

楽器音や動物の鳴声の音色と 音声の言語情報を保持したクロス合成 VOCODER

西 大輝^{1,a)} 西村 竜一^{1,b)} 入野 俊夫^{1,c)} 河原 英紀^{1,d)}

概要: 楽器音や動物の鳴声と、音声の2つの音源の特徴を併せ持つ合成音を作るクロス合成 VOCODER の検討をしている。クロス合成は、音声の狭帯域伝送技術である VOCODER を応用した技術で、現在では楽曲制作や Vocal エフェクター等、音楽の分野で広く用いられる。しかし、クロス合成でつくられる合成音は、楽器音等の音色の特徴が失われ、元の楽器の音が何か不明確になるという問題がある。本報告では、この問題を解決するため、変調周波数領域を帯域制限することにより、音声の言語情報だけを残したスペクトルを用いる新たなクロス合成を提案する。さらに、変調周波数領域を処理するフィルタにおける遮断周波数の設計を検討し、その効果を主観評価実験により明らかにした。

キーワード: クロス合成 VOCODER, 楽器音, 動物の鳴声, 言語情報, 変調周波数領域

Cross synthesis VOCODER which preserves linguistic information and characteristic timbre of musical instruments and animal voices

NISHI TAIKI^{1,a)} NISIMURA RYUICHI^{1,b)} IRINO TOSHIO^{1,c)} KAWAHARA HIDEKI^{1,d)}

Abstract: A new design method of cross synthesis VOCODER, which synthesizes sounds by mixing features of two input sounds, such as speech and musical instruments or animal voices, is proposed. Cross synthesis VOCODER is originated from a narrow-band transmission technology and currently widely used as an effector for musical performance and production. However, current cross synthesis effects tend to deteriorate original character of musical instruments and linguistic information of the processed sound is not always intelligible. The proposed method provide ways to alleviate these difficulties using two technique. One is spectral global shape removal from the speech spectral envelope and the other is band-pass filtering in the modulation frequency domain. Subjective test results indicated relevance of the proposed techniques and provide design guideline of new flexible cross synthesis VOCODERs.

Keywords: Cross Synthesis VOCODER, musical instruments, animal voice, linguistic information, modulation frequency domain

1. はじめに

音声の狭帯域伝送のために発明された技術である VOCODER[1] は、現在では、楽器が歌うような効果を与えるエフェクターである“VOCODER”として、音楽

の分野で広く用いられている。なお、「楽器が歌うような効果」を与える技術は、クロス合成 (cross synthesis) と総称されており、Channel VOCODERをはじめ、Phase VOCODER[2] や LPC[3], [4] などの様々な手段を用いた実装が試みられている [5]。クロス合成では、2つの音の性質を混合して新しい音を合成する。楽器音と音声のクロス合成では、楽器の音色で言葉が話されるような効果を狙い、動物の鳴声と音声のクロス合成では、動物が話しているような効果を狙う。このようなクロス合成が成功するた

¹ Wakayama University
Sakaedani 930, Wakayama 640-8510, Japan
a) s125038@center.wakayama-u.ac.jp
b) nisimura@sys.wakayama-u.ac.jp
c) irino@sys.wakayama-u.ac.jp
d) kawahara@sys.wakayama-u.ac.jp

めには、「楽器の音色」や「その動物らしさ」を保った上で音声の言語情報を伝える方法を明らかにする必要がある。

本報告では、クロス合成に含まれる「元の楽器らしさ」や「元の動物の鳴声らしさ」と「音声の言語情報」の割合を自由に調整できる、新たなクロス合成 VOCODER を提案する。提案法では、まず、従来のクロス合成で「元の楽器らしさ」や「元の動物の鳴声らしさ」を破壊する主な原因となっていた、スペクトルの大域的な形状を変化させる成分を音声から取り除く。さらに、音声スペクトルの時間変化に注目することで、言語情報を伝える上で重要ではない時間的な変化を取り除く。これらの操作に含まれるパラメータと、クロス合成音から知覚される「元の楽器らしさ」や「元の動物の鳴声らしさ」および、「音声の言語情報」との関係性を調べることで、目的に応じた設計法を明らかにした。

