

音声の多値化による影響

白 滝 順・石 原 学

Effect of Some Sliced Speech

Jun SHIRATAKI and Manabu ISHIHARA

Abstract

This paper describes three experiments of reducing the normally very complicated speech wave to simple geometrical form without impairing intelligibility, and also describes the results of articulation tests of thus simplified speech. The simplification of the speech wave is performed by slicing the original input wave using comparator (s).

(1) In this experiment the slicing level is varied using a single pair of comparators. Optimum noise reduction is achieved by a slice level that is higher than the noise amplitude. The articulation score of the sliced speech at various levels, Fig. 3, shows that -40.5 dB (dB below the average peak amplitude) are optimum for the slice level in order to keep the sufficient intelligibility and to avoid the noise.

(2) In this experiment, two pairs of comparators are used to make the complex slicing speech waveform. The slice level of the first pair of comparators is set at the optimum level obtained by experiment (1), and that of the second pair of comparators is varied to find the optimum level. The outputs of both pairs of comparators are combined together by an adder to give the test output, the combination ratio of each output being kept at a predetermined value. The articulation score of the complex slicing speech wave at various levels of the second pair of comparators, Fig. 6, shows that the range of -14.5 dB to -11.0 dB is optimum for complex slicing speech wave in order to keep the sufficient intelligibility.

(3) In this experiment, also complex slicing speech waveform is used and the circuit and the slicing levels are the same as experiment (2), but the combinations ratio of each comparator output is varied. It becomes clear that the intelligibility results of this experiment is better than (1) over the wide range of combinations ratio, but the ratio itself is not so critical to the intelligibility.

1. はじめに

人間の音声には、多量の情報が含まれている。その音声情報を考える手がかりとして、零交差波¹⁾が研究されてきた。

零交差波とは、音声波 $f(t)$ の振幅を $f(t) \geq 0$ のとき常に一定値 A , $f(t) < 0$ のときに $-A$ (または 0) となるようにした波をいう。これは、定振幅であるために、能率のよい伝送手段として、通信方式にも応用されている^{2,3)}。

しかし、実際には、混変調によるひずみや微小なノイズなどがクリッピングされることになり、音声品質

が阻害される。そこで筆者らは、音声波形の振幅に注目をした。従来、振幅に関しては、量子化問題として考えられている。その場合にアナログ信号を粗く量子化することが必要となり、渡辺らは⁴⁾量子化について検討を行った。また、量子化そのものについては、圧伸の方式が種々⁵⁾研究されている。今回、コンパレータを用いた回路において、振幅レベルをそれぞれ変えることにより、明瞭度への影響について実験を行ったので報告する。

2. 零レベル除去について

2.1 実験

本回路の構成図を Fig. 1 に示す。2 値化する場合に

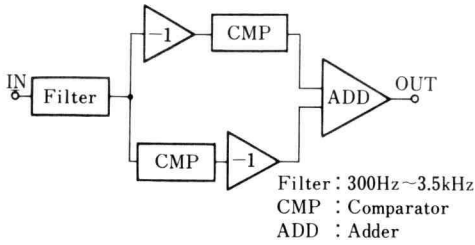
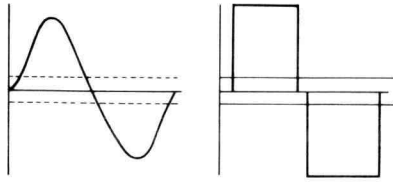


Fig. 1. Block diagram of 5 level



1) Speech wave 2) 3 level wave forms
Fig. 2. Schematic illustration of out put

は、ノイズも信号として処理されてしまう可能性があり、明瞭度などに影響をあたえてしまう。田宮らは⁶⁾、定レベル交差波として最適スライスレベルの実験を行ったが、これは零レベルでなく、零レベルより+側にスライドしたものであった。そこで零交差波に波形処理を行うとき、零レベルの信号を除去することとした(以後3値化“3 Level”と呼ぶ)。

Fig. 1に示すように、回路はFilter (300 Hz~3.5 KHz)の後、並列回路構成したコンパレータにより、正負でスライスし、加算回路により、Fig. 2の出力波形を得る。

