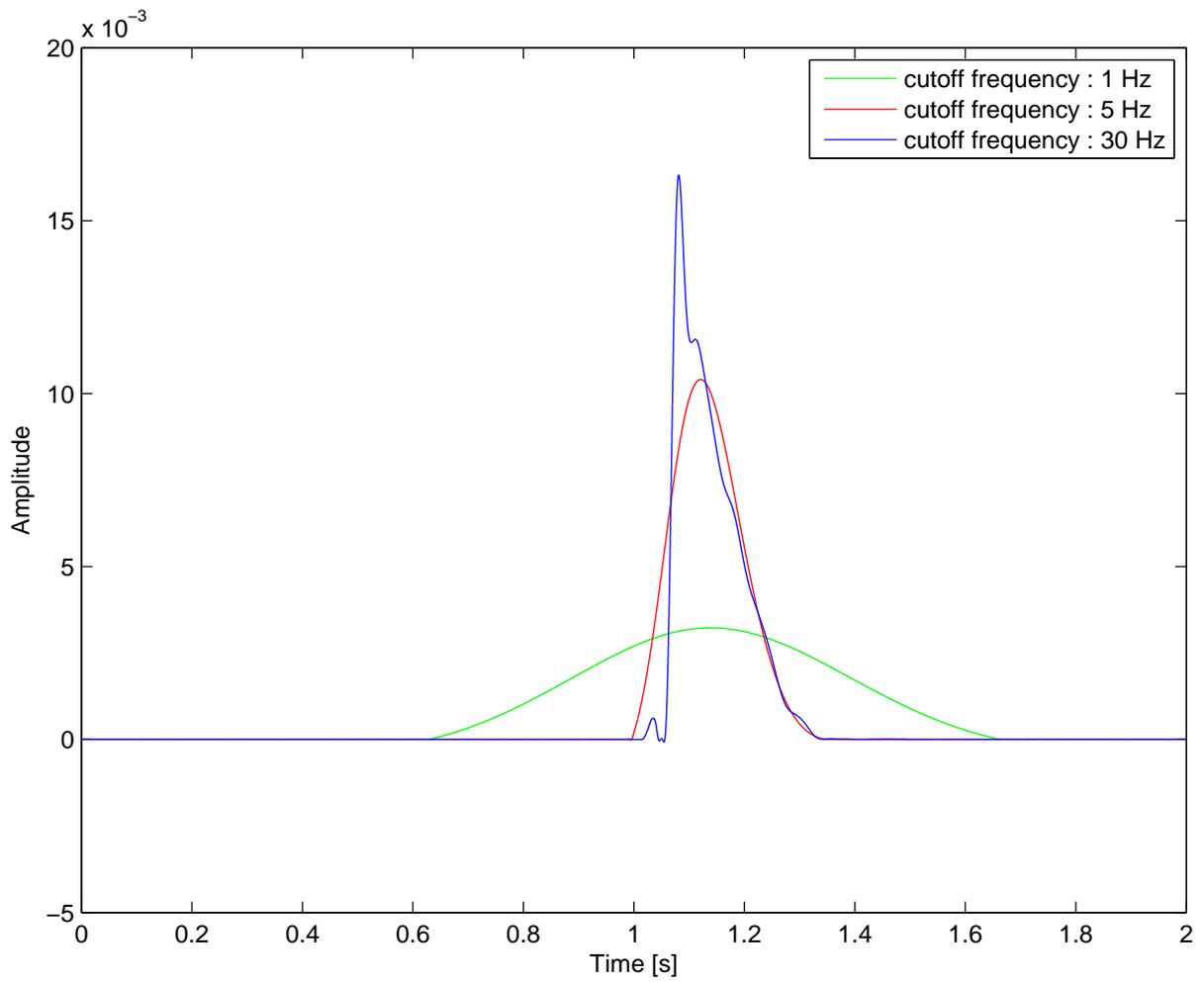


図 6.78: /za/ の第 3 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

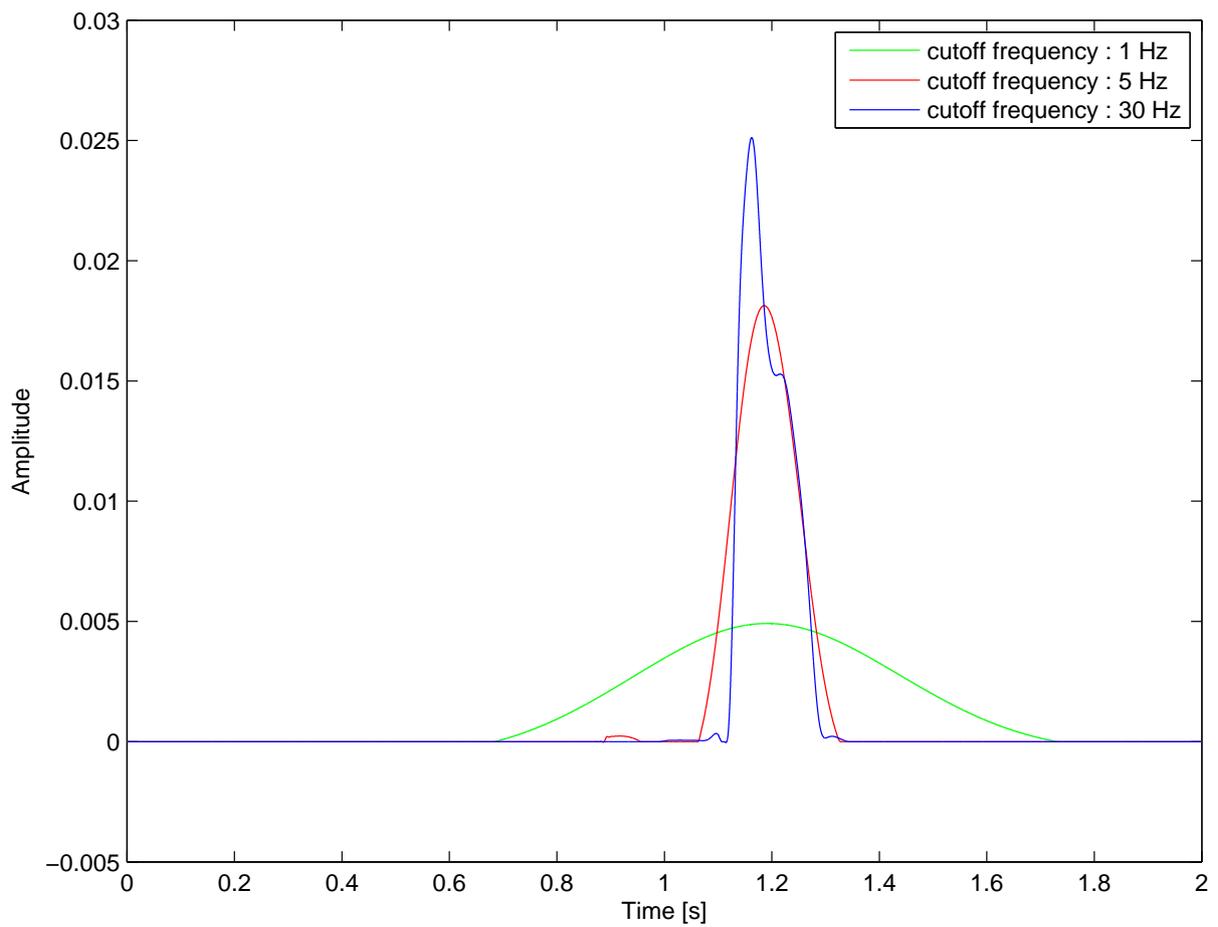


図 6.79: /sa/ の第 4 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

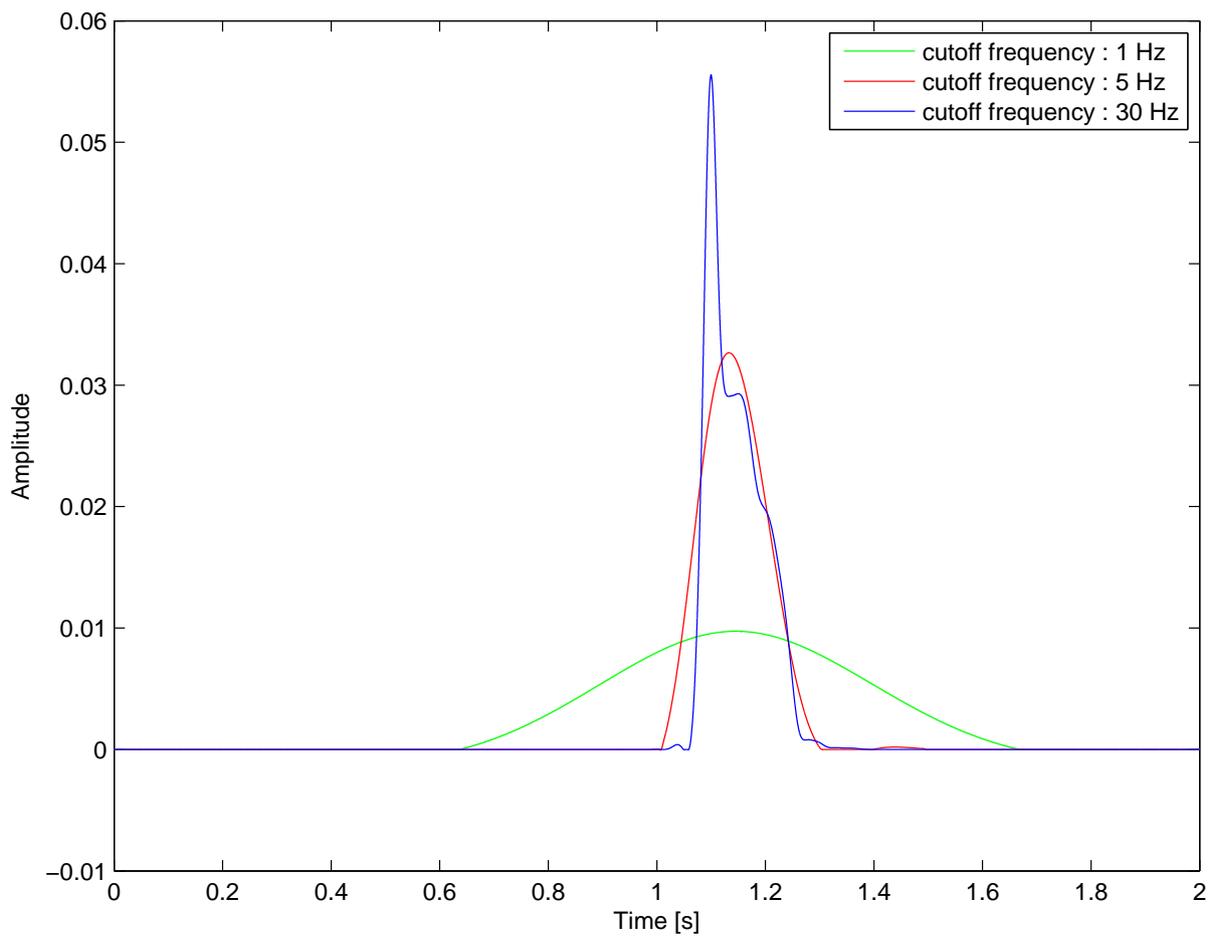


図 6.80: /za/ の第 4 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

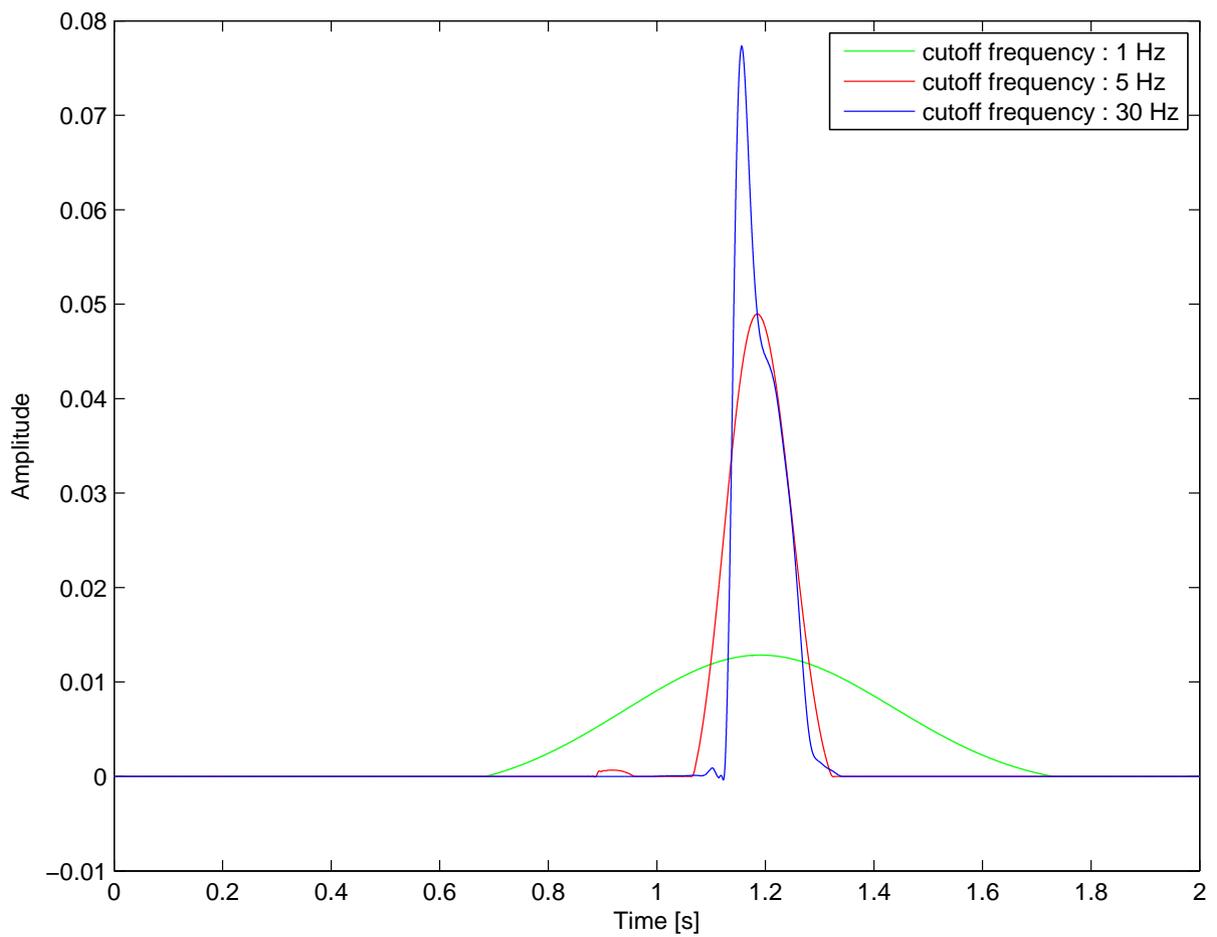


図 6.81: /sa/ の第 5 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

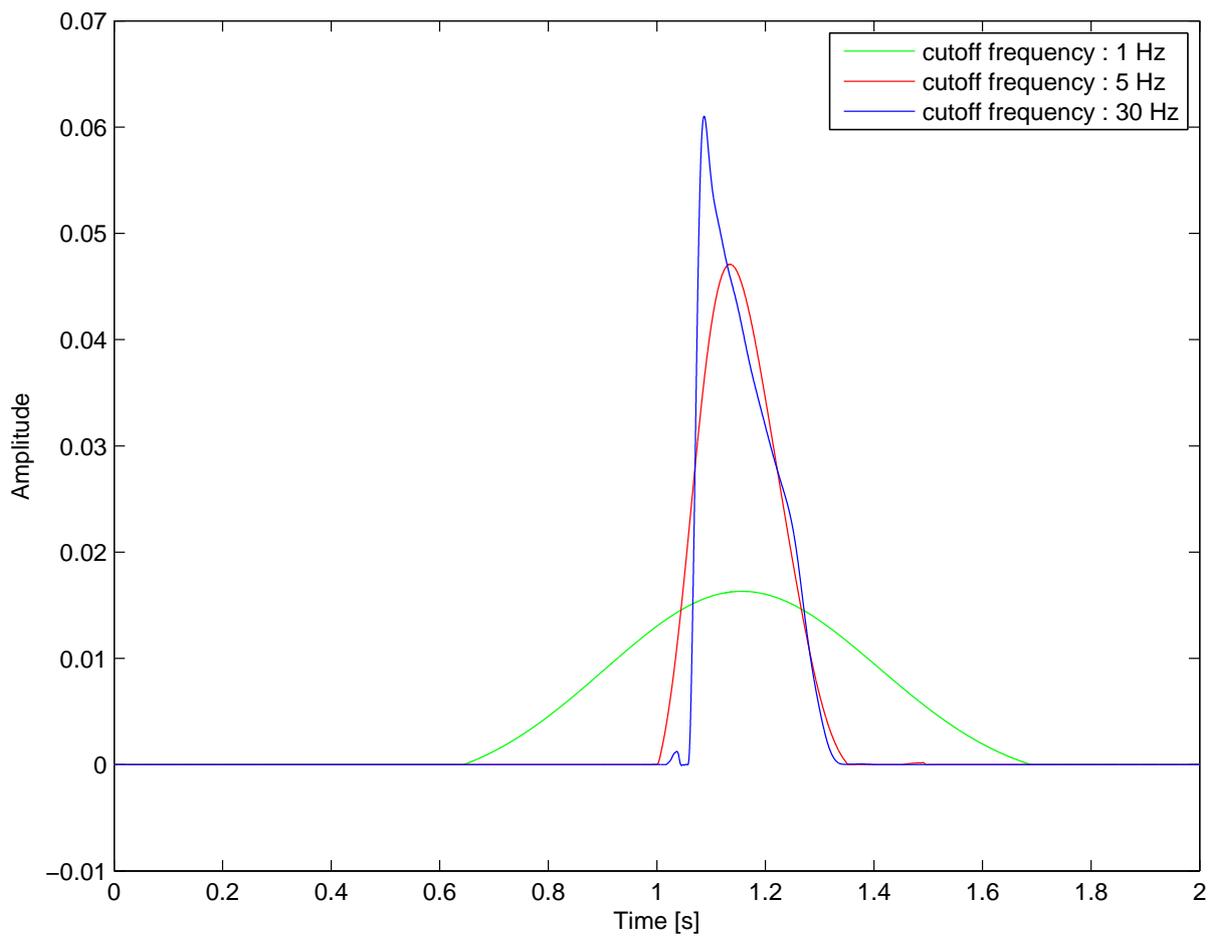


図 6.82: /za/ の第 5 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

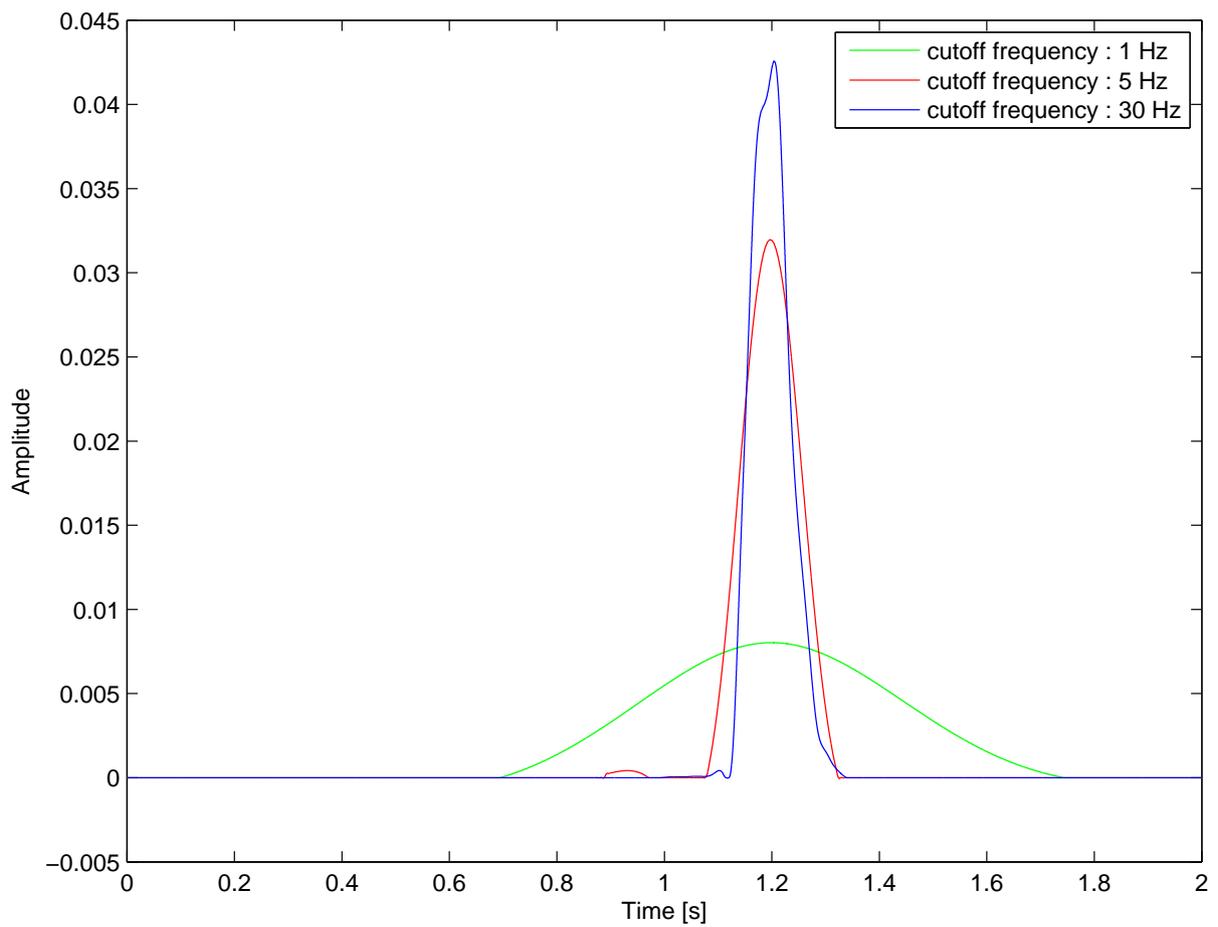


図 6.83: /sa/ の第 6 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

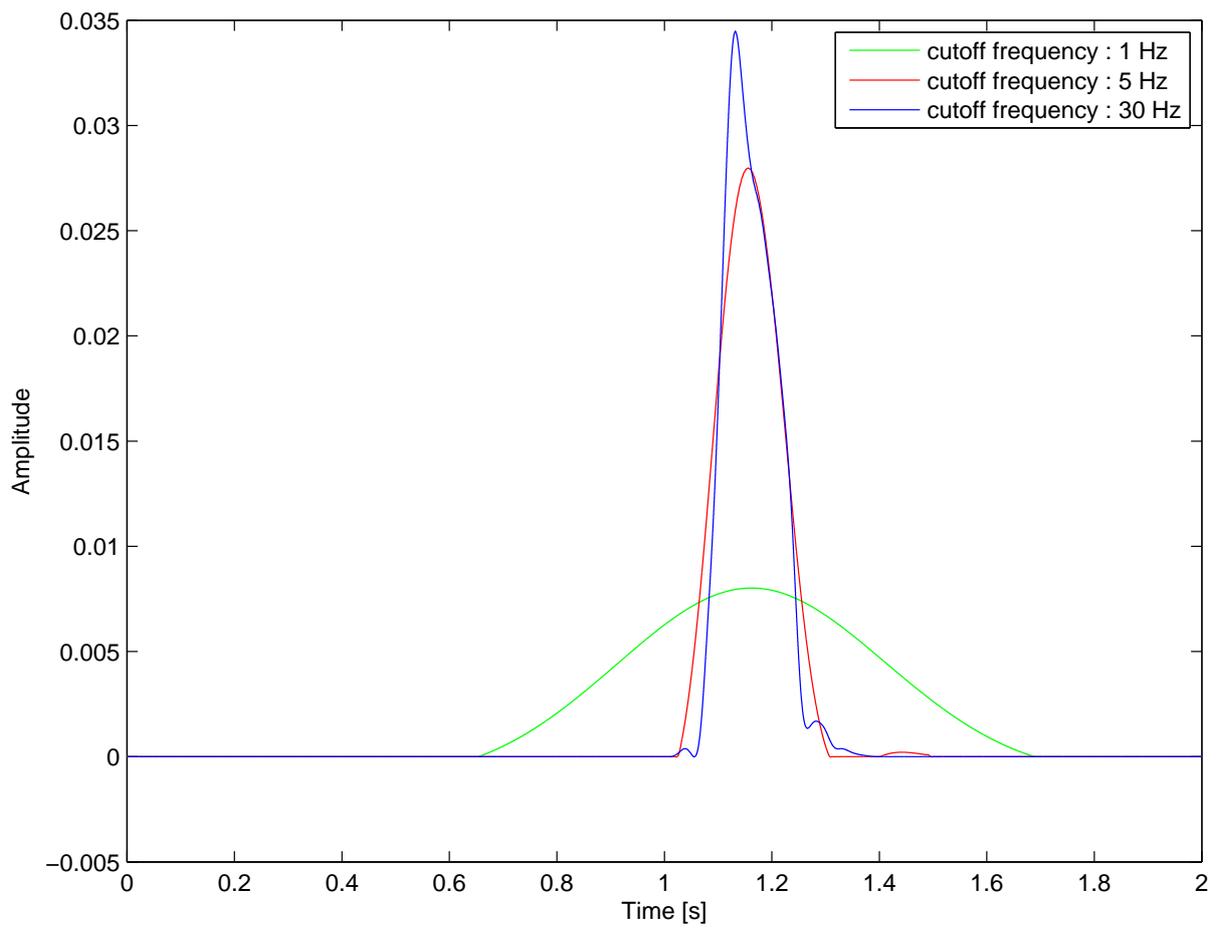


図 6.84: /za/ の第 6 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

