

システム基礎科学実験Ⅰ

「スペクトル分析」

実験実施日:平成16年7月27・29日

グループ: F

学生証番号: 40413

氏名: 諸町大地

共同実験者: 本間大士・三瓶雅迪

目的

音声を例にとって計算機によるスペクトル分析の手法とその物理的意味を学ぶ。

実験

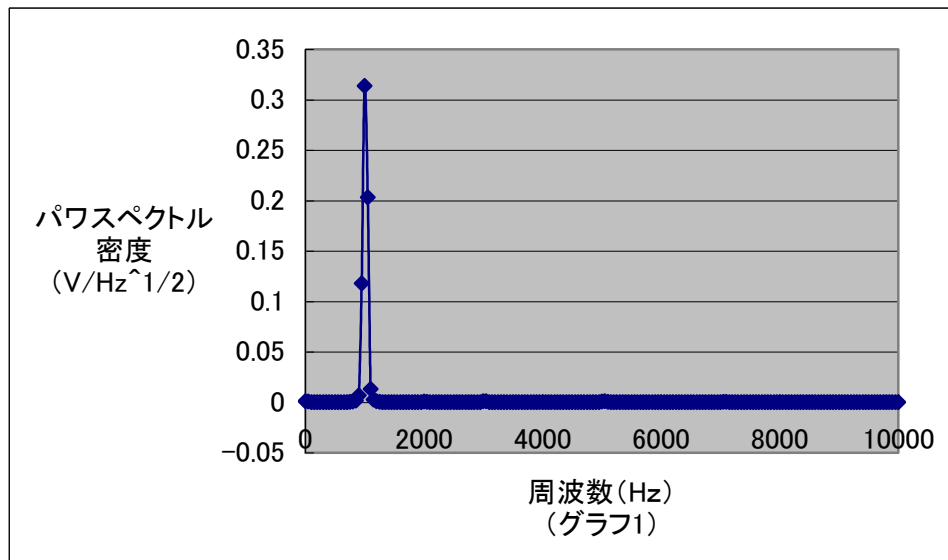
1-i

(方法)

FFT アナライザにファンクションジェネレーター(以下 FG.)から 1[kHz]の正弦波を入力し、パワスペクトルを求めた。ただし、 $N=512$ 、 $V_{pp}=4V$ 、時間窓はハニング窓とする。

(結果)

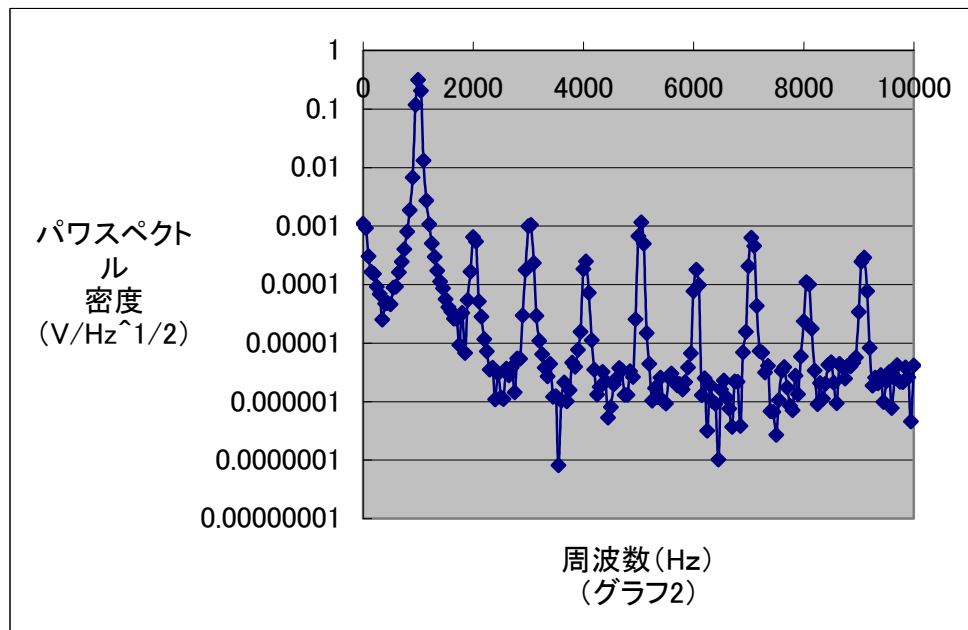
得られたスペクトラムをグラフ 1 に、このグラフのピーク付近での値を表 1 に示す。



周波数(Hz)	950	1000	1050
パワスペクトル密度(V/Hz ^{1/2})	0.11795	0.313727	0.203466
パワー	0.834032	2.218385	1.438722

(表1)

さらに、パワスペクトル密度を対数でプロットしなおしたものをグラフ 2 に示す。



このグラフでのピーク値を表 2 に示す。

周波数(Hz)	パワスペクトル密度(V/Hz ^{1/2})
1000	3.13727E-01
2000	6.45900E-04
3050	1.04000E-03
4050	2.46628E-04
5050	1.15500E-03
6050	1.78204E-04
7050	6.26770E-04
8050	1.08775E-04
9100	2.89909E-04

