twitterアカウント: @VocaListener
twitterハッシュタグ: #vocalis

VocaListener2（ぼかりす2）

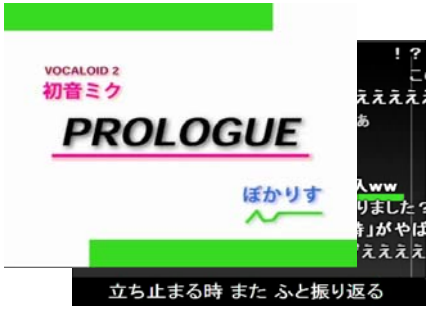
ユーザ歌唱の音高と音量だけでなく 声色変化も真似る歌声合成システムの提案

中野 倫靖, 後藤 真孝
（産業技術総合研究所）

2010年7月28日
第86回音楽情報科学研究会(SIGMUS)

VocaListener1（ぼかりす1）の実現

- 2008年4月28日
  - デモ動画をWeb掲載
    - 【初音ミク】PROLOGUE【ぼかりす】
      - http://staff.aist.go.jp/t.nakano/VocaListener/index-j.html
      - http://www.nicovideo.jp/watch/sm3128145



- 2008年5月28日
  - 中野, 後藤: **VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案**. 2008-MUS-75.

ぼかりす1で解決を目指した課題

- 2007年8月31日
  - VOCALOID2「初音ミク」発売

楽譜と歌声合成パラメータの入力を支援

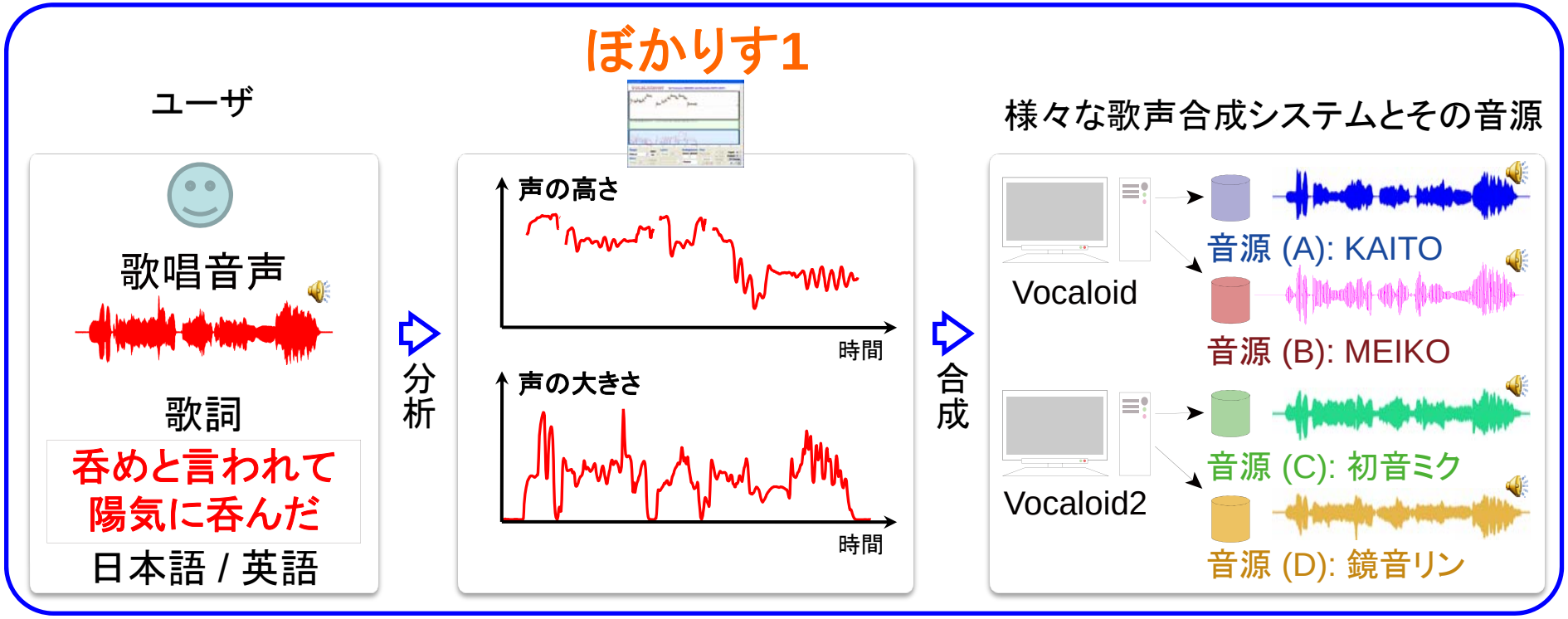
- 2007年12月27日
  - VOCALOID2「鏡音リン・レン」発売

歌声合成システムや音源（歌手の声）を変えた場合

歌声合成パラメータを自動的に再調整

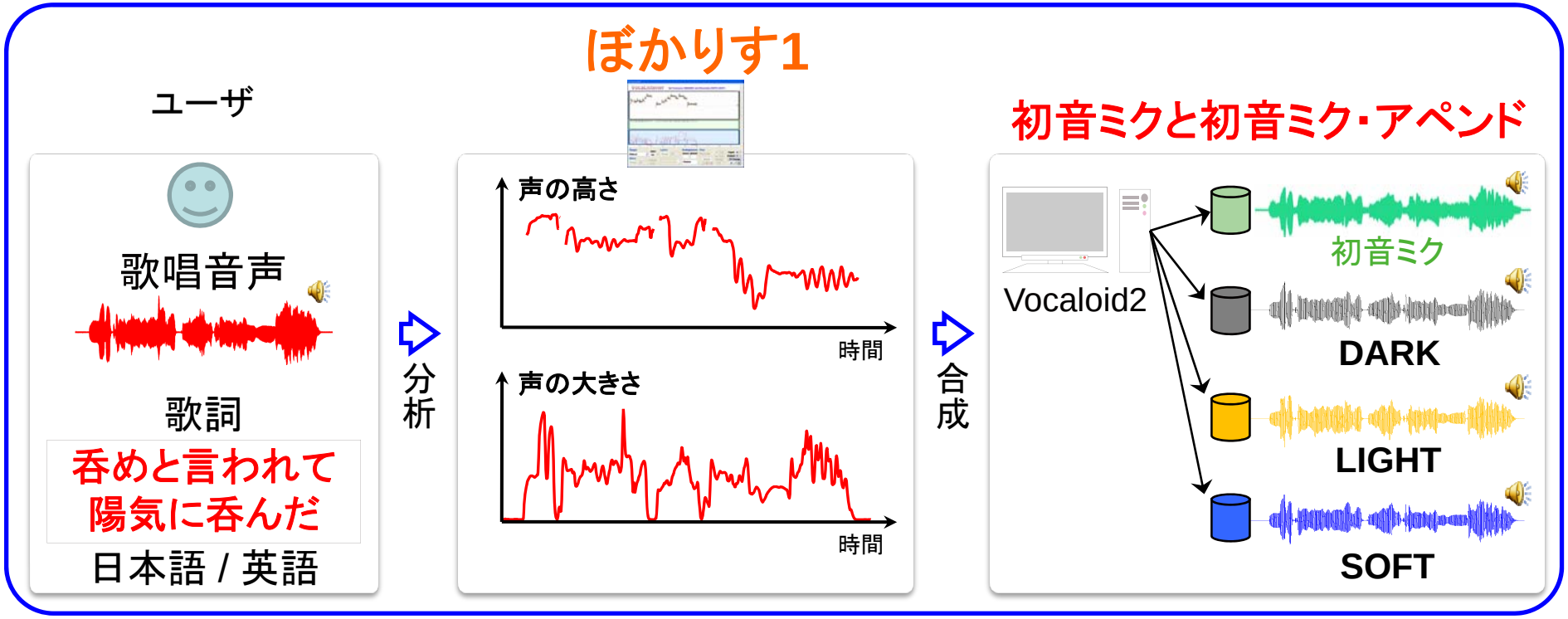
ぼかりす1で可能になったこと

- 歌うだけで楽譜・歌声合成パラメータ推定
  - 人間らしい歌い方で合成
- 歌声合成システムや音源を替えても同じ歌い方



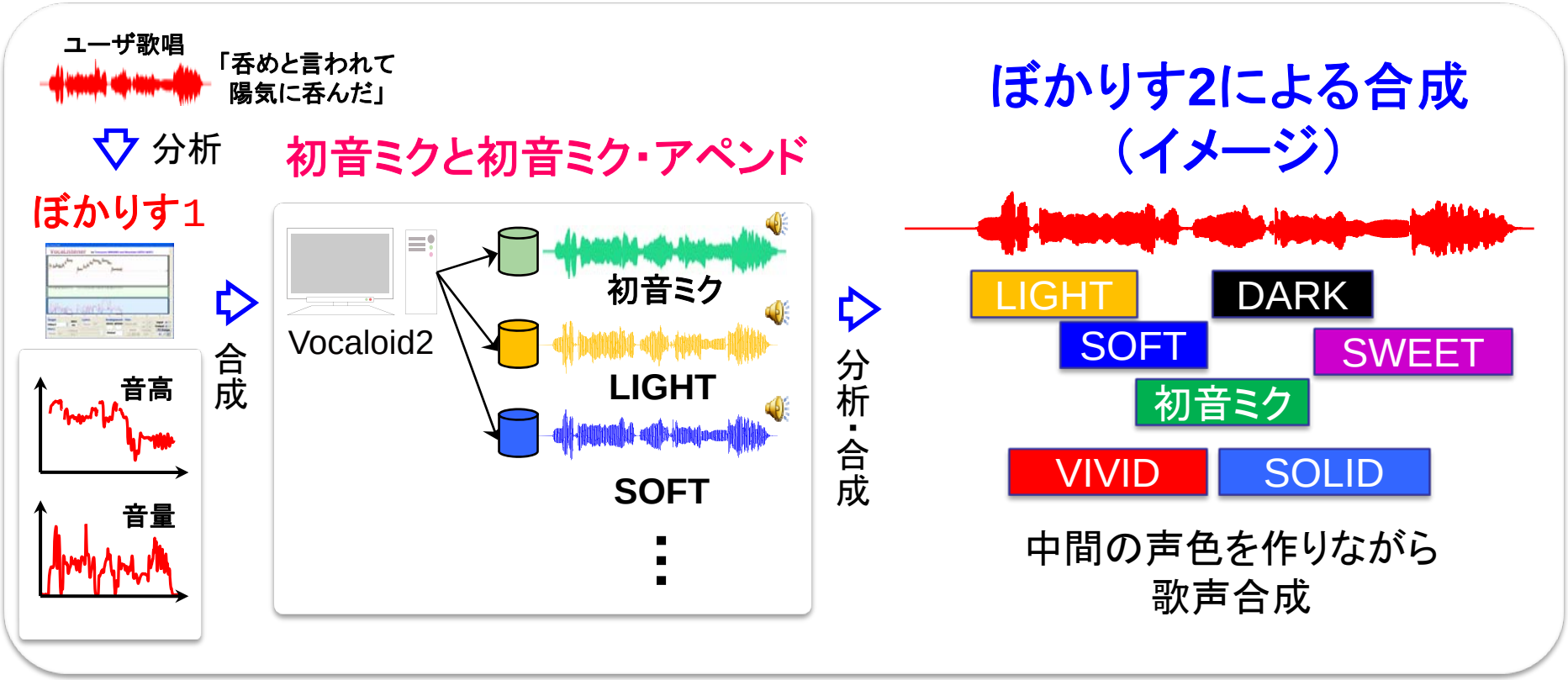
2010月4月30日:「初音ミク・アペンド」発売

- 初音ミクの声色(6種類)を替えて合成可能
  - DARK, LIGHT, SOFT, SWEET, SOLID, VIVID



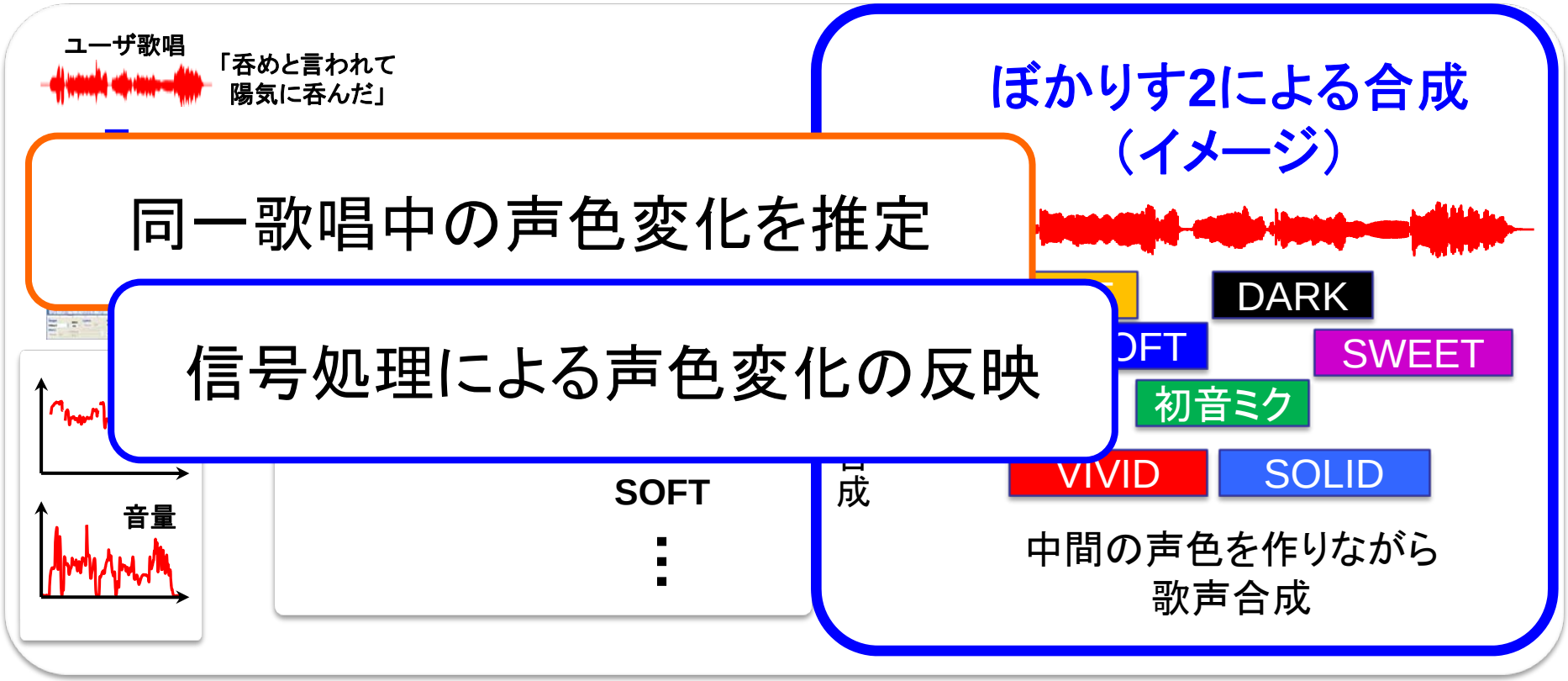
ぼかりす2で解決を目指す課題

- ユーザ歌唱の**声色変化も真似る**ように
  声色を滑らかに変化させながら歌声合成



ぼかりす2で解決を目指す課題

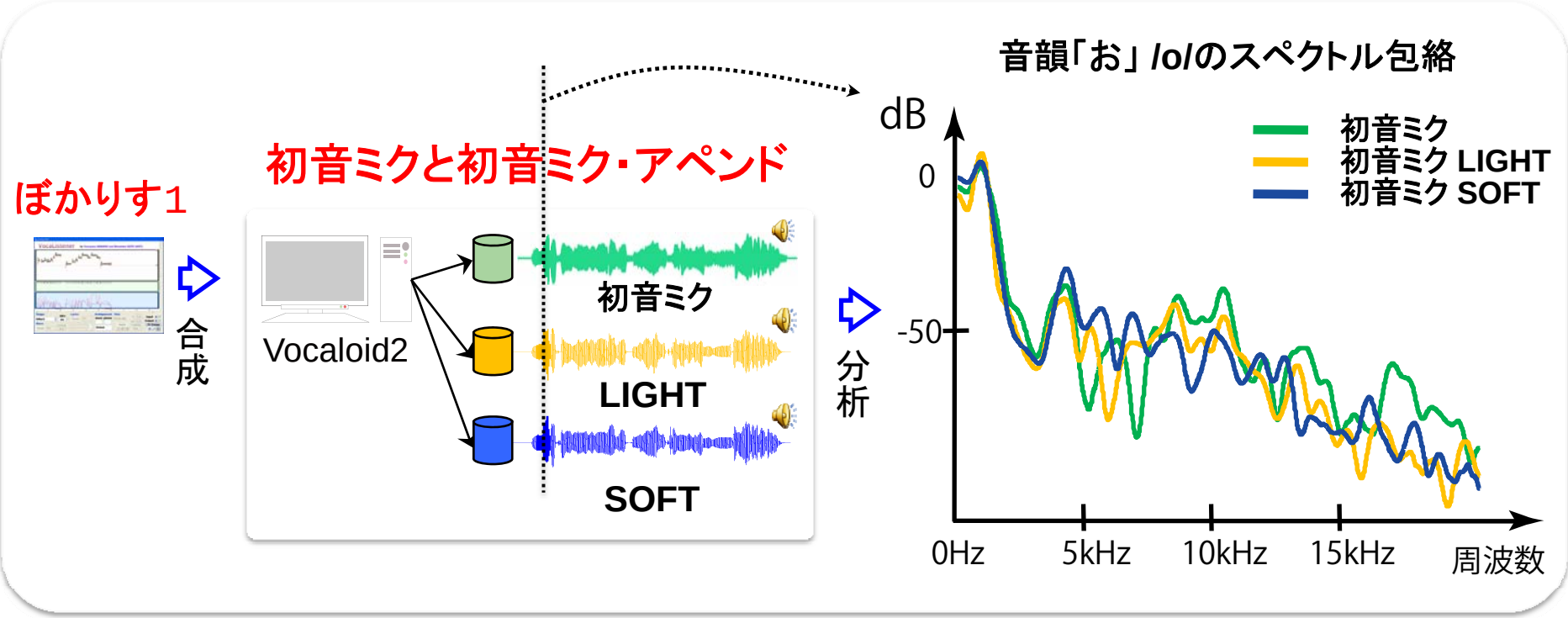
- ユーザ歌唱の**声色変化も真似る**ように
  声色を滑らかに変化させながら歌声合成