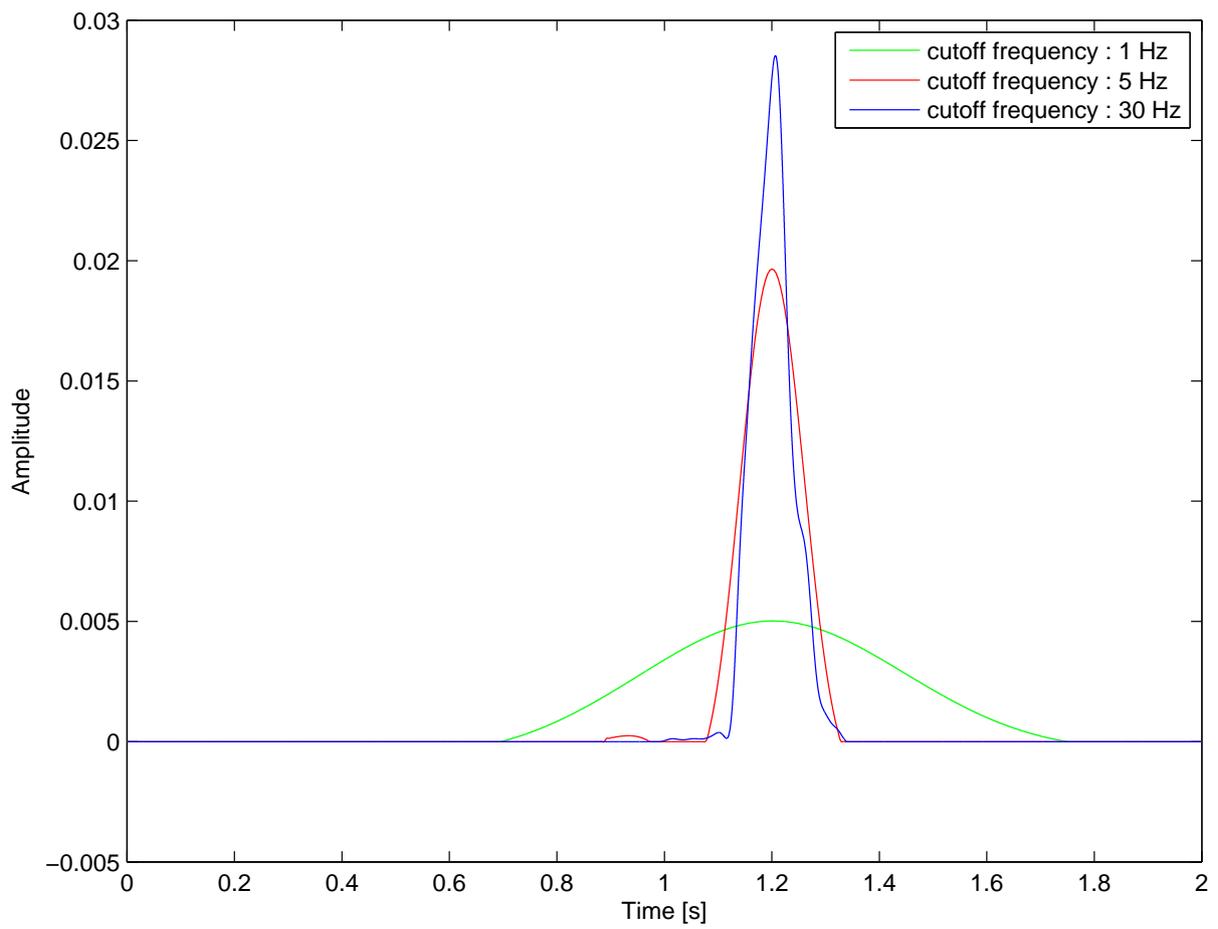


図 6.85: /sa/ の第 7 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

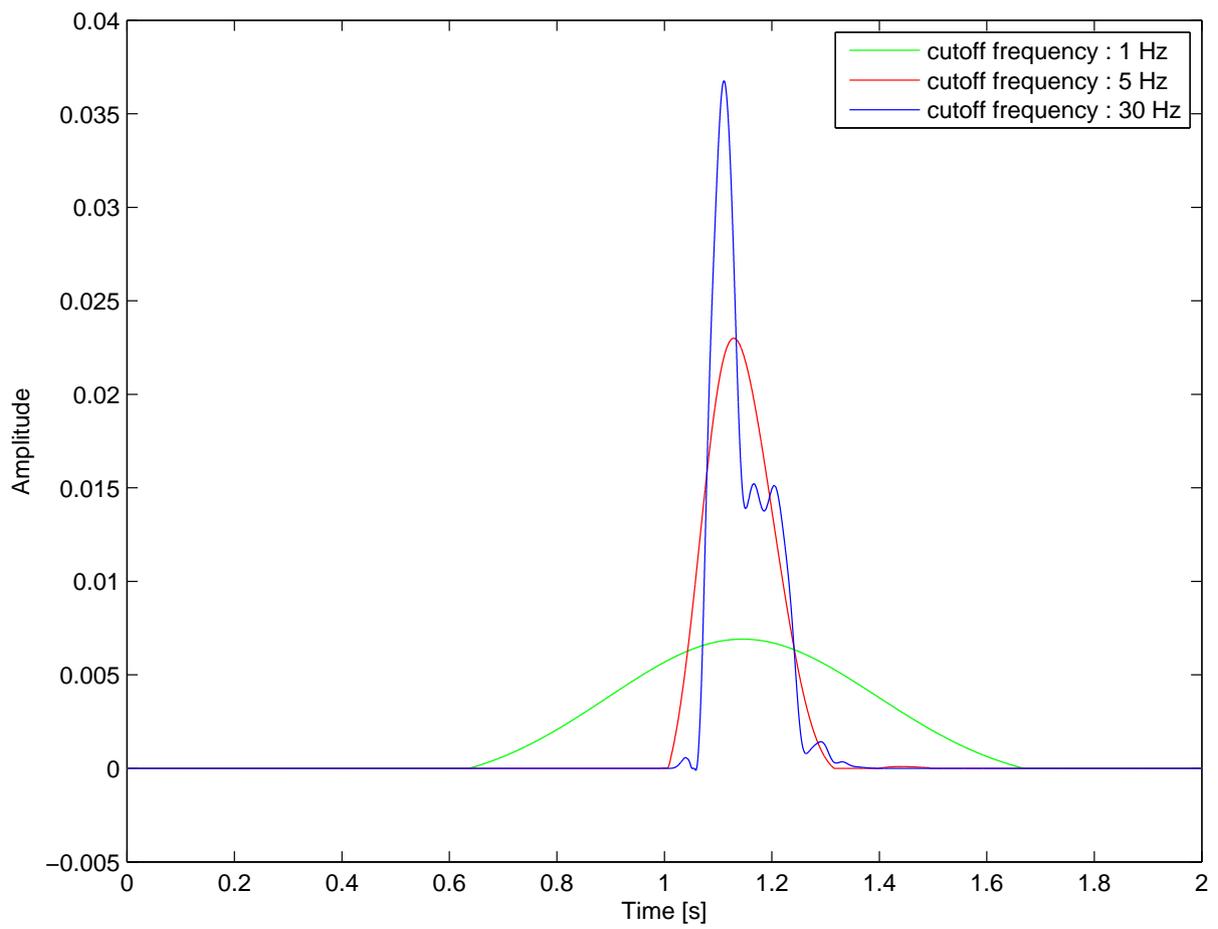


図 6.86: /za/ の第 7 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

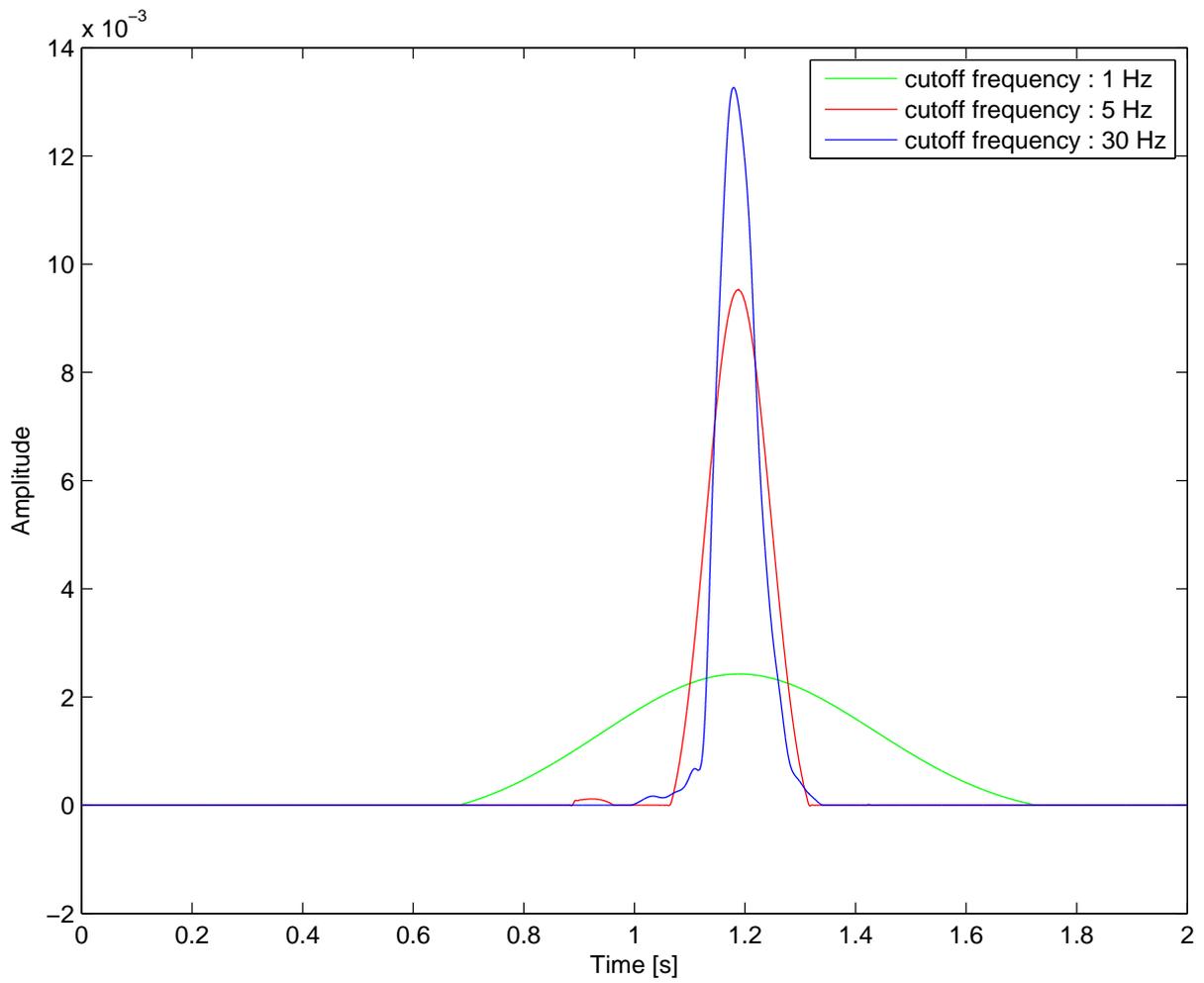


図 6.87: /sa/ の第 8 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

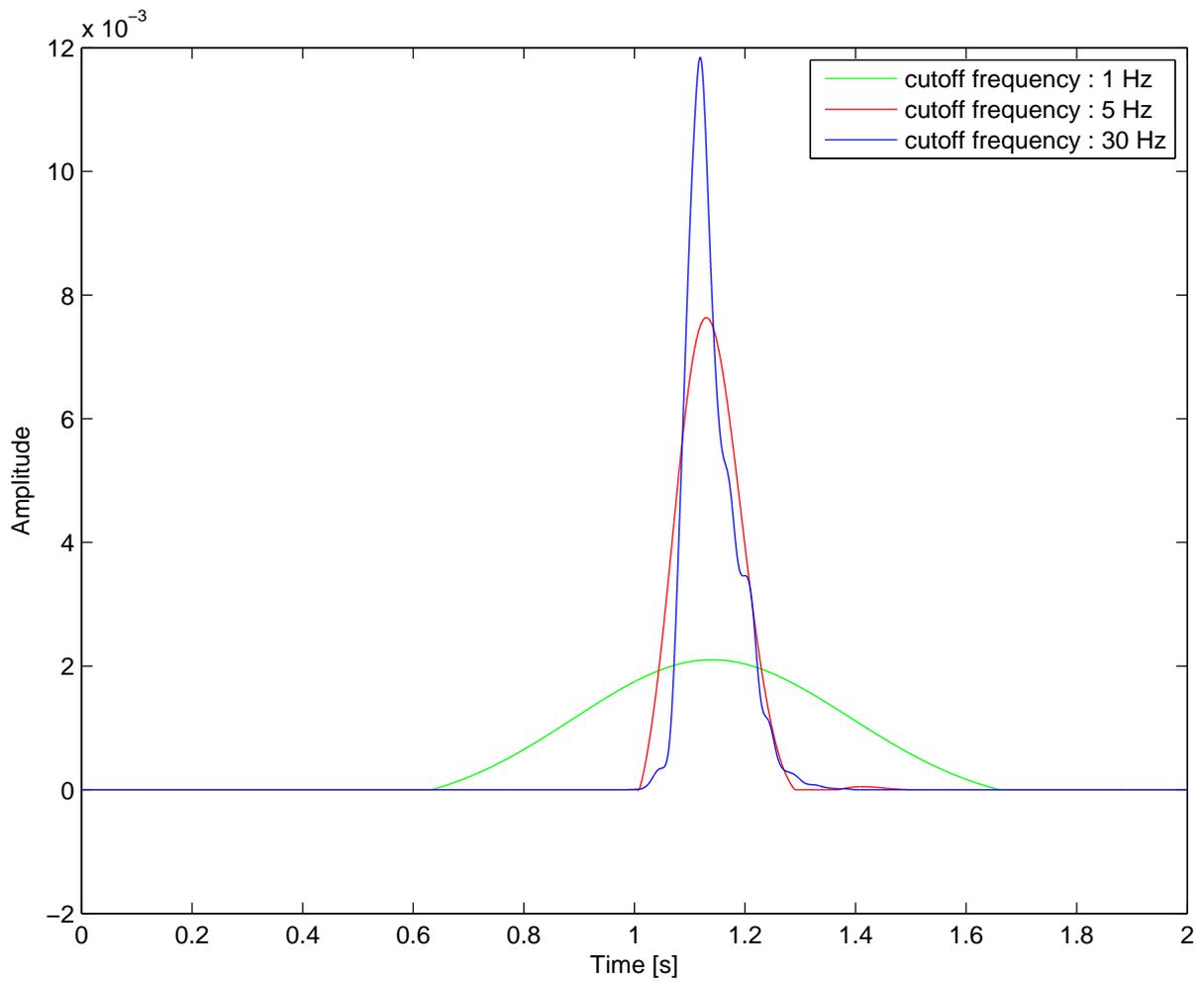


図 6.88: /za/ の第 8 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

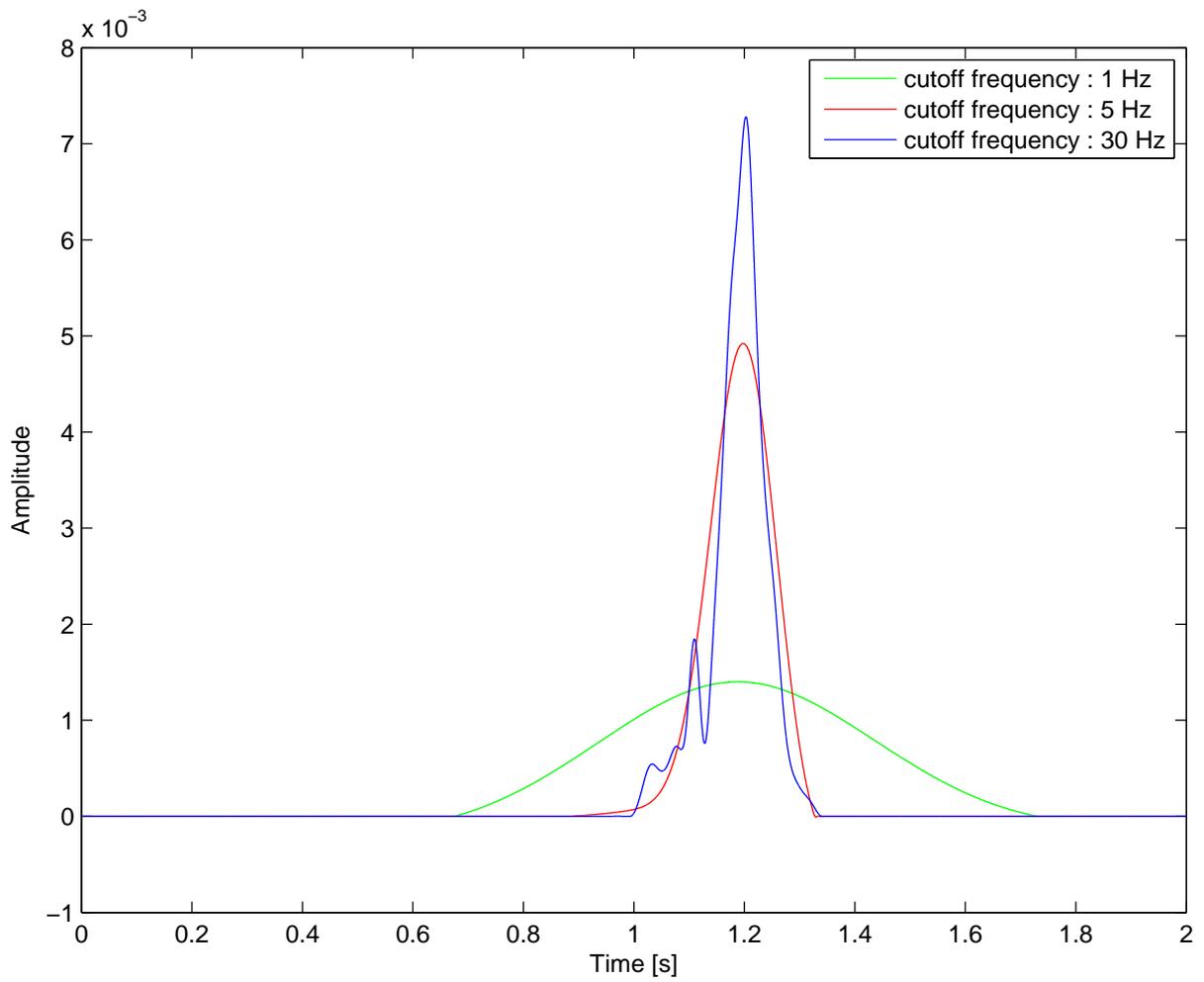


図 6.89: /sa/ の第 9 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

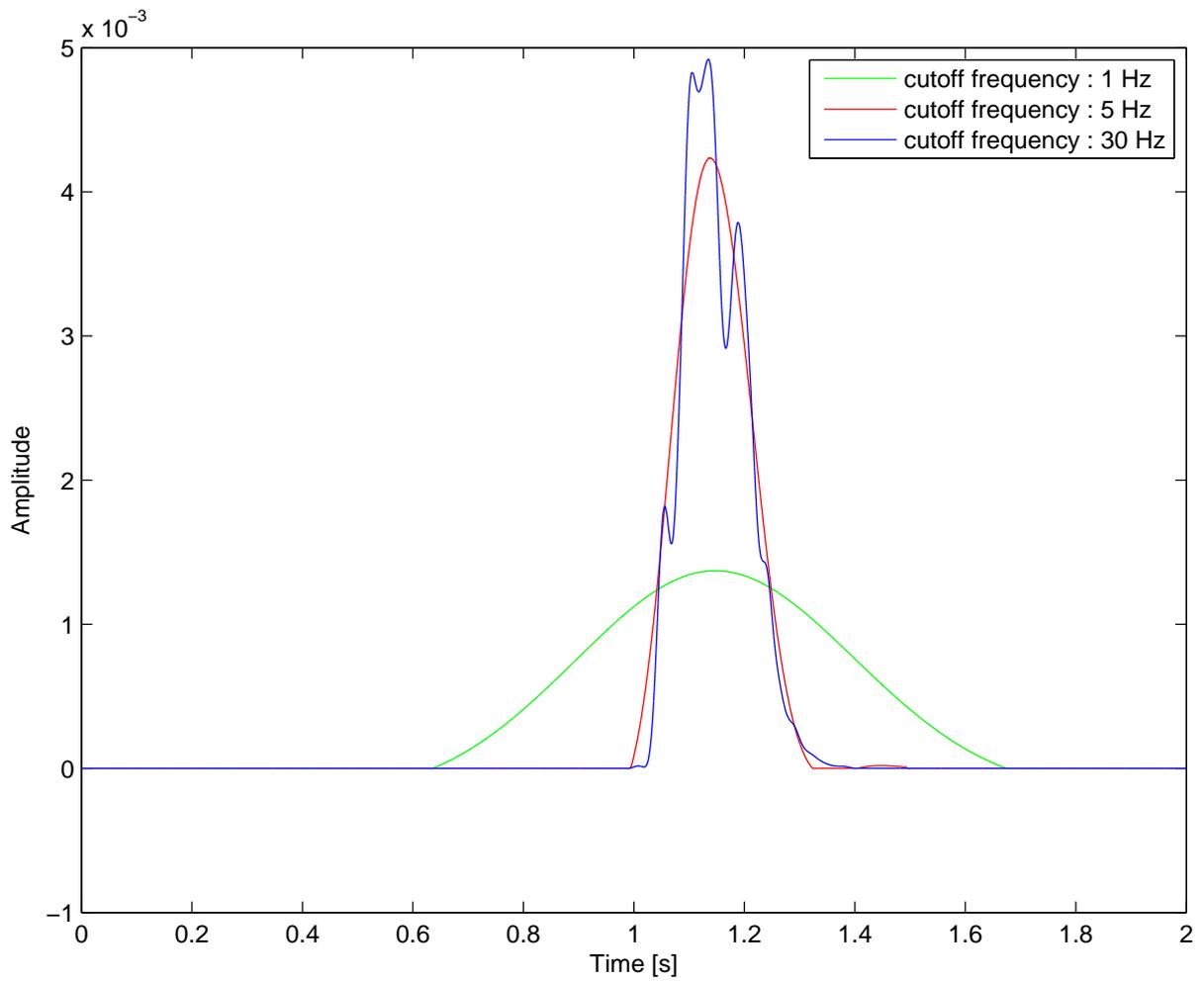


図 6.90: /za/ の第 9 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

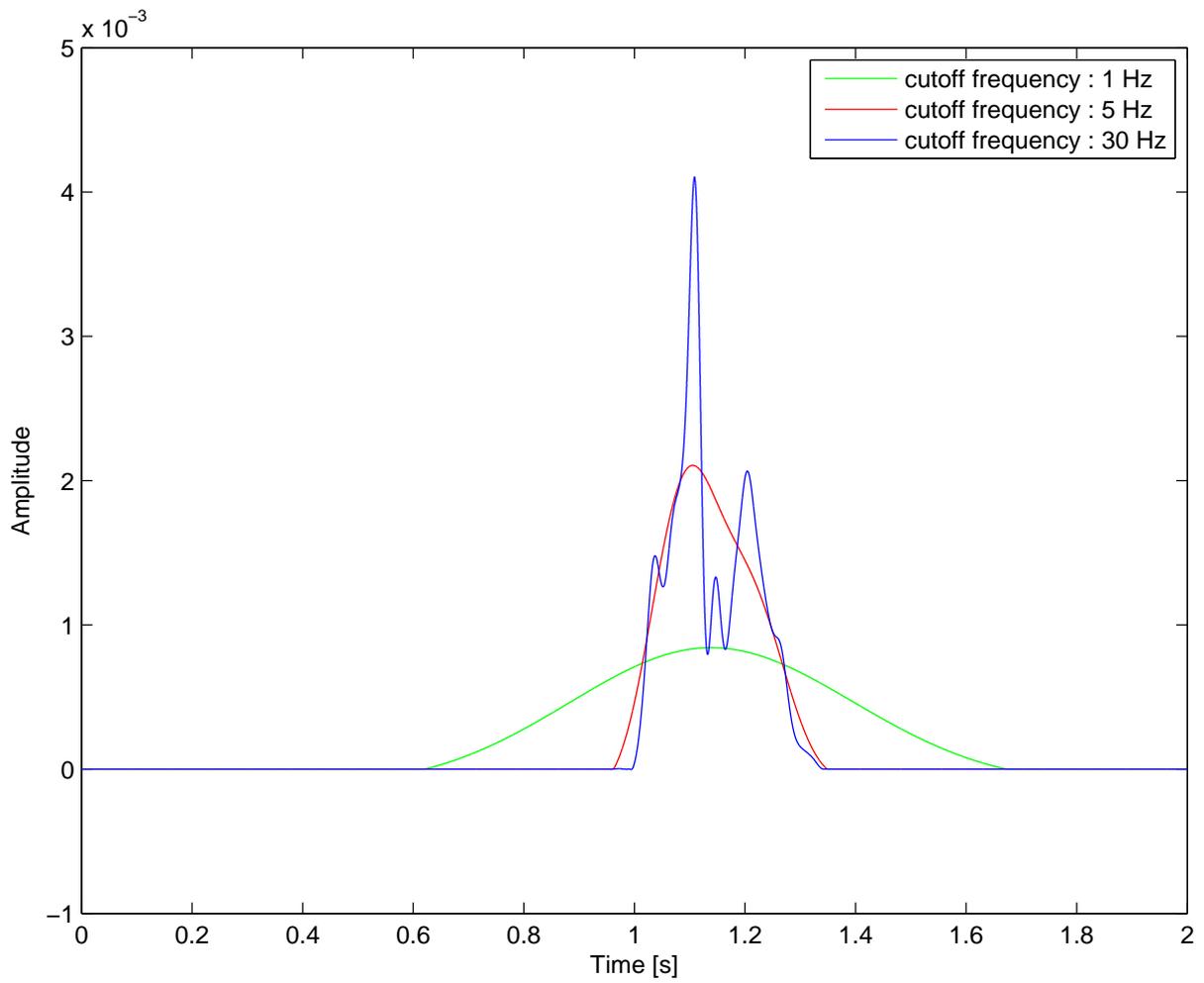


図 6.91: /sa/ の第 10 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