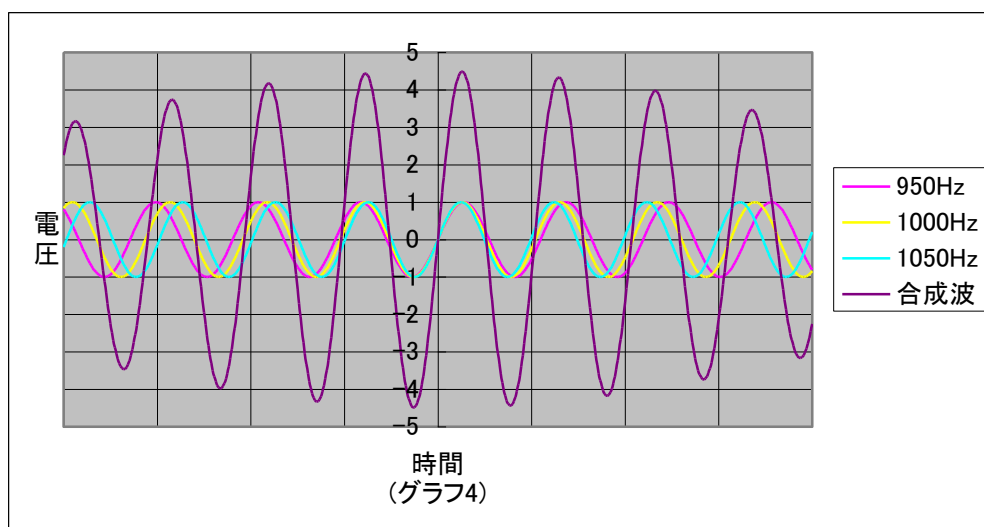
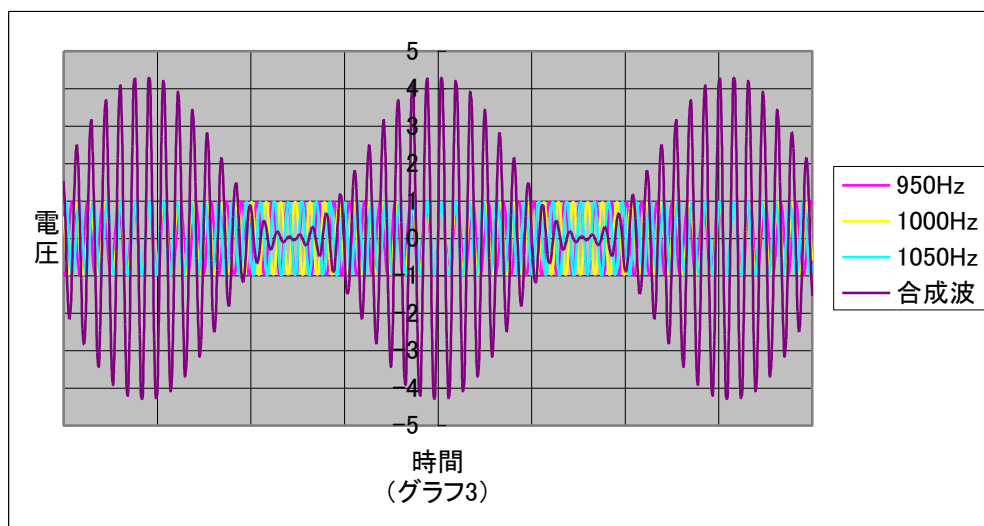
(表2)

(考察)

グラフ 1 より、FG.は 1 [kHz]の正弦波を正しく出力していること、また、FFT アナライザもきちんとフーリエ変換していることがわかる。

ただし、FG.は完璧な 1 [kHz]の正弦波を出力しているわけではないこともわかる。1 [kHz]の前後 50[Hz]でのパワスペクトル密度は 0.1～0.2[V/Hz^{1/2}]以上もあり、1 [kHz]でのパワスペクトル密度 0.31[V/Hz^{1/2}]に対して無視できない値である。周波数の近い波を重ね合わせると、うなりが生じる(グラフ 3)。この合成波の一部分だけを見ると、ほとんど正弦波といってよい形が生まれる(グラフ 4)。FFT アナライザで観察するのは短時間であるので、グラフ 4 のような合成波での近似は妥当であるといえる。この時、合成波の周波数は重ね合わせの時のパワーにより変化する。このことを考慮すると、950[Hz]・1050[Hz]でパワスペクトル密度が比較的大きいのは、実際の周波数が 1[kHz]からずれているためといえる。つまり、入力波形はほぼ 1[kHz]の正弦波で近似できるが、多少の誤差を補正するために、

ある程度のパワーで 950[Hz]・1050[Hz]の波を重ねあわせる必要があったということになる。これは実験時に FG.の表示が正確には 1[kHz]でなかったことから裏づけされる。
(グラフ 3・4 での重ね合わせのパワーは表 1 に従う)



また、グラフ 2・表 2 より、1[kHz]のほぼ整数倍の周波数でパワーが極大値を取ることがわかる。非常に小さな値ではあるが、対数表示してみると、その前後よりかなり大きな値をしていることがわかる。こちらは先のうなりを用いた近似では、振幅が変化するため、これを補正するためや波形の細かい揺らぎを修正するために値が大きくなっていると考えられる。

さらに、完全には 1[kHz]の整数倍ではないことから、入力周波数のより正確な値が予想できる。元の周波数を $1000 + \alpha$ [Hz]とすると、

$$(1000 + \alpha) \times 2 < 2025$$

$$(1000 + \alpha) \times 3 > 3025$$

$$(1000 + \alpha) \times 8 < 8075$$

$$(1000 + \alpha) \times 9 > 9075$$

という関係式が得られる。これより入力された元の周波数は 1008.33～1009.37[Hz]であることが予想される。実際の FG.の表示を記録しておかなかったが、確かに、これくらいの値であったと記憶している。