ぼかりす2
の
実現方法（概要）

物理的な現象としての声色の違い

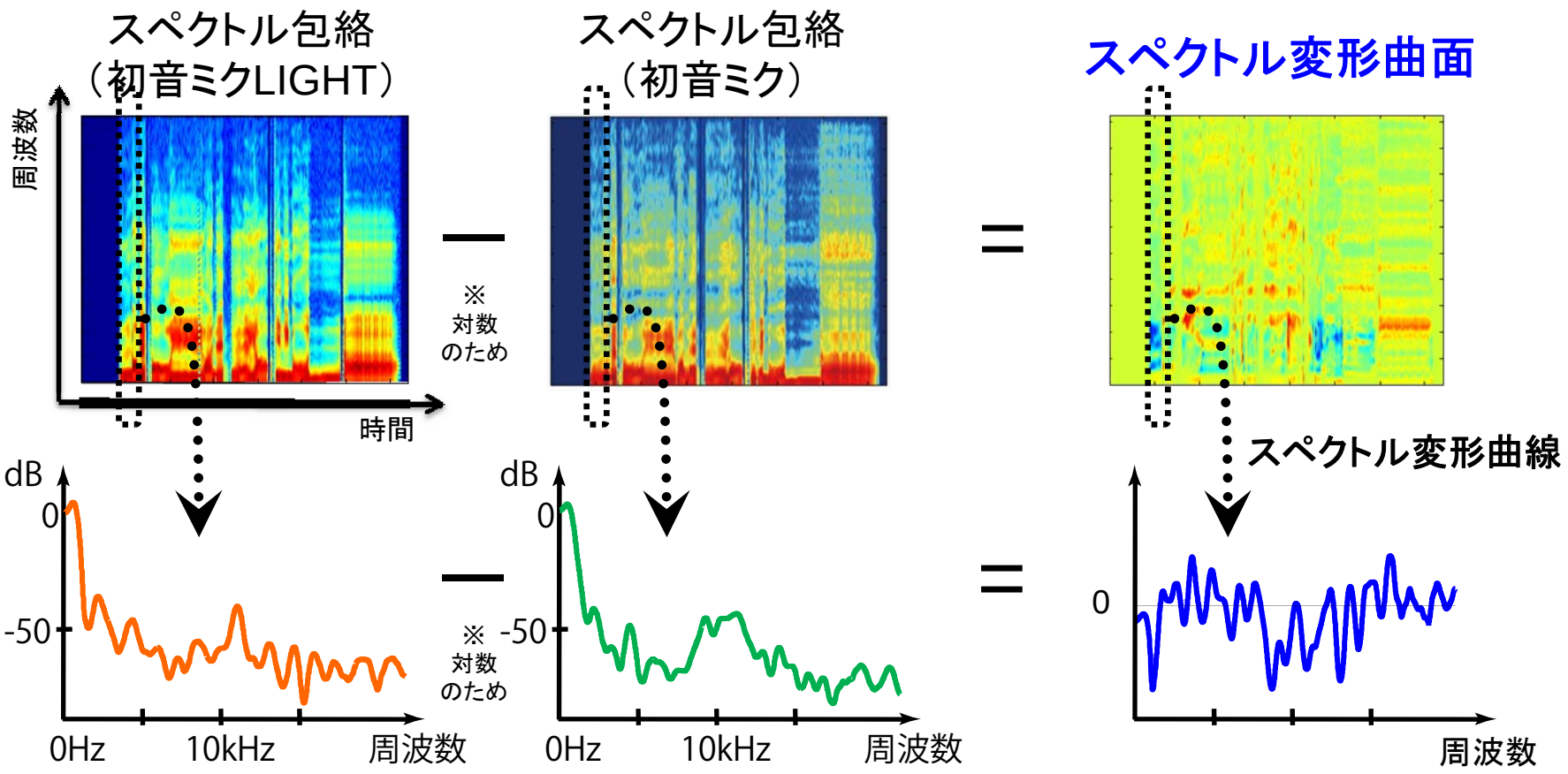
- スペクトル包絡（声道特性）形状の違い
  - ぼかりす1による時刻同期した歌唱の場合