具体的には、まず、スペクトルの大域的な形状を変化させる成分を音声から取り除くために、平均値正規化スペクトルを求める方法を提案する。次いで、言語情報を伝える上で重要ではない時間的な変化を取り除くために、音声のスペクトルの時間軌跡のフーリエ変換することで求められる、音声の変調スペクトルをフィルタリングする方法を提案する。これらの提案した方法に基づいて、主観評価実験を行い、フィルタリングのパラメータと「元の楽器らしさ」や「元の動物の鳴声らしさ」及び、「音声の言語情報」との関係性を明らかにする。本資料では、まず、音声のスペクトル包絡分析に用いた、音声分析変換合成方式 STRAIGHT[6]の後継技術である TANDEM-STRAIGHT[7]について紹介する。次いで、音声の全体の形状を取り除く平均値正規化スペクトルの求め方について、変調周波数領域におけるフィルタリングについて述べる。さらに、処理したスペクトル包絡から求められる最小位相応答を用いた時変フィルタによるクロス合成 VOCODER の実装について述べ、提案法の有効性を示す主観評価実験の手順と結果について報告する。

2. 音声のスペクトル分析

本資料で用いた TANDEM-STRAIGHT によるスペクトル分析法と、ピーク強調を利用した改良について説明する。周期信号を短時間 Fourier 変換して求めたパワースペクトルには、信号の周期性に起因する周期的な変動が、時間と周波数の両方向に生ずる。TANDEM-STRAIGHT[7]では、以下に示す2段階の処理により、この周期性の影響を除去したスペクトル包絡を求めている。

2.1 TANDEM スペクトル

基本周期 T_0 の半分の間隔を隔てた2つの時刻で求められるパワースペクトル $P(\omega, t)$ の平均として、時間的に変動しないスペクトル $P_T(\omega, t)$ (TANDEM スペクトル) が

求められる [8]。

$$P_T(\omega, t) = \frac{P(\omega, t - \frac{T_0}{4}) + P(\omega, t + \frac{T_0}{4})}{2} \quad (1)$$

2.2 STRAIGHT スペクトル

TANDEM スペクトルには、周波数軸上で基本周波数の周期で変動する成分が含まれる。この周期的変動は、基本周波数の幅の矩形の平滑化関数による周波数方向の移動平均により取り除くことができる。この平滑化と、時間窓による平滑化が重なることによる過剰な平滑化の影響は、consistent sampling の考え方をを用いることにより補償することができる。これらの2つの処理と、処理結果の正值性を保証するための処理は、以下のように cepstrum lifter による一段階の処理にまとめることができる [9]。

$$P_{TST}(\omega) = \exp(\mathcal{F}^{-1}[g_1(\tau)g_2(\tau)C_T(\tau)]) \quad (2)$$

$$\text{where } g_1(\tau) = \tilde{q}_0 + 2\tilde{q}_1 \cos(\frac{2\pi\tau}{T_0}) \quad (3)$$

$$g_2(\tau) = \frac{\sin(\pi f_0 \tau)}{\pi f_0 \tau} = \mathcal{F}[h_2(\omega)] \quad (4)$$

$$h_2(\omega) = \begin{cases} \frac{1}{\omega_0} & |\omega| \leq \frac{\omega_0}{2} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$C_T(\tau) = \int_{-\infty}^{\infty} \ln(P_T(\omega, t)) e^{j\omega\tau} d\omega \quad (6)$$

ここで、 $g_1(\tau)$ は、過剰平滑化の補償用のリフタを表し、 $g_2(\tau)$ は、移動平均用のリフタを表す。なお、 τ は、quefrency を表す。また、 \mathcal{F} は、Fourier 変換を表し、 $\omega_0 = 2\pi f_0$ は、基本周波数に対応する角周波数を表す。補償用のリフタの係数 \tilde{q}_0 と \tilde{q}_1 には、合成音声の品質の改良のために数値的に最適化されたものが用いられている [10]。