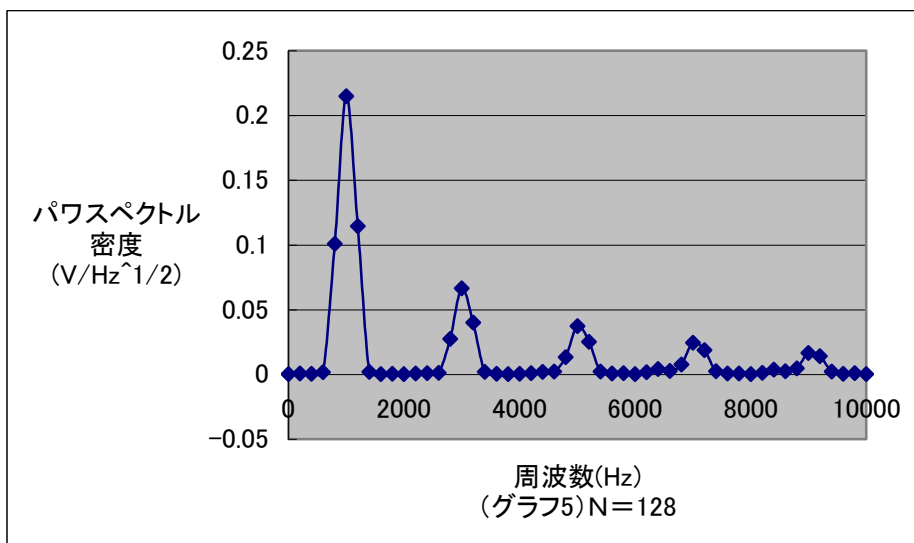
1- ii

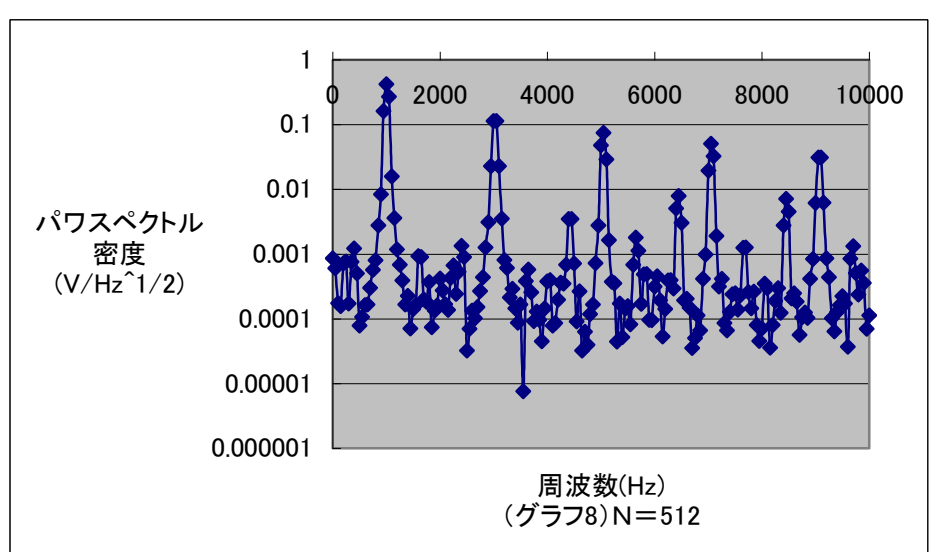
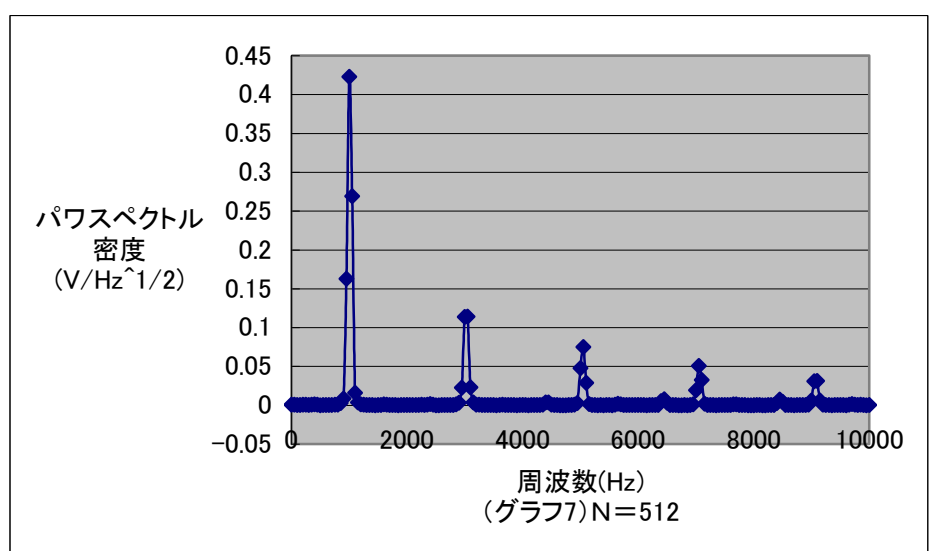
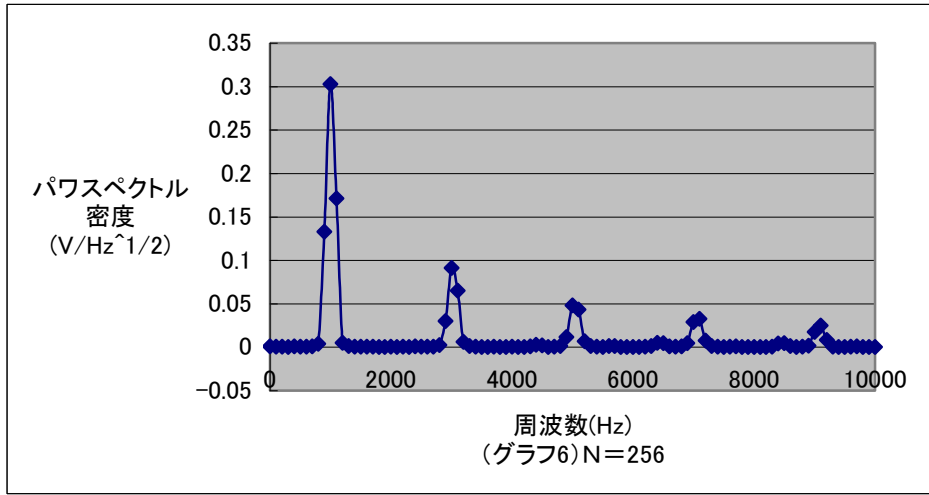
(方法)

FFT アナライザに FG.から 1[kHz]の矩形波を入力し、パワスペクトルを求めた。ただし、 $N=128, 256, 512$ 、 $V_{pp}=4V$ 、時間窓はハニング窓とした。

(結果)

このとき得られた $N=128, 256, 512$ でのスペクトラムをグラフ 5～7 に示す。また、縦軸を対数でプロットしてみたところ、特徴ある形は得られなかった。念のため、グラフ 8 に $N=512$ の時の縦軸を対数でプロットしたグラフを示すが、パワスペクトル密度 0[V/Hz^{1/2}]付近での誤差が拡大されているだけだともわれる。