※スペクトル包絡はSTRAIGHT [Kawahara *et al.*, 1999] により推定

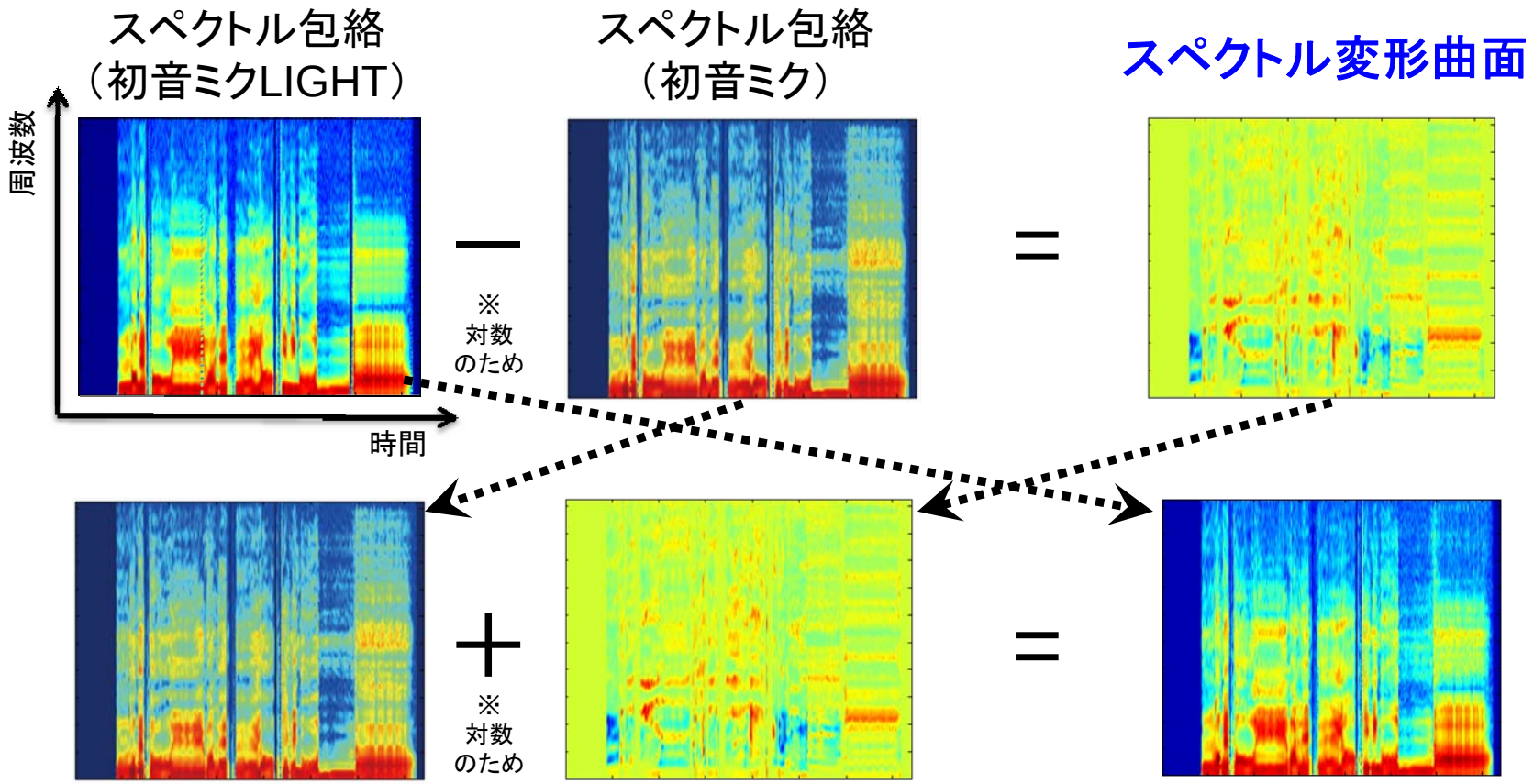
スペクトル包絡変形に基づく初音ミクLIGHT合成

スペクトル変形曲面（相対的な違い）の導入



スペクトル包絡変形に基づく初音ミクLIGHT合成

スペクトル変形曲面（相対的な違い）の導入

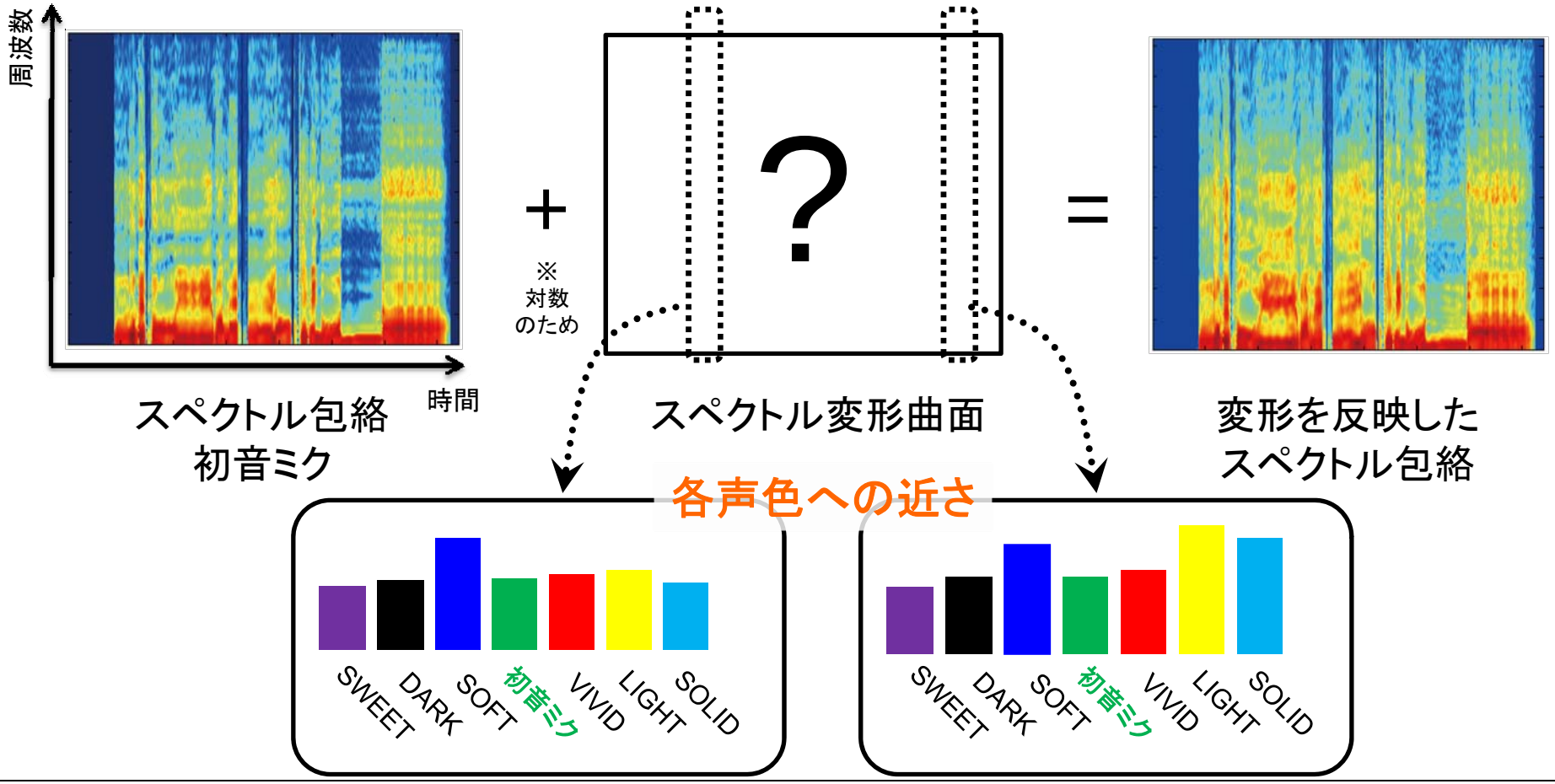


スペクトル包絡変形に基づく初音ミクLIGHT合成

ロ スペクトル変形曲面（相対的な違い）の導入

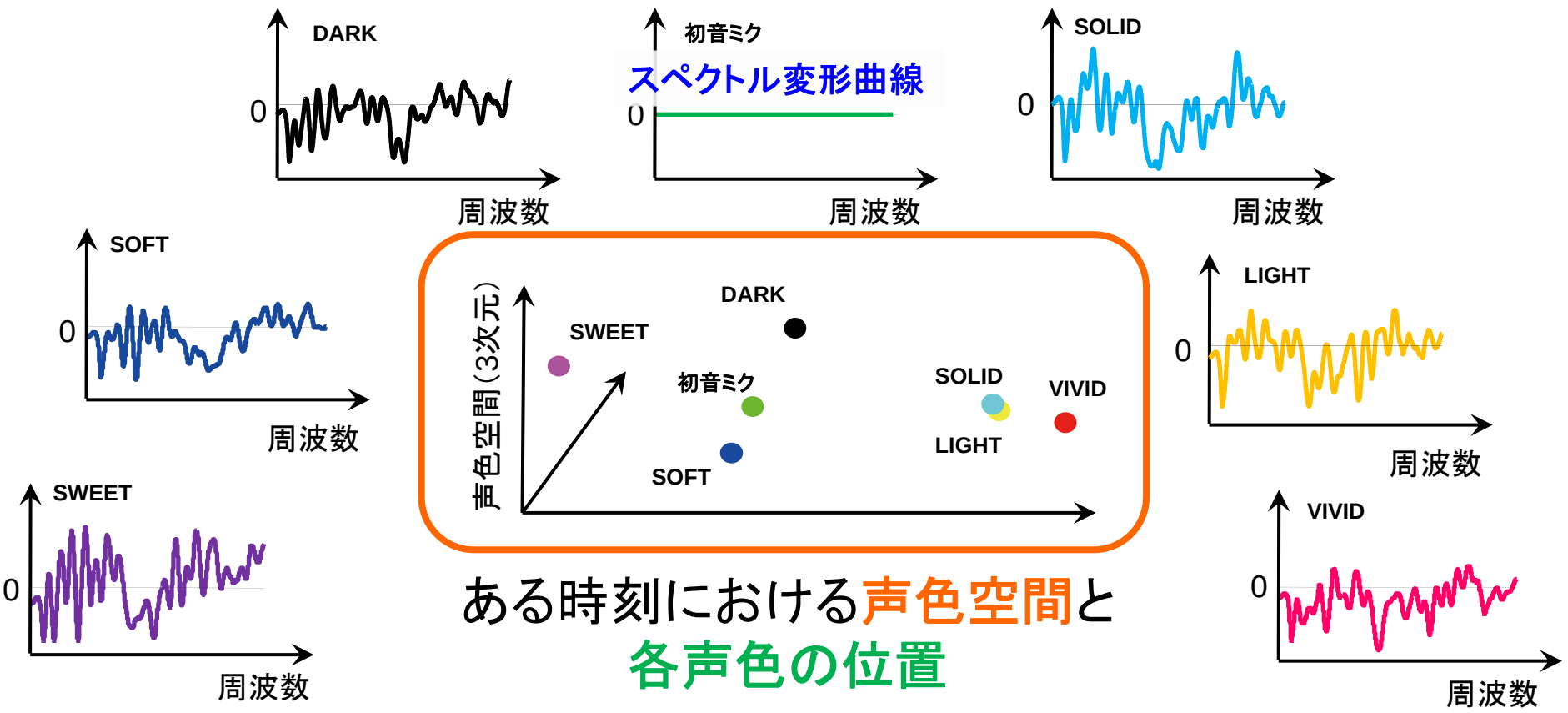
初音ミクを基準として滑らかに変形して合成

- ぼかりす2ではスペクトル変形曲面を推定



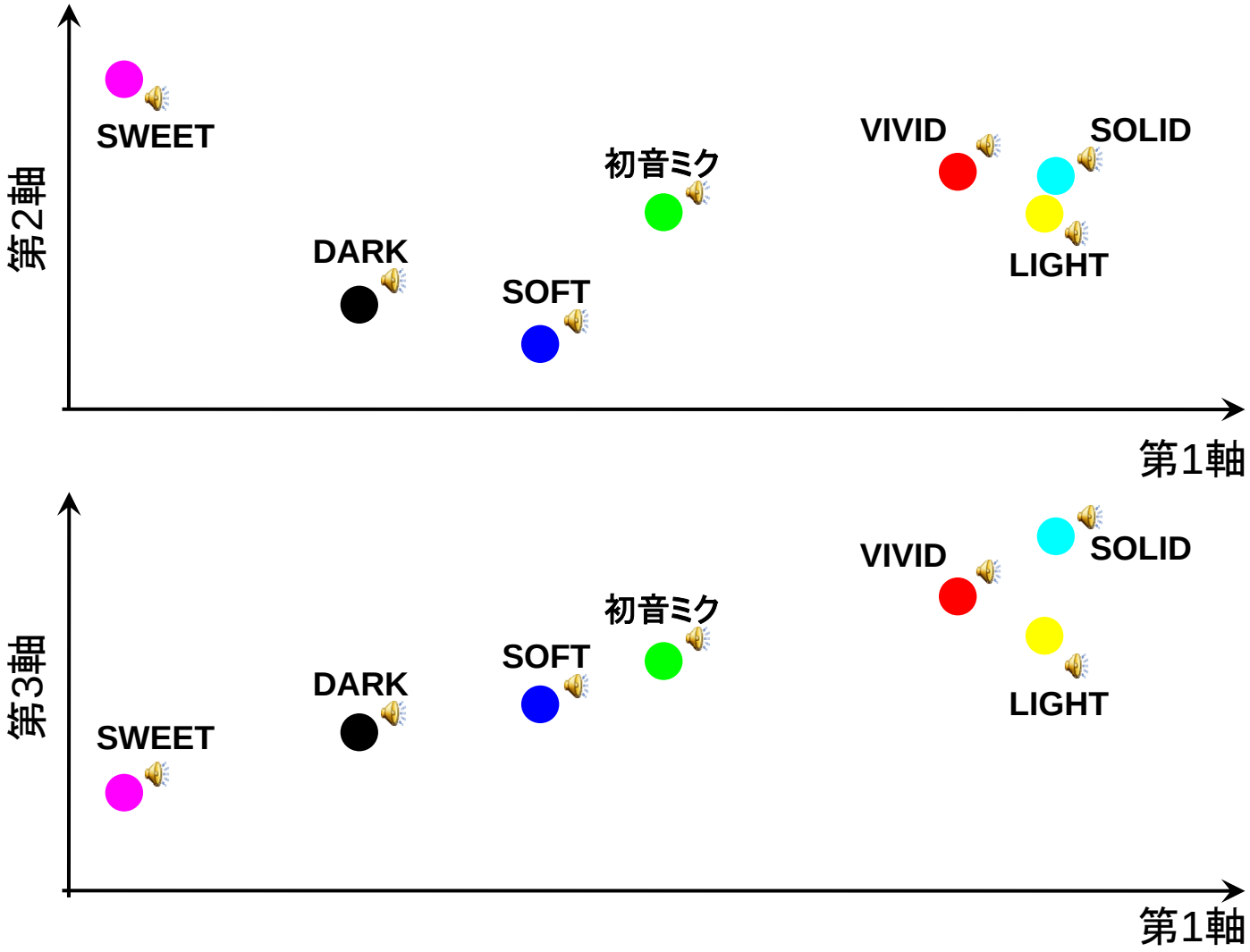
声色空間に基づくスペクトル変形曲線の推定

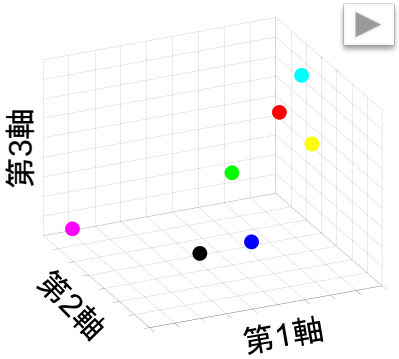
- スペクトル包絡を分析して声色空間（3次元）へ射影



ある時刻における声色空間と
各声色の位置

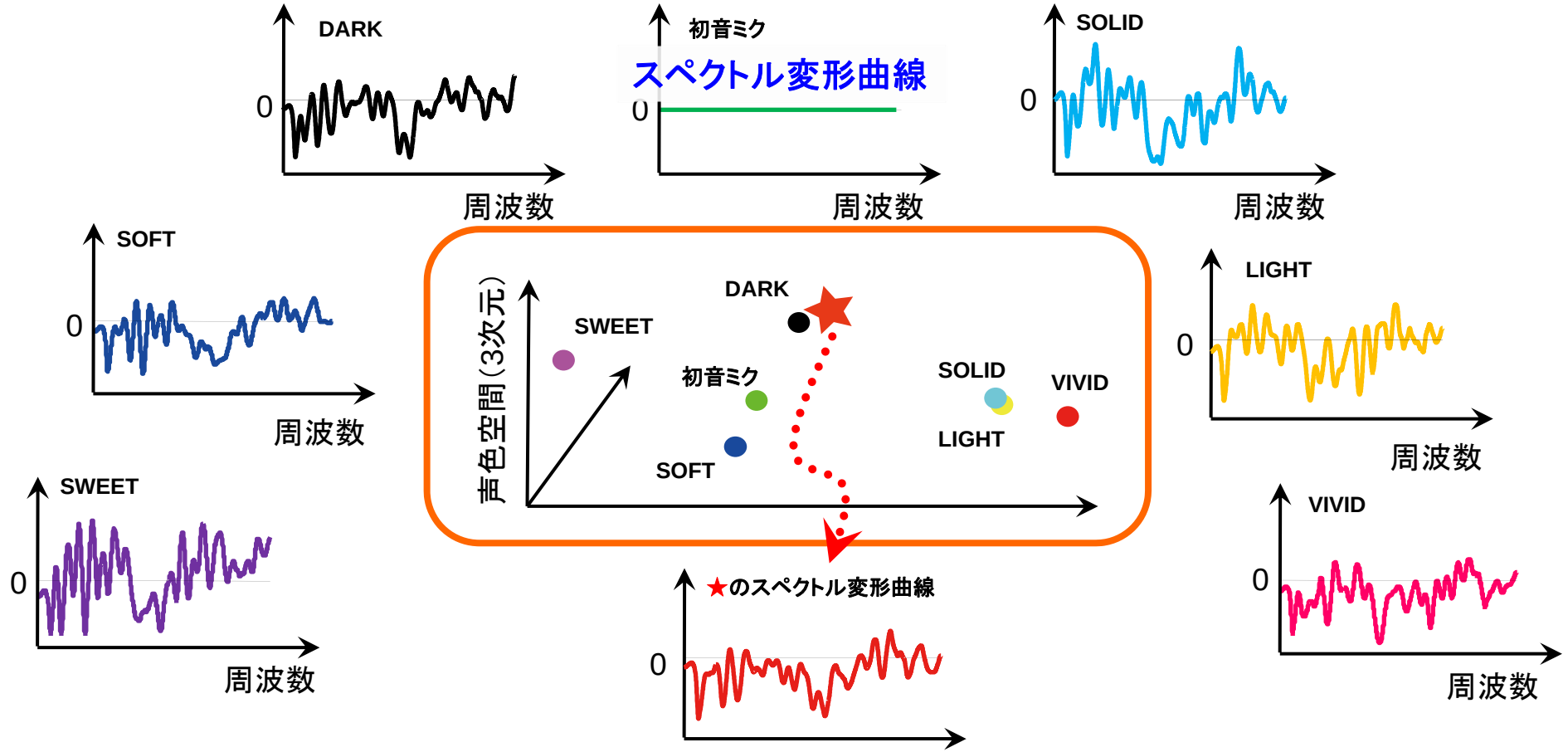
声色空間の構成結果（全時刻の平均）





声色空間に基づくスペクトル変形曲線の推定

- スペクトル包絡を分析して声色空間（3次元）へ射影

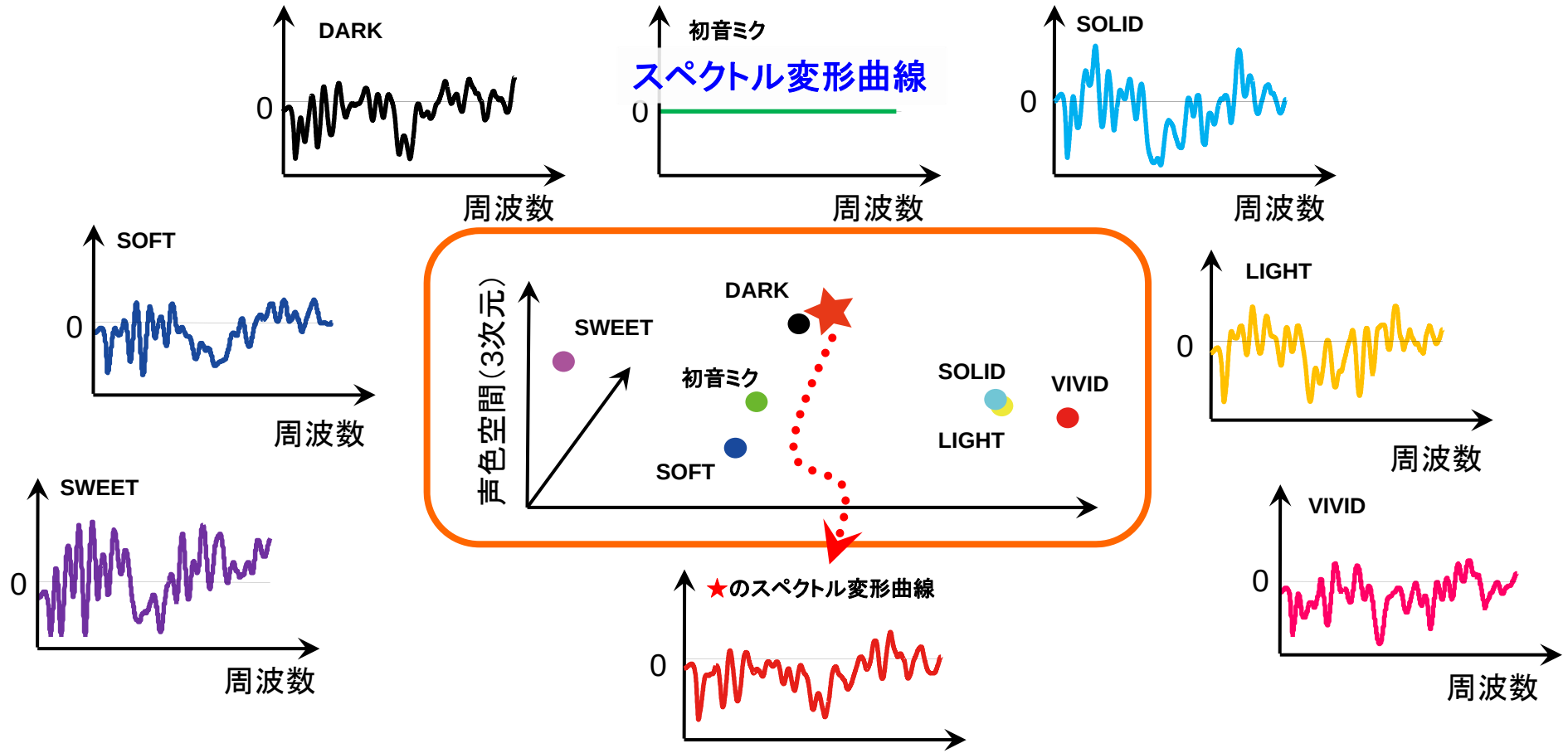


デモ：合成結果と声色空間

- 【初音ミク】大漁船【ぼかりす2】
  - http://staff.aist.go.jp/t.nakano/VocaListener2/index-j.html
  - http://www.nicovideo.jp/watch/sm11523161