3. 楽器の音色の保存に適した音声スペクトル

楽器音等と音声をクロス合成すると、音声のスペクトルの影響によって、元の楽器音等の音色が変化してしまう。そこで、楽器音等の音色を変化させる原因と考えられる、音声の全体の形状を取り除いた平均値正規化スペクトルを求める。これにより、楽器音の音色を保存するのに適した、音声の周波数特性を求める。

3.1 長時間平均スペクトル

まず、取り除くべき音声スペクトルの全体的な形状の表現について、検討する。出発点は、音声の長時間平均スペクトルである。スペクトログラムを $P(\omega, t)$ としたとき、これを時間方向に平均したものが長時間平均スペクトルである。対象とする音声 $(0, T_s)$ の範囲に存在するとする。このとき、長時間平均スペクトル $\overline{P_t(\omega)}$ は、以下により定義される。

$$\overline{P_t(\omega)} = \frac{1}{T_s} \int_0^{T_s} P(\omega, t) dt \quad (7)$$

しかし、聴覚における時間方向の積分がパワーの積分である必然性は無い。ここでは、ある単調な（非線形を含む）変換関数 $g(x)$ と、その逆関数を用いて、以下のように一般化した長時間平均スペクトル $\overline{P_{t,g}(\omega)}$ を導入する。

$$\overline{P_{t,g}(\omega)} = g^{-1} \left(\frac{1}{T_s} \int_0^{T_s} g(P(\omega, t)) dt \right) \quad (8)$$

ここで、 $g(x)$ を恒等写像とすれば、 $\overline{P_{t,g}(\omega)}$ は、 $\overline{P_t(\omega)}$ と一致する。ここで用いる $g(x)$ としては、ラウドネス感覚を近似する冪乗関数や、Weber-Fechner の法則に対応する対数関数などが候補となる。

本資料では、まず、 $g(x)$ を、対数関数とした場合について検討した。

$$\overline{P_{t,\ln}(\omega)} = \exp \left(\frac{1}{T_s} \int_0^{T_s} \ln(P(\omega, t)) dt \right) \quad (9)$$

3.2 スペクトルの全体的な形状

素材として用いる音声为数秒程度の短いものである場合、こうして求めた一般化された長時間平均スペクトルをそのままスペクトルの全体的な形状として用いることはできない。ここでは、聴覚末梢形における周波数配置（具体的には ERB_N number[11] を用いる）に基づいて周波数方向にスペクトルを平滑化することにより、全体的な形状を求めることとする。あるパワースペクトルを $S(\omega)$ とする。 $\lambda(\omega)$ を、角周波数を ERB_N number に変換する関数とし、 $\omega(\lambda)$ を、 ERB_N number を角周波数に変換する逆関数とする。ここで、スペクトルを ERB_N number 軸上で、 a の幅で矩形の平滑化を行なうものとする。平滑化されたスペクトル $\tilde{S}_a(\omega)$ は、以下により求められる。

$$\tilde{S}_a(\omega) = \frac{1}{C} \int_{\omega(\lambda(\omega)-a/2)}^{\omega(\lambda(\omega)+a/2)} S(q) dq \quad (10)$$

$$C = \omega(\lambda(\omega) + a/2) - \omega(\lambda(\omega) - a/2)$$

ここで、 $S(\omega)$ に、 $\overline{P_{t,\ln}(\omega)}$ を代入することにより、対数スペクトルの長時間平均に基づくスペクトルの全体的な形状 $\tilde{P}_{a,\ln}(\omega)$ が求められる。

3.3 平均値正規化スペクトル

こうして求められたスペクトルの全体的な形状を用いて、各フレームの STRAIGHT スペクトル $P_{ST}(\omega, t)$ を正規化することにより、楽器音の音色を変化させる音声の周波数特性を改善した平均値正規化スペクトル $D(\omega, t)$ が求められる。対数スペクトルの長時間平均に基づく平均値正規化スペクトルを $D_{a,\ln}(\omega, t)$ とする。

$$D_{a,\ln}(\omega, t) = \frac{P_{ST}(\omega, t)}{\tilde{P}_{a,\ln}(\omega)} \quad (11)$$

