データベースシステム 117－10
（1999．1．23）

Glue/Filter演算によるビデオ検索モデルと質問処理について

十河 孝至† プラダン スジット††
田島 敬史††† 田中 克己††

1本のビデオストリームにおいて、キーワードによる区間検索を行うとき、和(Union)や共通部分(Intersection)だけでは断片的な区間しか表現できず、意味的にまとまりのある区間を検索することが難しい。そのため本論文では，インデックス付けされたビデオユニットから区間を動的に合成できるGlue演算について述べ、その解集合から不適切な区間を取り除くためのタイム ウィンドウ フィルタに加えて、最長ノイズフィルタを提案する。また、そのフィルタとグルー(Glue)演算を組み合わせた場合の問い合わせ処理の効率化についても述べる。

Towards an Efficient Query Processing for Video Retrieval Model based on Glues and Filters

TAKASHI SOGO ,† SUJEET PRADHAN ,†† KEISHI TAJIMA ††† and KATSUMI TANAKA ††

While issuing keywords-based queries to fragmentarily indexed video databases, interval opertaions such as union and intersection can produce only fragmentary intervals that not necessarily be meaningful to the end users. To a certain extent, recently proposed query mechanism based on glue operations and filter[2] has been successful in dynamically computing meanigful intervals from such a database. However, the semantics of noise filter, which they have defined to eliminate irrelevant intervals from the answer set, does not justify the actual definition of the noise in an answer interval. In this paper, we review the proposed query mechanism and define a new semantics for the noise filter. We also show how our proposed methodology can be integrated with their groundwork and lay a big foundation for an efficient query processing.

1. はじめに

1つのビデオデータにおいて、幾つかのキーワードによりそれらのキーワードが出現する区間を問い合わせるとき、ユーザとしては意味的にまとまりのある区間を答えとして望むのが一般的である。それは、ユーザがただ単にキーワードが出現している断片を検索したいのではなく、意味のある一連のシーン単位での解を望んでいるからである。そのため、ビデオの区間検索において和(Union)や共通部分(Intersection)だけでは、そのような意味のある区間を解として返すことができない。これは、実際のビデオデータでは、その意味的にまとまりのある一連のシーンの中に、それらのキーワードが全く出現しない区間が存在したり、キーワード全てが同時に出現する区間が存在しない可能性が高いためである。

そこで本論文では、ビデオ検索モデルとして、Sujeetらにより提案されたグルー(Glue)演算[2]を用いることにする。このグルー演算を用いることで、答えとなる区間を予めデータベースに格納しておく[3]のではなく、ある問い合わせに対して答えとなり得るすべての区間を動的に生成することができる。また、このグルー演算のみでは不適切な解も生成してしまうので、そのような区間を解から取り除くフィルタも提案されている。しかし、彼らのノイズの定義は制限された形で定義されており、このままではフィルタの一般化が難しい。そのため本論文では、ノイズの定義を修正することによりフィルタを一般化できること、そのことにより質問処理をさらに効率化できることについて述べる。

以下2章では、本論文のビデオ検索に対する基本的

† 神戸大学大学院自然科学研究科情報知能工学専攻
†† 神戸大学大学院自然科学研究科情報メディア科学専攻
††† 神戸大学工学部情報知能工学科

な考え方について述べ、3章では、Sujeetらにより提案されたグルー演算とフィルタの定義について述べる。また4章では、本論文におけるノイズの定義について述べ、5章では、その定義を用いた最長ノイズ フィルタの提案を行う。そして、6章では、グルー演算における重複解について議論する。