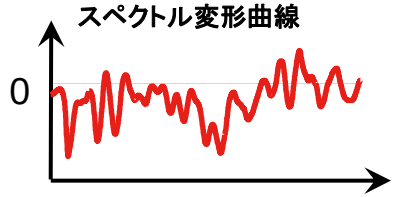
声色空間に基づくスペクトル変形曲線の推定

- スペクトル包絡を分析して声色空間（3次元）へ射影

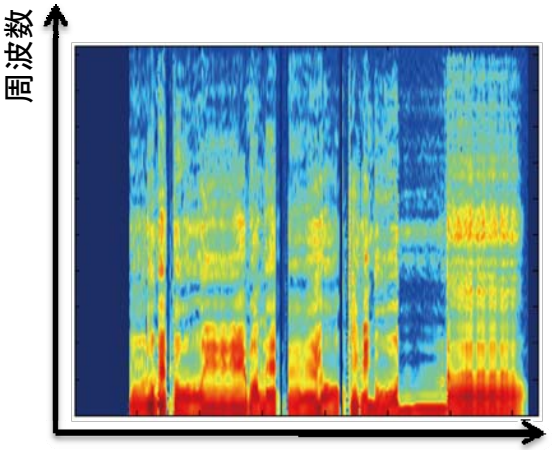


スペクトル変形曲面の生成と適用

- 全時刻のスペクトル変形曲線を合わせて生成

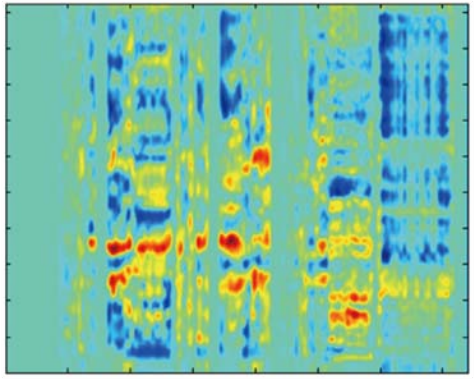


全時刻で推定



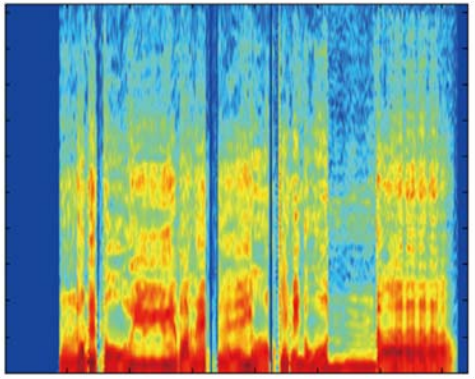
スペクトル包絡
「大漁船」
初音ミク

＋
※
対数
のため



スペクトル変形曲面

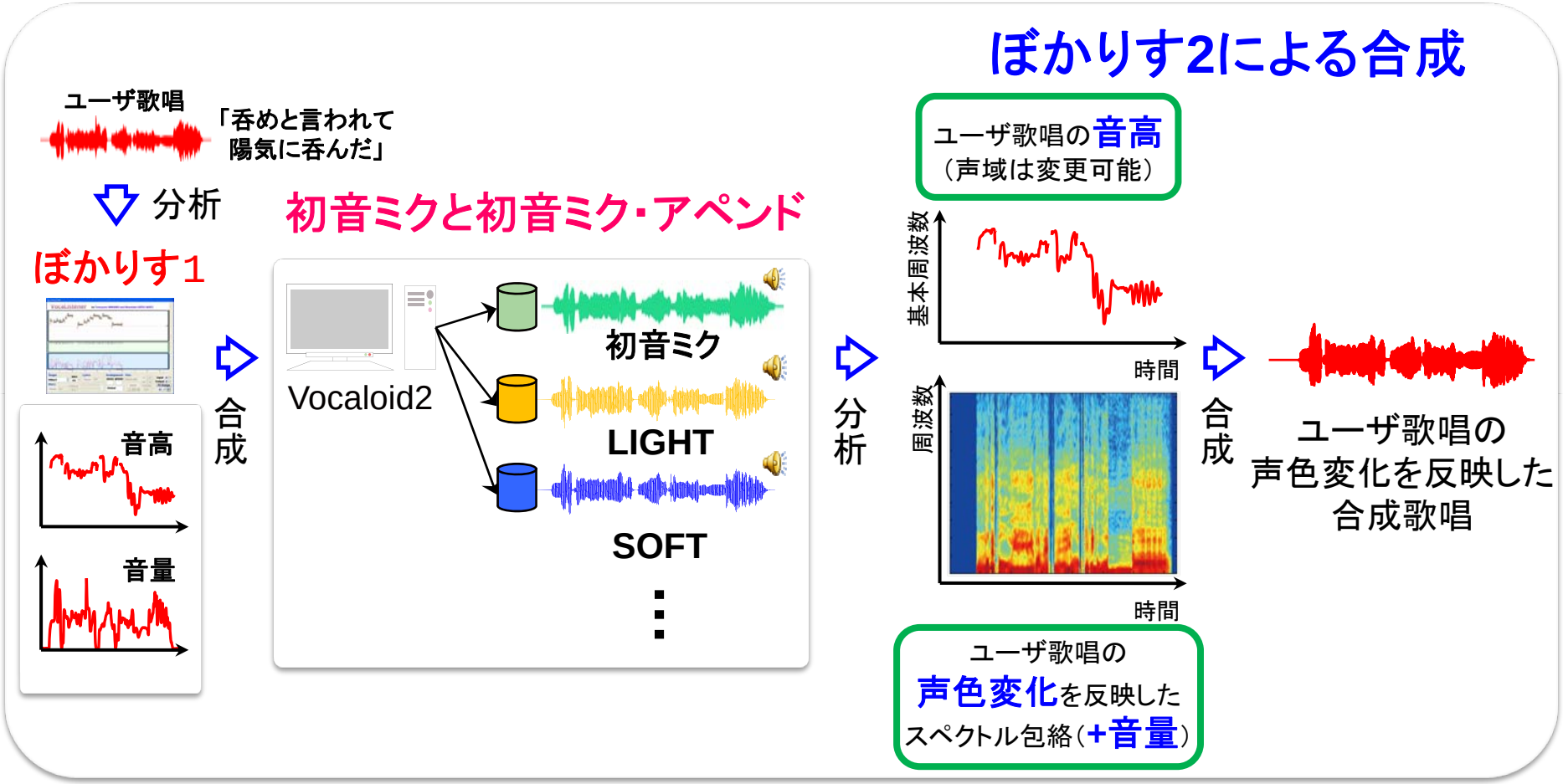
＝



ユーザ歌唱の
声色変化を反映した
スペクトル包絡

ぼかりす2による歌声合成

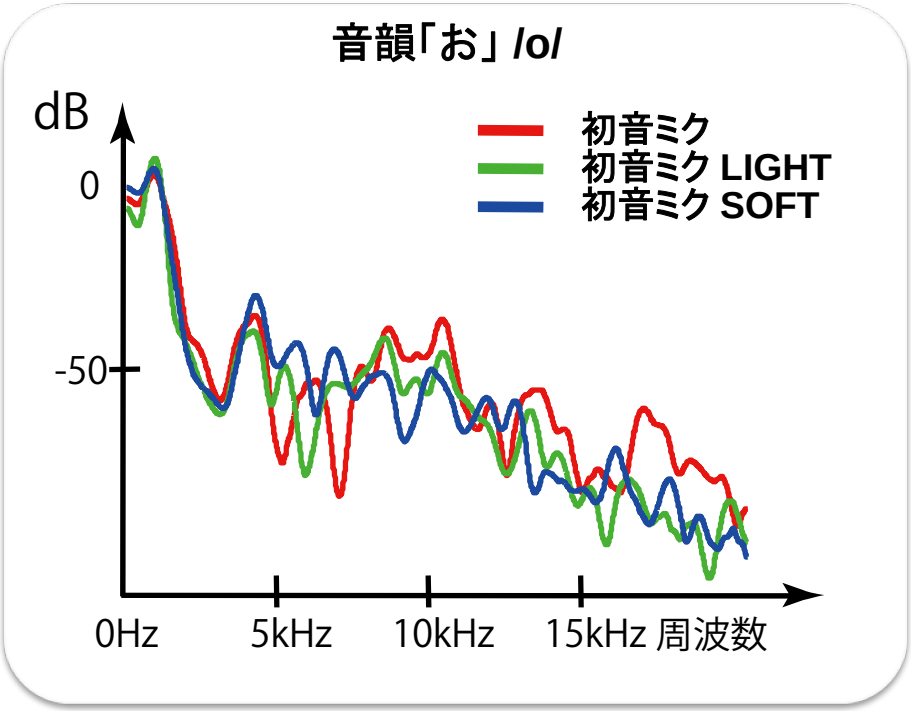
ユーザ歌唱の音高・音量・声色変化を真似る



ぼかりす2
の
実現方法（詳細）

ユーザ歌唱の声色変化を真似る

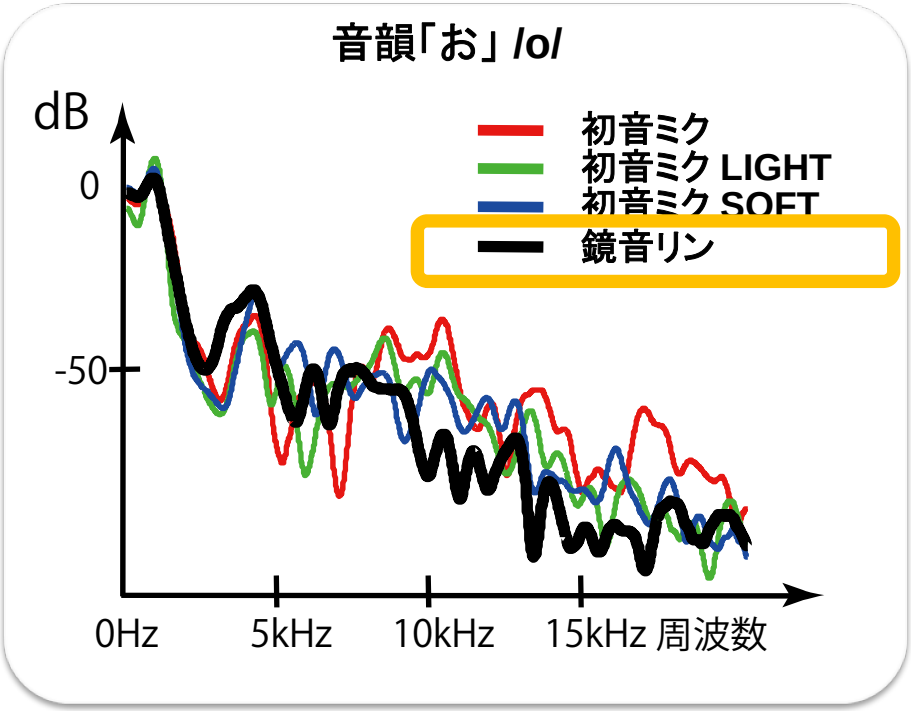
- ユーザ歌唱のスペクトル包絡を
  合成対象と同じ声色空間へ射影する必要がある



※スペクトル包絡は [Kawahara *et al.*, 1999] により推定

スペクトル包絡形状は歌唱者にも依存

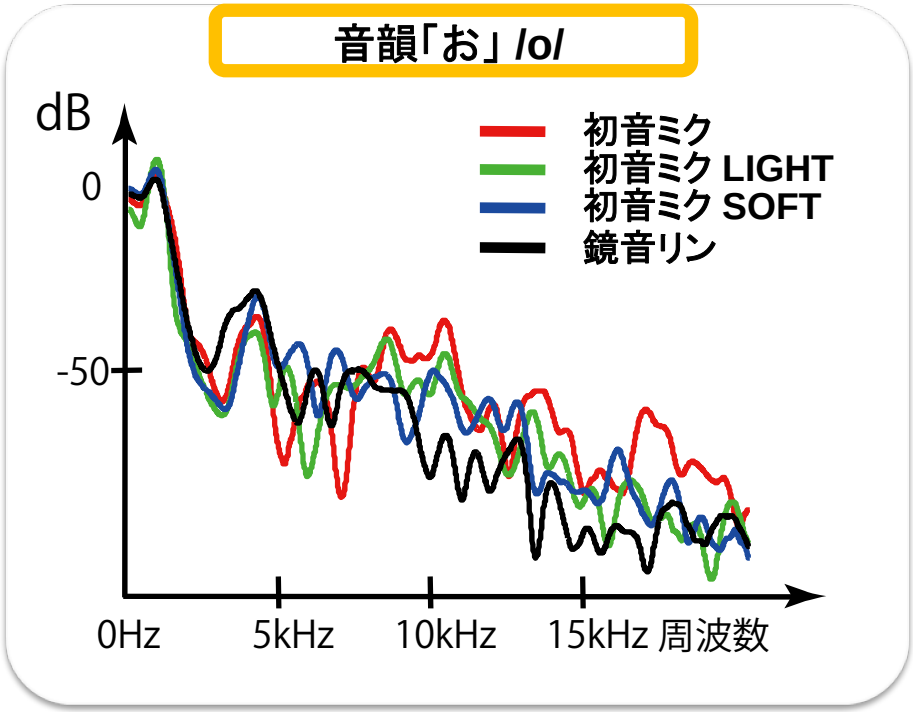
- ユーザ歌唱のスペクトル包絡を
  合成対象と同じ声色空間へ射影する必要がある

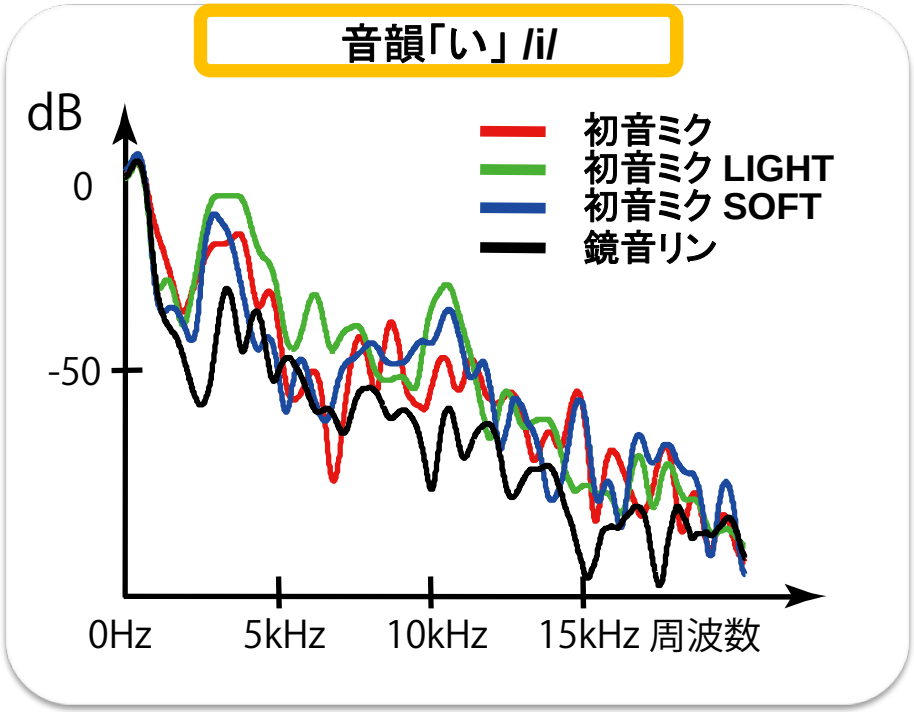


※スペクトル包絡は [Kawahara *et al.*, 1999] により推定

スペクトル包絡形状は音韻（歌詞）にも依存

- ユーザ歌唱のスペクトル包絡を
  合成対象と同じ声色空間へ射影する必要がある





※スペクトル包絡は [Kawahara *et al.*, 1999] により推定

スペクトル包絡形状は音韻（歌詞）にも依存

- ユーザ歌唱のスペクトル包絡を
  合成対象と同じ声色空間へ射影する必要がある

スペクトル包絡形状
＝ 音韻性＋個人性**＋声色成分**（＋ノイズ等）

声色成分
＝ スペクトル包絡**－音韻性－個人性**

※スペクトル包絡は [Kawahara *et al.*, 1999] により推定

時刻が同期した多様な歌声を合成する

- **ぼかりす1を最大限活用**

[RWC-MDB-G-2001 No.91]

  - **全日本語VOCALOID17種類の「大漁船」を利用**
    - KAITO, MEIKO, 初音ミク, 鏡音リン, 鏡音レン, がくっぽいど, 巡音ルカ, メグッポイド, 氷山キヨテル, 歌愛ユキ, SF-A2開発コードmiki
    - 初音ミクAppend (DARK, LIGHT, SOFT, SOLID, SWEET, VIVID)
      - http://staff.aist.go.jp/t.nakano/VocaListener/index-j.html
      - http://www.nicovideo.jp/mylist/7012071

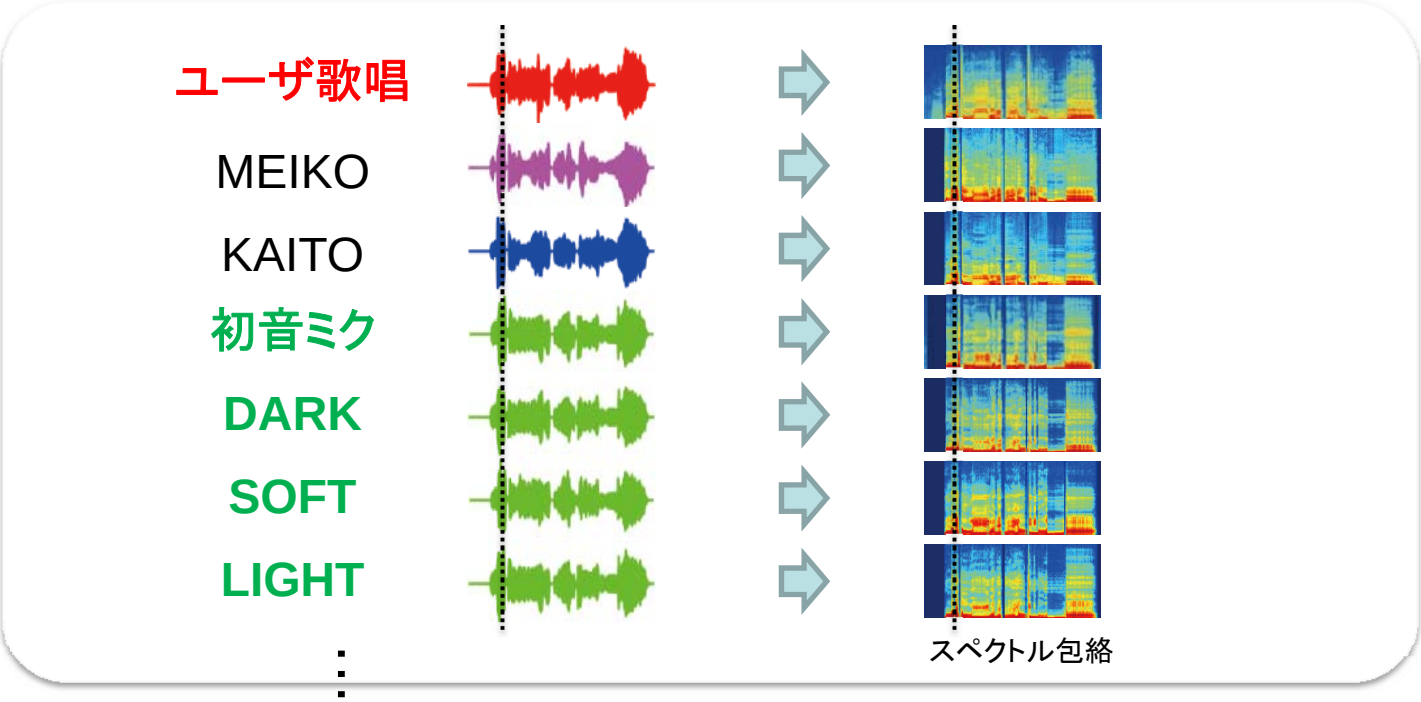
計17種類の合成歌唱音声

多様な歌声からの声色空間の構成

- 時刻が同期した多様な歌声

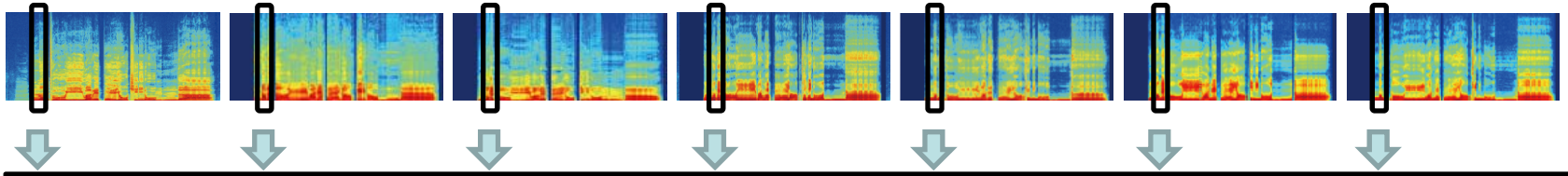
（ぼかりす1による17種類＋ユーザ歌唱）

のスペクトル包絡を推定 [Kawahara *et al.*, 1999]