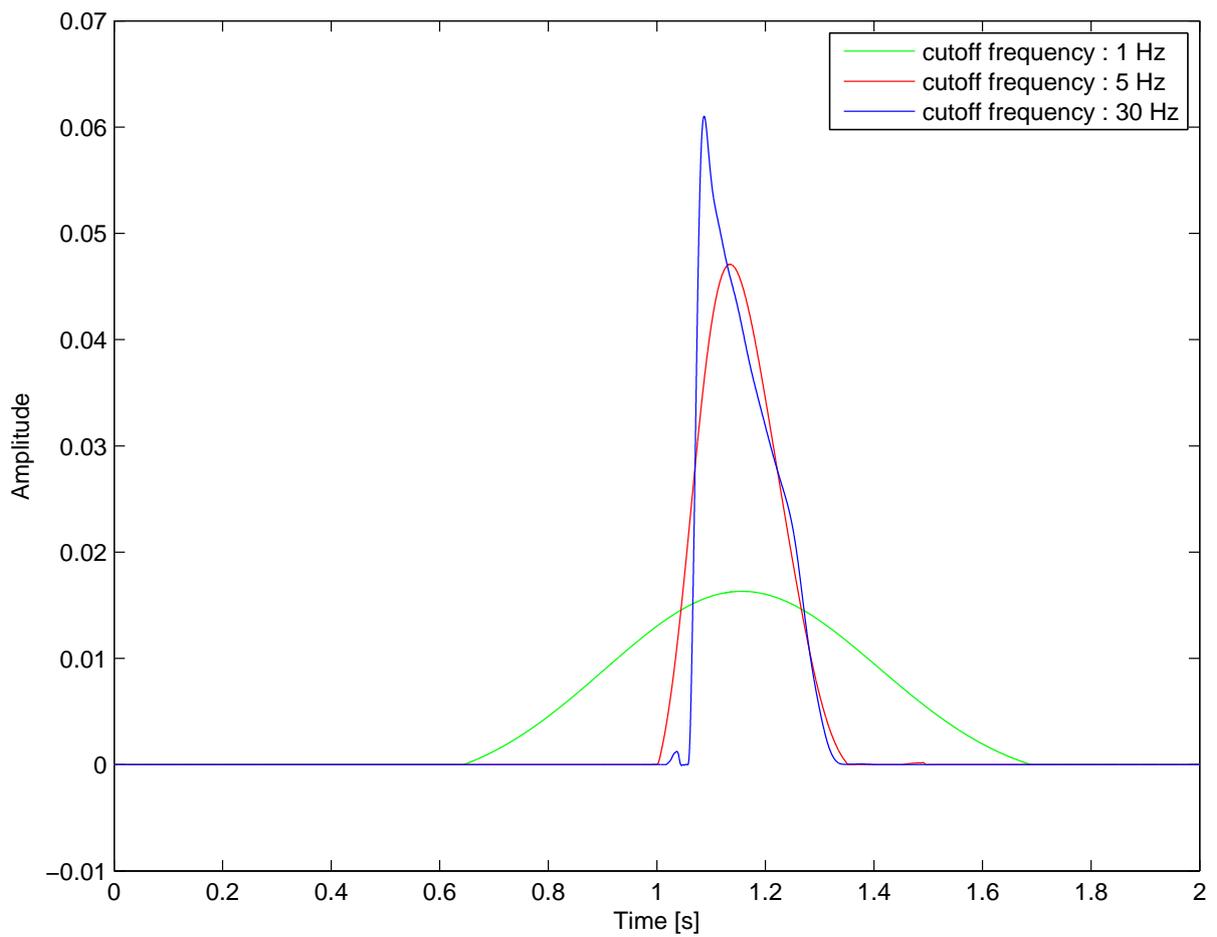


図 6.92: /za/ の第 10 周波数帯域の振幅包絡を低域通過フィルタのカットオフ周波数で制御した場合の変化

単音節音声（ /a/ , /sa/ , /ha/ , /za/ ）を用いた実験より 各被験者の正答率

単音節音声（ /a/ , /sa/ , /ha/ , /za/ ）を用いた実験より，被験者 8 名の正答率をそれぞれ下記に示す．

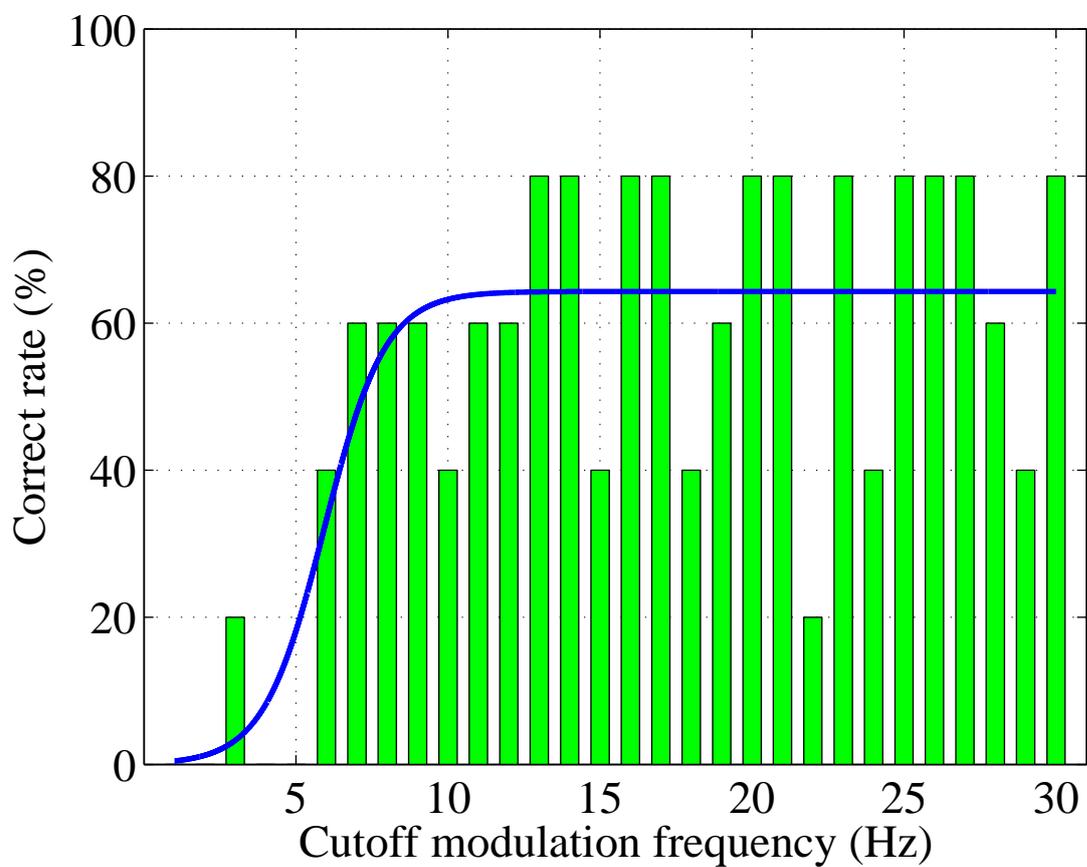


図 6.93: 被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率（原信号 /a/）

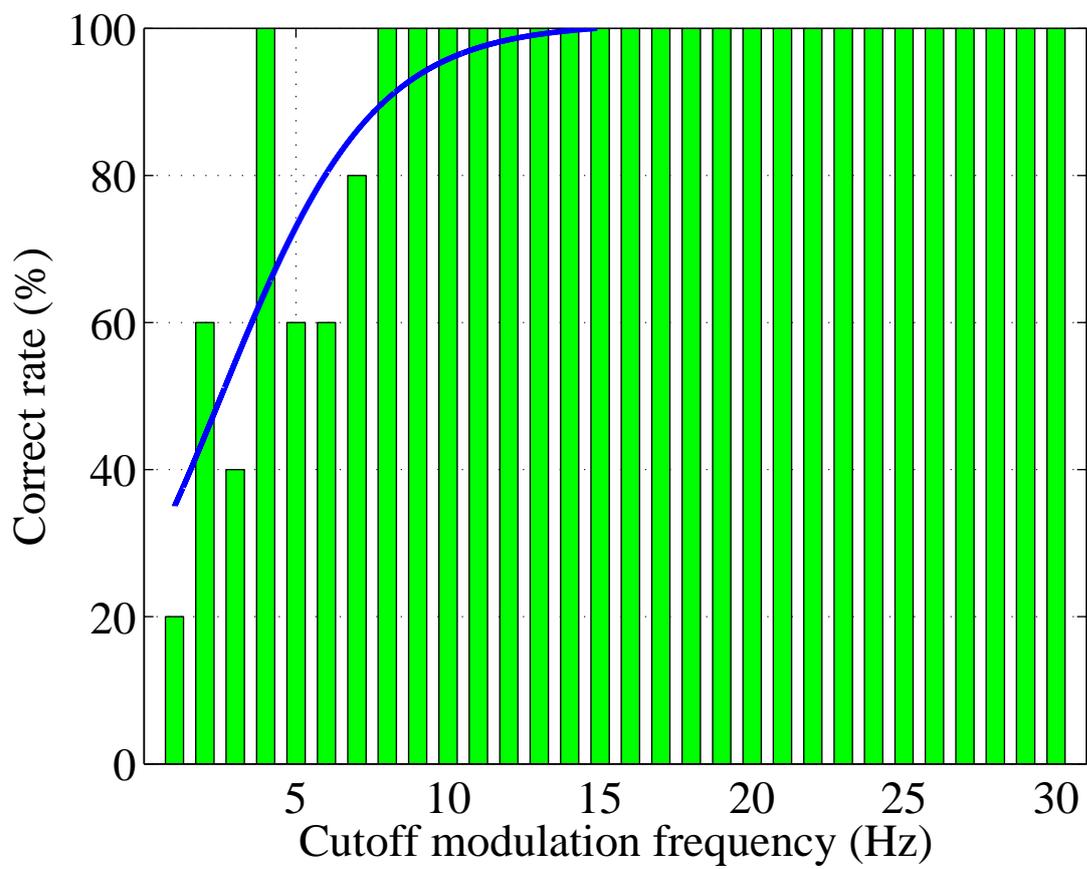


図 6.94: 被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

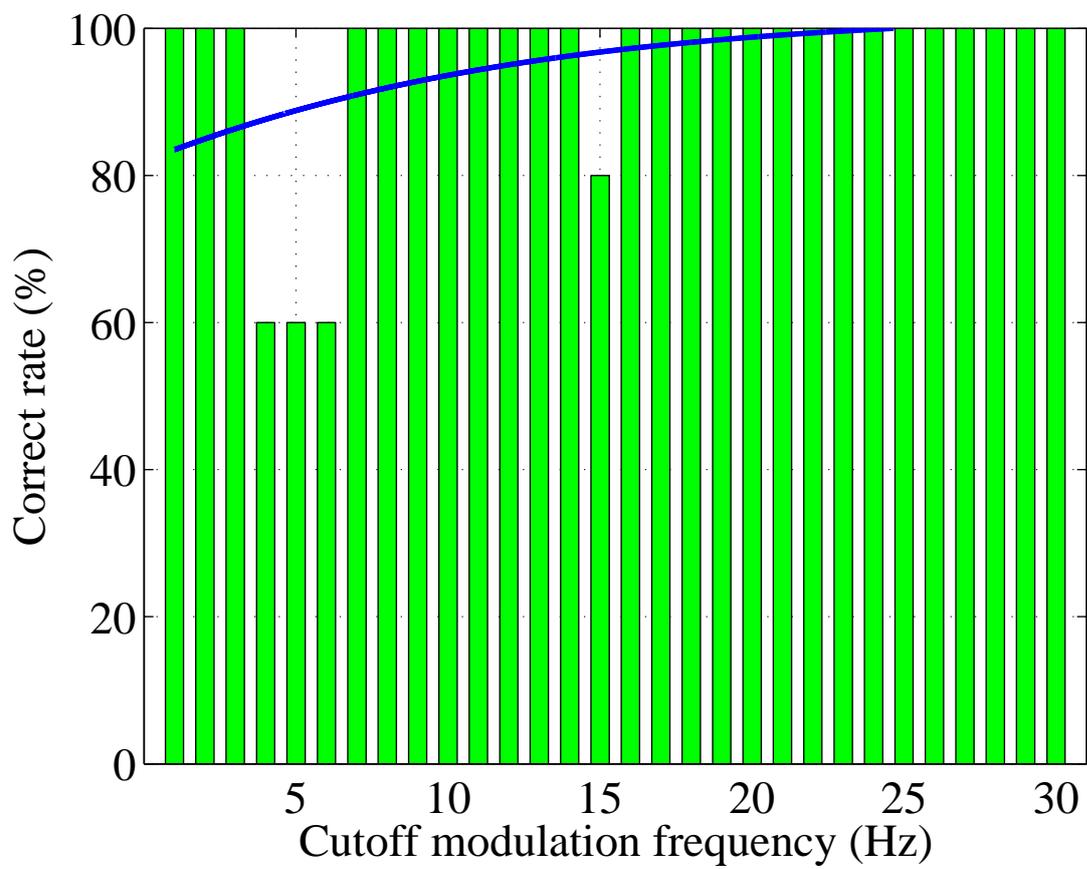


図 6.95: 被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

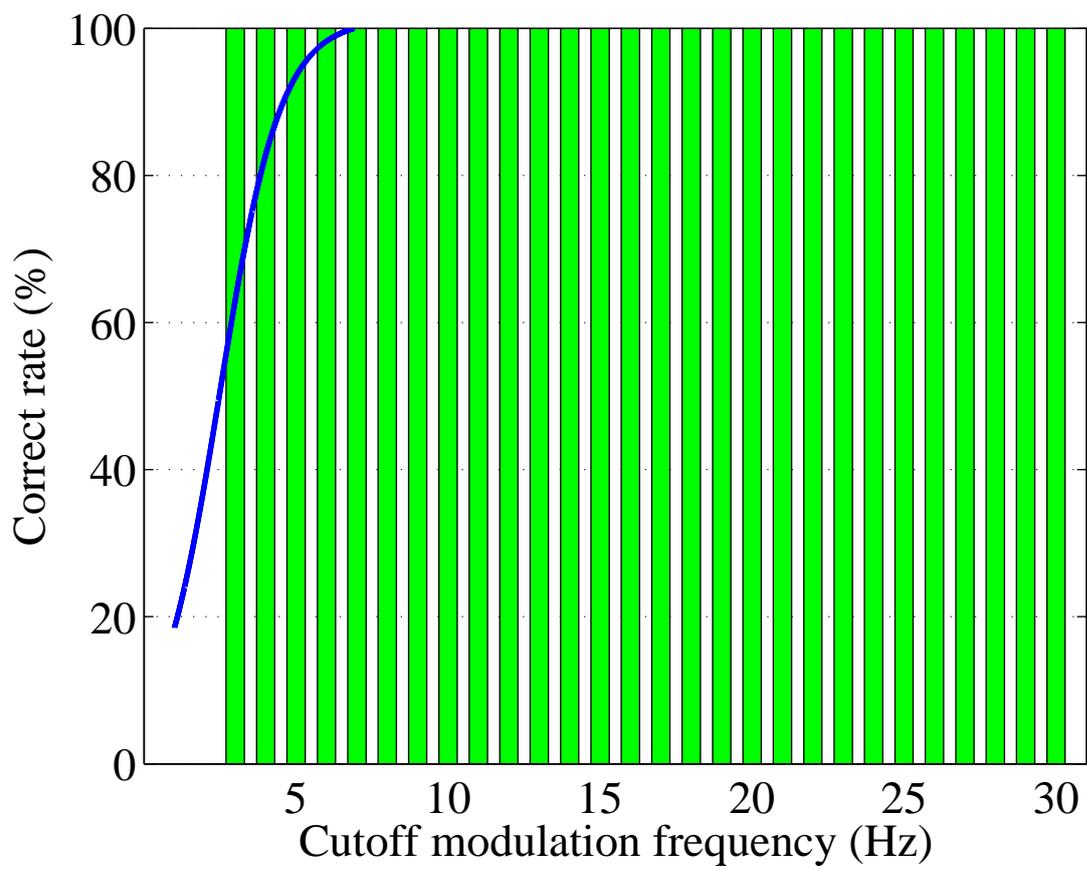


図 6.96: 被験者 A : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

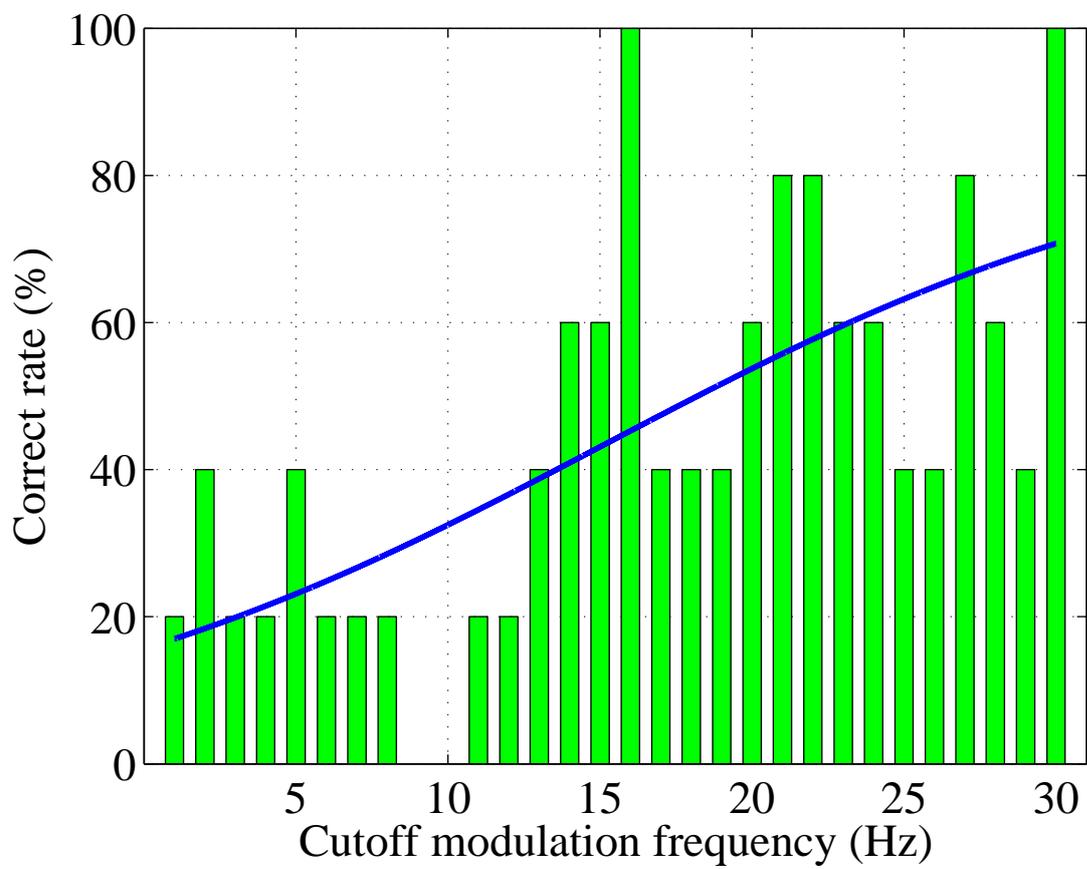


図 6.97: 被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

