2N-9

直感的な音探索のための 音響データベースシステムと音響モーフィング

魚田　慧†　　橋本　周司†
早稲田大学大学院　理工学研究科†

１．まえがき

現在よく使われている音響データベースは，作業者の手作業によりテキストのインデックスを音響に付加したものであり，適切なキーワードや音源名を関連付ける作業が煩雑である．また，複数の作業者がキーワードを付加するため，キーワードの付け方が個々の主観的な評価に大きく依存してしまう．これに対して，筆者らは音響波形から自動抽出した特徴量をインデックスとして，音響と擬音語を検索キーとするデータベースシステムを開発してきた[1]-[3]．このシステムはユーザに複数の候補音を提示し，ユーザが選択した音響をもとに検索を繰り返すことで対話的な音探索が可能である．しかしながらユーザが望む音響がデータベース内に存在しない場合，検索を繰り返しても所望の音響を得ることはできない．この問題点を解決するために，我々は音響モーフィングによる音響合成手法に着目した．音響モーフィングによって新規音響を生成することで，データベース内にない音響をも探索範囲に含めることができ，ユーザは望む音響にさらに近づくことが可能となる．

本稿では提案システムの概要とモーフィング手法について述べ，導入したモーフィング音響の音質評価実験について報告する．

２．システム概要

図１に提案システムの概要を示す．図中の a～f は音響探索の流れを示している．

まず，ユーザは擬音語（入力部 a）あるいは音響を検索キーとしてシステムに入力する．入力された検索キーが擬音語である場合，処理部 b によって検索キーとなる音響を生成する．音響だけでなく擬音語を検索キーとすることで，より直感的かつ簡単に検索キーの入力を行える[2][3]．擬音語での入力は検索キーとなる類似音を用意することが困難な場合，非常に有力な手段である．

入力された検索キーは処理部 c によって自動的に特徴量抽出される．特徴量は音高・音量・音色の時間変化とその時間に関する微分値に加え[2]，全周波数帯域のパワースペクトルのうち

Sound Database System and Audio Morphing for Intuitive Sound Retrieval
† Kei Uota　Dept. of Applied Physics, Waseda University
† Shuji Hashimoto　Dept. of Applied Physics, Waseda University

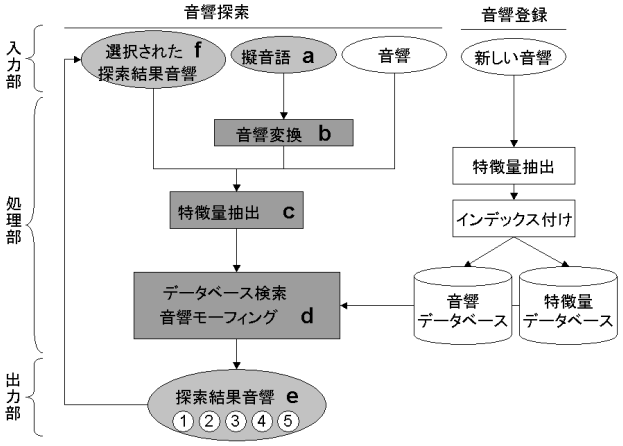

図１．システム概要

基本周波数のパワーの占める割合を採用している[1]．この特徴量はピッチの抽出度を表しており，この値が高ければ単一のピッチを有する音響（例：楽器音）になり，この値が低ければ単一のピッチを抽出できない音響（例：ホワイトノイズ）を表す．

抽出された特徴量をもとに，システムはデータベースを検索し，探索結果音響として最も特徴の類似した５つの音響を出力する（処理部 d）．ユーザは 5 つの音響のうち，所望する音響に近いと感じた音響を一つ選択する（出力部 e）．それに対してシステムは選択された音響を次の検索キー（入力部 f）として検索しユーザに結果を提示することを繰り返す．これにより，ユーザはシステムが提示した音響を試聴し選択するだけで，音響や信号処理の専門的知識と経験を必要とすることなく直感的にシステムを利用することができる．

システムは検索を繰り返すうちに，探索される特徴量空間の範囲が閾値よりも小さくなったら，探索結果音にモーフィング音響を加えてユーザに提示する（処理部 d）．モーフィングの対象となる 2 音は，過去にユーザが選択した複数の探索結果音の中から，特徴量空間内で現在の探索範囲の中心位置にモーフィング音響がくるように選ばれる．これにより，データベースに存在しない音響を含めた，より小さい特徴量空間内での音響探索が可能となる（図 2）．