2. ビデオ検索における前提

ビデオ検索には、Video On Demandのようなビデオ自体を検索する場合とビデオの中のシーンを検索する場合がある。本論文においては、後者のシーン検索の場合を想定している。そのような場合に、断片的にキーワードがふられている一つのビデオストリームから、幾つかのキーワードにより区間検索を行うと、従来の和演算や共通部分演算では必ずしもユーザが求める区間を生成することができるとは限らない。例えば、図1のように「dog」,「man」などのキーワードがふられているビデオデータがあるとする。ここで、ユーザが「人」と「犬」で区間検索を行うと、man ∧ dogというAND検索では区間aだけが解として返される。また、man ∨ dogというOR検索では区間a,b,c,dを解として返すことができる。ここで、OR検索による区間a,b,cを和演算することにより区間1は生成できるが、区間2はノイズが含まれているため生成できない。しかし、区間2も意味的にまとまりのある区間である可能性がある。このように、検索対象のキーワードが全く現れていない区間を含んでいても、問い合わせの解となり得る。グルー演算を用いることによって、そのような解も求めることが可能となる。我々は、この検索対象のキーワードが全く現れない区間を「ノイズ」と呼ぶことにする。

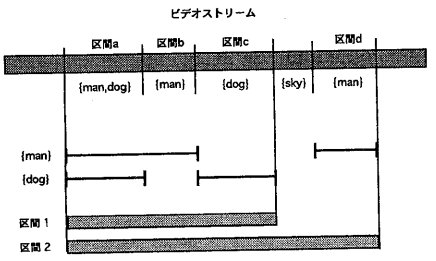

図1　区間演算

3. Sujeetらによるグルー／フィルタ演算の定義

グルー演算[2]とは、Sujeetらにより提案された区間同士もしくは、区間の集合同士から新たな区間や区間の集合を作り出す演算である。また、グルー演算の解の中から不適切な区間を取り除く演算がフィルタ演算である。

3.1 グルー演算

定義1 (区間グルー) 区間グルー演算⊕は任意の区間x,yに対して、以下のように1つの区間iを作り出す。ここで、f_s, f_eはそれぞれ区間のStart,Endフレームを求める関数とする。

$$x \oplus y = i[s,e] \quad \text{ただし、} s = \min(f_s(x), f_s(y)),\ e = \max(f_e(x), f_e(y))$$

ここで、$i[s,e]$は区間iがsフレームからeフレームまでの一連のフレーム列であることを示している。すなわち、区間グルー演算$x \oplus y$は、区間x,yをともに含んだ最小の区間iを解として返す。例えば、$x[10,20], y[30,40]$のとき、$x \oplus y = i[10,40]$となる。

区間グルー演算の主な代数的性質を以下に示す。

- 可換則：$x \oplus y = y \oplus x$
- 結合則：$(x \oplus y) \oplus z = x \oplus (y \oplus z)$
- べき等則：$x \oplus x = x$

定義2 (ペアワイズグルー) ペアワイズグルー演算$\bigoplus$は任意の区間集合X,Yに対して、以下のように区間集合を作り出す。

$$X \bigoplus Y = \{i | x \in X, y \in Y, i = x \oplus y\}$$

すなわち、演算結果は区間集合X,Yから要素を一つずつ取り、すべての組合わせにおいて区間グルー演算を適用して得られる区間集合である。図2は、それぞれk_x, k_yというキーワードがふられている区間の集合X,Yのペアワイズグルー演算の例である。

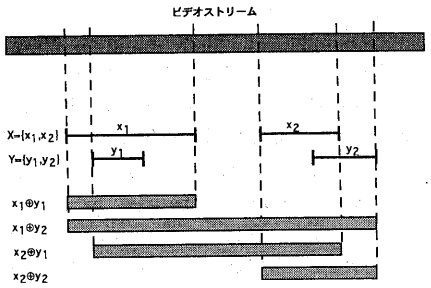

図2　ペアワイズグルー演算

ペアワイズグルー演算の性質を以下に示す。

- 可換則：$X \bigoplus Y = Y \bigoplus X$
- 結合則：$(X \bigoplus Y) \bigoplus Z = X \bigoplus (Y \bigoplus Z)$

ただし、べき等則は成り立たないことに注意を要する。

定義3 (パワーセットグルー) パワーセットグルー演算$\bigotimes$は任意の区間集合X,Yに対して、以下のように区間集合を作り出す。

$$X \bigotimes Y = \{i | X' \subseteq X, Y' \subseteq Y, X' \neq \phi, Y' \neq \phi, i = \bigoplus(X' \cup Y')\}$$