主成分分析(PCA)に基づく声色空間の構成

- **歌詞が同一**の**時刻同期した**スペクトル包絡



各時刻において音韻性は同じ

=

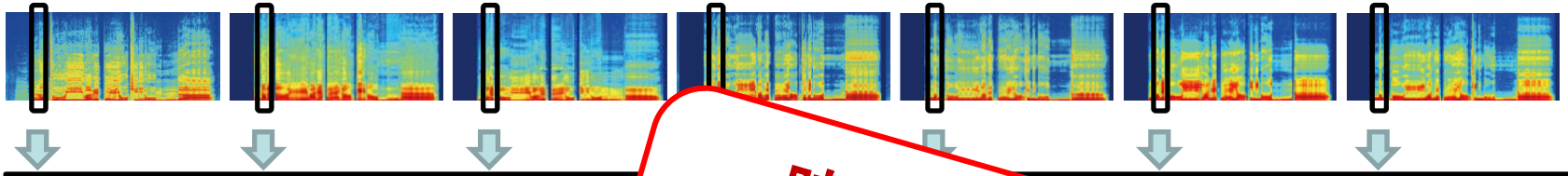
音韻性以外の個人性・**声色成分**による**変動が大きい**

=

低次主成分は個人性・**声色成分**の寄与が大きい

主成分分析(PCA)に基づく声色空間の構成

- 歌詞が同一の時刻同期したスペクトル包絡



各時刻

音韻性以外の個人性・声色成分

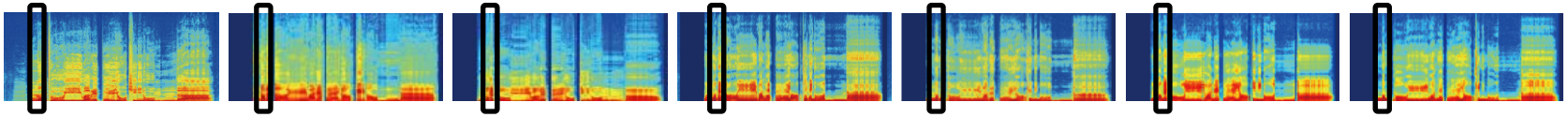
=

低次主成分は個人性・声色成分の寄与が大きい

時刻毎にPCAしただけでは
時刻毎に異なる声色空間
ができてしまう

主成分分析(PCA)に基づく声色空間の構成

- **歌詞が同一**の**時刻同期した**スペクトル包絡



各時刻における**全歌唱者**の全スペクトル包絡でPCA

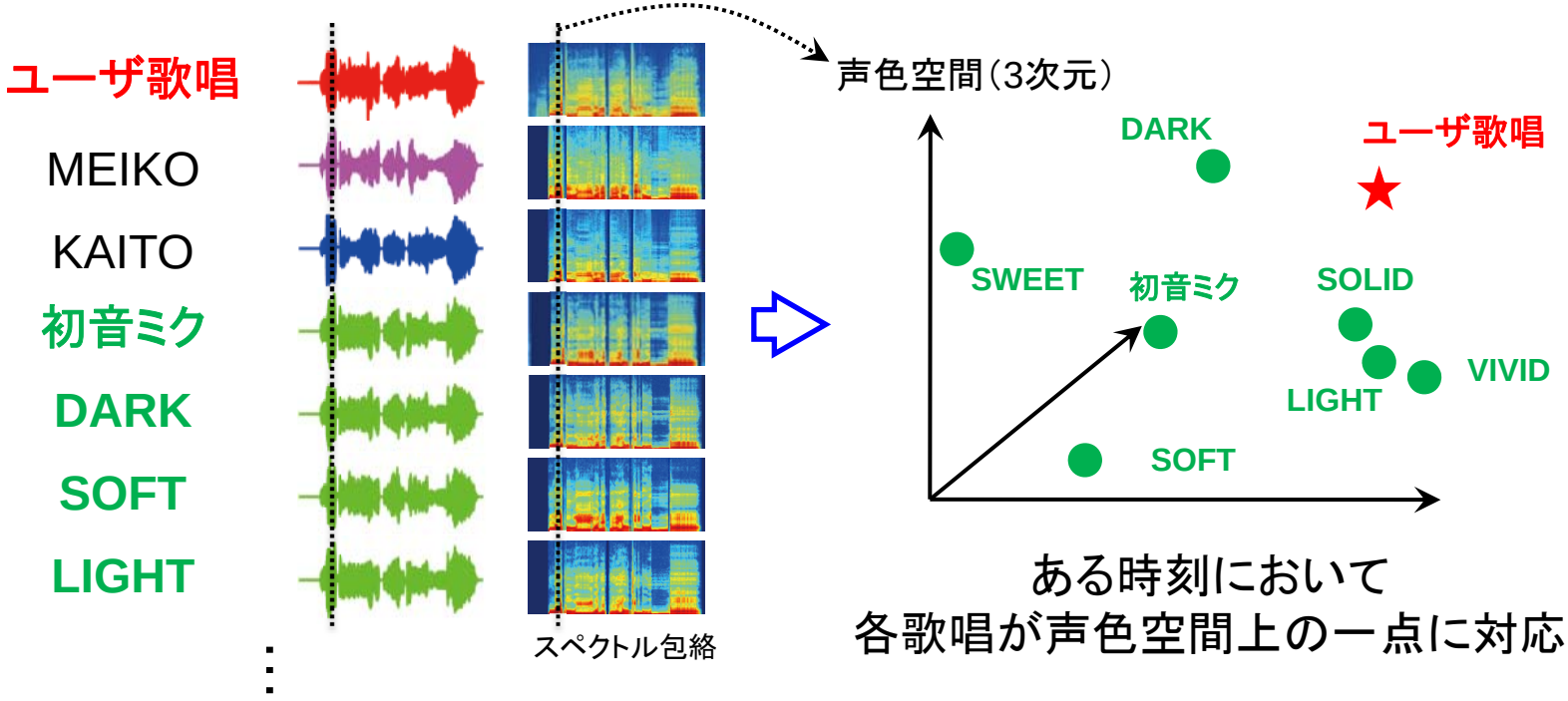
低次主成分※を保存してスペクトル包絡に逆射影

※平均5.04次元を使用

全時刻の**全歌唱者**の全スペクトル包絡でPCA

主成分分析(PCA)に基づく声色空間の構成

- 時刻が同期した多様な歌声のスペクトル包絡を
  （ぼかりす1による17種類＋ユーザ歌唱）
  PCAに基づいて3次元の**声色空間**へ射影



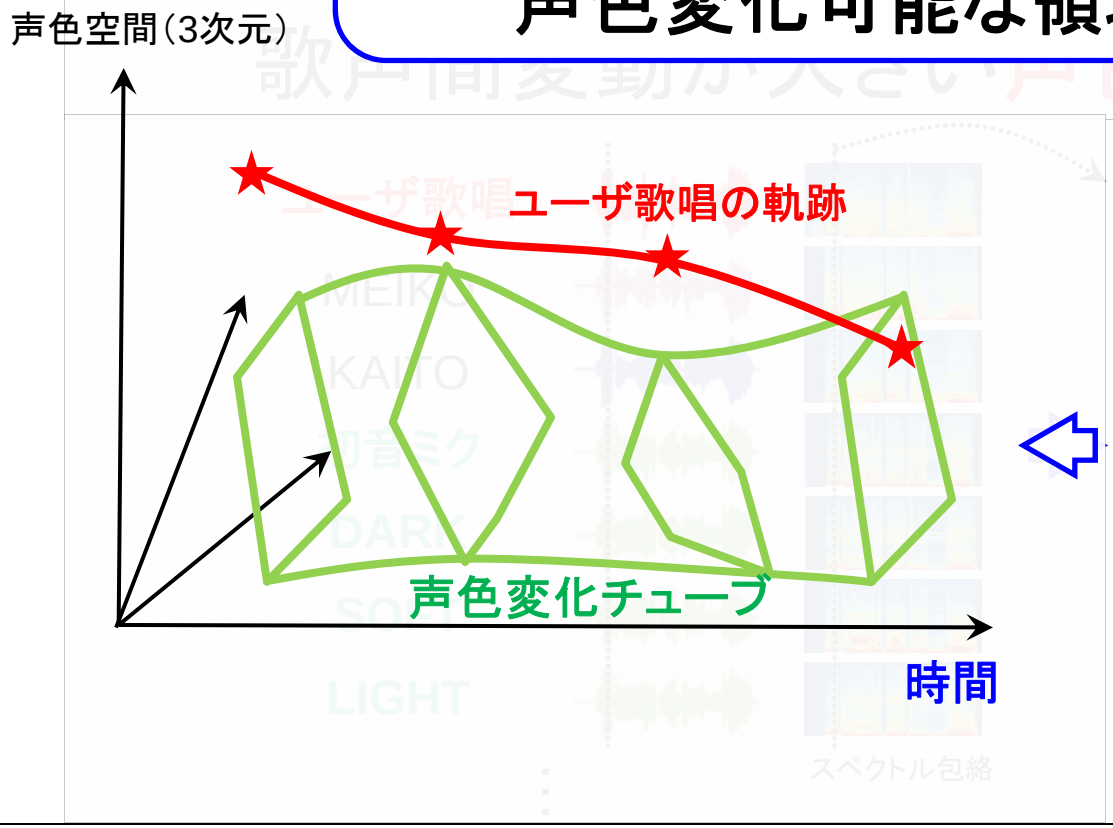
声色変化可能な領域

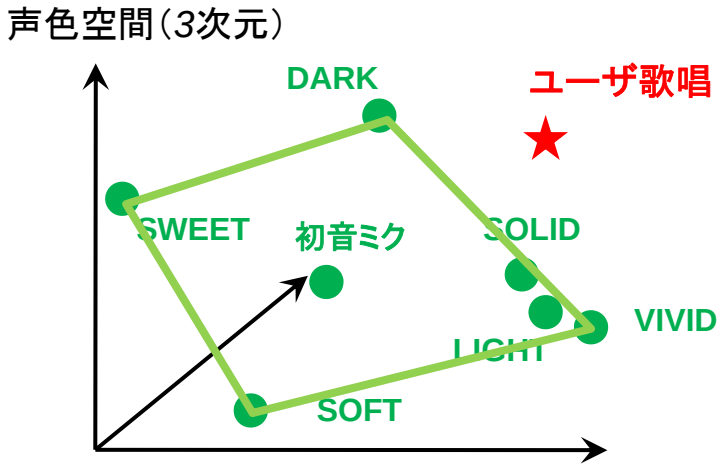
各声色に囲まれた
多面体（ポリトープ）を想定
その内側が変形可能な領域と仮定

ある時刻において
各歌唱が声色空間上の一点に対応

声色変化可能な領域

多面体(ポリトープ)の
時間軌跡を声色変化チューブと呼び
声色変化可能な領域と考える

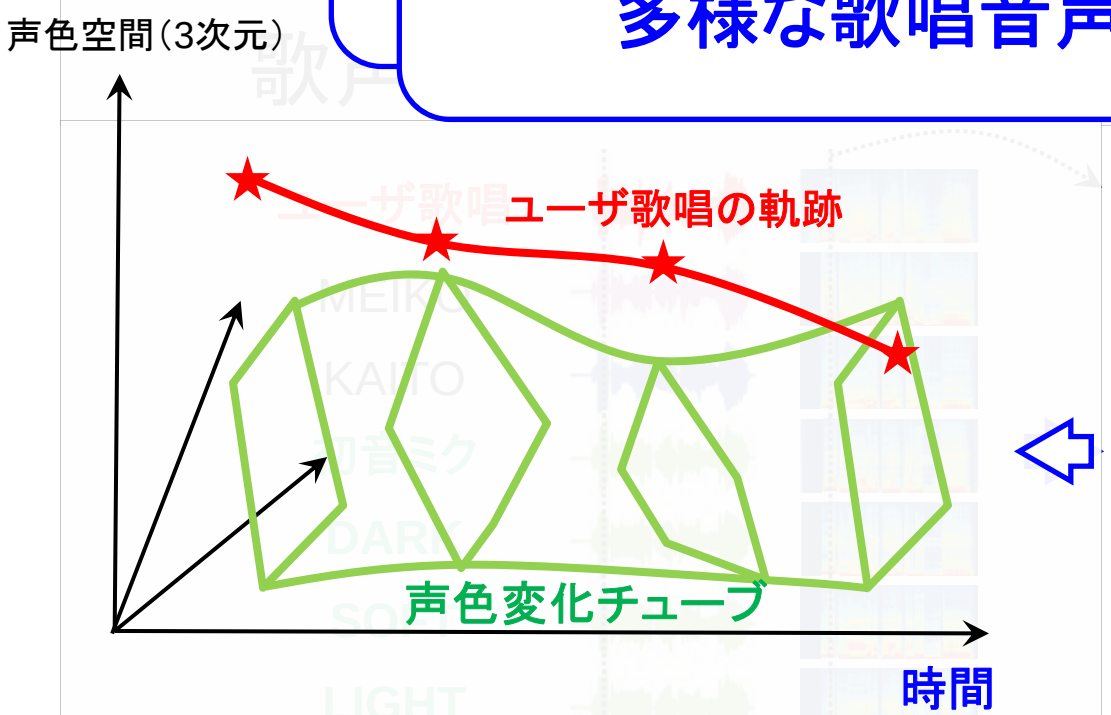




ある時刻において
各歌唱が声色空間上の一点に対応

声色変化可能な領域