まず、入力波形を

$$v(t) = \begin{cases} -4[V](-L < t < 0) \\ +4[V](0 < t < +L) \end{cases}$$

$(L = 1/2000[s])$

として、区間 $[-L, +L]$ でフーリエ変換する。

この時、この波形が奇関数であることに着目すれば、 \cos の係数は 0 であり、 \sin の係数だけを求めればいいことがわかる。

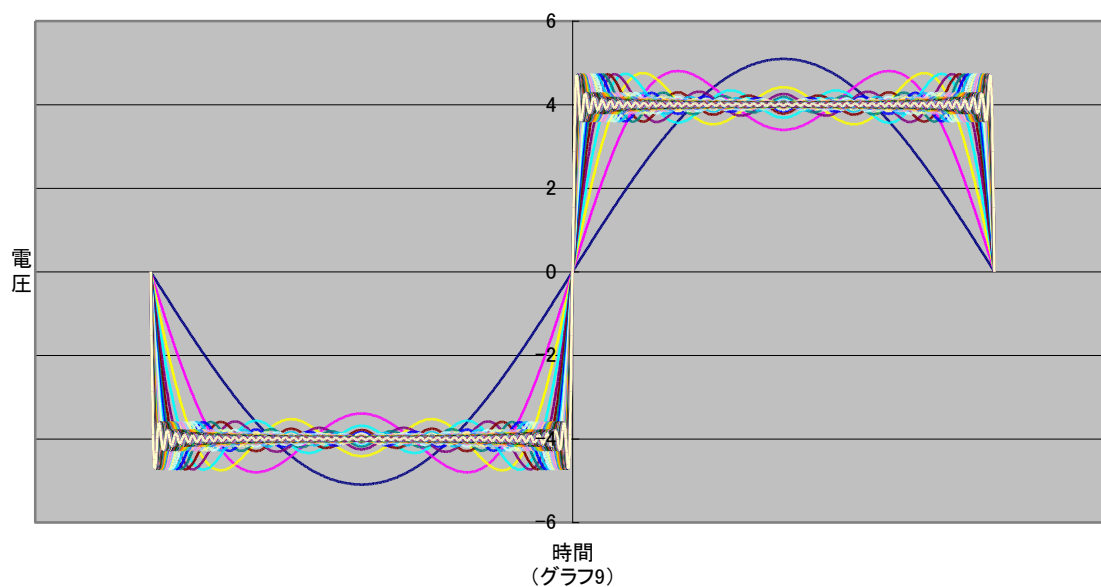
\sin の係数を求めると、

$$\begin{aligned} b_n &= \frac{1}{L} \int_{-L}^L v(t) \sin \frac{n\pi}{L} t dt \\ &= \frac{1}{L} \int_{-L}^0 -4 \sin \frac{n\pi}{L} t dt + \frac{1}{L} \int_0^L 4 \sin \frac{n\pi}{L} t dt \\ &= \frac{8}{L} \int_0^L \sin \frac{n\pi}{L} t dt = \frac{8}{L} \left[-\frac{L}{n\pi} \cos \frac{n\pi}{L} t \right]_0^L \\ &= \frac{8 \{1 - (-1)^n\}}{n\pi} \end{aligned}$$

となる。これが正しいかを確かめるため、

$$v(t) = \sum_{n=1}^k b_n \sin \frac{n\pi}{L} t$$

のグラフを $k=1,3,5\cdots 9$ で描いてみると、グラフ 9 のようになり、確かに正しいことがわかった。また、このグラフより、不連続点付近でのぶれは k をどんなに大きくしても消えないことがわかる。



周波数と b_n との関係を表 3 に示す。

周波数	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
b_n	5.093	0	1.698	0	1.019	0	0.728	0	0.566	0

(表3)

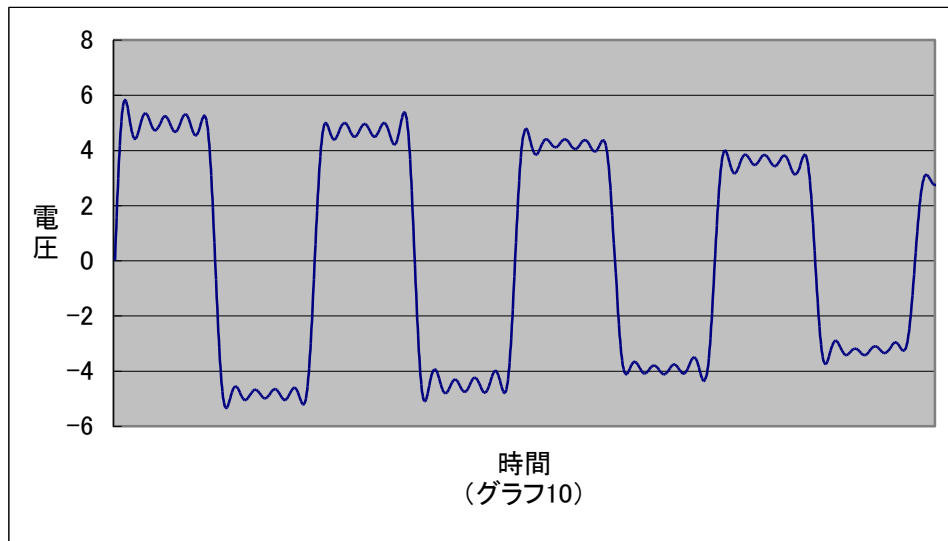
これに対し、 $N=128,256,512$ でのパワースペクトルのピークとその前後での値、及びその合計値は表 4 のようになっている。

周波数(Hz)	スペクトル密度(V/√Hz)	パワー	周波数(Hz)	スペクトル密度(V/√Hz)	パワー	周波数(Hz)	スペクトル密度(V/√Hz)	パワー
800	0.100916	1.427168	900	0.133037	1.33037	950	0.16251	1.149119
1000	0.214981	3.04029	1000	0.302925	3.02925	1000	0.422696	2.988912
1200	0.114379	1.617563	1100	0.171364	1.71364	1050	0.269136	1.903079
計		6.085022	計		6.07326	計		6.04111
2800	0.027415	0.387707	2900	0.030119	0.30119	2950	0.022744	0.160824
3000	0.066522	0.940763	3000	0.091018	0.91018	3000	0.113783	0.804567
3200	0.040037	0.566209	3100	0.065178	0.65178	3050	0.113969	0.805883
計		1.894678	計		1.86315	計		1.771274
4800	0.013238	0.187214	4900	0.011583	0.11583	5000	0.047574	0.336399
5000	0.037195	0.526017	5000	0.048136	0.48136	5050	0.074923	0.529786
5200	0.025241	0.356962	5100	0.043109	0.43109	5100	0.028886	0.204255
計		1.070192	計		1.02828	計		1.070439
6800	0.007637	0.108003	7000	0.029052	0.29052	7000	0.019409	0.137242
7000	0.024438	0.345606	7100	0.032551	0.32551	7050	0.050638	0.358065
7200	0.018643	0.263652	7200	0.007687	0.07687	7100	0.032326	0.228579
計		0.717261	計		0.6929	計		0.723886
8800	0.004595	0.064983	9000	0.017623	0.17623	9000	0.006145	0.043452
9000	0.016626	0.235127	9100	0.024837	0.24837	9050	0.030944	0.218807
9200	0.014172	0.200422	9200	0.00829	0.0829	9100	0.031091	0.219847
計		0.500533	計		0.5075	9150	0.006235	0.044088
						計		0.526194