但し、$\bigoplus(\{i_1, \ldots, i_n\}) = i_1 \oplus \ldots \oplus i_n$

これは、区間集合X,Yのそれぞれから1つ以上の要素を取りだし、それらに区間グルー演算を適用して作り出した区間集合である。すなわち、パワーセットグルー演算は次式のように表すことができる。

$$X \bigotimes Y = (X \bigoplus Y) \cup (X \bigoplus X \bigoplus Y) \cup (X \bigoplus Y \bigoplus Y) \cup (X \bigoplus X \bigoplus X \bigoplus Y) \cup \vdots$$

上式は複雑に見えるが、以下のように3つのペアワイズグルー演算に変形できる。

定理1 [2] X,Yを区間の集合とすると、次式が成り立つ。 $X \bigotimes Y = (X \bigoplus X) \bigoplus (Y \bigoplus Y)$

基本的にグルー演算は、2つの区間をつなぎ合わせる演算である。よって、その解である区間の両端(Startフレーム,Endフレーム)は、もとの2つの区間によって決定づけられる。パワーセットグルー演算では、2つの区間集合からそれぞれ1つ以上の区間を取りだし、それらをグルー演算でつなぎ合わせる。その取りだし方は、多数あるが解となる区間の両端の組合わせは、定理1の3つのペアワイズグルー演算ですべて表すことができる。パワーセットグルー演算の例を図3に示す。

3.2 タイム ウィンドウ フィルタ

一般的に、ユーザがシーンを検索したいとき、ビデオ全体にわたるような長い解は不適切である。そこで、そのような長い解を取り除くために、指定された長さより長い区間を解から排除するフィルタがタイム ウィンドウ フィルタである。

定義4 (タイム ウィンドウ フィルタ) $|i|$を区間iの長さ、Wを指定されたタイム ウィンドウとすると、区間iに対するタイム ウィンドウ フィルタ$F_W(i)$を次のように定義する。

$$F_W(i) = \begin{cases} i, & \text{if } |i| \leq W \\ undefined, & \text{otherwise} \end{cases}$$

また、区間の集合Xに対して、$F_W(X)$を次のように定義する。

$$F_W(X) = \{i | i \in X \text{かつ} |i| \leq W\}$$

すなわち、タイム ウィンドウ フィルタ$F_W(X)$は任意の区間集合XからXの部分集合への関数である。

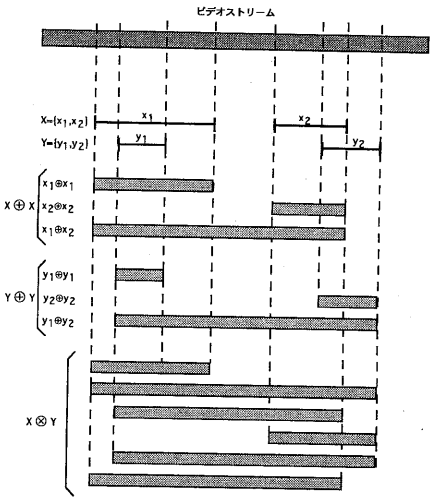

図3 パワーセットグルー演算

3章の**定理1**で述べたグルー演算の性質に対して、タイム ウィンドウ フィルタF_Wは次の**定理2**のように適用できる。

定理2 [2] X,Yを任意の区間集合とすると以下の式が成り立つ。

$$F_W(X \bigotimes Y) = F_W(F_W(X \bigoplus X) \bigoplus F_W(Y \bigoplus Y))$$

但し、F_Wはタイム ウィンドウ フィルタである。

ここで、**定理2**の左辺はパワーセットグルー演算を行った後にフィルタをかけているが、右辺は演算の途中の結果に対してもフィルタをかけている。このように、演算の過程でフィルタをかけることにより、解の候補となる区間を演算の初期の段階で減らすことが可能となる。これにより、問い合わせ処理の効率化が図れる。