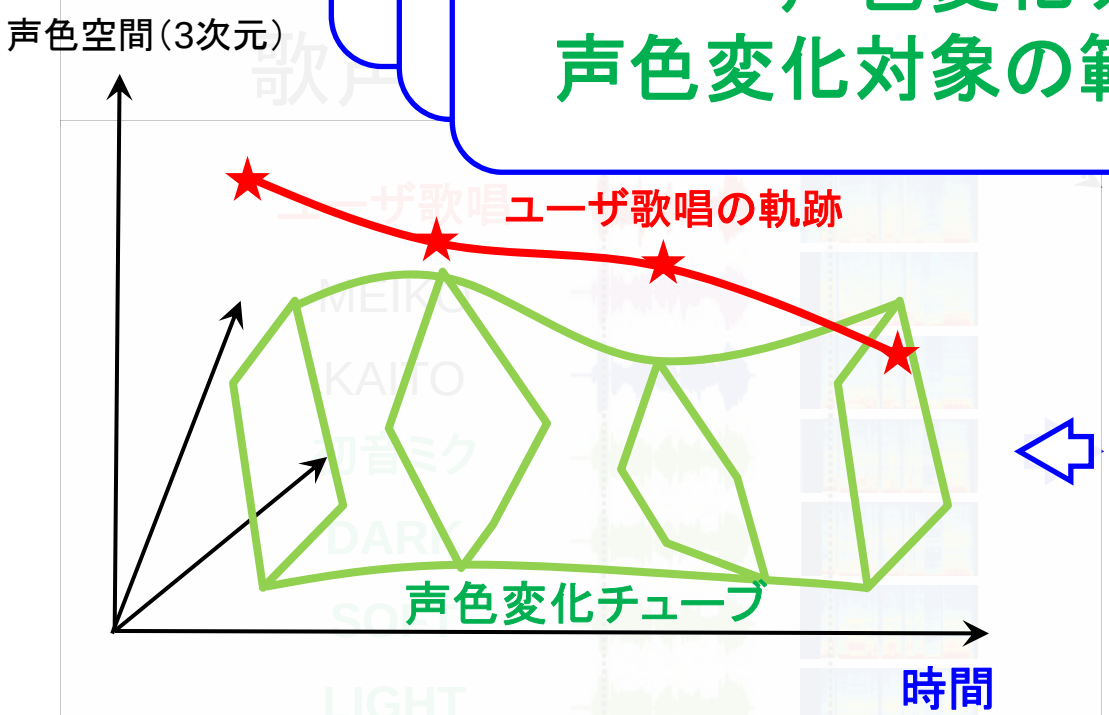
声色空間はユーザ歌唱を含めた多様な歌唱音声を表現でき



ある時刻において
各歌唱が声色空間上の一点に対応

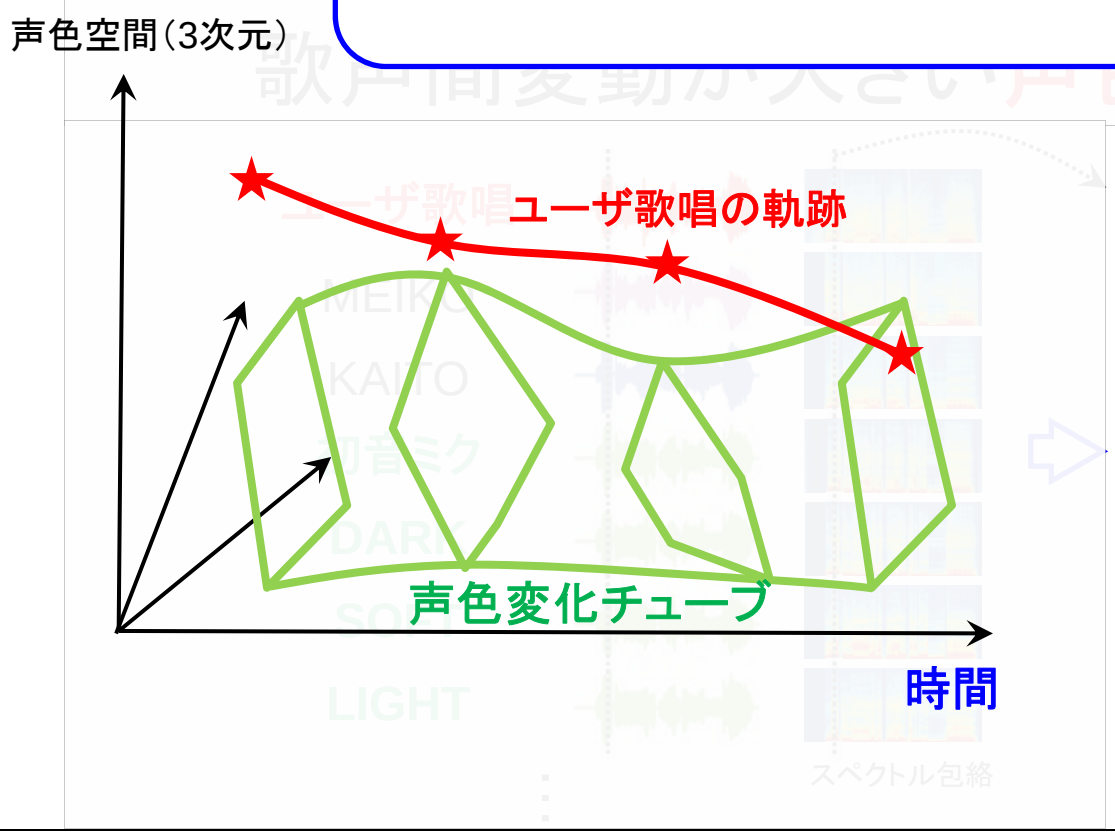
声色変化可能な領域

声色変化チューブで
声色変化対象の範囲を決定できる



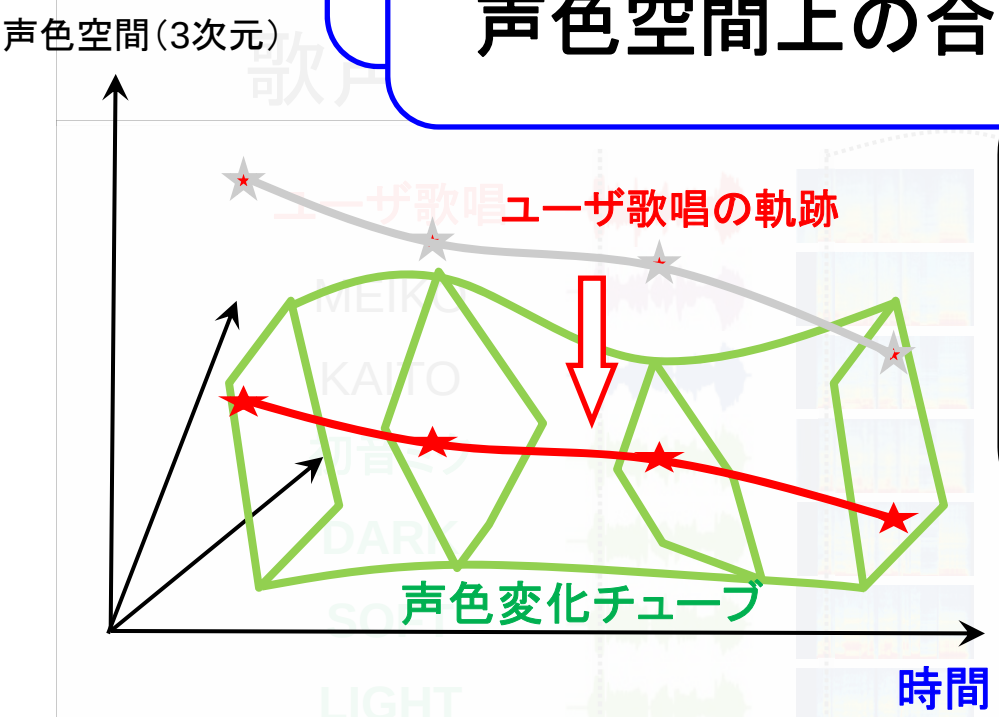
ユーザ歌唱の軌跡のシフト・スケーリング

ユーザ歌唱の軌跡は同じ空間の別の場所にあると考えられる



ユーザ歌唱の軌跡のシフト・スケーリング

シフト・スケーリング操作により
声色空間上の合成目標位置を決定

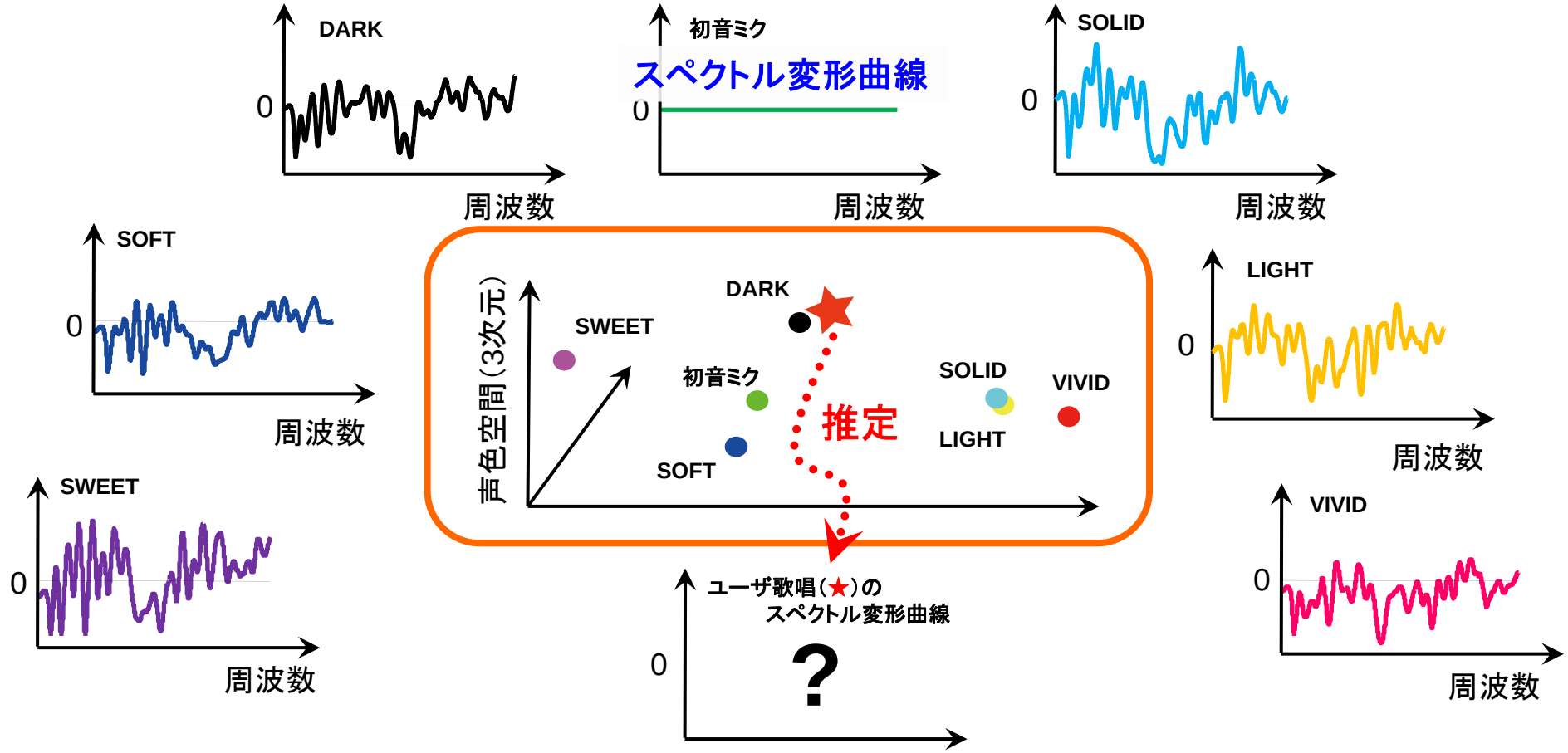


ユーザ歌唱を
各声色それぞれの次元の
平均と分散で正規化

声色空間に基づくスペクトル変形曲線の推定

[Turk et al., 2008]

- Radial Basis Functionを用いたVariational Interpolation



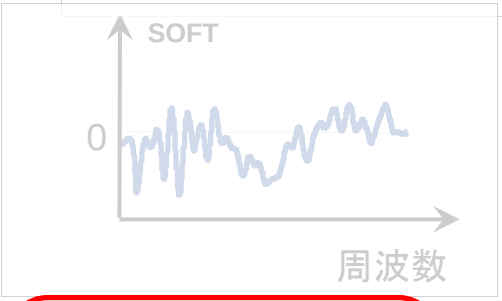
声色空間に基づくスペクトル変形曲線の推定

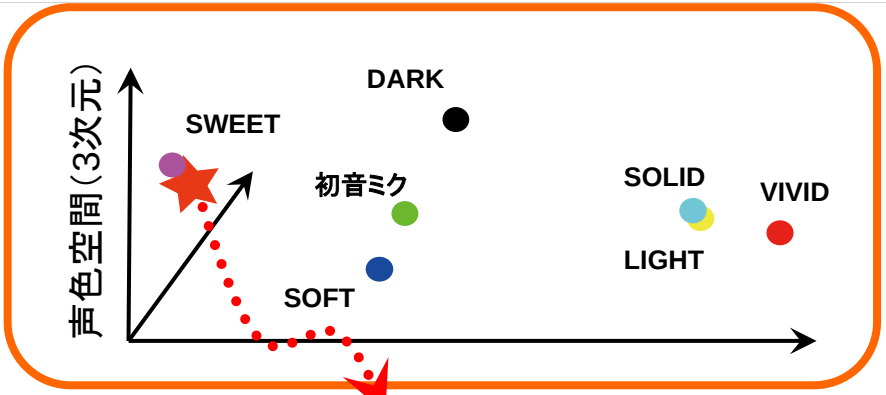
[Turk et al., 2008]

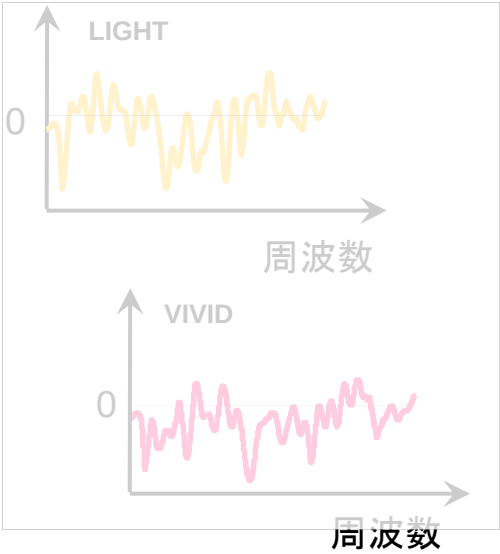
- Radial Basis Functionを用いたVariational Interpolation

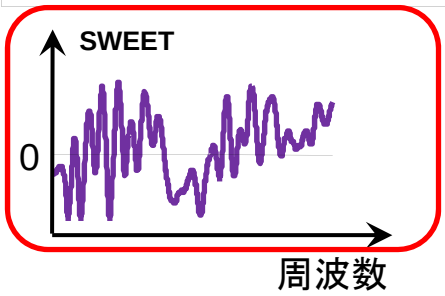
制約

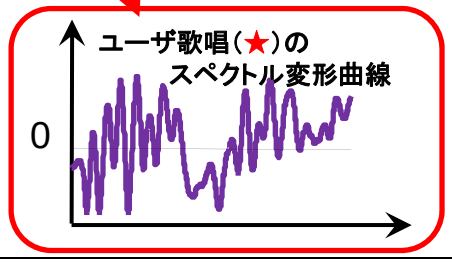
ユーザ歌唱が各声色と重なった場合
それと同じスペクトル変形曲線を生成











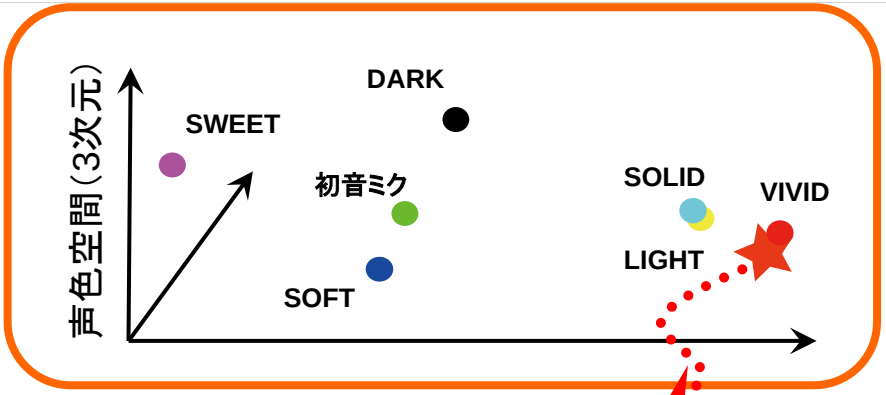
声色空間に基づくスペクトル変形曲線の推定

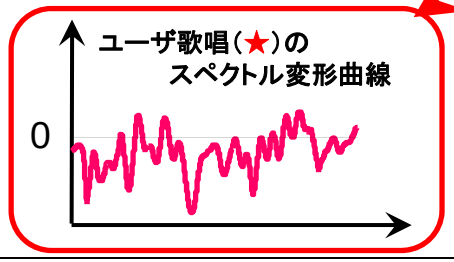
[Turk et al., 2008]