図1に TANDEM-STRAIGHT によって求めたスペクトログラムを、図2に音声スペクトルから、音声の全体の形状を除いた、平均値正規化スペクトルを示す。

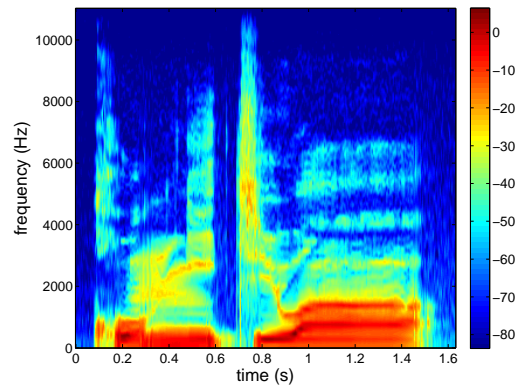


図1 音声「こんにちは」のスペクトログラム
Fig. 1 spectrogram /koNnichiwa/

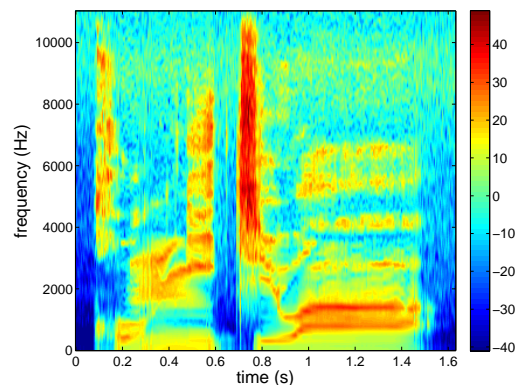


図2 音声「こんにちは」の平均値正規化スペクトログラム
Fig. 2 normalize spectrogram / koNnichiwa/

4. 変調周波数領域でのフィルタリング

音声スペクトルの変調周波数領域において、音声の言語情報を担う帯域は限られている [12][13]。求めた平均値正規化スペクトルを対数変換し、変調周波数領域でフィルタリングする。これらの処理により、クロス合成に含まれる「元の楽器らしさ」や「元の動物の鳴らしさ」と「音声の言語情報」の割合を自由に調整できる。新たなクロス合成 VOCODER の実現を試みた。

4.1 変調周波数領域でのフィルタリング

文献 [14] で指摘されているように、明瞭度への悪影響を避けるために、直線位相の FIR フィルタを用いた。さらにフィルタはインパルス応答をできるだけ短くし、不要な極性の振動を避けるよう設計した。高域通過フィルタは、原点に置いた高さ1のインパルスと、積分値が-0.9となるように負の係数を掛けた Hanning 窓の和を、インパルス応答としたものを用いた。低域通過フィルタは、原点に置いたインパルスの高さを調整した信号と、Blackmann 窓の和を、インパルス応答としたものを用いた。どちらも零周波数での利得を-20dB 付近になるようにした。音声の言語情報は、変調周波数領域で1~16Hzにあるとされ、特に