4. ノイズの定義

本論文においてノイズとは、検索対象のキーワードが全く現れない区間であると定義する。1つのキーワードに対するノイズの定義は、以下のようになる。

定義5 (ノイズ) kというキーワードがふられている区間の集合をXとすると、kのノイズを求める関数は、

$$Noise(k) = \max(\overline{X})$$

と表される。但し、

$$\overline{X} = \{i[f_s, f_e] | (\forall i'(\neq i) \in X)(i' \cap i = \phi)\}$$

また、

$$\max(Z) = \{i | i \in Z \text{かつ} (\forall i'(i' \neq i) \in Z)(i \supseteq i' \text{または} i \cap i' = \phi)\}$$

すなわち、キーワードkに対するノイズとは、ビデオストリーム全体Vのうちキーワードkがふられていない区間（ノイズ区間）の集合である。

さらに、複数のキーワード$K=\{k_1,\ldots,k_n\}$に対するノイズとは、それらのキーワードがいずれも全く現れない区間の集合である。すなわち、それらのキーワードがふられている区間の集合をそれぞれ$X_1,\ldots,X_n$とすると、以下のようにそれぞれのノイズ区間集合の共通部分(Interval Set Intersection)演算[1]により表すことができる。

$$Noise(K)=\overline{X_1}\odot\ldots\odot\overline{X_n}$$

ここで、共通部分演算の定義は以下の通りである。

定義6（区間の共通部分演算） $(\exists f)(f\in i_1\wedge f\in i_2)$であるような区間$i_1,i_2$に対して、区間の共通部分演算$\odot$を行うと次のような区間iが得られる。

$$i_1\odot i_2=i[s,e]\quad \text{ただし、}\ s=\max(f_s(i_1),f_s(i_2))$$
$$e=\min(f_e(i_1),f_e(i_2))$$

すなわち、区間の共通部分演算とは、2つの区間i_1,i_2が共通のフレームを含んでいるとき、その共通フレームからなる最大の区間iである。

定義7（区間集合の共通部分演算） 2つの区間の集合X,Yに区間集合の共通部分演算$\bigodot$を行うと、2つの集合X,Yのそれぞれの要素間の共通部分演算で構成される区間集合を返す。

$$X\bigodot Y=\{i|x\in X,y\in Y,i=x\odot y\}$$

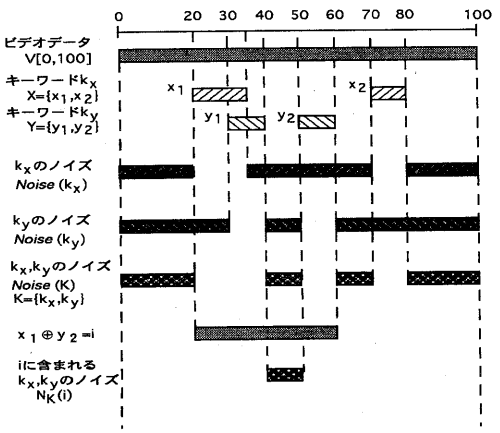

図4　ノイズの定義

このことにより、どのようなキーワードの組合わせに対するノイズもそれぞれのキーワードのノイズから容易に作ることが可能である。例として、ビデオ全体が$V[0,100]$で、k_xというキーワードがふられている区間の集合が$X=\{x_1[20,35],x_2[70,80]\}$、同様に$k_y$というキーワードがふられている区間の集合が$Y=\{y_1[30,40],y_2[50,60]\}$である図4のようなときを考える。すると、$k_x,k_y$それぞれのノイズは、図のようにビデオ全体$V$の中でそれぞれのキーワードが出現していない区間の集合である。また、k_x,k_yのノイズは、それぞれのノイズ区間集合の共通部分演算により図のようになる。

さらに、ある区間iに含まれるキーワード集合Kに対するノイズを求めるには、以下のように区間iとKに対するノイズ区間集合$Noise(K)$の共通部分演算を行えばよい。ここで、$N_K(i)$は区間iに含まれるKに対するノイズ区間の集合を求める関数とする。

$$N_K(i)=i\bigodot Noise(K)$$

$N_K(i)$の例として、図4のように$K=\{k_x,k_y\},i=x_1\oplus y_2$のときを考えると、$i$と$Noise(K)$の共通部分演算により図のように求めることができる。