零レベル電圧の変化は、-45.0 (dB), -40.5 (dB), -37.5 (dB), -34.5 (dB), -31.0 (dB), -29.5 (dB), -28.5 (dB)の7通りである。

実験に用いた音声資料は⁷⁾、NHK アナウンサーによる100単音節の発声テープを使用し、被験者は21歳~35歳までの健康な男性5名である。

2.2 結果及び考察

実験結果をFig. 3に示す。これによると、スライスレベルが-40.5 (dB)の場合が明瞭度が一番良い。このことは、-40.5 (dB)の位置がノイズ成分などを除去し、さらに、子音などの小さな音声成分を保存しうる値であると考えられる。

子音別にみると、/h/, /y/, /my/, /hy/, /my/, /gy/

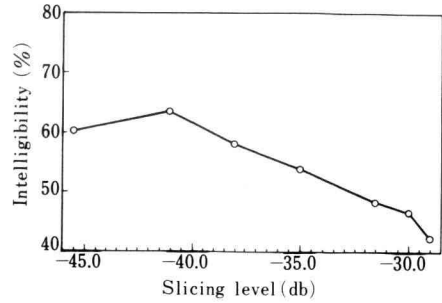


Fig. 3. The articulation score of the sliced speech various levels

などが、全体的に50%以下の低い明瞭度を示している。これらは、3値化する時点ですでに子音成分がくずれていると考えられる。また、スライスレベルで顕著な、傾向を示すものとして/s/, /b/, /g/, /z/などがある。

3. 5 値化について

3.1 S₂の可変について

3.1.1 実験

本実験の回路構成をFig. 4に示す。

2.1において得られた-40.5 (dB)のスライスレベル値に設定し、もう一組のコンパレータの比較電圧S₂を可変させることにより、最適値を求める。このときの出力電圧の関係をFig. 5に示す(以後、5値化“5 Level”と呼ぶ)。S₂の値は、-28.5 (dB), -20.5 (dB), -14.5 (dB), -11.0 (dB), -8.46 (dB), -6.52 (dB)の6通りである。音声資料、被験者などの条件は3値化の実験(零レベル除去)と同じである。

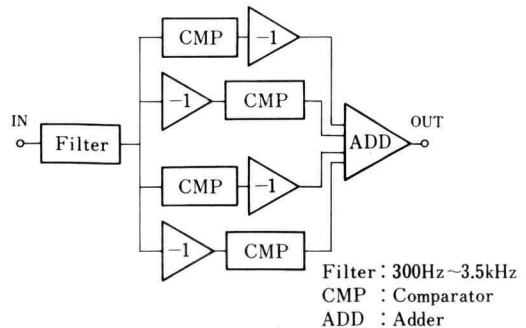
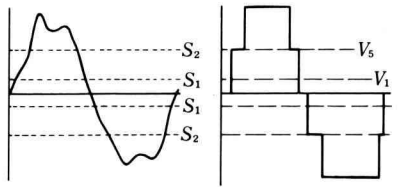


Fig. 4. Block diagram of 5 level



1) Speech wave 2) 5 level wave forms
Fig. 5. Schematic illustration of output

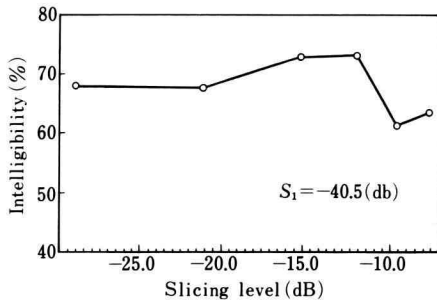


Fig. 6. The articulation score of the complex slicing speech wave at various levels of second pair comparators

3.1.2 結果及び考察

実験結果を Fig. 6 に示す。可変範囲が、 $-14.5 \sim -11.0$ (dB) で明瞭度が $72.8 \sim 73.0\%$ と一番高い。しかし、 $-28.5 \sim -11.0$ (dB) の明瞭度は $68 \sim 73\%$ と、 5% 程度の差である。 -11.0 (dB) を超える電圧から明瞭度は悪化する。すなわち、 S_2 のスライスレベルについては、電圧をある値 (しきい値) より大きくすることに問題があると考えられる。