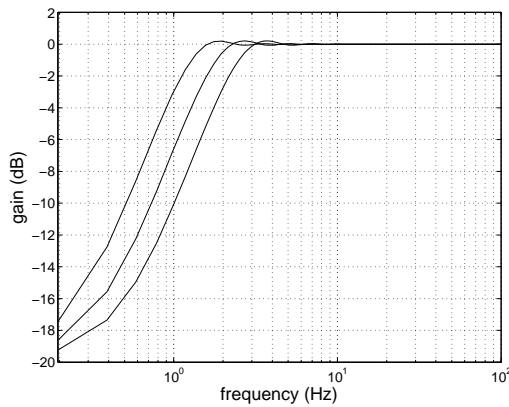


図 3 高域通過フィルタの変調周波数伝達特性

Fig. 3 Modulation transfer function of the high-pass filter

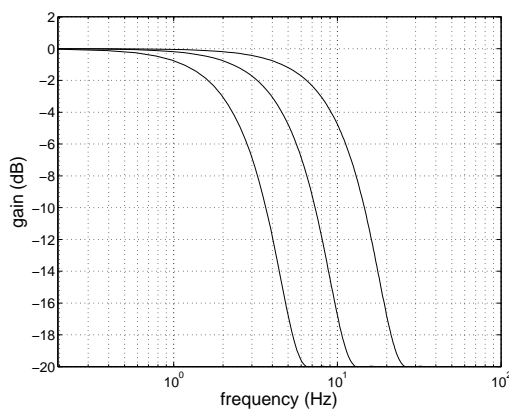


図 4 低域通過フィルタの変調周波数伝達特性

Fig. 4 Modulation transfer function of the low-pass filter

4Hz 付近が重要とされている。高域通過フィルタの遮断周波数は $1, \sqrt{2}, 2$ Hz, 低域通過フィルタの遮断周波数は $2, 4, 8$ Hz で検討した。図 3 に、こうして設計した高域通過フィルタの変調周波数伝達特性を示し、図 4 に、低域通過フィルタの変調周波数伝達特性を示す。これらのフィルタでは、利得が 3dB 低下する周波数を遮断周波数として定義している。図に示されているように、遮断周波数から 1 オクターブはなれた周波数での利得は、低域通過フィルタで約 -12dB, 高域通過フィルタで約 -11dB であり、フィルタとしては、かなり緩やかな遮断特性のであることが分かる。これは文献 [14] で用いている緩やかな遮断特性に相当しており、従来の急峻な遮断特性を用いた知覚に関する知見がそのまま成立するかは明らかではない。本資料では、主観評価実験により、遮断周波数による知覚への影響を明らかにする、また、3 章で求めた平均値正規化スペクトルの変調周波数領域を高域通過フィルタで処理したスペクトログラムを図 5 に、低域通過フィルタで処理したスペクトログラムを図 6 に示す。

5. クロス合成の実装

FFT を用いた高速畳込みに必要なバッファ長を短くす

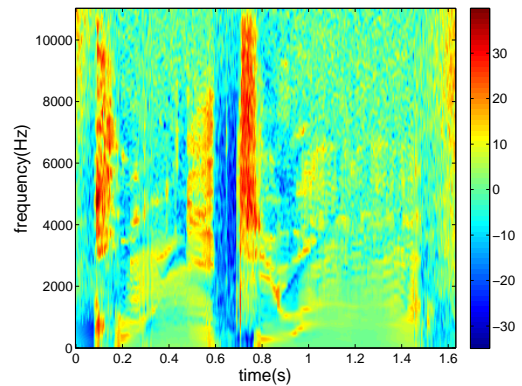


図 5 変調スペクトルをハイパス処理した音声「こんにちは」のスペクトログラム

Fig. 5 spectrogram /koNnichiwa/ by filtering modulation spectrum

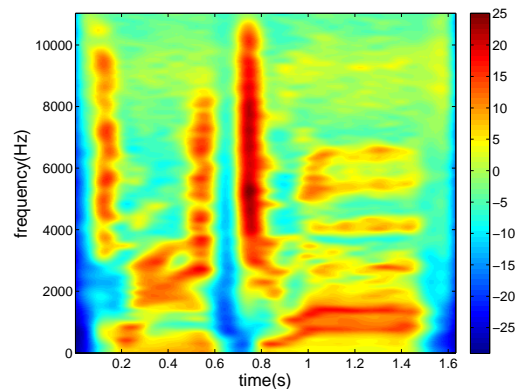


図 6 変調スペクトルをローパス処理した音声「こんにちは」のスペクトログラム

Fig. 6 spectrogram / koNnichiwa/ by filtering modulation spectrum

るため、以上のような処理を施した音声のスペクトルから、複素 cepstrum を介して求められる最小位相インパルス応答 [15] を用いて、楽器音等をフィルタリングする。音声スペクトルはフレーム毎に変化するため、最小位相インパルス応答も毎回変化する。なお、時変フィルタとしての実装では、フレームの接続時の不連続による応答を避けるため、窓長 10 ms の Hanning 窓を用いて切出した楽器音を 5 ms 毎に処理し、OLA(overlap and add) を行った [16]。