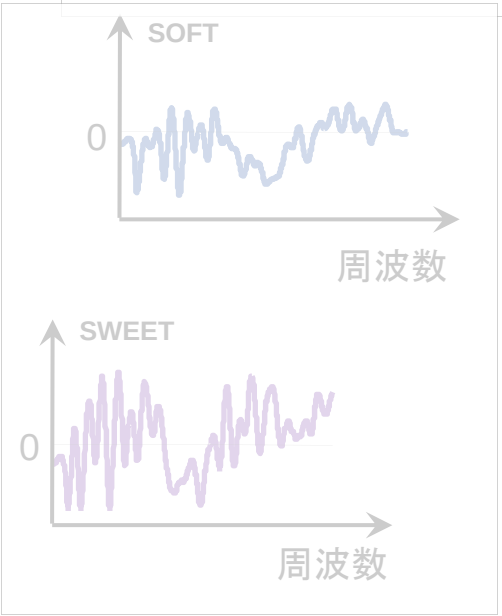
- Radial Basis Functionを用いたVariational Interpolation

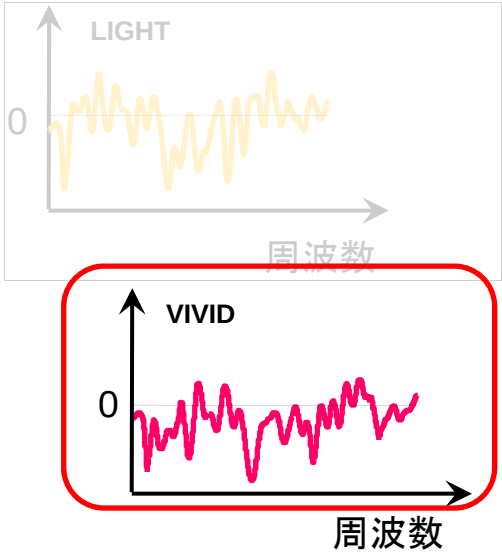
制約

ユーザ歌唱が各声色と重なった場合
それと同じスペクトル変形曲線を生成





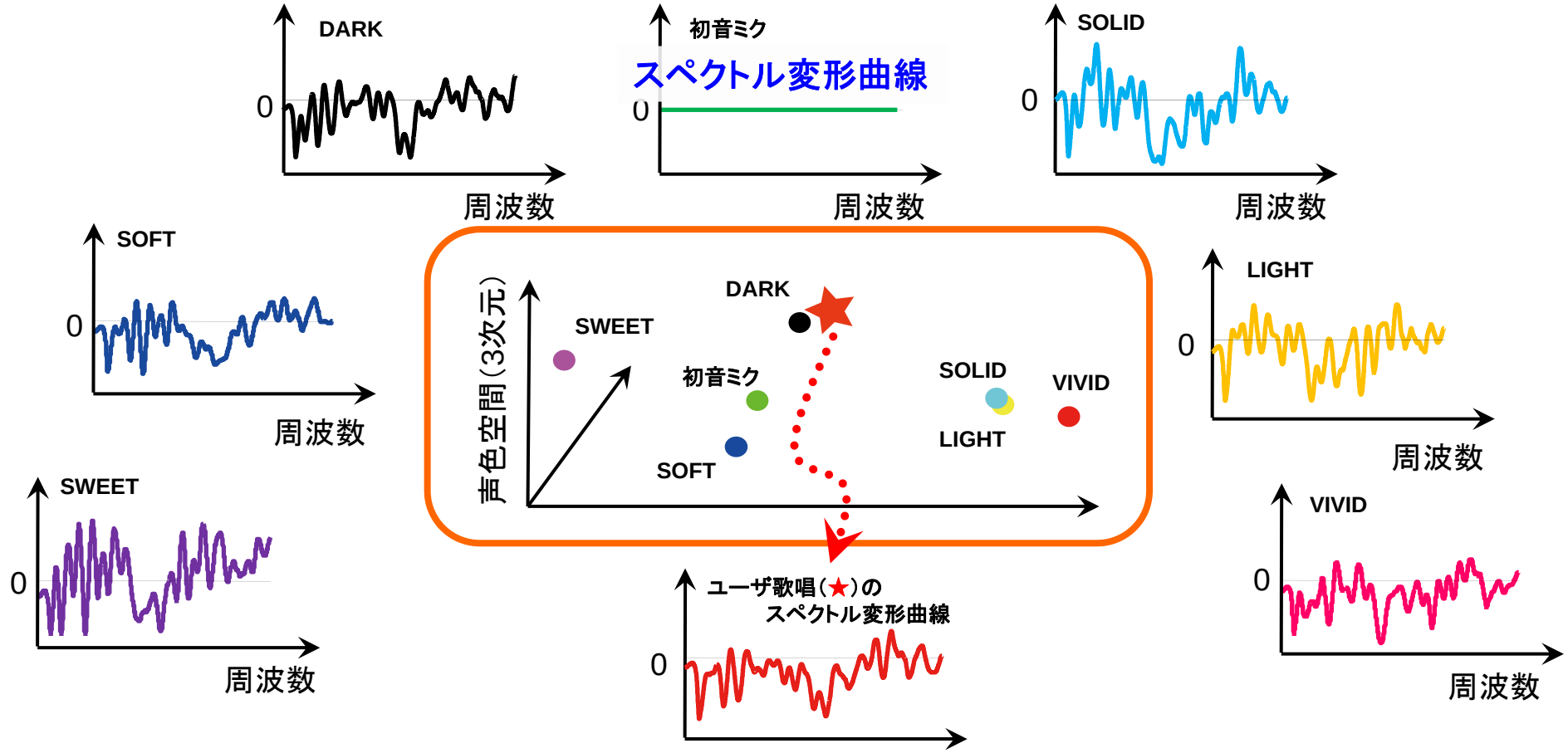




声色空間に基づくスペクトル変形曲線の推定

[Turk et al., 2008]

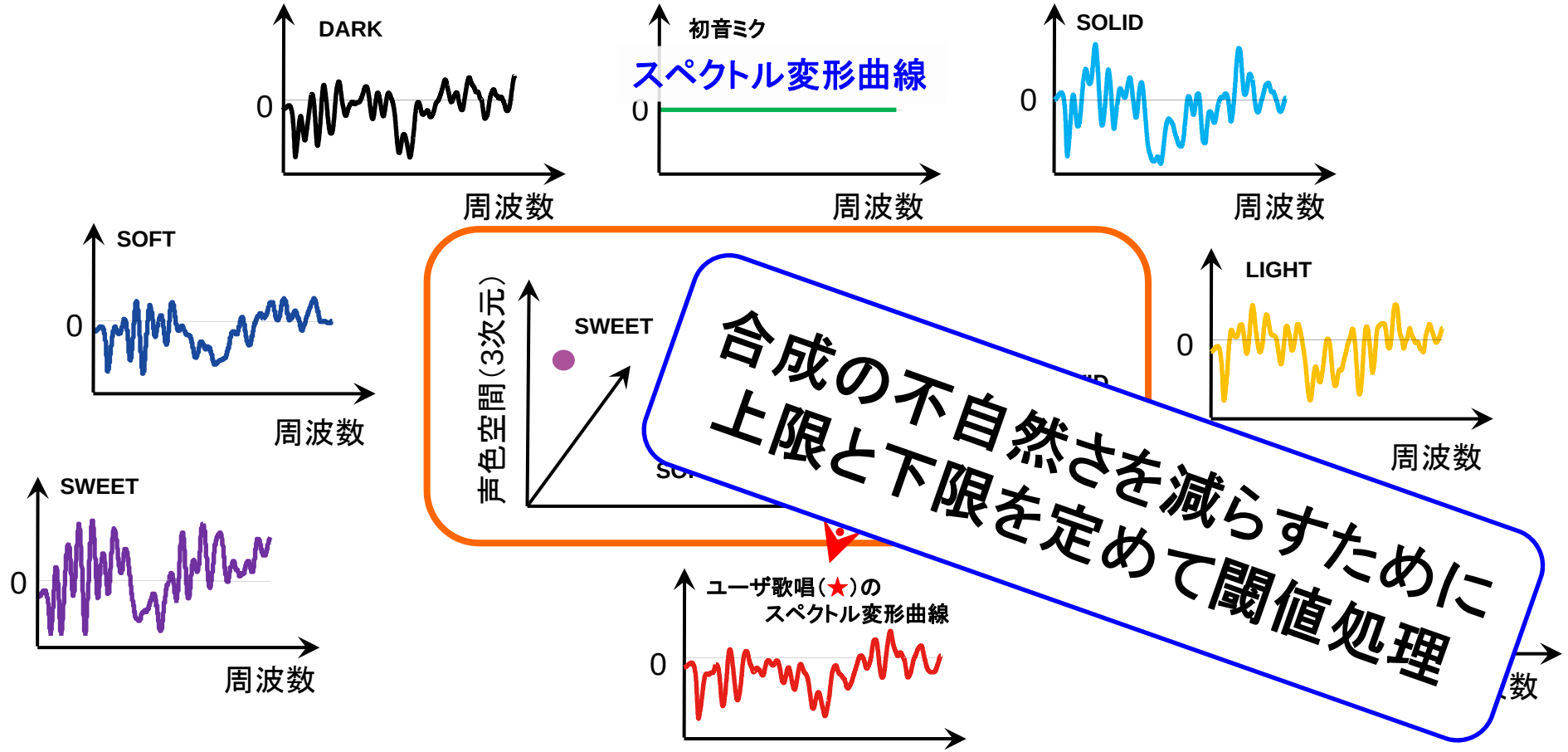
- Radial Basis Functionを用いたVariational Interpolation



声色空間に基づくスペクトル変形曲線の推定

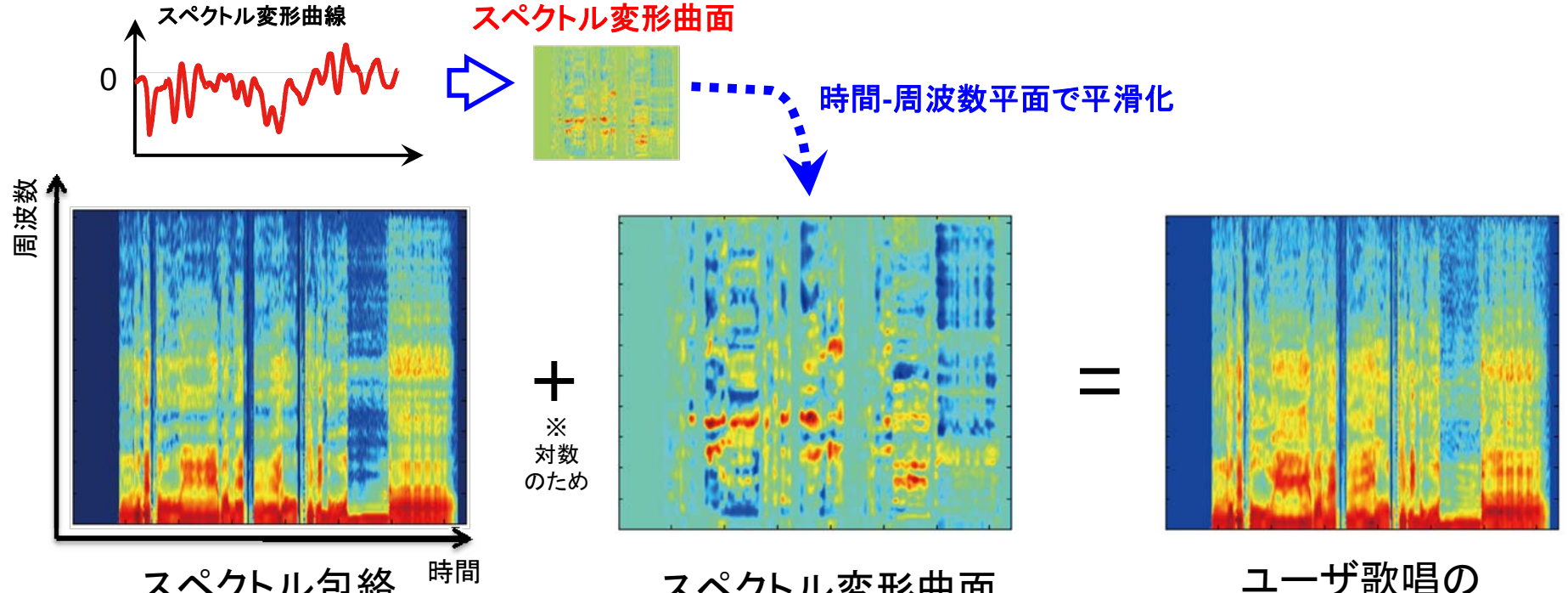
[Turk et al., 2008]