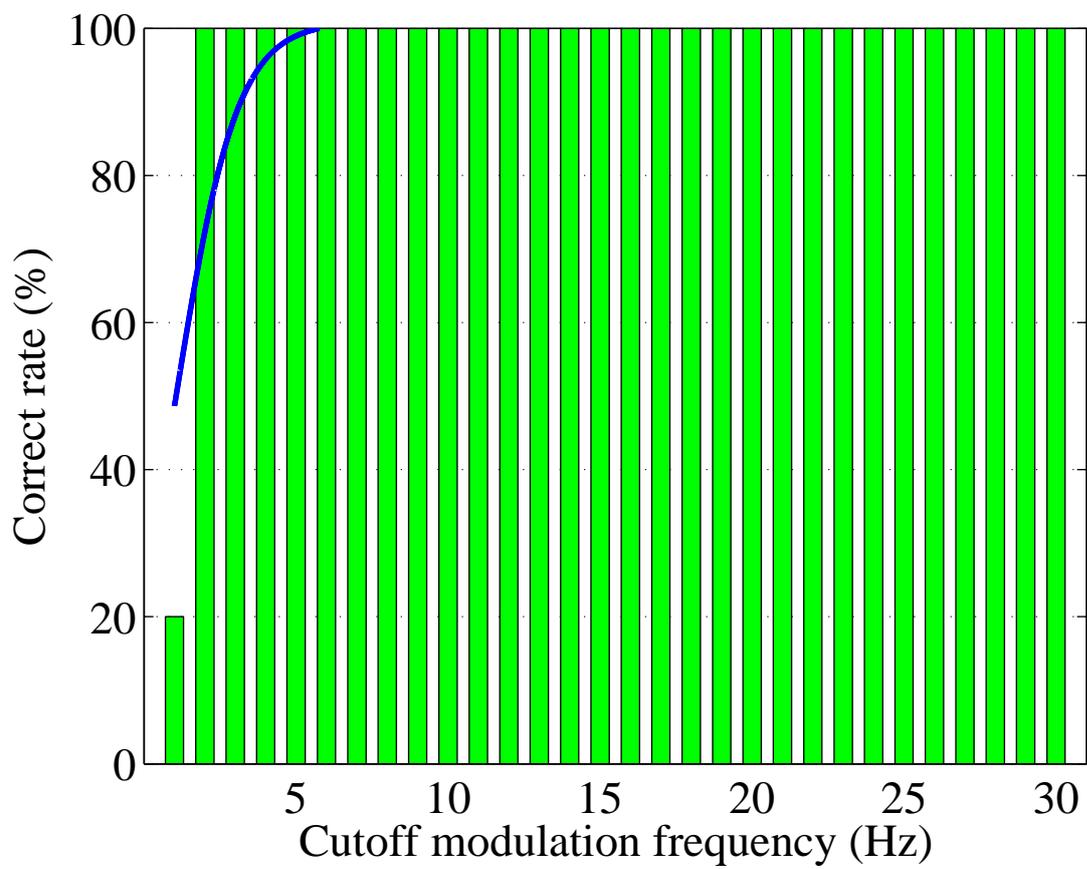


図 6.98: 被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

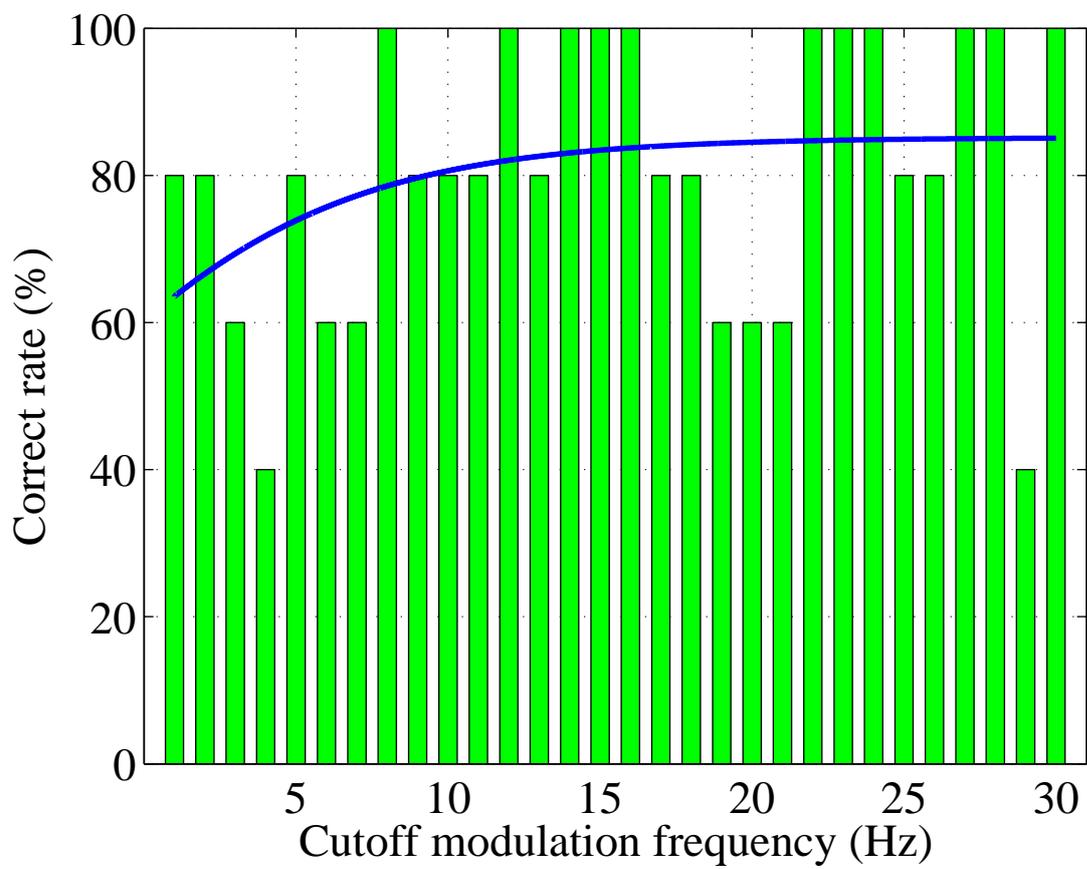


図 6.99: 被験者 B : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

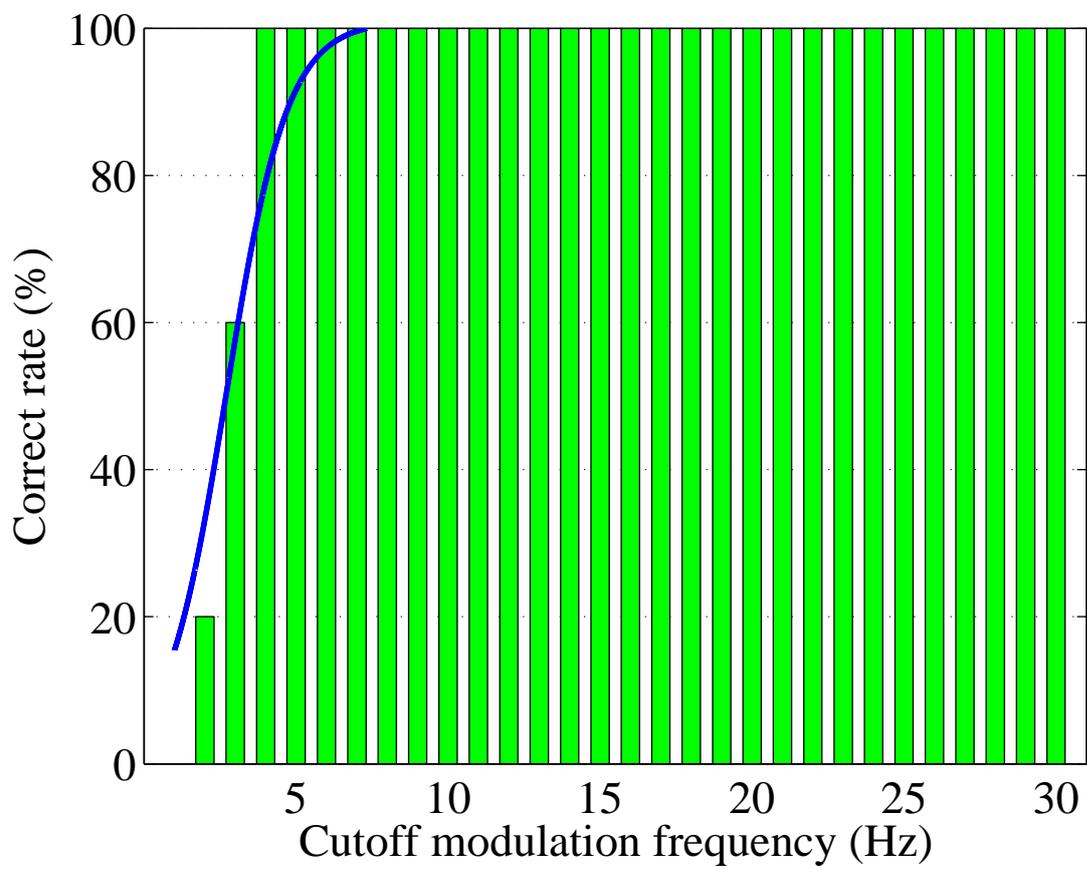


図 6.100: 被験者 B: 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

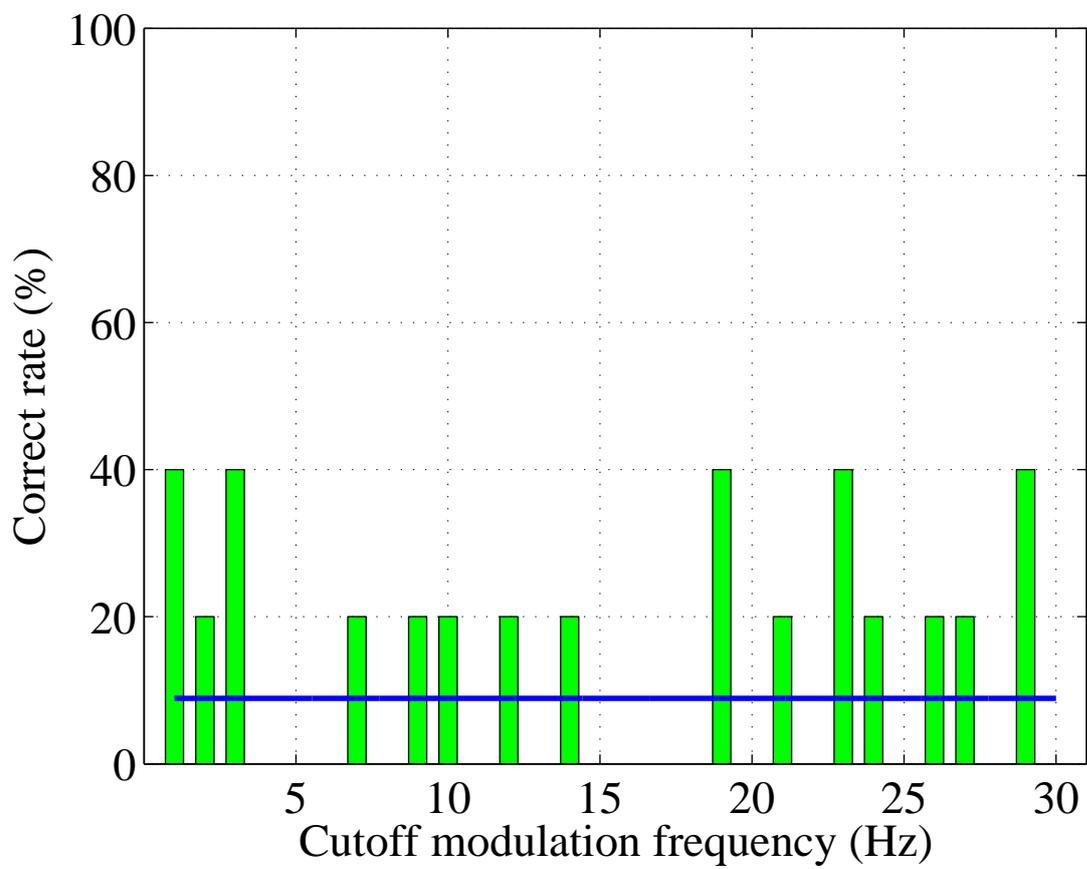


図 6.101: 被験者 C : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

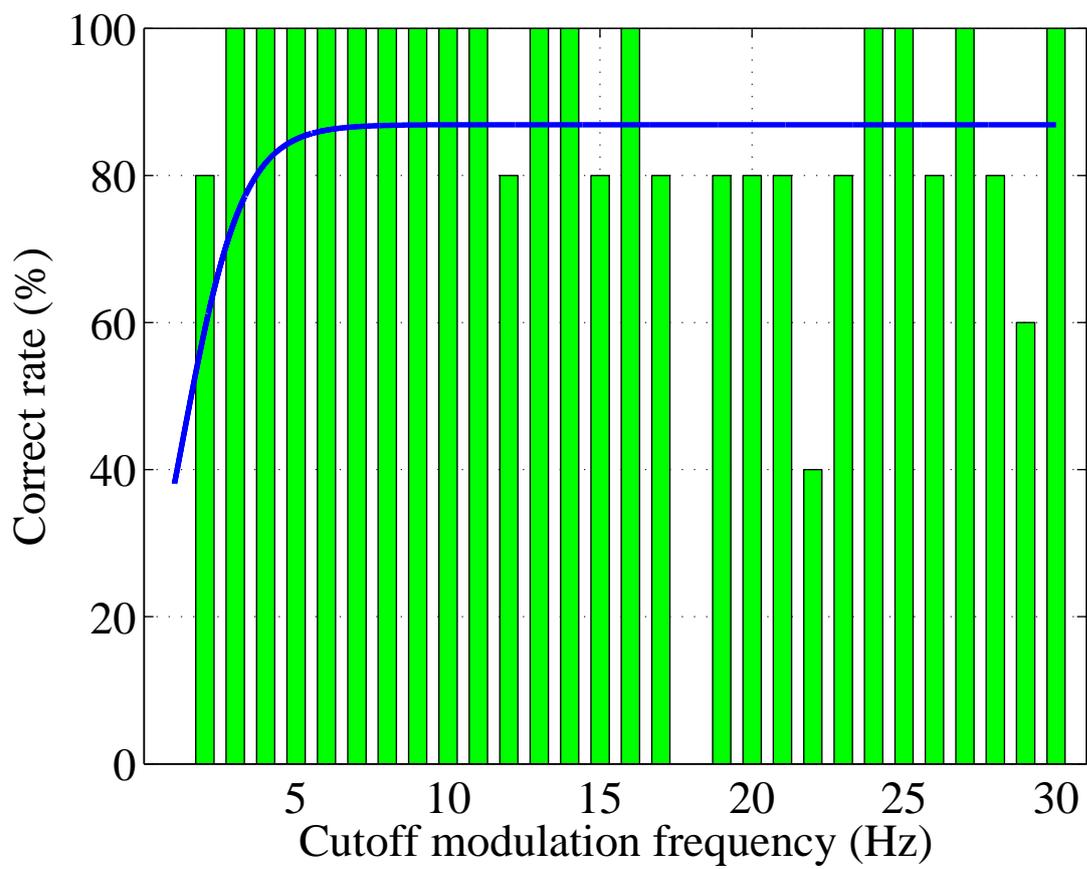


図 6.102: 被験者 C : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

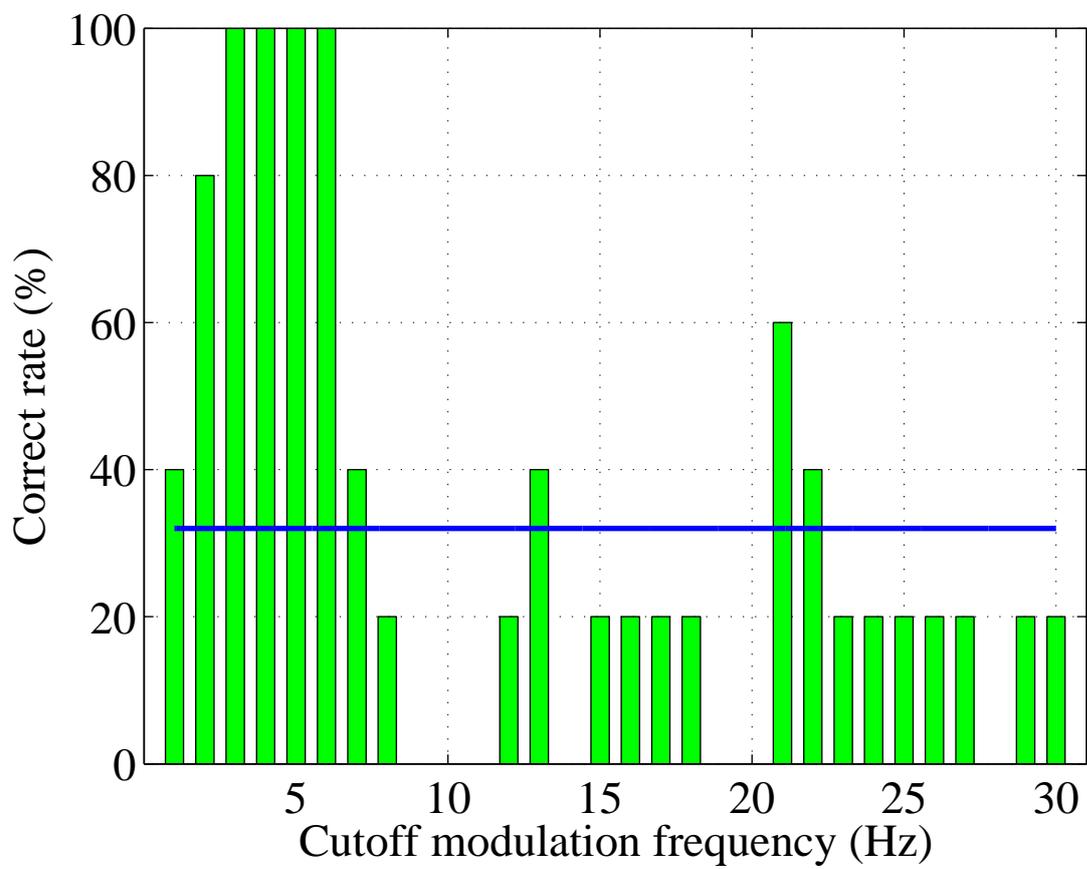


図 6.103: 被験者 C : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

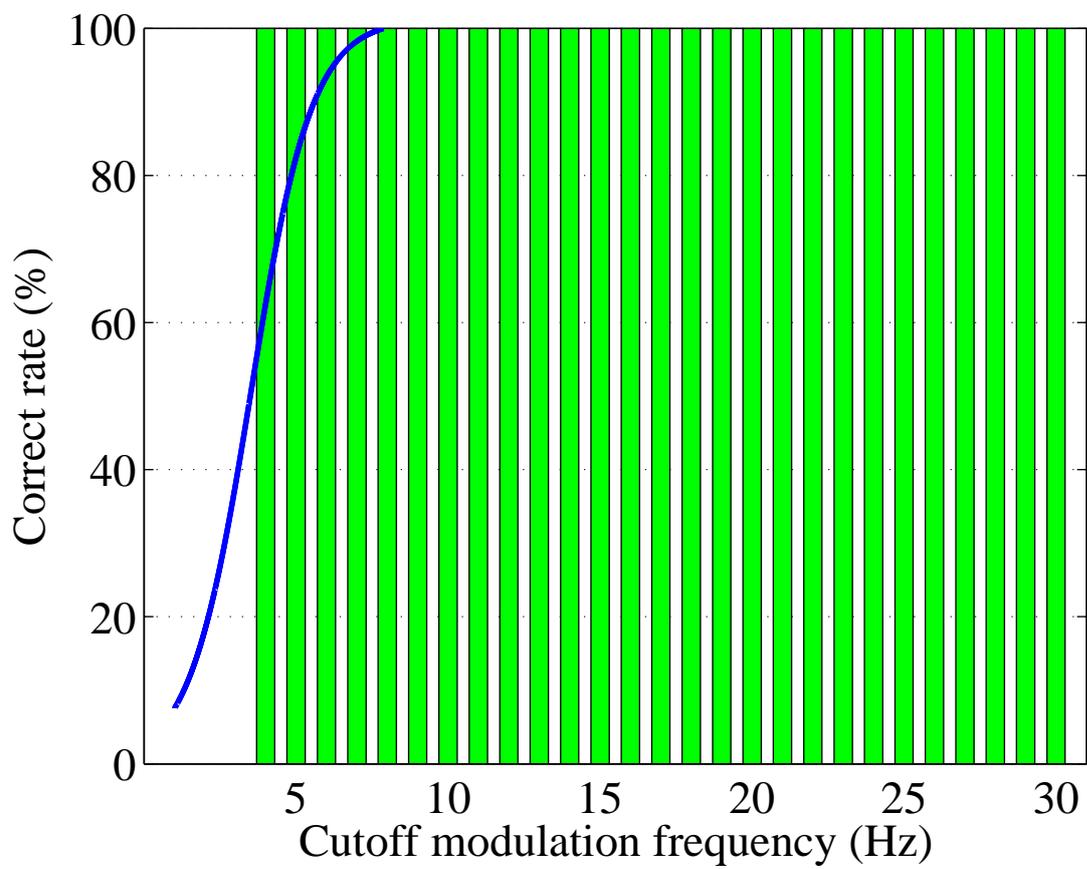


図 6.104: 被験者 C : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

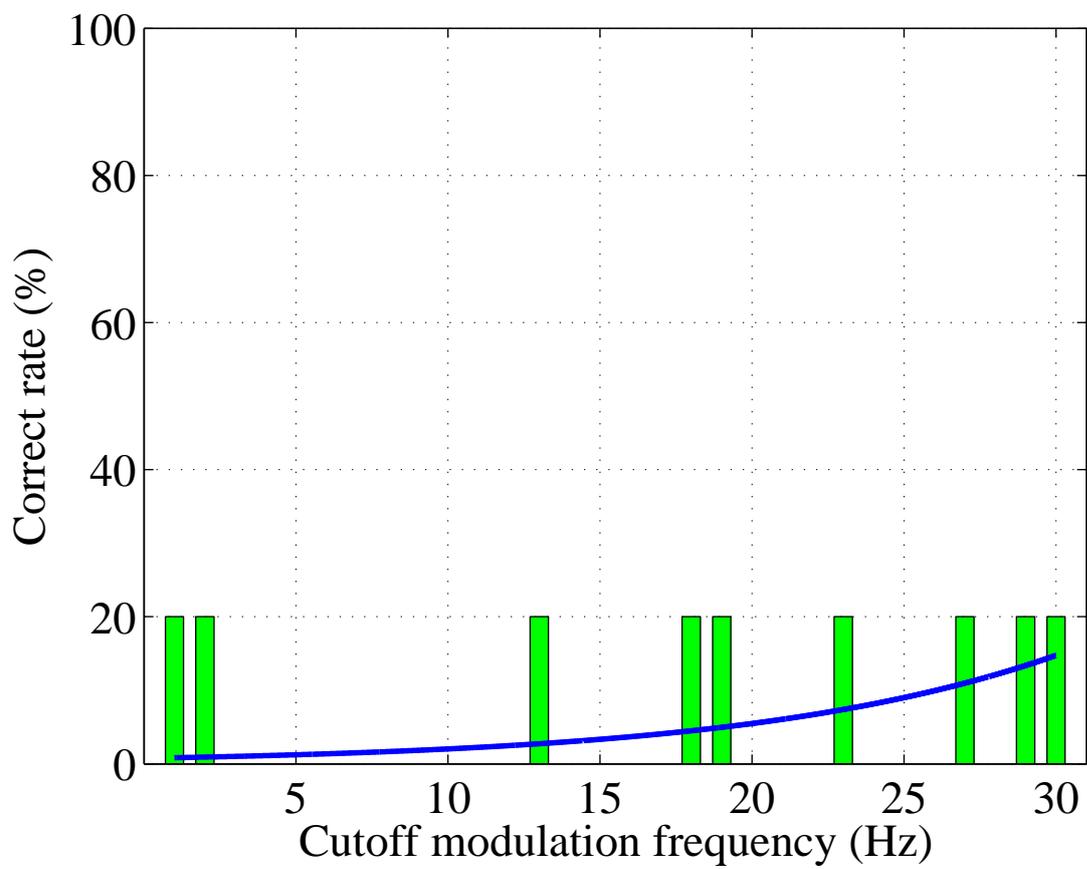