6. クロス合成 VOCODER の評価

音声の変調周波数領域をフィルタリングする際の、遮断周波数の設計による効果を、主観評価実験によって調べた。評価対象は、楽器音の音色と、言語情報の保存性についてである。

6.1 音声および楽器音、動物の鳴声の試料

音声試料として、男性 1 名が発声した “ohayogozaimasu” を用いた。楽器音には、「RWC 研究用音楽データベース：

表 1 (行) 高域通過フィルタの遮断周波数,
 (列) 低域通過フィルタの遮断周波数

Table 1 (row)high-pass cut-off filter,
 (column)low-pass cut-off filter

(Hz)	1	$\sqrt{2}$	2
2	a	d	g
4	b	e	h
8	c	f	i

楽器音 [17] から、「C3」「E3」「G3」を組み合わせて作成した、ピアノ (No.01 楽器記号 PF) とエレキギター (No.13 楽器記号 EG) の C3 コードの和音を用いた。動物の鳴声には、「効果音全集」動物. 鳥. 蛙 から「15. ウシ」「17. ウマ」を用いた。いずれの試料も標準化周波数は 44.1kHz である。

6.2 音声の処理条件

音声の分析法として、TANDEM-STRAIGHT を用いた。次いで、提案法である平均値正規化スペクトルを求めた。このようにして求めたスペクトル包絡を変調周波数領域でフィルタリングする。用いたフィルタは 9 通り「a~i」(表 1) で、遮断周波数 1Hz, $\sqrt{2}$ Hz, 2Hz の高域通過フィルタ (図 3) と、遮断周波数 2Hz, 4Hz, 8Hz の低域通過フィルタ (図 4) の組み合わせで検討した。

6.3 被験者

被験者の年齢は 21~24 歳で、男性 6 名、女性 3 名の、計 9 名である。被験者はいずれも日本語を母国語としており、聴力、発話に関する病歴のない健常者である。各被験者には文章にて実験参加の同意を得た後、実験を行った。

6.4 実験方法

実験環境は、防音室 (YAMAHA AVITECS) における MacBook Pro から D/A converter (EDIROL UA-101)、ヘッドフォン “SENNHEISER HD-580” を介して両耳での受聴である。実験刺激である合成音のレベルは、刺激の全長で求めた自乗平均値が 16bit 信号のフルスケールの -26dB となるように設定した。再生システムの音圧レベル (A 特性) は、刺激音圧が HATS (B&K4128) を用いて 60~70dB となるように調整して呈示した。聴取実験は、楽器音と動物の鳴声の音色の保存性についての評価と、音声の言語情報の保存性についての評価を分けて行った。サーストンの一対比較法により、元の楽器音または動物の鳴声、合成音 S1、合成音 S2 の順に呈示した。ここでは、合成音 S1 と合成音 S2 のどちらが、元の楽器音または動物の鳴声の音色に近いかを、評価した。次いで、音声、合成音 S1、合成音 S2 の順に呈示した。ここでは、合成音 S1 と合成音 S2 のどちらが、言葉が聞き取り易かったかを、評価した。

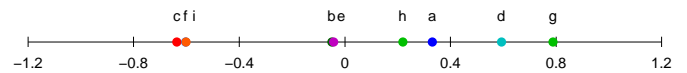


図 7 楽器音と動物の鳴声の音色の保存性についての実験結果

Fig. 7 result about preserves characteristic timbre of musical instruments and animal voices

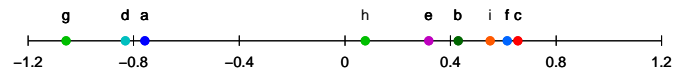


図 8 言語情報の保存性についての実験結果

Fig. 8 result about preserves linguistic information of musical instruments and animal voices