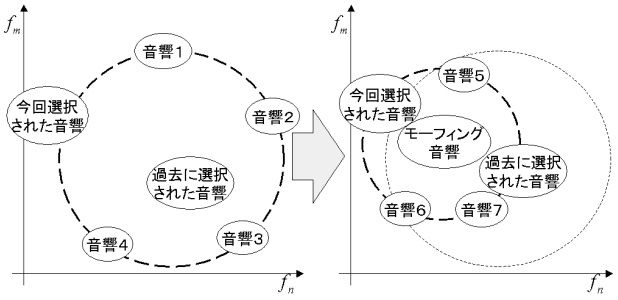

図 2. 探索範囲の変化

3. モーフィング

本システムの音響モーフィングには，代表的なモーフィング手法である短時間スペクトル上での補間を用いている[4]．システムによって選ばれた 2 音に対し，以下の処理によって音響モーフィングを行う．

まず，対象となった 2 音に対してそれぞれ短時間スペクトルを計算する．窓関数に三角窓を使い，オーバーラップを大きくすることにより，FFT・IFFT 処理によって生じる雑音を最小限にしている．次に，音量の時間変化に対して DP マッチングを用いて時間軸のマッチングをとる．周波数軸は，基本周波数とその倍音成分を抽出し，マッチングをとる．システムは補間係数αを指定し，時間長，音量，基本周波数，音色を補間する．モーフィング音響は補間係数αが 0 なら音響 A と等しく，1 なら音響 B と等しくなる．時間長 T は式（1）によって補間される．

$$T_m = (1-\alpha)T_a + \alpha T_b \qquad (1)$$

音量 A，基本周波数 F の補間に関しては，時間軸でマッチングされたフレーム毎にそれぞれ式（2），式（3）によって求められる．

$$A_m = (1-\alpha)A_a + \alpha A_b \qquad (2)$$

$$F_m = (1-\alpha)F_a + \alpha F_b \qquad (3)$$

音色は時間軸でマッチングされたフレーム毎，正規化周波数毎に式（4）によって補間される．

$$S_m = (1-\alpha)S_a + \alpha S_b \qquad (4)$$

最後に得られたスペクトルを逆フーリエ変換して事でモーフィング音響を得る．

4. 実験

モーフィング音響の音色評価を試みた．実験に用いた音響はモノラル音源で，サンプリングレートは 22050Hz，量子化ビットは 16bit，FFT フレーム長は 1024 点，オーバーラップは 128 点である．これにより，周波数分解能は 22Hz，オーバーラップによる見かけ上の時間分解能は 46ms となる．

図 3 にモーフィング音響のスペクトルを示す．

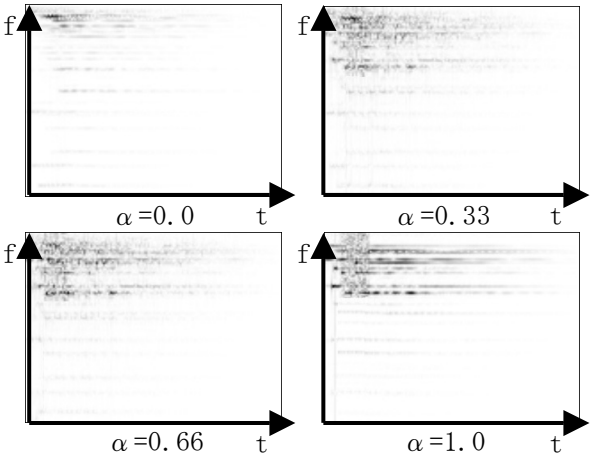

図 3. モーフィング音響のスペクトル

αの変化に伴い，スペクトルが滑らかに変化していることがわかる．

本手法では，音量変化を用いて時間軸マッチングしているため，モーフィング対象の 2 音の音量変化がある程度類似している音響でないとモーフィングが滑らかに変化しない．また，また，うまく基本周波数がとれない音響に対するモーフィングも雑音が大きくなるという問題もあるが，検索結果を元に新しい音響の生成が可能であることが判った．

5. まとめ

音響モーフィングを用いて，ユーザが音響を探索する際にデータベースに存在しない音響をも提示できる音響データベースシステムを構築した．また，そのモーフィング部分についての音質評価実験を行った．今後は，探索システムについてのユーザの主観評価実験を行い，より少ない探索回数で所望の音響を提示できるシステムを構築していきたい．

6. 参考文献

[1] Qi Hai, Pitoyo Hartono, Kenji Suzuki and Shuji Hashimoto, “Sound Database retrieved by sound,” Acoust. Sci. Tech., Vol. 23, No. 6, pp. 293-300, 2002.

[2] 魚田 慧，鈴木 健嗣，Pitoyo Hartono，橋本 周司，“擬音語と音響を用いた音響データベースの検索，” 電子情報通信学会 HCS 研究会，2004.11.

[3] 魚田 慧，橋本 周司，“擬音語と音響を用いた音響データベースの直感的な音探索，” 電子情報通信学会全国大会，2006.03.

[4]Slaney, M., Covell, M., and Lassiter, B., “Automatic Audio Morphing,” Proceedings IEEE International Conference Acoustics, Speech and Signal Processing, Vol.2, 1996.