図 6.105: 被験者 D : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

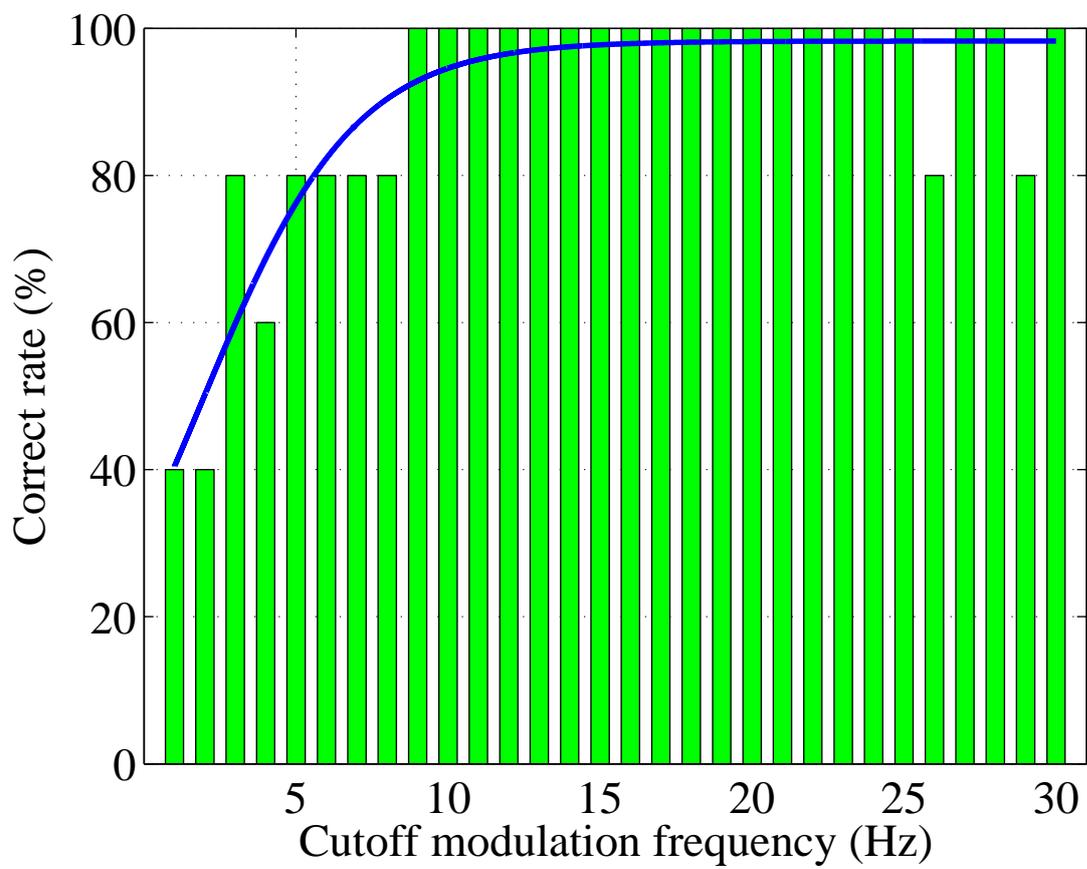


図 6.106: 被験者 D : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

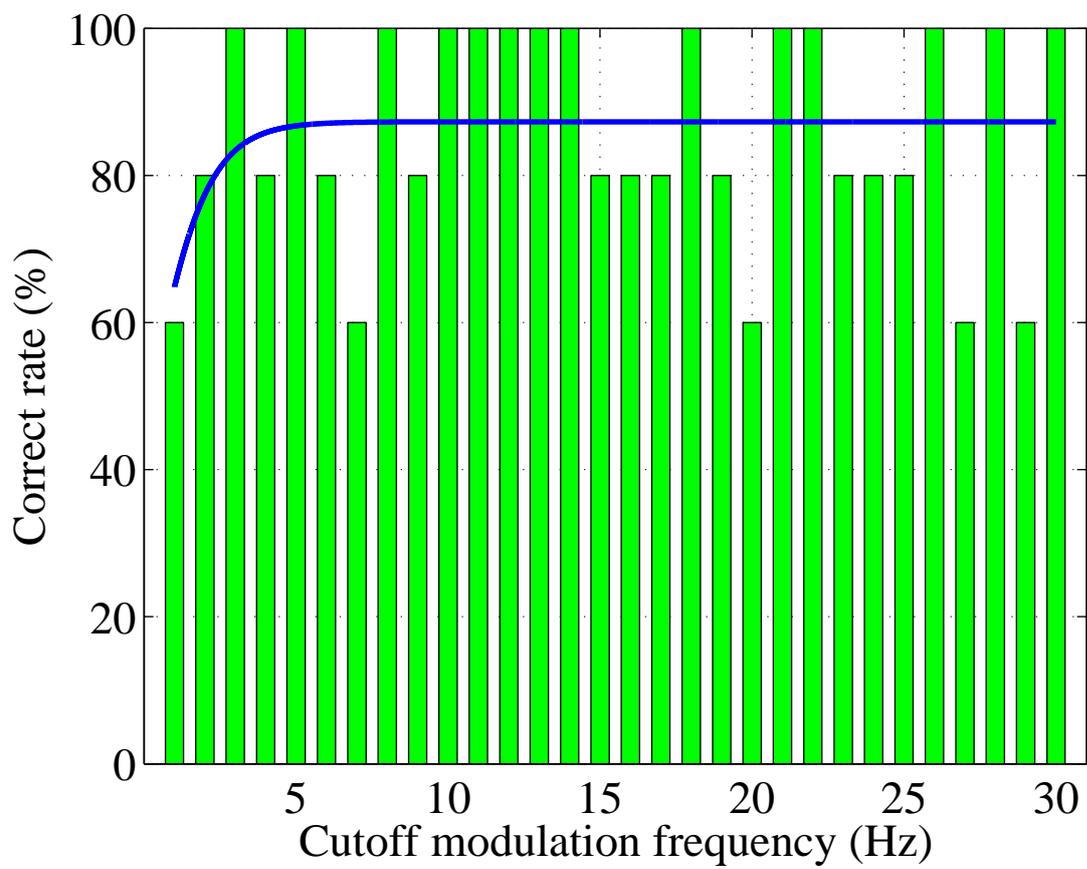


図 6.107: 被験者 D : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

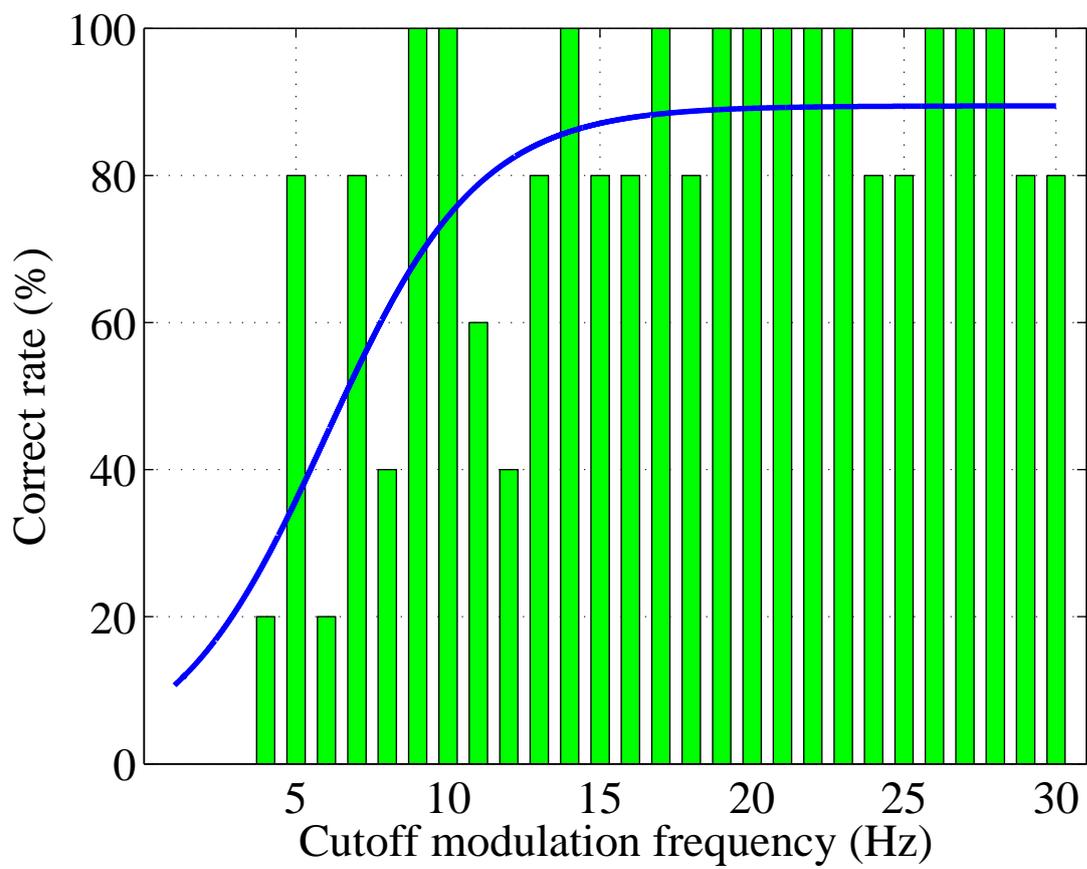


図 6.108: 被験者 D : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

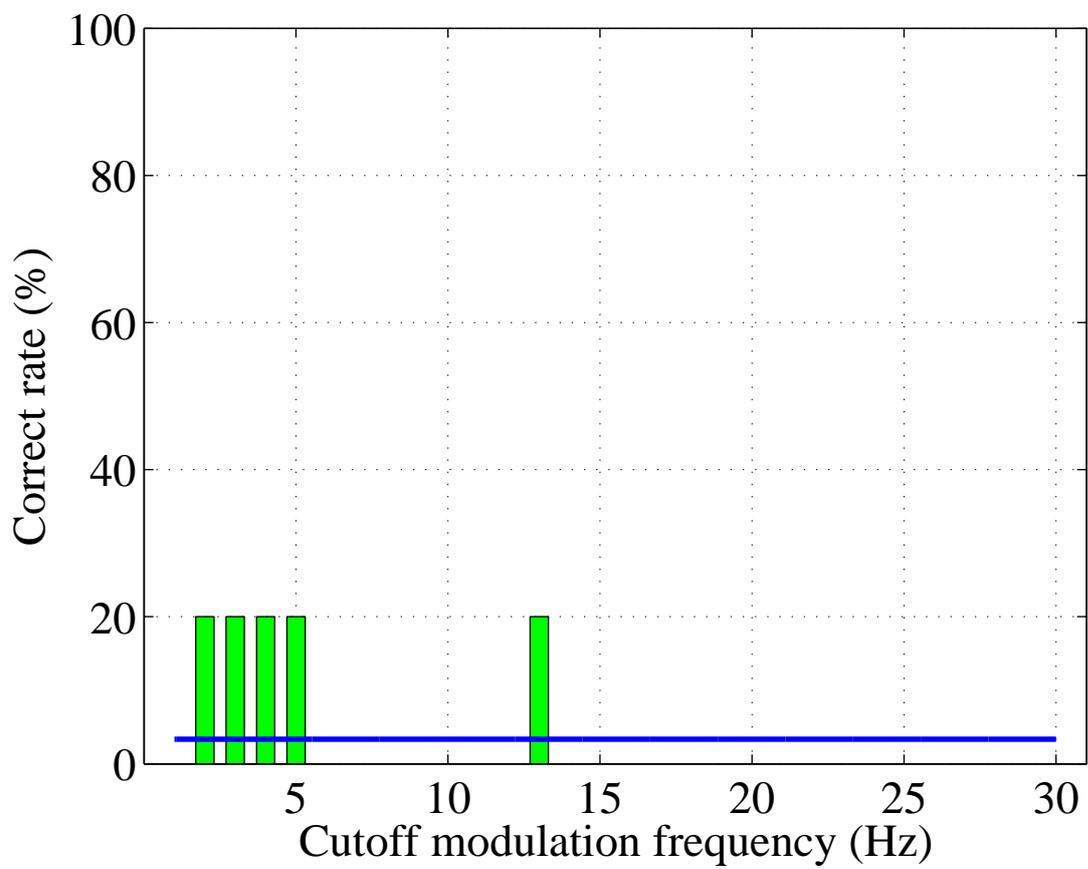


図 6.109: 被験者 E : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

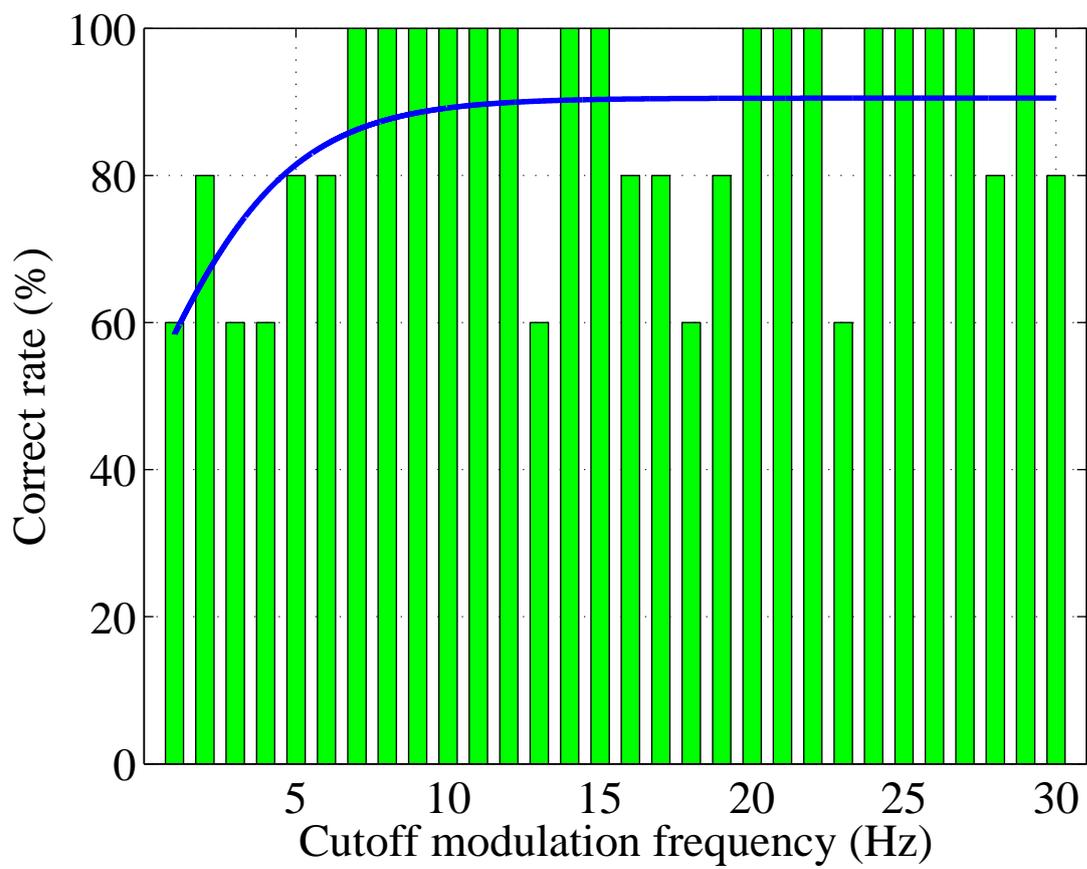


図 6.110: 被験者 E : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

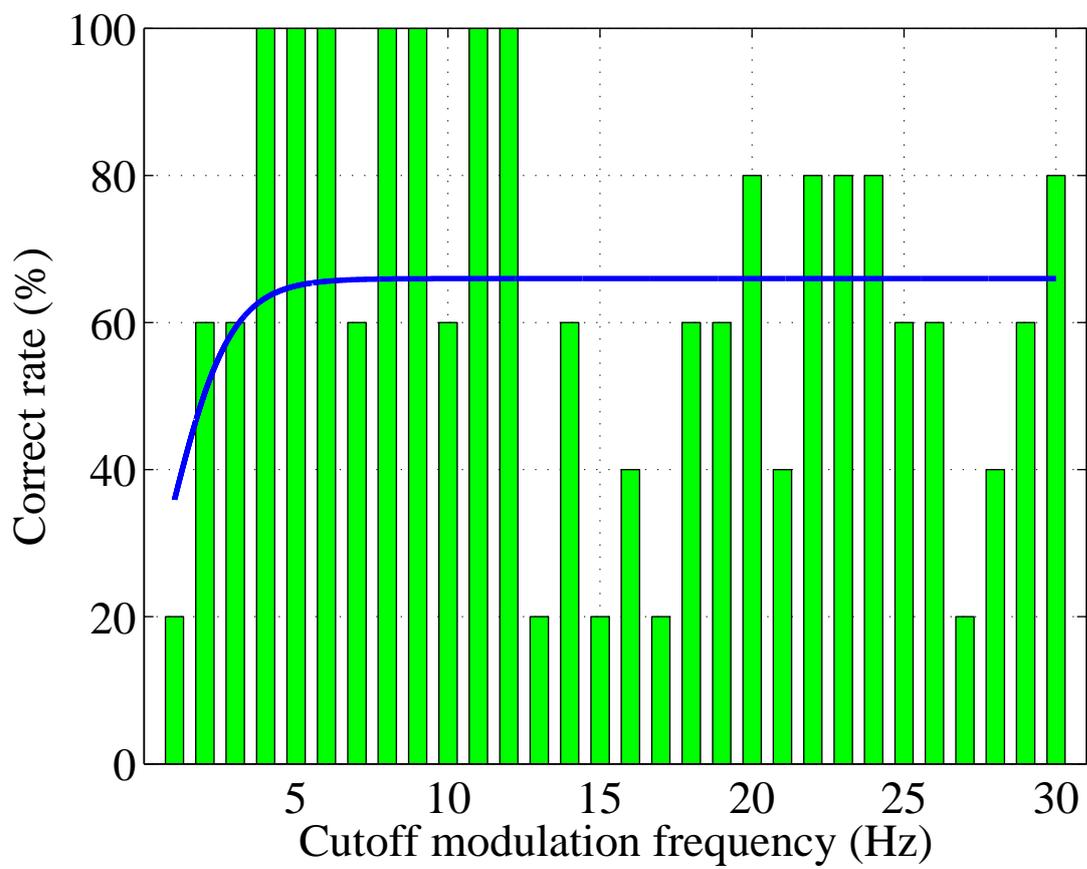


図 6.111: 被験者 E : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

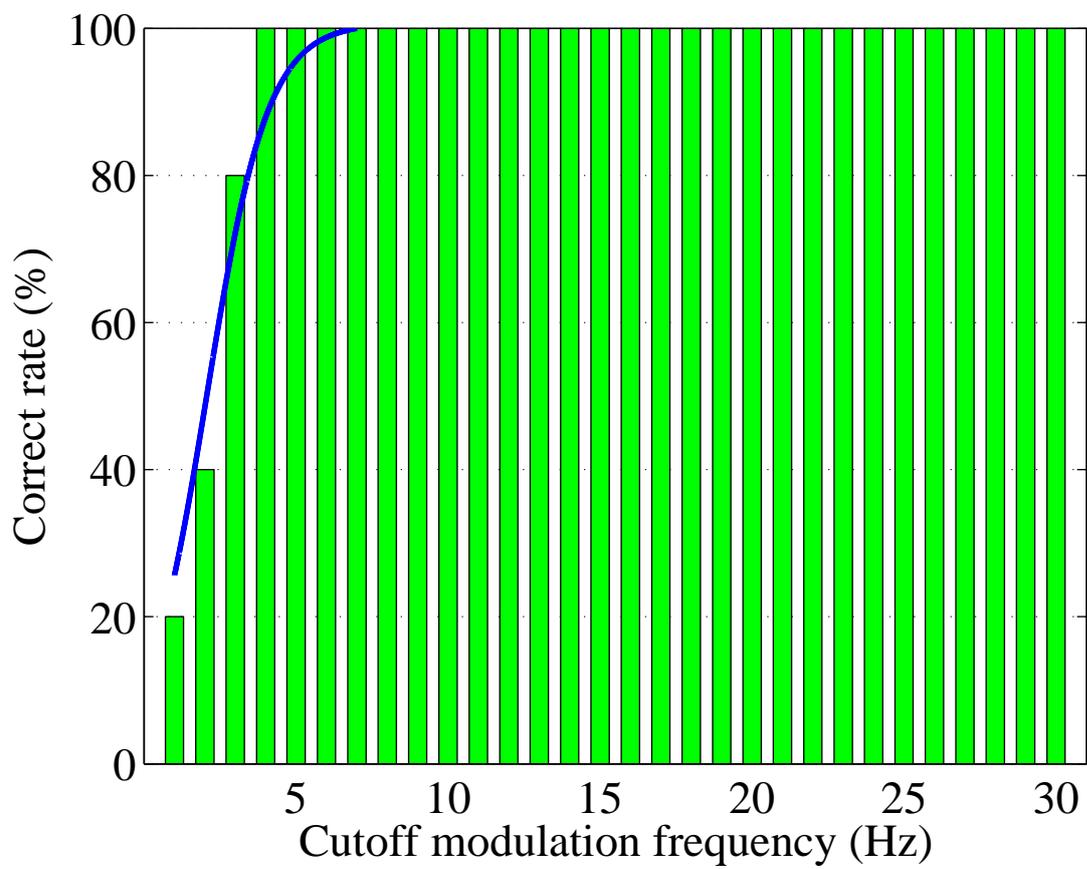