6.5 主観評価実験結果

図 7 に聴衆実験の楽器音の音色の保存性についての評価結果を示す。図の横軸は、楽器音の音色の保存性に関する間隔尺度を示す。スコアが大きいほど、音色がより元の楽器音の音色に近いことを表している。図 8 に聴衆実験の言語情報の保存性についての評価結果を示す。図の横軸は、言語情報の保存性に関する間隔尺度を示す。スコアが大きいほど、合成音の言葉が聞き取り易いことを表している。これにより、フィルタの遮断周波数により、楽器音や鳴声の音色の保存性と、音声の言語情報の保存性を制御できることが示された。

また、図 9, 図 10 は、音声と合成する楽器音や鳴声ごとに結果を示している。図の横軸は設定したフィルタ a~i、縦軸は間隔尺度を示す。ここでは、低域通過フィルタの遮断周波数よりも、広域通過フィルタの遮断周波数の効果が大きい事が明らかになった。また、音色の保存性と言語情報の保存性は、楽器音や動物の鳴声に関わらない同様の傾向を示した。

さらに、図 11 は、横軸はフィルタ a~i、縦軸は間隔尺度として、楽器音の音色と言語情報の保存性の関係性を示している。楽器音や鳴声の保存性と、音声の言語情報の保存性は、遮断周波数の設定に対して、逆の傾向を示した。これらの結果は、クロス合成で、言語情報を重視したいか、元の楽器や鳴声の音色を重視したいか、目的に応じて設計するための基礎資料として用いることができる。

7. まとめ

本報告では、音楽の分野の “VOCODER” として、「元の楽器らしさ」や「元の動物の鳴声らしさ」と「音声の言語情報」の割合を自由に調整できる、新しいクロス合成 VOCODER を提案した。提案法では、まず、「元の楽器らしさ」や「元の動物の鳴声らしさ」を維持するには、如何にして音声のスペクトルから楽器音等の音色を変化させてしまう成分を取り除くかを明らかにする必要があった。そこで、音声スペクトルの大局的な形状を変化させる成分を音声から取り除き、次いで、音声スペクトルの時間変化に

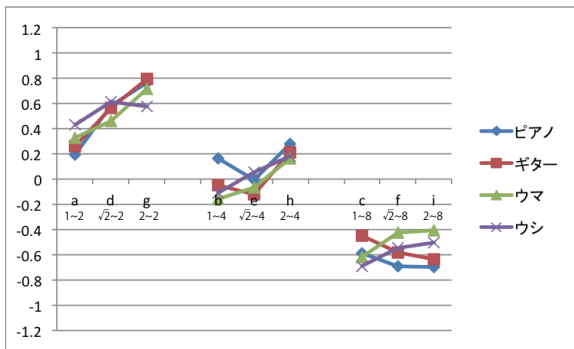


図 9 楽器と鳴声の音色の保存性についての実験結果

Fig. 9 result about preserves characteristic timbre of Instruments and animal voice

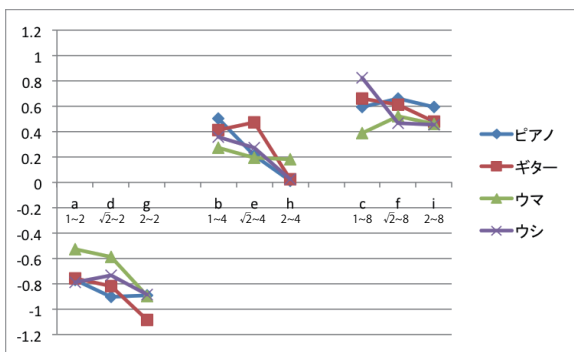


図 10 言語情報の保存性についての実験結果

Fig. 10 result about preserves linguistic information

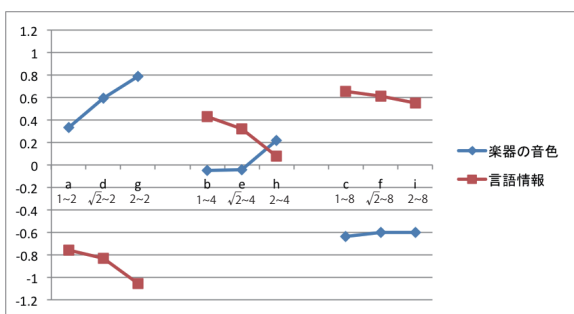


図 11 楽器音等の音色と言語情報の保存性についての実験結果

Fig. 11 preserves linguistic information and characteristic timbre of musical instruments and animal voices