子音別についてみると、 $/s/$, $/h/$, $/\gamma/$, $/z/$, $/sh/$ などが特にその影響を受けるようである。すなわち、この子音の特徴が、このスライスレベルに依存していると考えられる。

3.2 出力電圧比の可変について

3.2.1 実験

3.1.1 の回路構成で、スライスレベル S_1 は、 -40.5 (dB)、 S_2 は -11.0 (dB) に設定し、出力電圧振幅比を可変する。振幅比は、Fig. 5 の出力波形のピーク電圧と V_5 との比で得られる。その振幅比を可変し、明瞭度の測定を行った。

3.2.2 結果及び考察

実験結果を Fig. 7 に示す。振幅電圧比 V_5 を可変し

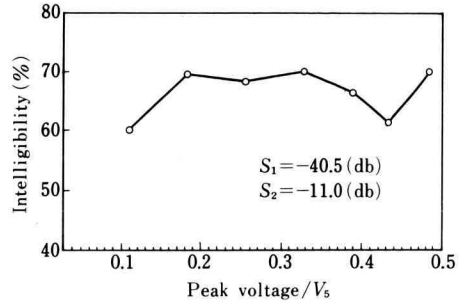


Fig. 7. The articulation score of the combinations ratio of each comparator output is varied.

た場合は、スライスレベル S_2 を可変したように、特別な傾向が見られない。すなわち、出力振幅比については、変動のすくない、 $0.185 \sim 0.390$ 程度の振幅比を与えれば良いと考えられる。

4. むすび

波形処理を行い、音声の明瞭度の測定を行った結果、次のことが明らかとなった。

- (1) ノイズレベル位置が明瞭度に及ぼす影響が大きい。これは、一般的に子音は小さく、その子音レベルとノイズレベルとのレベル差が重要な位置を占めていることを示している。
- (2) 零レベル付近での除去は、田宮らの実験において -30 dB \sim -40 dB であった。本実験で、ノイズを除去し、明瞭度の良好なレベルは $-45.0 \sim -40.5$ (dB) となり同様な結果が得られた。
- (3) 3 値化、5 値化を比較すると、明瞭度に対する貢献は、3 値化が大きい。5 値化は音声の品質 (受聴のしやすさ) に貢献すると考えられる。
- (4) 子音別についてみると、3 値、5 値の順のように高い明瞭度が得られる。しかし、 $/ny/$, $/hy/$, $/my/$, $/gy/$ については、多少上昇するものの、明瞭度が 50% 以下である。これは 3 値化の段階で子音成分に影響を与えるためと考えられる。そのために、5 値化においても明瞭度が上昇しないのであろう。
- (5) 異聴⁹⁾ の点からみると、 $/m/$, $/n/$ についての異聴が多い。これは、鼻音系の子音であるためと考えられる。しかし、これも子音成分への影響が含まれていると考えられる。

最後に、単音節の明瞭度が 60% あれば、文章了解度

は90%以上得られることが知られている。このことから、音声情報を十分に伝えうると考えられる。

今後の課題としては、鼻音系の異聴，[C+C+V]系単音節の明瞭度の改善が必要である。

参 考 文 献

- 1) たとえば，電子通信学会編：聴覚と音声，(昭和57)。
- 2) 田宮・平松：定レベル交叉位置伝送パルス方式，電気三学会連合大会予稿，No. 942 (昭和32-4)。
- 3) 片山泰男：自己相関一致化によるビット列音声の変調法，電子通信学会総合全国大会，1-85 (昭和57)。
- 4) 渡辺・勝野・鈴木・大泉：音声波形の量子化に伴う要因効果の分析，電気通信学会雑誌，第46巻11号，p. 1664～p. 1671 (昭和38-11)。
- 5) たとえば，宮川・小泉：現代デジタル通信方式，オーム社，第2章 (昭和59)。
- 6) 田宮・平松：定レベル交叉波の特質とその応用，日本音響学会誌，第14巻第2号，p. 143～p. 150 (昭和33)。
- 7) 日本音響学会明瞭度委員会編：明瞭度試験法の規準 (昭和31)。
- 8) 落合・福村：音声通信における基本品質について，研究実用化報告，第8巻，第12号 (昭和35)。