図 6.112: 被験者 E : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

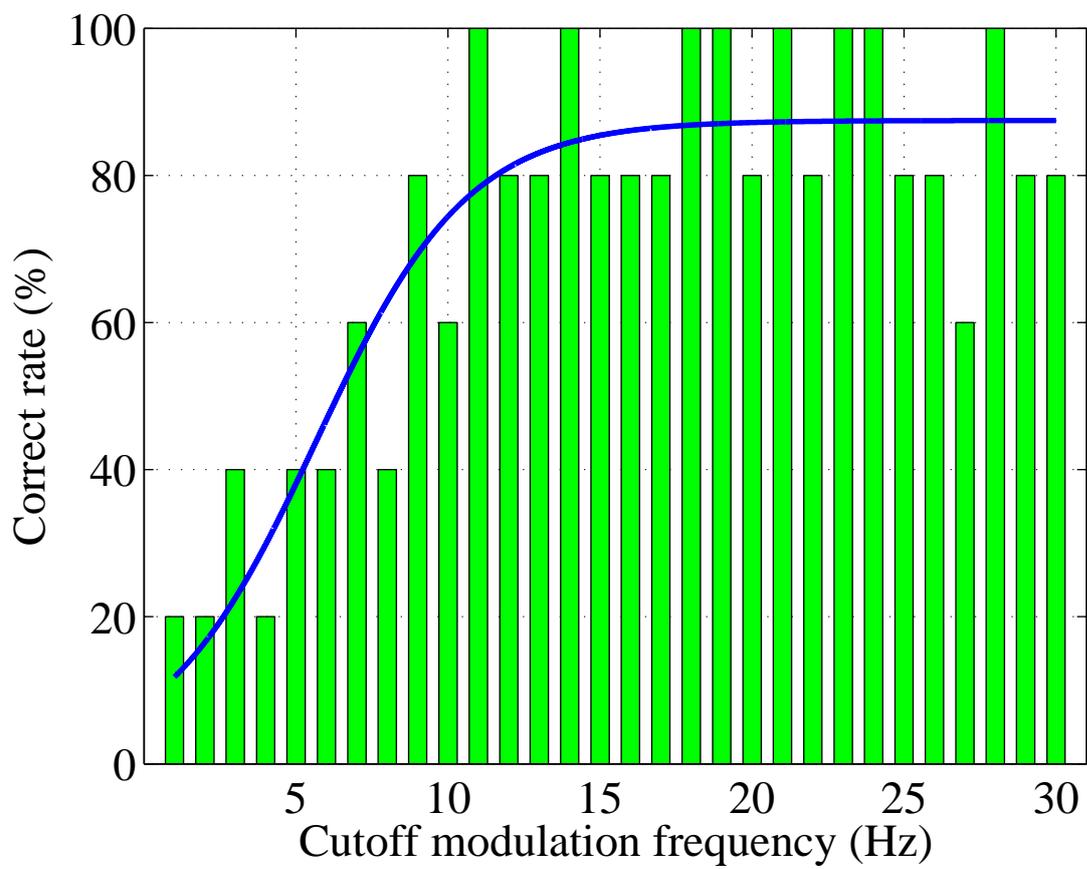


図 6.113: 被験者 F : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

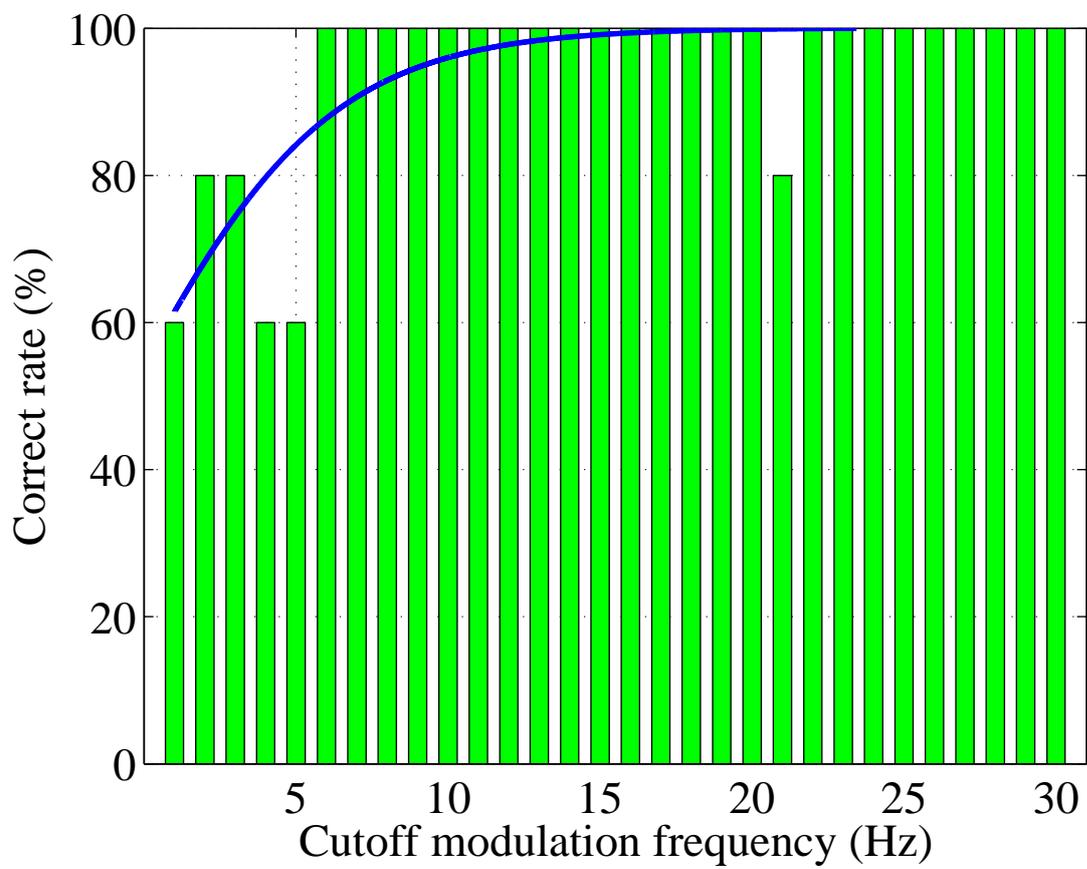


図 6.114: 被験者 F : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

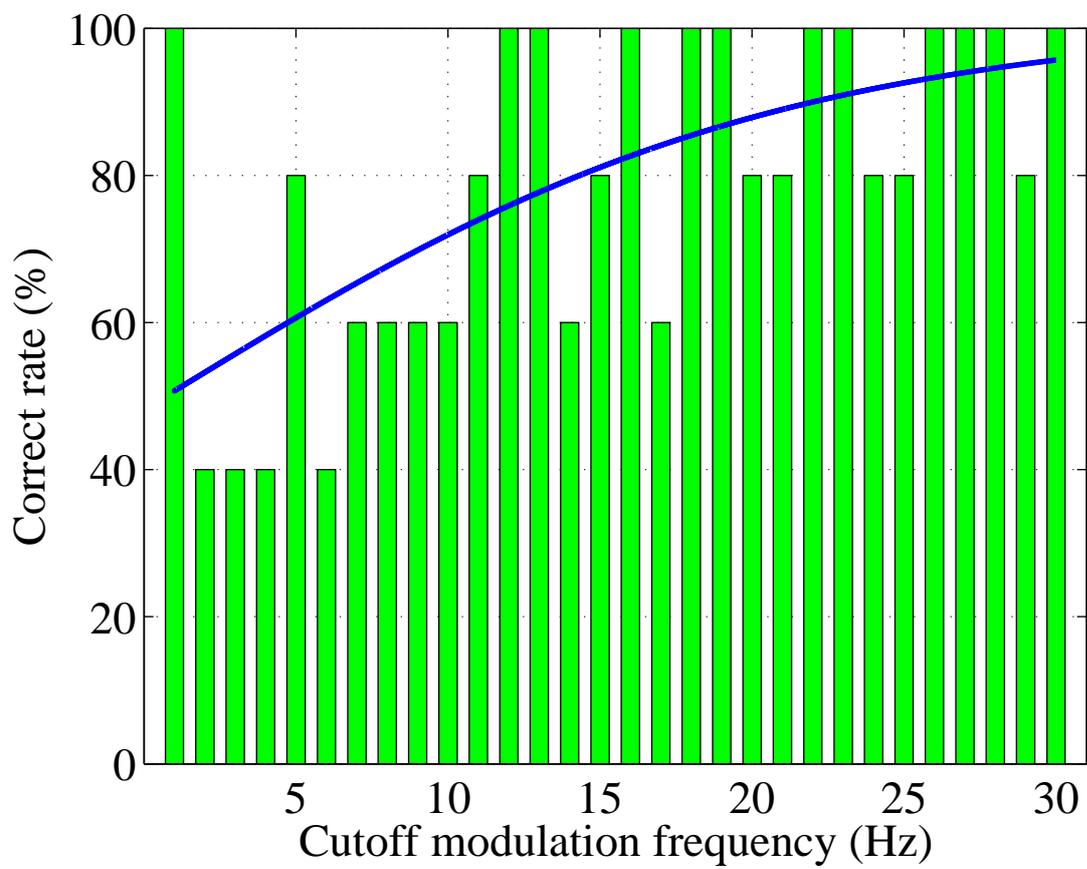


図 6.115: 被験者 F : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

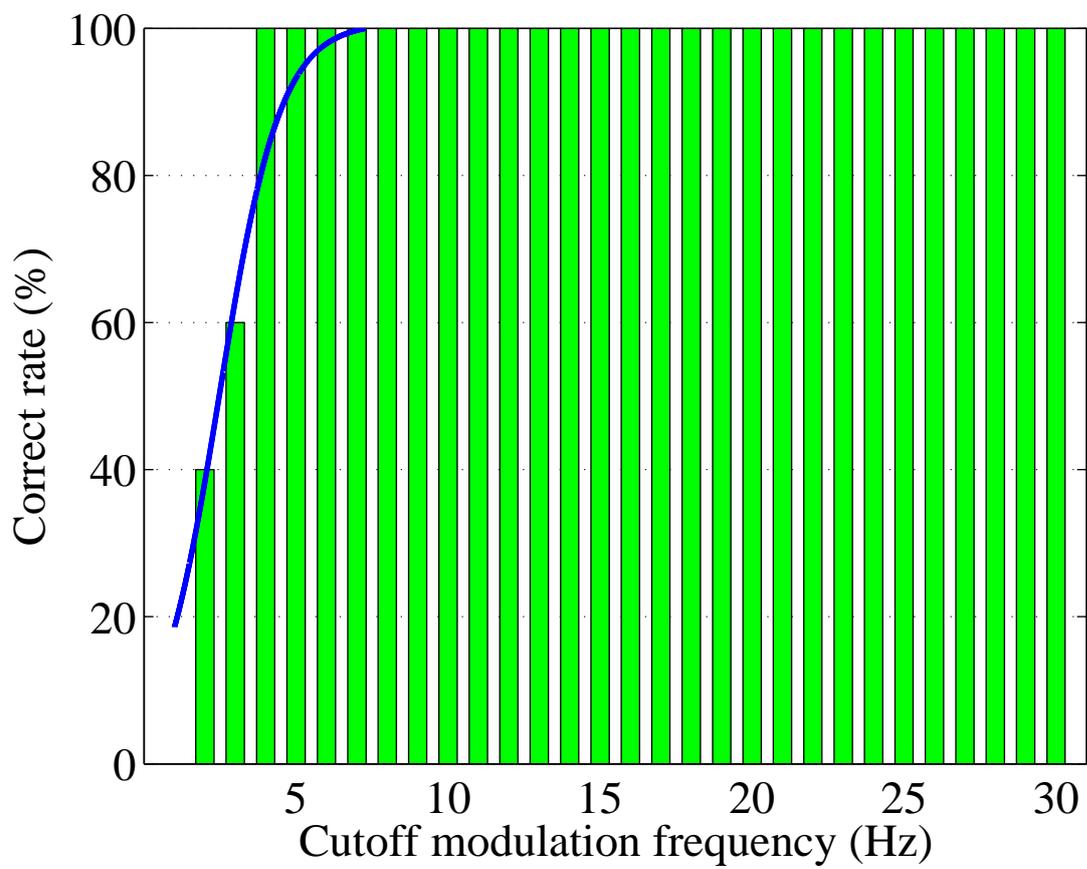


図 6.116: 被験者 F : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

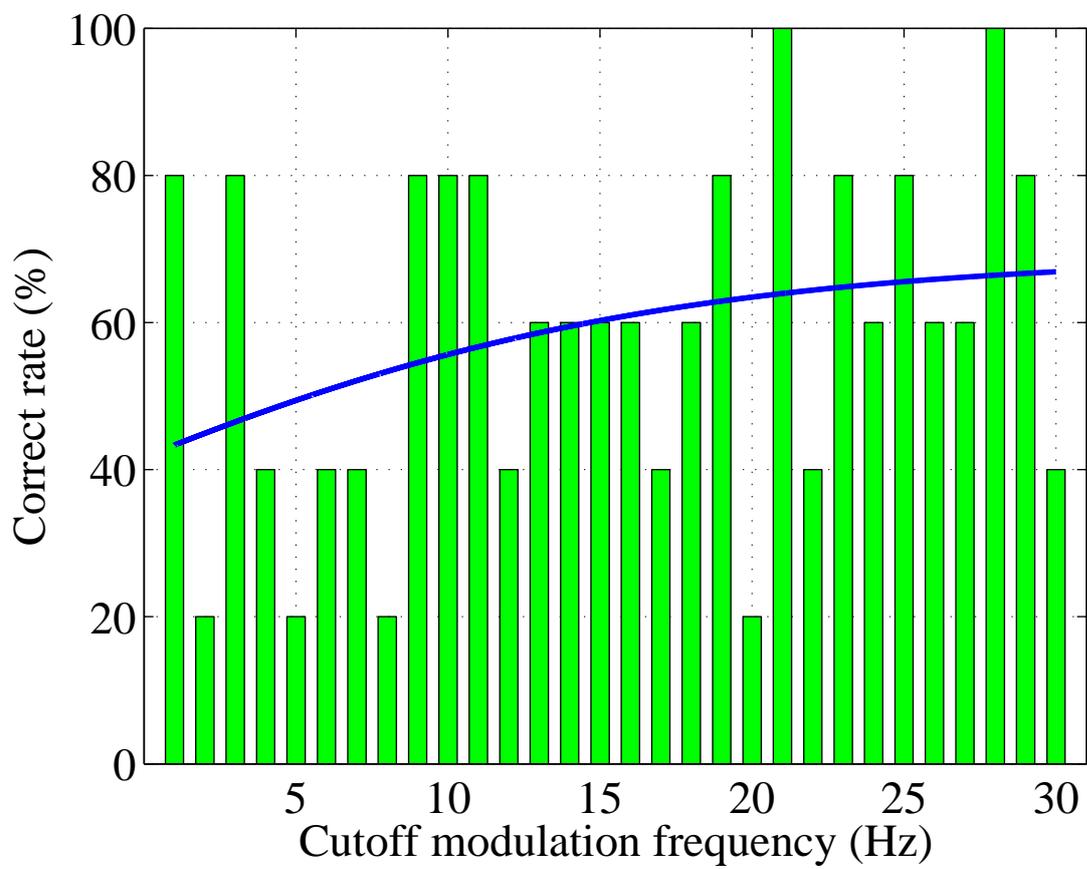


図 6.117: 被験者 G : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

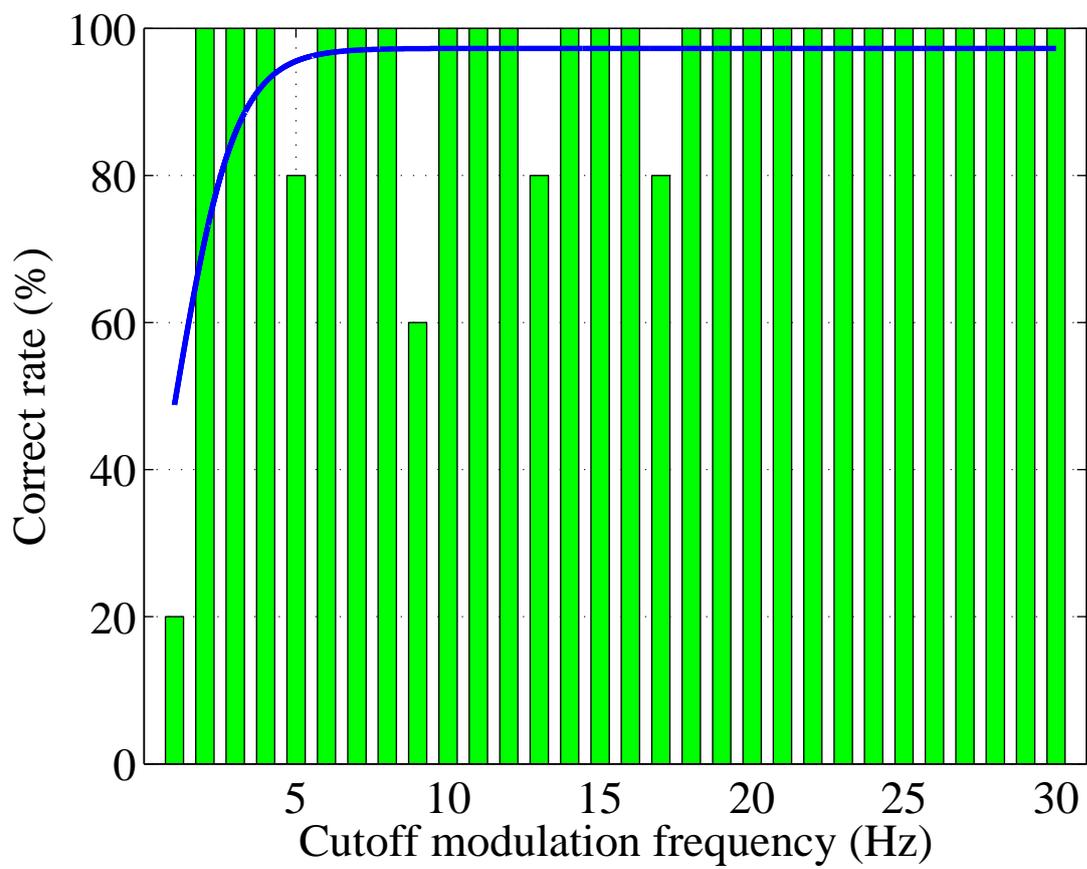


図 6.118: 被験者 G : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

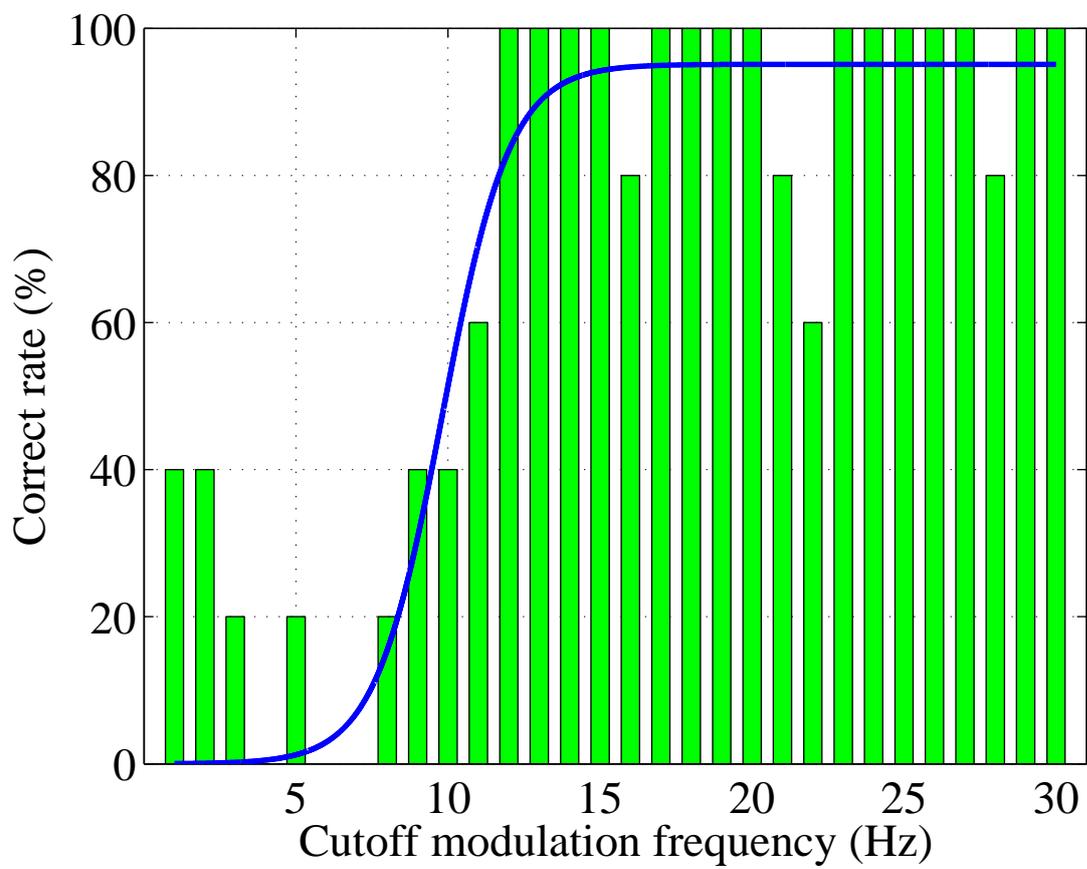