注目することで、言語情報を伝える上で重要ではない時間的な変化を取り除いた。聴衆実験では、これらの操作に含まれるパラメタと、クロス合成音から知覚される「元の楽器らしさ」や「元の動物の鳴声らしさ」および、「音声の言語情報」との関係性を調べた。その結果、クロス合成で、言語情報を重視したいか、元の楽器や鳴声の音色を重視したいか等、目的に応じた設計法を明らかにした。

謝辞 本研究の一部は、和歌山大学平成 23 年度独創的研究支援プロジェクトの支援を受けた。また、併せて科学研究費補助金の支援を受けた。

参考文献

- [1] H. Dudley: Remaking speech, J.Acoust.Soc.Am., vol.11, no.2, pp.169-177, 1939.
- [2] J. L. Flanagan: Phase vocoder, the Bell System Technical Journal, pp.1493-1509, 1966.
- [3] 板倉文忠, 齋藤収三: 統計的手法による音声スペクトル密度とホルマント周波数の推定, 電子通信学会論文誌. A 53(1), pp.35-42, 1970.
- [4] B. S. Atal, S. L. Hanauer: Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoust. Soc. Am., vol.50, 2B, pp.637-655, 1971.
- [5] C. Roads: The Computer Music Tutorial, The MIT Press, 2002.
- [6] H. Kawahara, I. Masuda, and A. deCheveigné: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, Speech Communication, vol. 27, no. 3-4, pp.187-207, 1999.
- [7] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, H. Banno, "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," Proc. ICASSP 2008. Las Vegas., pp.3933-3936, 2008.
- [8] 森勢将雅, 高橋 徹, 河原 英紀, 入野 俊夫. 窓関数による分析時刻の影響を受けにくい周期信号のワースペクトル推定法, 電子情報通信学会 論文誌 D, vol.J 90-D, No.12, pp.3265-3267, 2007.
- [9] Hideki Kawahara and Masanori Morise, Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, Sadhana, Vol. 36, Part 5, October 2011, pp. 713-727
- [10] 赤桐隼人, 森勢将雅, 入野俊夫, 河原英紀. スペクトルピークを強調した F0 適応型スペクトル包絡抽出法の最適化と評価. 電子情報通信学会 論文誌 A, Vol. J94-A, No. 8, pp. 557-567, 2011.
- [11] B.C.J. Moore, "聴覚心理学概論," 誠信書房, 1994.
- [12] R. Drullman, J.M. Festen, and R.Plomp: Effect of temporal envelope smearing on speech reception, J.Acoust. Soc. Am.,vol.95, no.2, pp.1053-1064, Feb,1994.
- [13] R. Drullman, J.M. Festen, and R.Plomp:Effect of reducing slow temporal modulations on speech reception., J.Acoust.Soc.Am., vol.95, no.5, pp.2670-2680, May,1994.
- [14] 金寺 登, 荒井隆行, 船田哲男. 変調スペクトルの重要な成分のみを選択的に用いた雑音に強い音声認識, 信学論 (D-II), Vol.J84-D-2, No.7, pp.1261-1269, 2011.
- [15] 今井 聖: 音声信号処理-音声の性質と聴覚の特性を考慮した信号処理-, 森北出版, (1996).
- [16] 河原英紀, 和田芳佳, 西大輝, 森勢将雅, 西村竜一, 入野俊夫. "音声分析変換成系における時変フィルタの実装と駆動情報の表現について," 日本音響学会聴覚研究会資料, Vol.41, No.7, pp.561-566, 2011.
- [17] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, Vol.45, No.3, pp.728-738, March 2004.