(表4)

まず、グラフ 5～7 をみると、データ数 N が増えるほどパワースペクトルのピークは高く、鋭くなることからわかる。FG.からの出力が完璧な $1[\text{kHz}]$ の矩形波であればパワーは表 3 で示される b_n の値のようになり、ピーク幅は 0 となるはずである。しかし、実際には完璧な $1[\text{kHz}]$ の矩形波ではないため、微妙にピーク位置が異なる。これを離散的な周波数の正弦波の重ねあわせで表そうとしているので、ピーク幅を持つことになる。よって、サンプル数を増やし、周波数分解能が良くなるほどピークは細くなる。

また、表 3・4 ではピーク付近でのパワーの合計を出した。これはパワースペクトル密度を積分したものに当たる。これがサンプル数に関わらずほぼ一定であり、 b_n にほぼ等しいことから、本来 b_n であるはずのパワーが、前後に広がったと見ることができる。最後に $N=512$ の時のパワースペクトル密度を元に作成した入力波をグラフ 10 に示す。周波数が $1[\text{kHz}]$ からずれているため、ウナリが生じさせ、その一部を使っている。 b_5 までしかないので、あまり良い近時とはいえないが、大体の形は表されている。



2

(方法)

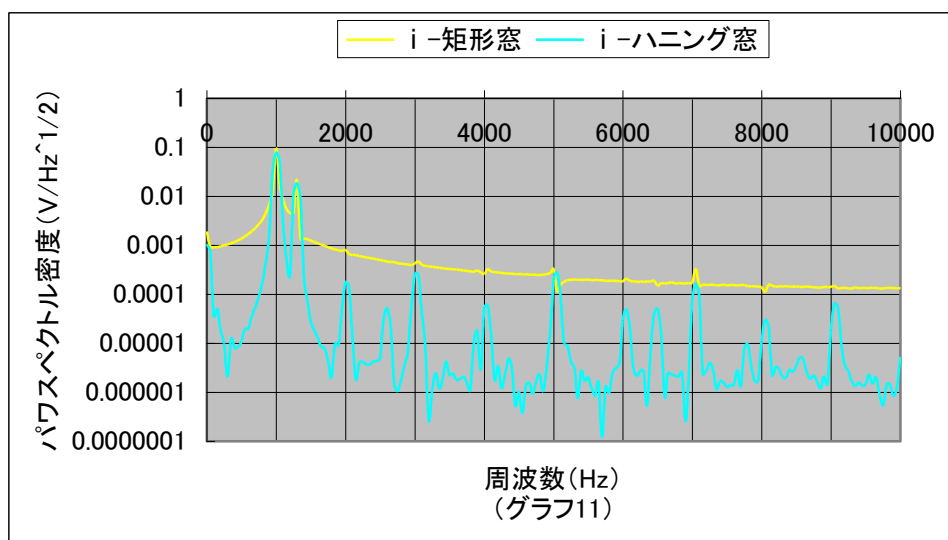
2 台の FG. から出力した表 5 に上げる異なる振幅・周波数を持つ正弦波を合成し、矩形窓・ハニング窓をそれぞれ用いてパワースペクトルを求めた。