ノイズを以上のように定義することにより、検索対象のキーワードが決定すれば、そのノイズも絶対的に決定し、グルー演算中においてもノイズは変化しない。それに対し、Sujeetらが定義したノイズ[2]では、グルー演算ごとにノイズが変化してしまう。このようなノイズ定義の違いが最長ノイズ フィルタにおいてどのように影響するかを次章で述べる。

5. 最長ノイズ フィルタ

ノイズを含んだ解を返すことができるのがグルー演算の特徴の一つであるが、あまりに長いノイズを含んだ解は不適切であると言える。それは、あまりに長いノイズを含んだ区間は、意味的につながっている区間である可能性が低いためである。そのため、ある長さ以上のノイズを含んでいる解を検索結果から取り除くフィルタが必要となる。そこで、最長ノイズ フィルタを提案する。

定義8（最長ノイズ フィルタ） 検索対象のキーワード集合をK,指定された最長ノイズをNとすると、区間iに対する最長ノイズ フィルタ$F_{N,K}(i)$を次のように定義する。

$$F_{N,K}(i)=\begin{cases} i, & \text{if } \max|N_K(i)|\le N \\ undefined, & \text{otherwise}\end{cases}$$

但し、$\max|N_K(i)|$は、iに含まれるノイズ区間の集合$N_K(i)$の要素のうち最長のノイズ区間の長さを求めるものとする。さらに、区間集合Xに対して、$F_{N,K}(X)$を次のように定義する。

$$F_{N,K}(X)=\{i|i\in X\text{かつ}\max|N_K(i)|\le N\}$$

最長ノイズ フィルタとグルー演算の間にも、タイム ウィンドウ フィルタのときと同様に次のような性質が成り立つ☆。

補題1　与えられたキーワード集合$K=\{k_1,\ldots,k_n\}$に対して、どのような2つの区間a,bにおいても次式が成り立つ。

$$F_{N,K}(a\oplus b)=F_{N,K}(F_{N,K}(a)\oplus F_{N,K}(b))$$

補題2　与えられたキーワード集合$K=\{k_1,\ldots,k_n\}$に対して、どのような2つの区間集合A,Bにおいても次式が成り立つ。

$$F_{N,K}(A\bigoplus B)=F_{N,K}(F_{N,K}(A)\bigoplus F_{N,K}(B))$$

これら**補題1，2**から、**定理1**のグルー演算の性質に対して、最長ノイズ フィルタ$F_{N,K}$は次の**定理3**のように適用できる。

定理3　キーワード集合$K=\{k_1,\ldots,k_n\}$について検索を行うとき、X,Yを任意の区間集合とすると以下の式が成り立つ。

$$F_{N,K}(X\bigotimes Y)=$$
$$F_{N,K}(F_{N,K}(X\bigoplus X)\bigoplus F_{N,K}(Y\bigoplus Y))$$

但し、$F_{N,K}$は最長ノイズ フィルタである。

定理3は、タイム ウィンドウ フィルタと同様に、最長ノイズ フィルタにおいても演算処理の初期においてフィルタをかけることにより、問い合わせ処理の効率化が図れることを示している。これら**補題1，2**、**定理3**が成り立つのは、ノイズの定義において、検索対象のキーワードが決まれば、ノイズも絶対的に決まるように定義しているためである。Sujeetらのノイズ定義では、グルー演算ごとにノイズが変化するので、**補題1，2**および**定理3**が成り立たず、最長ノイズ フィルタをタイム ウィンドウ フィルタと同様に扱うことができない。

6. 議　　論

上記ようなフィルタを演算過程に挿入することにより、解の候補を演算の初期段階で減らすことは問い合わせ処理の効率化において有効である。これらのフィルタは、不適切であると思われる解を取り除くフィルタである。一方、パワーセットグルー演算やペアワイズグルー演算を行う場合を考えてみる。これらの演算は、区間集合同士の演算であり、それぞれの集合から要素を取りだし、すべての組み合わせで区間グルー演算を行うものである。このとき、異なった要素の組合わせであっても、ある条件を満たせば同じ解が発生する。例としては、図5のような場合である。図5のように重複した解が生まれる条件は、以下のように3つの場合に分けることができる。