図 6.119: 被験者 G : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

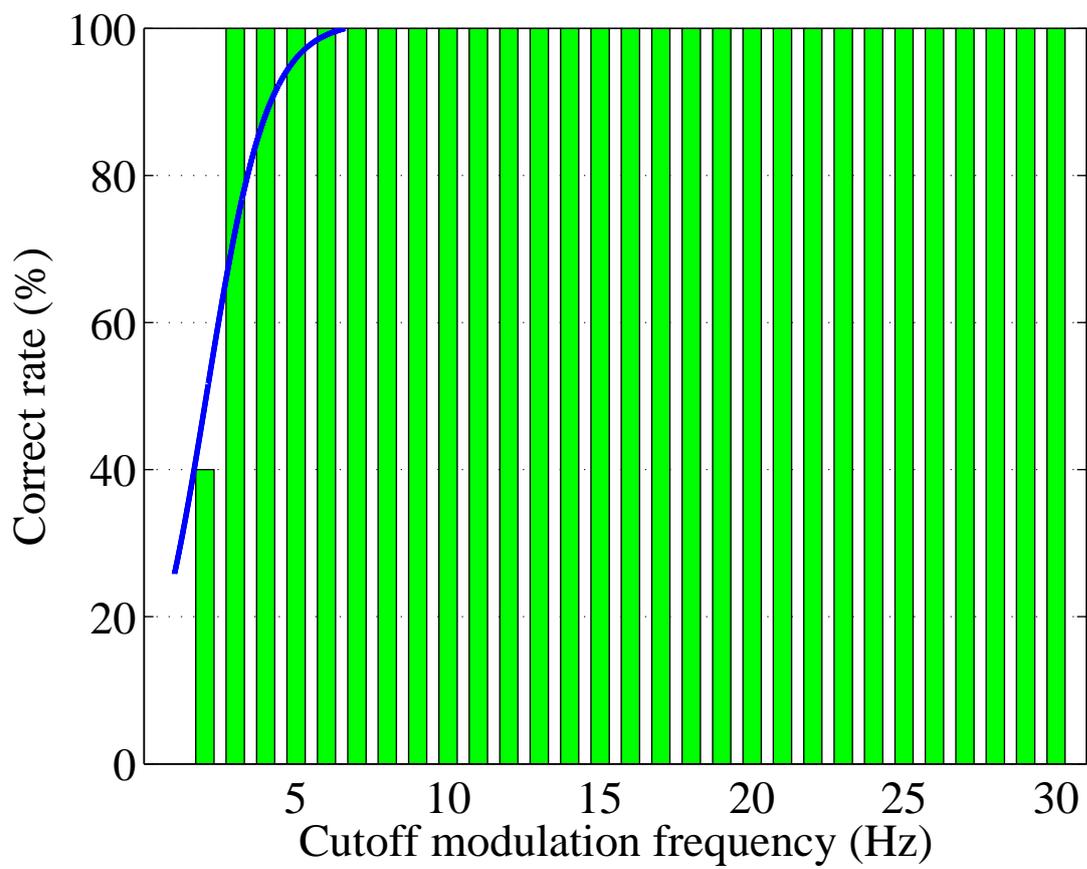


図 6.120: 被験者 G : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

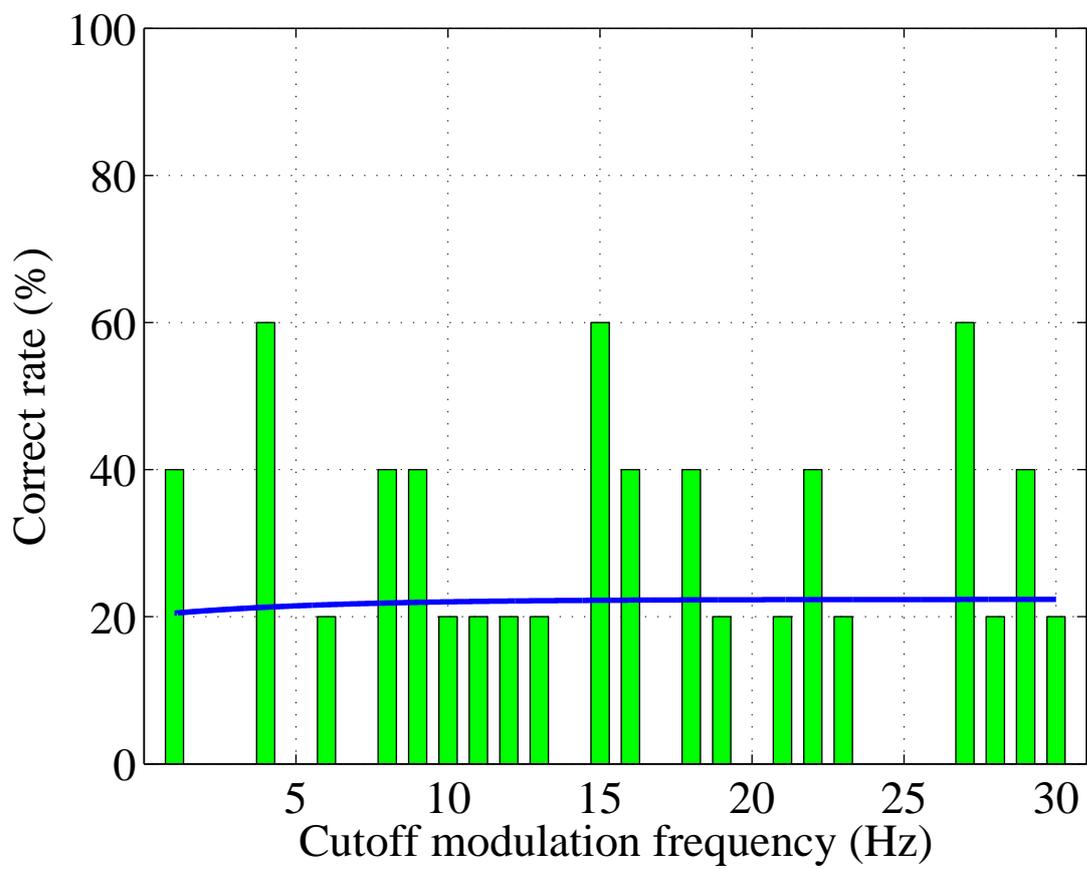


図 6.121: 被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /a/)

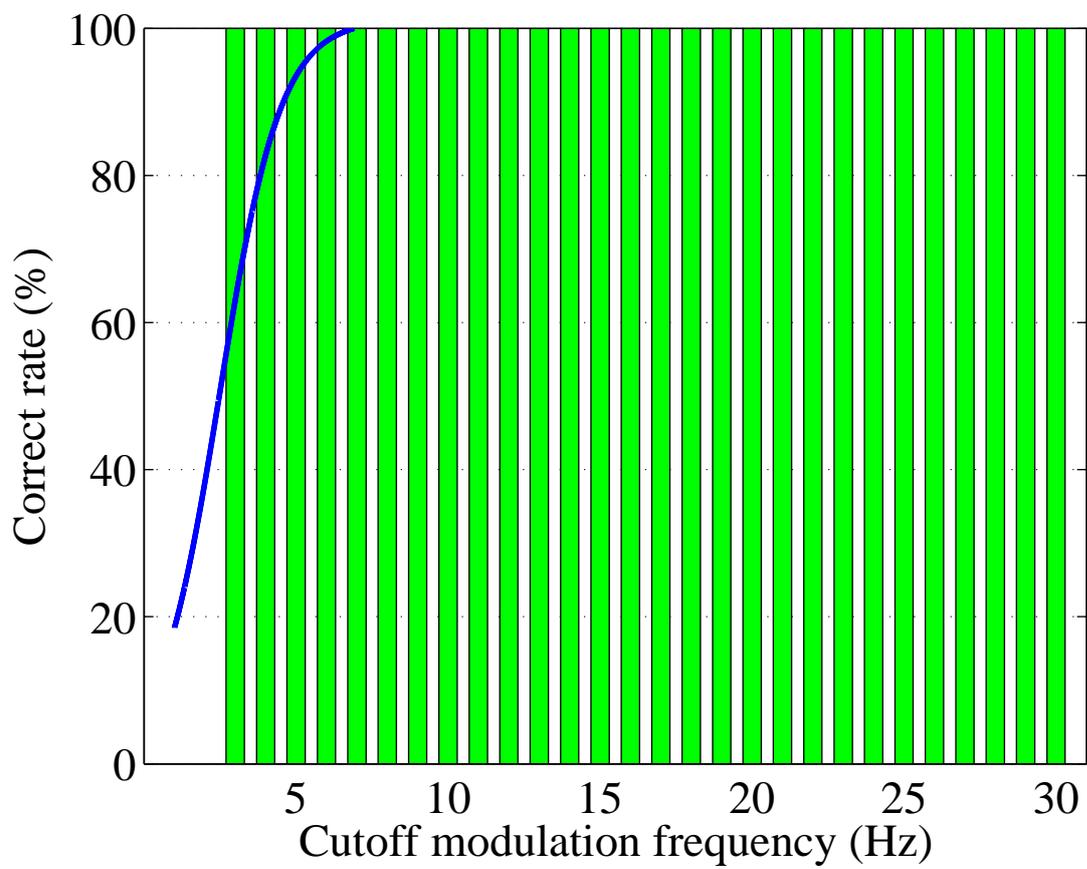


図 6.122: 被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /sa/)

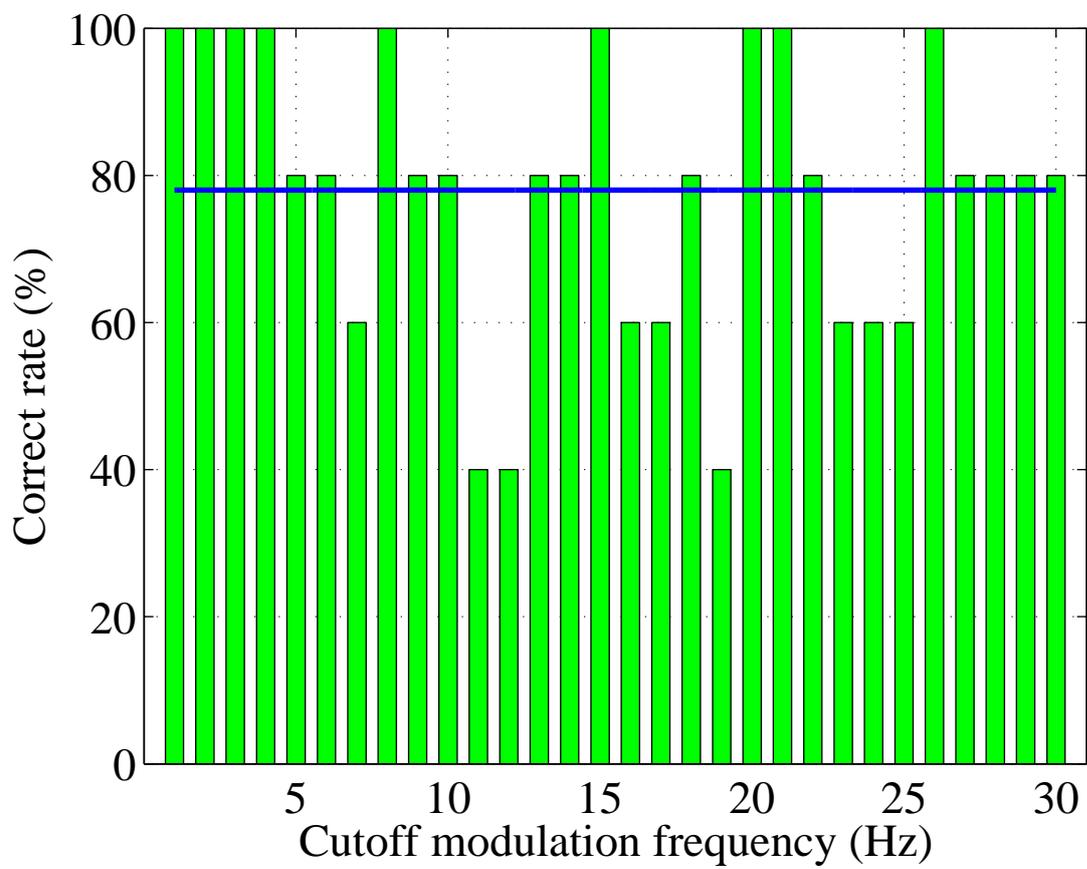


図 6.123: 被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /ha/)

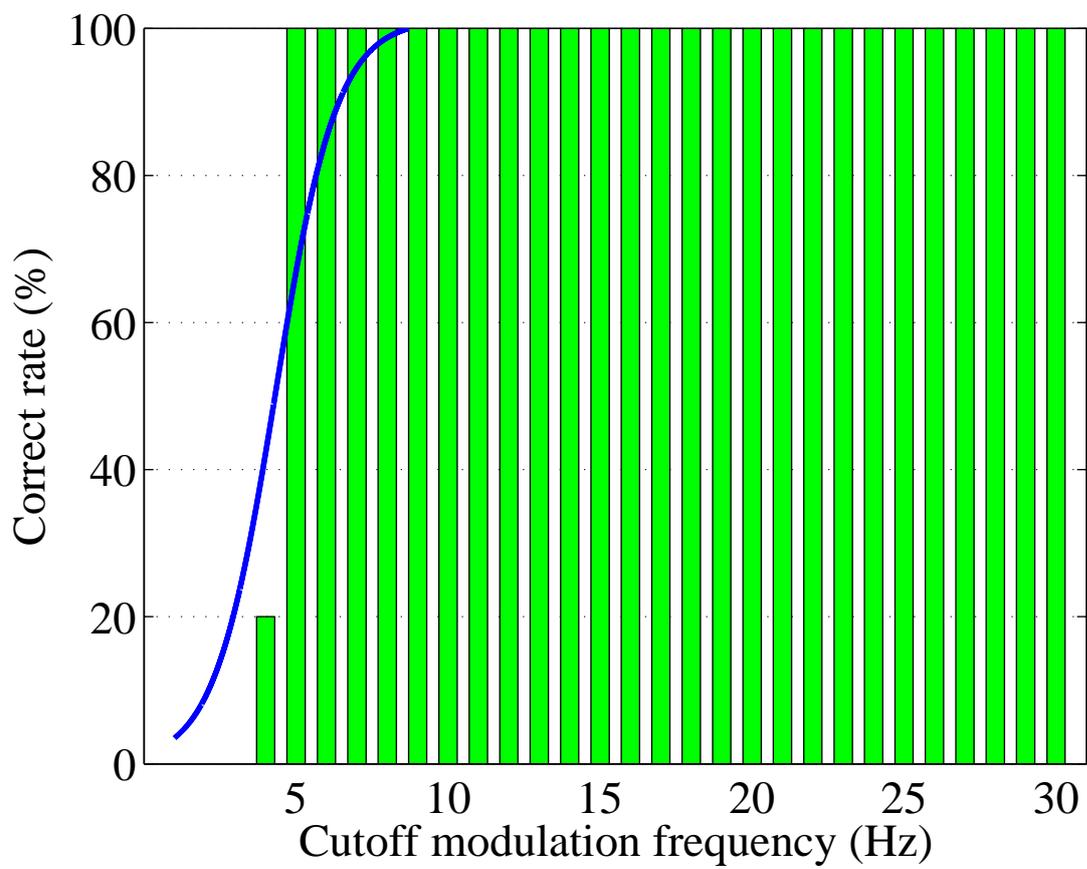


図 6.124: 被験者 H : 低域通過フィルタのカットオフ周波数に対する正答率 (原信号 /za/)

使用したヘッドホン

本研究で使用したヘッドホン , SENNHEISER HDA 200 の周波数特性は下記の表に示す .

表 6.1: ヘッドホンの周波数特性

Frequency [Hz]	Response [dB] right	Response [dB] left
125	73.2	72.7
177	79.7	80.4
250	83.7	85.0