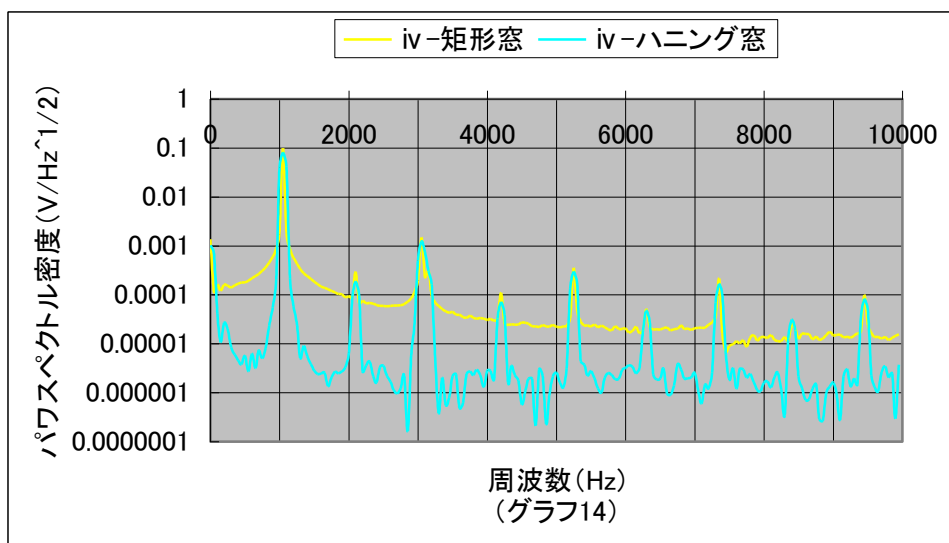
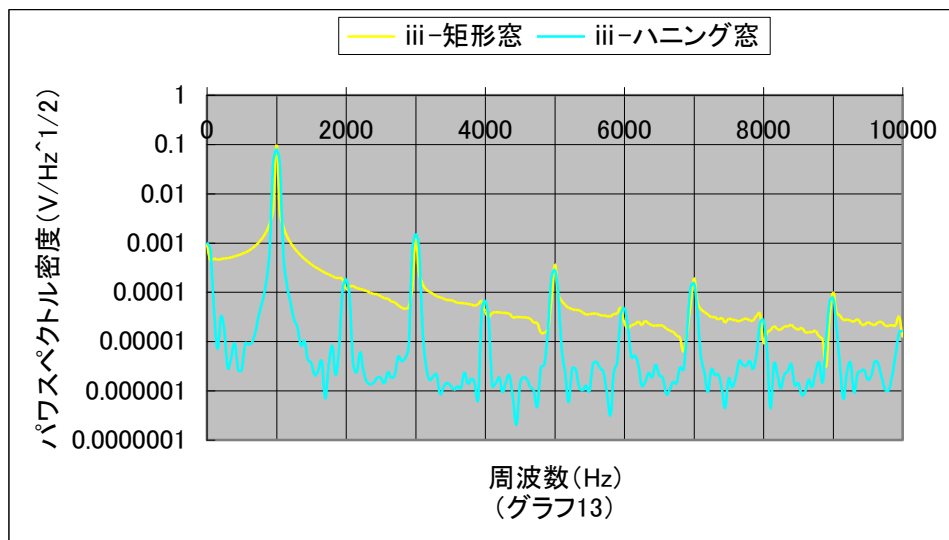
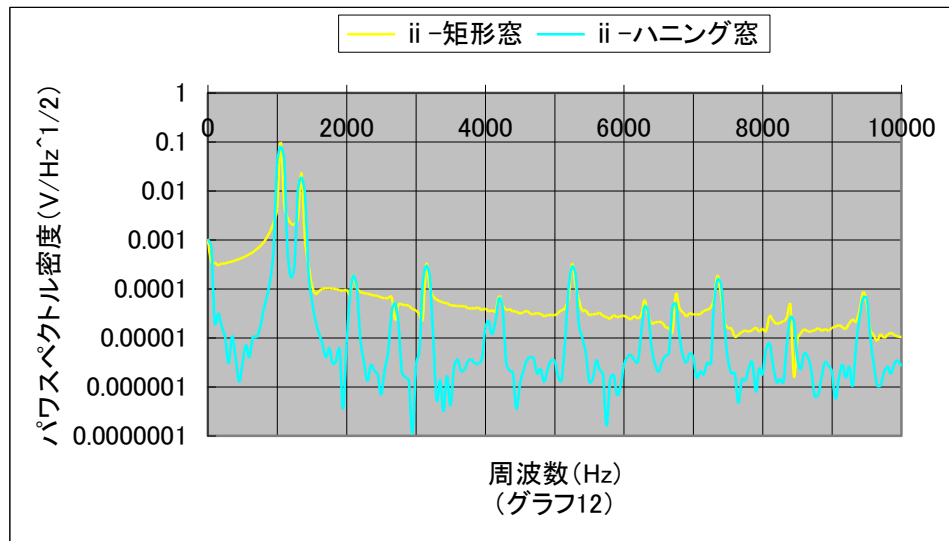
	FG.1		FG.1	
	電圧(V)	周波数(kHz)	電圧(V)	周波数(Hz)
i	1.00	1.00	0.25	1.30
ii	1.00	1.05	0.25	1.35
iii	1.00	1.00	0.25	3.00
iv	1.00	1.05	0.25	3.05

(表5)

(結果)

得られたスペクトラムを表 11～14 に示す。





(考察)

まず、縦軸を対数で表示しなかった場合、0.01 以下はほとんど目視できない。このため、矩形窓であろうが、ハニング窓であろうが、ほとんど違いのないグラフとなる。この時のグラフは入力波である複合正弦波の各成分を表すものとなっている。つまり、各成分の周波数でピークを迎えている。

一方、対数表示したものは、矩形窓・ハニング窓で明らかな違いが出ている。矩形窓を用いたものでは先に述べたピーク以外はなだらかであり、ハニング窓に比べ全体的に高い値となっている。それに対し、ハニング窓を用いたものは先に述べたピーク以外にも明確なピークがいくつも出ている。このピーク位置は先のピーク位置の整数倍となっている。

矩形窓を用いたものでピークが2つしか出なかったのは、矩形窓を通すと両端が急に 0 となり、これを近似するためにはほぼ全ての周波数の正弦波をある程度のパワーで重ね合わせる必要があるためと思われる。この2つのピーク以外は非常に小さなものであり、両端を表すための正弦波のパワーでかき消されてしまったのであろう。

よって、入力周波数の細かな誤差を無視したい時には矩形窓のほうがきれいで見やすいが、スペクトラムを細かく見たい場合には、両端の影響を抑えるために、ハニング窓を用いるほうが良いことがわかった。