(1)　$f_s(x)\le f_s(y_1), f_s(x)\le f_s(y_2)$,
$f_e(y_1)\le f_e(x), f_e(y_2)\le f_e(x)$のとき
解iは、$i[f_s(x),f_e(x)]$となる。

(2)　$f_s(y_1)<f_s(x), f_s(y_2)<f_s(x)$,
$f_e(y_1)\le f_e(x), f_e(y_2)\le f_e(x)$,
$s=f_s(y_1)=f_s(y_2)$のとき
解iは、$i[s,f_e(x)]$となる。

(3)　$f_s(x)\le f_s(y_1), f_s(x)\le f_s(y_2)$,
$f_e(x)<f_e(y_1), f_e(x)<f_e(y_2)$,
$e=f_e(y_1)=f_e(y_2)$のとき
解iは、$i[f_s(x),e]$となる。

パターン1は、ある区間に対して、その区間に完全に含まれる区間であれば、どんな区間であれグルー演算の解は同じになることを示している。また、パターン2、3は、それぞれある区間よりも前もしくは、後にあり、Startフレームもしくは、Endフレームが等しければ、グルー演算の解が同じになることを示している。このようなとき、事前に重複した解を作る要素の組合わせが分かっていれば、重複した解を発生させる組合わせについては1度の演算を行うだけでよい。このことによって、質問処理の効率化が可能である。このような、重複解を生む演算を避けるためのフィルタも考える必要がある。

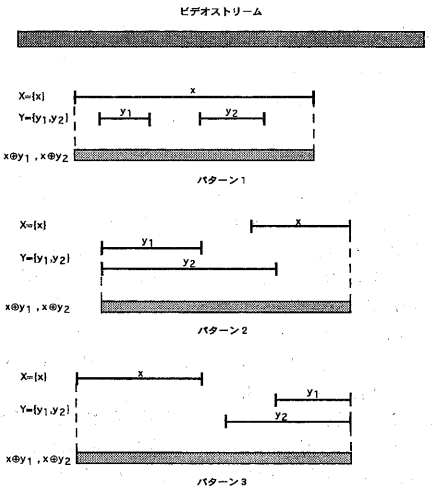

図5　重複解の発生

7. 結　　論

断片的にインデックス付けされたビデオデータに対

☆ 補題1，2、定理3の証明は付録を参照

して、キーワードにより一連のシーンを検索するとき、ただ単にキーワードがふられている区間の和や共通部分をとっても検索結果は断片的なものしか得られない。一般的にユーザは、意味的にまとまりのある一連のシーンとしての結果を求めている。そのようなとき、断片的に付けられたキーワードだけを考えるのではなく、キーワードが付けられていない区間、すなわちノイズについても考慮する必要がある。Sujeetらにより提案されたグルー演算は、そのノイズを含んだ区間も解として返すことができるので、よりユーザが求めている検索結果に近いものを返すことが可能である。さらに、明らかに不適切と思われる解を取り除くフィルタを演算途中の解候補に適用することによって、問い合わせ処理の効率化が可能となる。しかしながら、彼らによって提案されたフィルタのうち、最長ノイズフィルタにおいては、ノイズ定義の不十分さから問い合わせ処理の効率化が十分に発揮できていない。そこで、本論文ではノイズの定義に修正を加え、最長ノイズフィルタにおいてもタイム ウィンドウ フィルタと同様に問い合わせ処理の効率化が可能となることを示した。

謝　辞

この研究は部分的に文部省科学研究費特定領域研究「高度データベース」の援助を受けています。また，部分的に日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」によっています。ここに記して謝意を表すものとします。

参 考 文 献

1) S. Pradhan, K. Tajima, and K. Tanaka. "A Query Model to Synthesize Answer Intervals from Indexd Video Units". *Submitted to Journal.*
2) S. Pradhan, K. Tajima, and K. Tanaka. "Interval Glue Operations and Answer Filtering for Video Retrieval". *To Appear in IPSJ Transactions on Databases*, vol.40, no.SIG3(DBS & FI 1), January 1999.
3) R. Weiss, A. Duda, and D. Gifford. "Composition and Search with a Video Algebra". *IEEE MultiMedia*, 2(1):12–25, Spring 1995.