354	85.9	87.3
500	87.8	89.0
707	88.2	89.7
1000	90.0	90.8
1414	90.9	90.5
2000	95.7	95.0
2828	100.6	101.2
4000	92.2	91.7
5657	86.5	85.7
8000	85.7	86.2

学会発表リスト

1. 西野恭生, 宮内良太, 鷓木祐史, “音声の振幅包絡の変調成分が言語情報の取得に与える影響,” 平成 25 年度電気関係学会北陸支部連合大会, September 2013.
2. 西野恭生, 宮内良太, 鷓木祐史, “音声の各周波数帯域の振幅包絡に含まれる言語情報,” 日本音響学会聴覚研究会資料, 43(7), 547–552, October 2013.
3. 西野恭生, 宮内良太, 鷓木祐史, “日本語音声の子音知覚は振幅包絡の変調成分の制御により変化するのか?,” 日本音響学会聴覚研究会資料, 44(1), 53–58, February 2014.
4. Nishino, Yasutaka., Miyauchi, Ryota., and Unoki, Masashi., “Study on Linguistic Information Contained in Temporal Amplitude Envelope of Japanese Speech Signals,” 2014 RISP International Workshop on Nonlinear Circuits and Signal Processing (NCSP'14), March 2014.
5. 西野恭生, 宮内良太, 鷓木祐史, “振幅包絡の変調成分の制御が日本語音声の子音知覚に与える影響の検討,” 日本音響学会 2014 年 春季研究会発表会, March 2014.