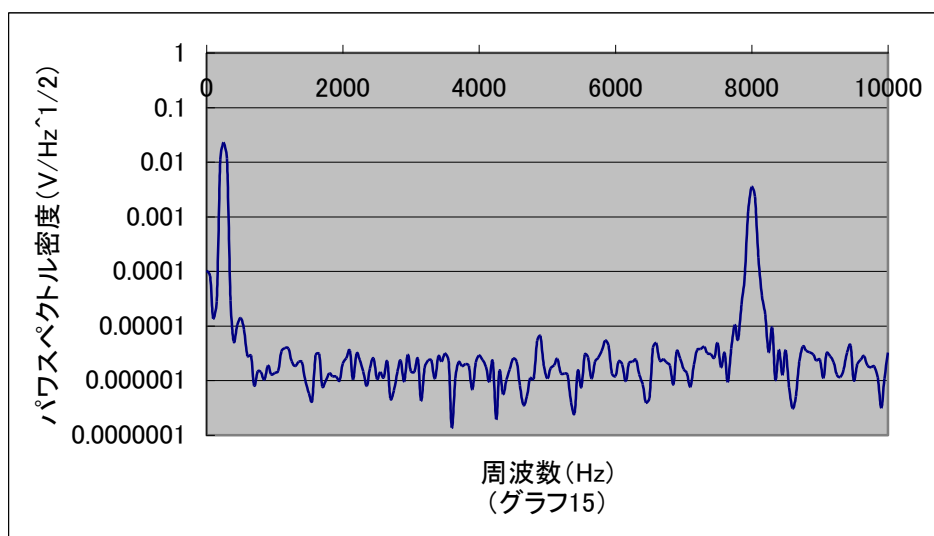
3

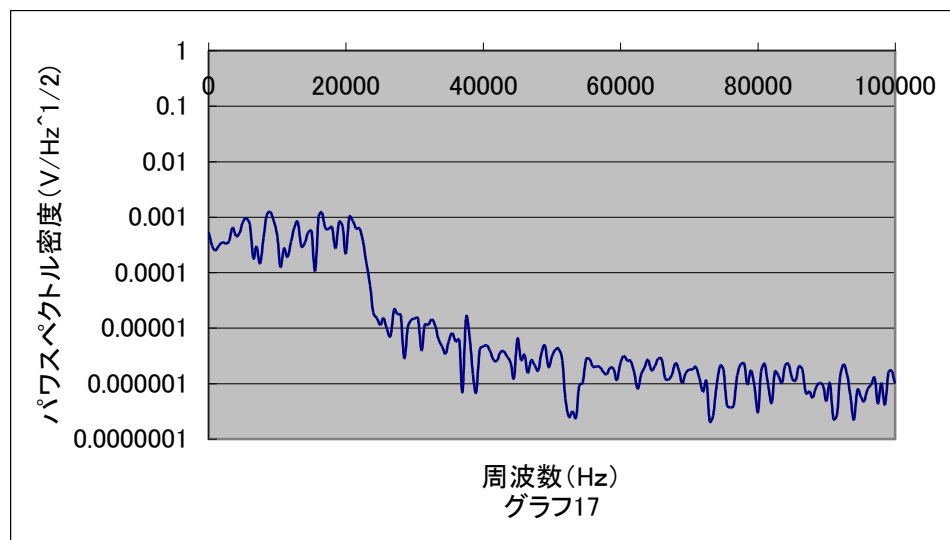
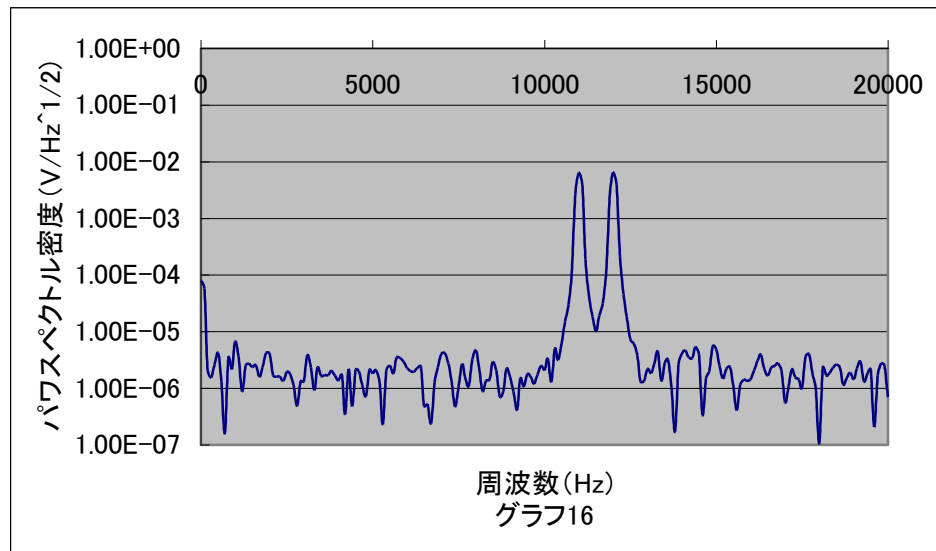
(方法)

テクニカル CD の No.40・43・79 の音を CD プレーヤーからの出力し、アンプを通してから FFT アナライザに入力し、パワースペクトルを観察した。ただし、この時の時間窓はハニング窓とした。

(結果)

No.40・43・79 のスペクトラムをそれぞれグラフ 15・16・17 に示す。





(考察)

No.40 は 250[Hz]と 8020[Hz]の波の合成波であるから、それがきちんとピークとなっている。ただし、250[Hz]のほうが8020[Hz]にくらべ細く鋭いので、250[Hz]のほうが正確に周波数の音を出しているようである。それ以外の部分では多少の上下はあるが、ノイズであると考えられる。

No.43 は 11[kHz]と 12[kHz]の波の合成波である。こちらもきちんとその周波数でピークを迎えている。こちらの二つの周波数の正確さは同じぐらいであるといえる。

No.79 はホワイトノイズである。20[kHz]までのパワースペクトル密度は 0.01[V/Hz^{1/2}]ではほぼ等しい。さらに 20[kHz]以上でのパワースペクトル密度は 0.000001[V/Hz^{1/2}]はであり、非常に小さくなったが、ほぼ等しい。人間の耳に聞こえるのは大体 20[Hz]～20[kHz]であり、この位置でのパワーが高くなっている。CD は人間の耳で聞こえない部分をカットしているので、もともと 20[kHz]以上の音は出力されない。よって、20[kHz]以上の部分は CD

からの出力のホワイトノイズではなく、真のノイズである。これにより 20[kHz]前後でのパワーの違いを説明できた。

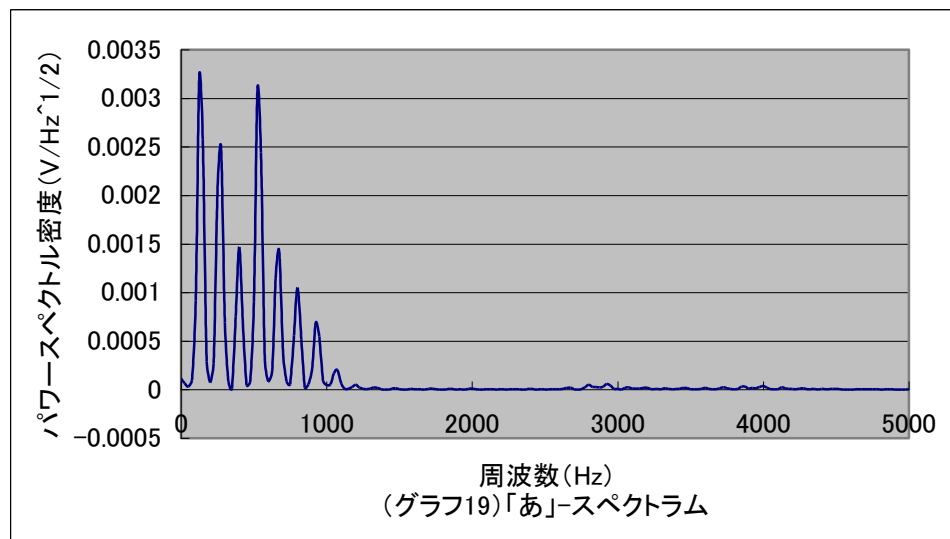
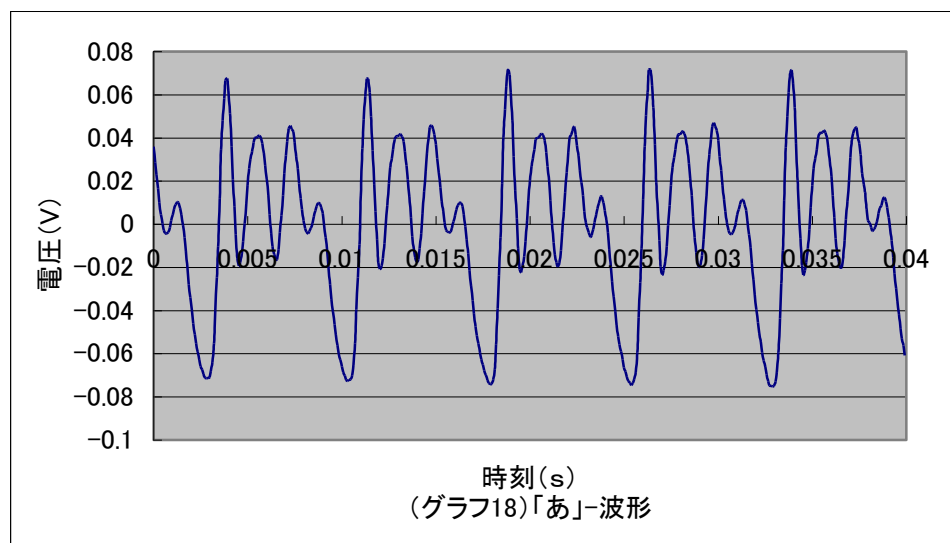
4