付　録

A.1 補題1の証明

証明: $\max|N_K(a)| \leq N$ かつ $\max|N_K(b)| \leq N$ のときは明らかに成り立つ。$\max|N_K(a)| > N$ または $\max|N_K(b)| > N$ のとき、右辺はundefinedとなる。ここで、区間グルーの定義から $a \oplus b$ は、区間aを区間の一部として含んでおり、同様に区間bも区間の一部として含んでいる。よって、区間a,bの持っているノイズも $a \oplus b$ は持っている。また、ノイズは検索対象のキーワード集合 K が決まった時点で絶対的に決まるので、演算中でノイズの大きさは変化しない。

以上から、$\max|N_K(a)| > N$ または $\max|N_K(b)| > N$ のときは、左辺も $\max|N_K(a \oplus b)| > N$ となり、undefinedとなる。よって、上式は成り立つ。∎

A.2 補題2の証明

証明: $F_{N,K}(A \bigoplus B)$ をU,$F_{N,K}(F_{N,K}(A) \bigoplus F_{N,K}(B))$ をVとし、$U = V$ を証明するために $U \supseteq V$ と $U \subseteq V$ を示す。

$F_{N,K}(A)$ の任意の要素をaと考え、$a \in F_{N,K}(A)$ と表す。当然、$a \in A$ であるから、$A \supseteq F_{N,K}(A)$ となる。同様にして、$B \supseteq F_{N,K}(B)$ となる。よって、$A \bigoplus B \supseteq F_{N,K}(A) \bigoplus F_{N,K}(B)$ となり、$F_{N,K}(A \bigoplus B) \supseteq F_{N,K}(F_{N,K}(A) \bigoplus F_{N,K}(B))$ となる。したがって、$U \supseteq V$ となる。

次に、Uの任意の要素cを考え、$c \in U$ と表す。ここで、$\max|N_K(c)| \leq N$ であり、また $a \oplus b = c$ であるようなAの要素aとBの要素bが存在する。$\max|N_K(c)| \leq N$ であるので、$\max|N_K(a)| \leq N$ となり、$a \in F_{N,K}(A)$ と書ける。同様にして、$b \in F_{N,K}(B)$ となる。また、$\max|N_K(a \oplus b)| \leq N$ であるので、$(a \oplus b) \in F_{N,K}(F_{N,K}(A) \bigoplus F_{N,K}(B))$ である。したがって、$c \in F_{N,K}(F_{N,K}(A) \bigoplus F_{N,K}(B))$ となり、$U \subseteq V$ となる。以上より、証明終わり。∎

A.3 定理3の証明

証明: 定理1より、$X \bigotimes Y = (X \bigoplus X) \bigoplus (Y \bigoplus Y)$ である。この両辺に最長ノイズ フィルタ $F_{N,K}$ をかけると、$F_{N,K}(X \bigotimes Y) = F_{N,K}((X \bigoplus X) \bigoplus (Y \bigoplus Y))$ になる。ここで、$X \bigoplus X = X', Y \bigoplus Y = Y'$ とする。また補題2より、

$$F_{N,K}(X' \bigoplus Y') = F_{N,K}(F_{N,K}(X') \bigoplus F_{N,K}(Y'))$$

X', Y' を置き換えて、

$$F_{N,K}((X \bigoplus X) \bigoplus (Y \bigoplus Y)) = F_{N,K}(F_{N,K}(X \bigoplus X) \bigoplus F_{N,K}(Y \bigoplus Y))$$

したがって、

$$F_{N,K}(X \bigotimes Y) = F_{N,K}(F_{N,K}(X \bigoplus X) \bigoplus F_{N,K}(Y \bigoplus Y))$$

以上、証明終わり。∎