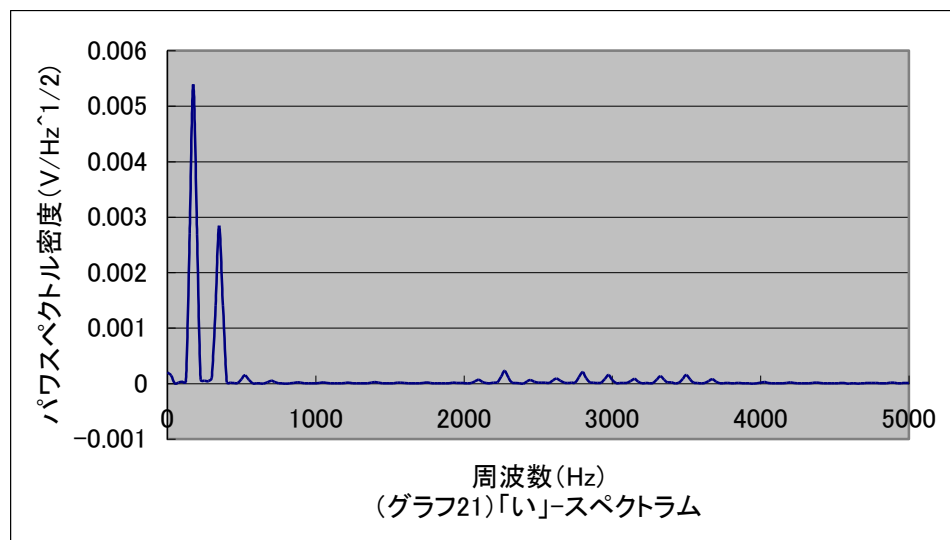
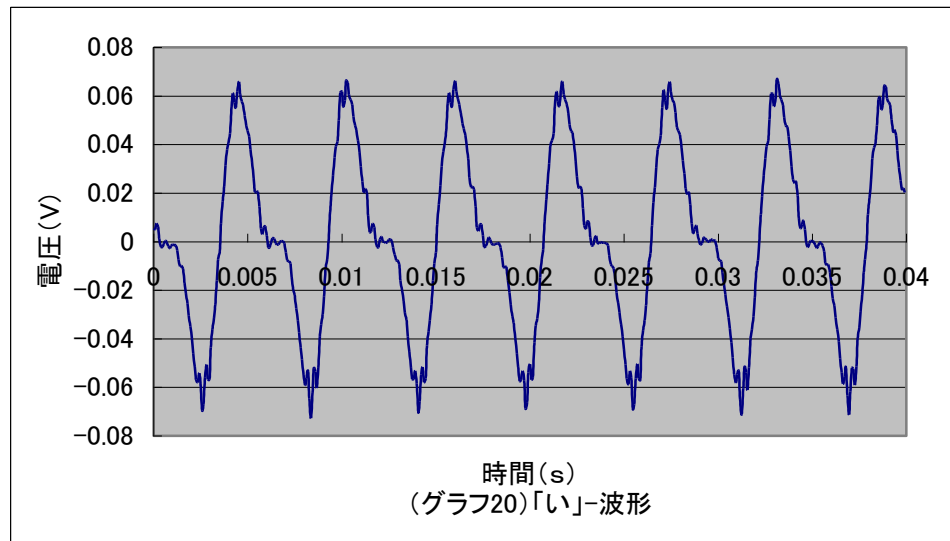
(実験)

マイクに「あ」「い」の声を入力し、パワーアンプで増幅した後に、FFT アナライザに入力し、パワースペクトルを観測した。

(結果)

自分の声の「あ」・「い」それぞれの波形およびスペクトラムをグラフ 18～21 に示す。





(考察)

グラフ 18・20 より、「あ」のピッチ周波数は 133.3336[Hz]、「い」のピッチ周波数は 174.9432[Hz]であった。また、第一・二・三ホルマントの周波数とピッチ周波数に対するエネルギーは表 6 のようになった。

		あ	い
第一ホルマント	周波数 (Hz)	125	175
	エネルギー (%)	21.02	50.60
第二ホルマント	周波数 (Hz)	275	350
	エネルギー (%)	16.36	26.45
第三ホルマント	周波数 (Hz)	400	525
	エネルギー (%)	9.199	1.389

(表6)

以上より、「あ」は第八ホルマントまでの寄与が非常に大きい。これにより、「あ」の波形はピッチの中にいくつもの大きな山ができています。これに対し、「い」は第二ホルマントま

での寄与で 75%以上を占めていて、第三ホルマント以上の寄与は小さい。このため、ピッチ内に大きな山は 1 つしかなく、細かなぎざぎざが目につく。