- Radial Basis Functionを用いたVariational Interpolation



スペクトル変形曲面の生成と適用

- 全時刻のスペクトル変形曲線を合わせて生成
  - 平滑化処理により急峻な変化を低減



スペクトル包絡
「大漁船」
初音ミク

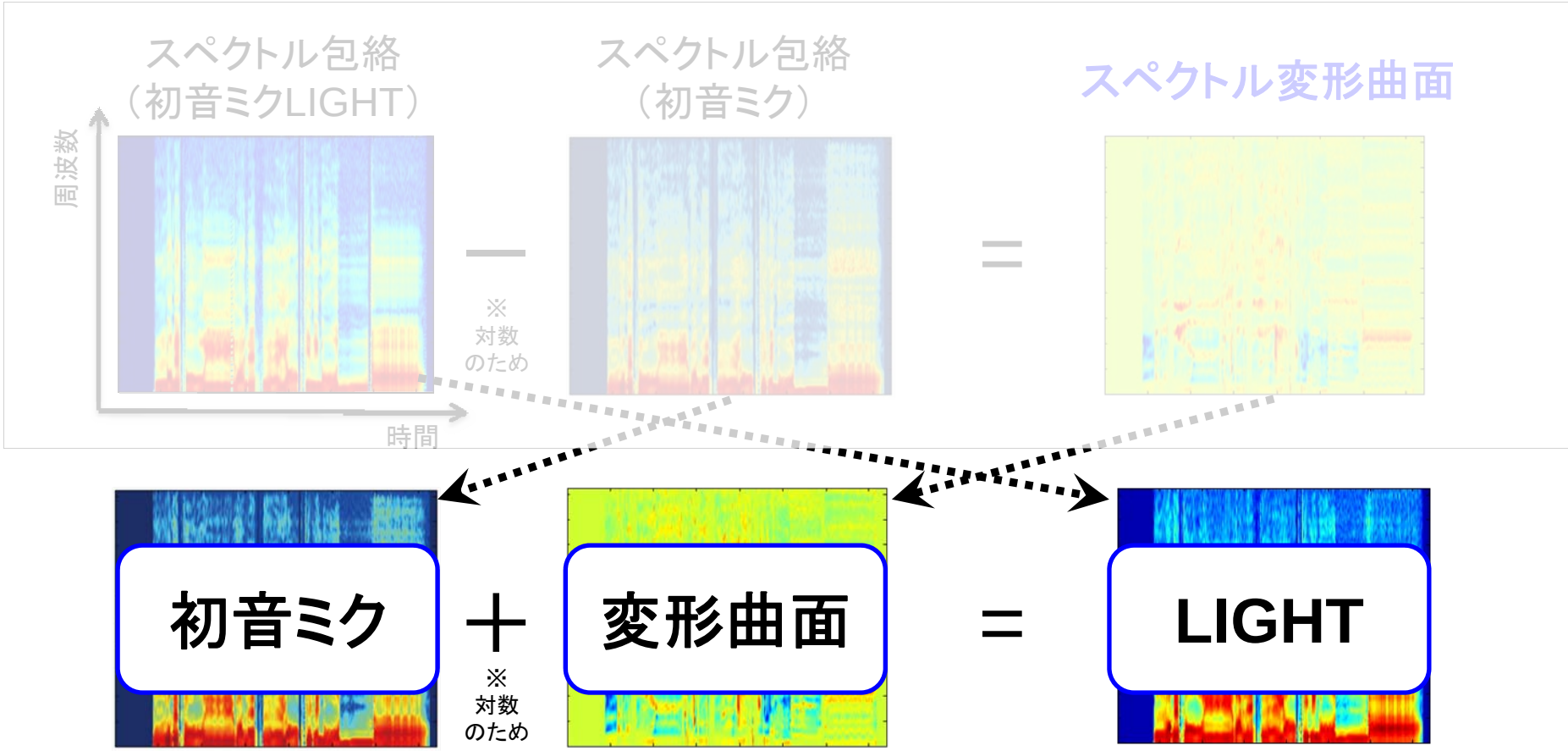
スペクトル変形曲面
（平滑化後）

ユーザ歌唱の
声色変化を反映した
スペクトル包絡

ぼかりす2
の
応用

スペクトル包絡変形に基づく初音ミクLIGHT合成

ロ スペクトル変形曲面（相対的な違い）の導入



スペクトル包絡変形に基づく初音ミクLIGHT合成

スペクトル変形曲面（相対的な違い）の導入

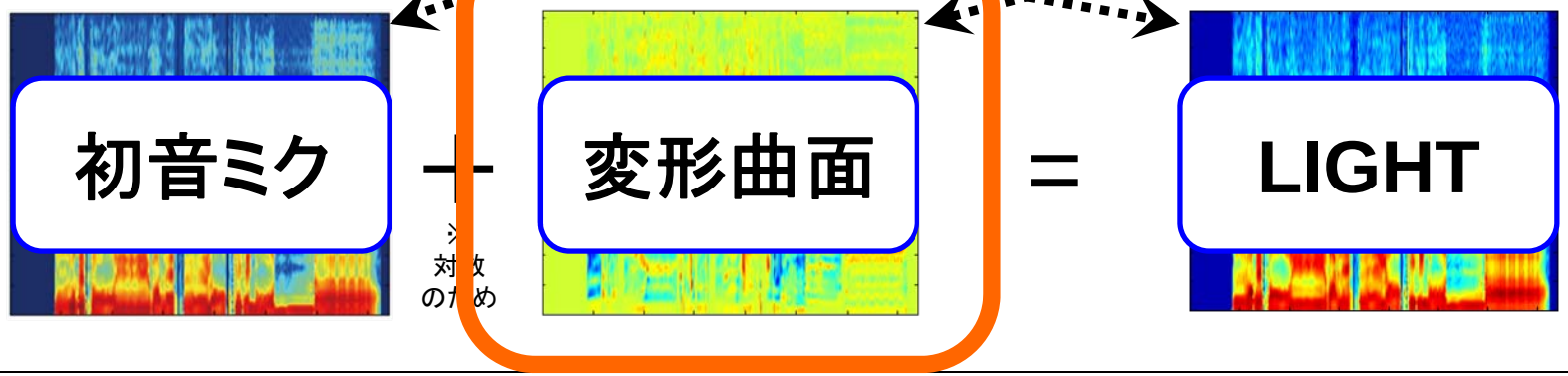
スペクトル包絡
（初音ミクLIGHT）

スペクトル包絡
（初音ミク）

スペクトル変形曲面

スペクトル変形曲面を替えると
初音ミクから初音ミク・アペンドが合成可能

時間



スペクトル包絡変形に基づく初音ミクLIGHT合成

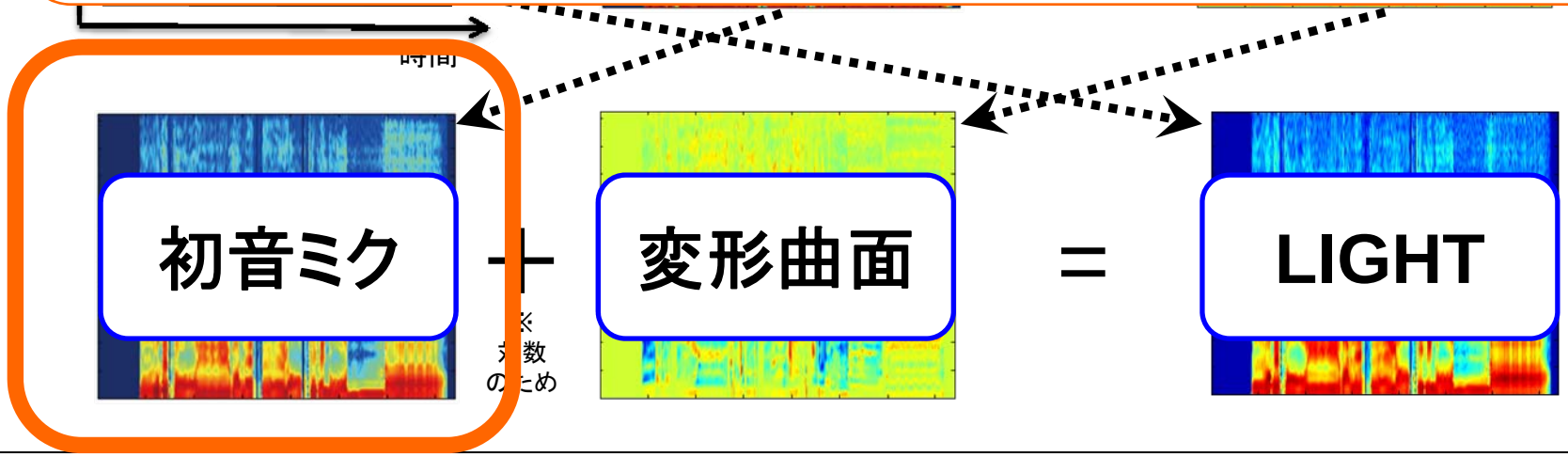
スペクトル変形曲面（相対的な違い）の導入

スペクトル包絡
（初音ミクLIGHT）

スペクトル包絡
（初音ミク）

スペクトル変形曲面

今度は「基準となるスペクトル」を替える



スペクトル包絡変形に基づく初音ミクLIGHT合成

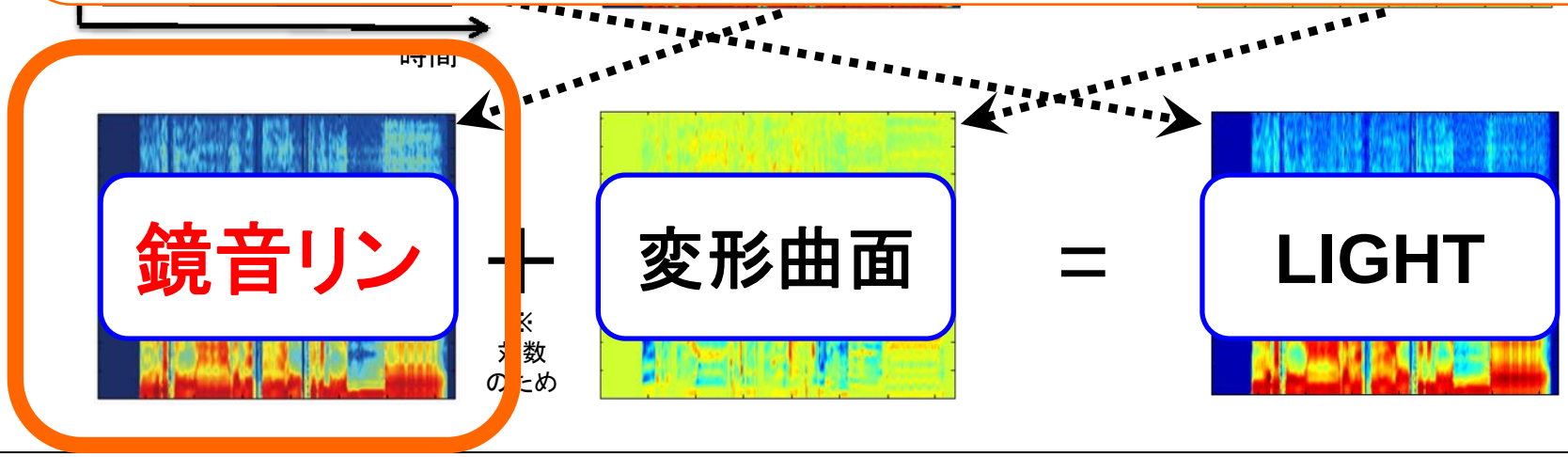
スペクトル変形曲面（相対的な違い）の導入

スペクトル包絡
（初音ミクLIGHT）

スペクトル包絡
（初音ミク）

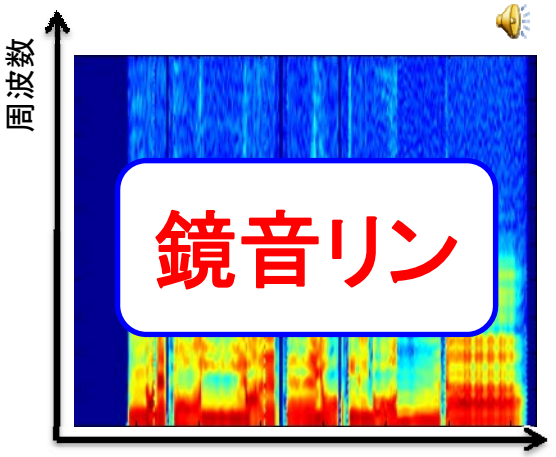
スペクトル変形曲面

今度は「基準となるスペクトル」を替える



声色転写による鏡音リン・擬似アペンド合成

- スペクトル変形曲面を別の音源に適用する

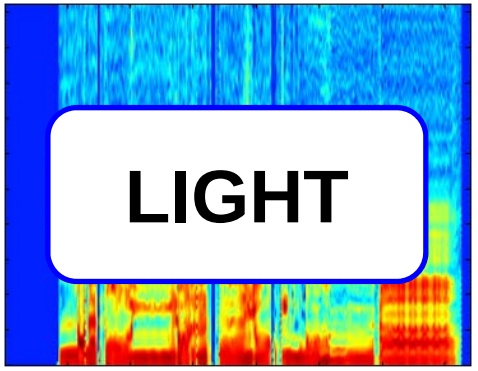


スペクトル包絡
「大漁船」
鏡音リン

＋
※
対数
のため

スペクトル変形曲面
初音ミク
⇒初音ミクLIGHT

＝

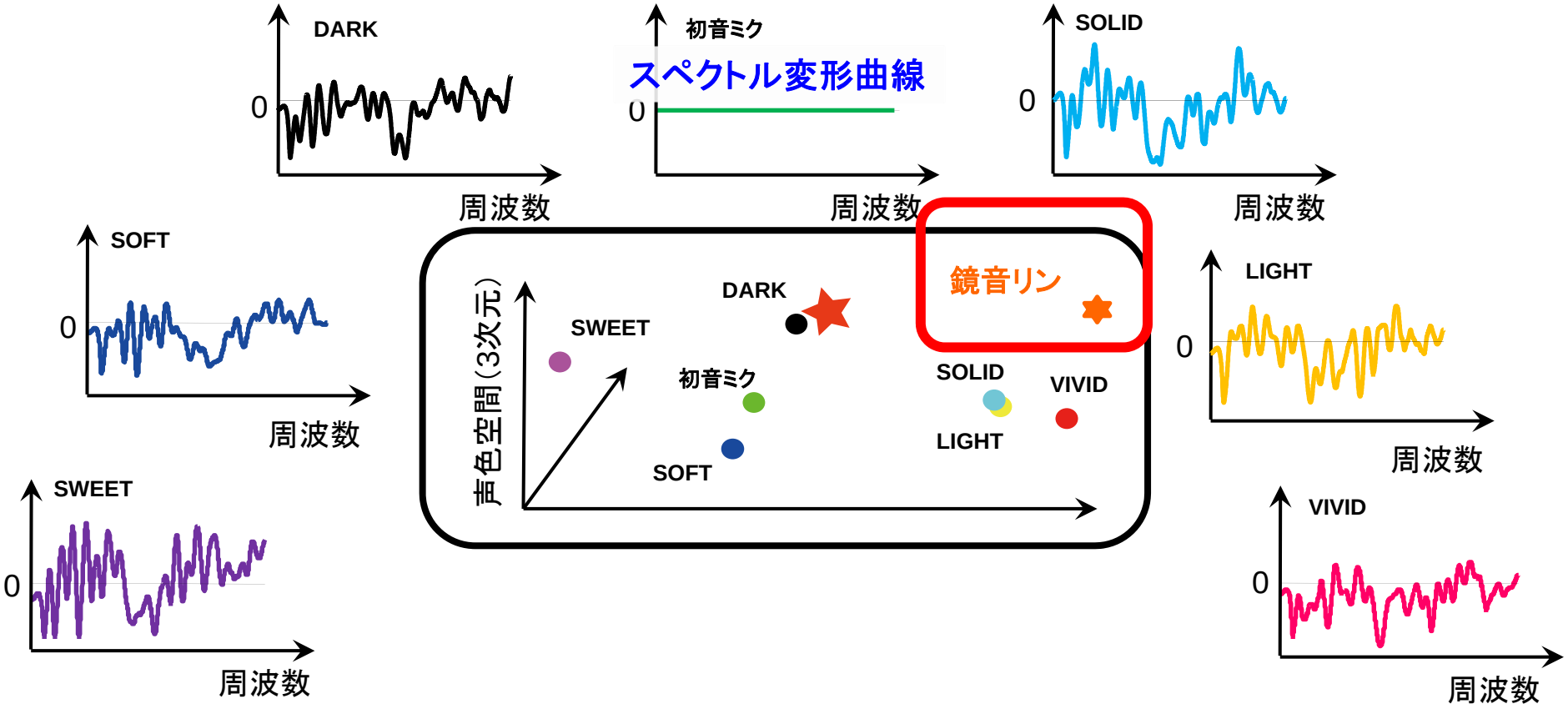


鏡音リン・擬似LIGHT
の
スペクトル包絡

※合成結果の具体例はhttp://staff.aist.go.jp/t.nakano/VocaListener2/index-j.html
http://www.nicovideo.jp/mylist/7012071で視聴可能

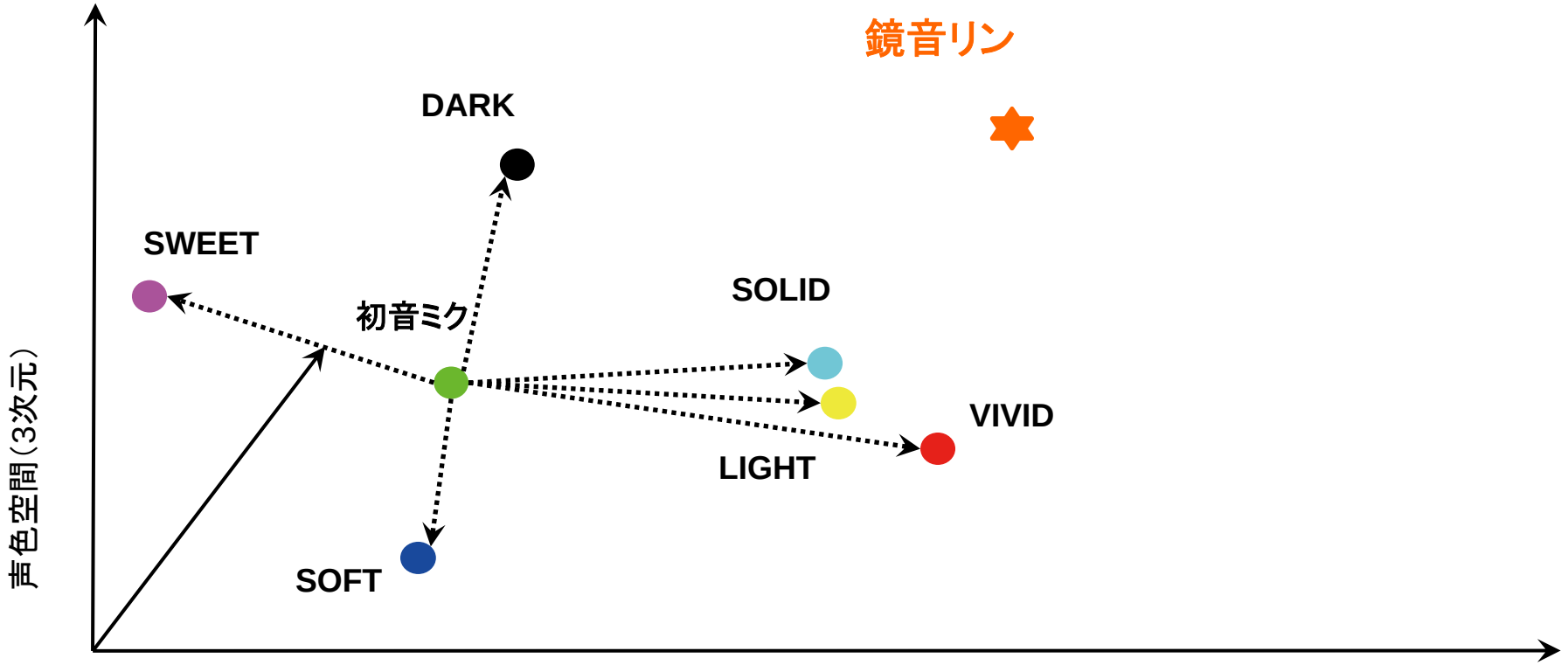
鏡音リン・擬似アペンドで声色変化を真似る

- 声色空間上で鏡音リン・擬似アペンドを生成



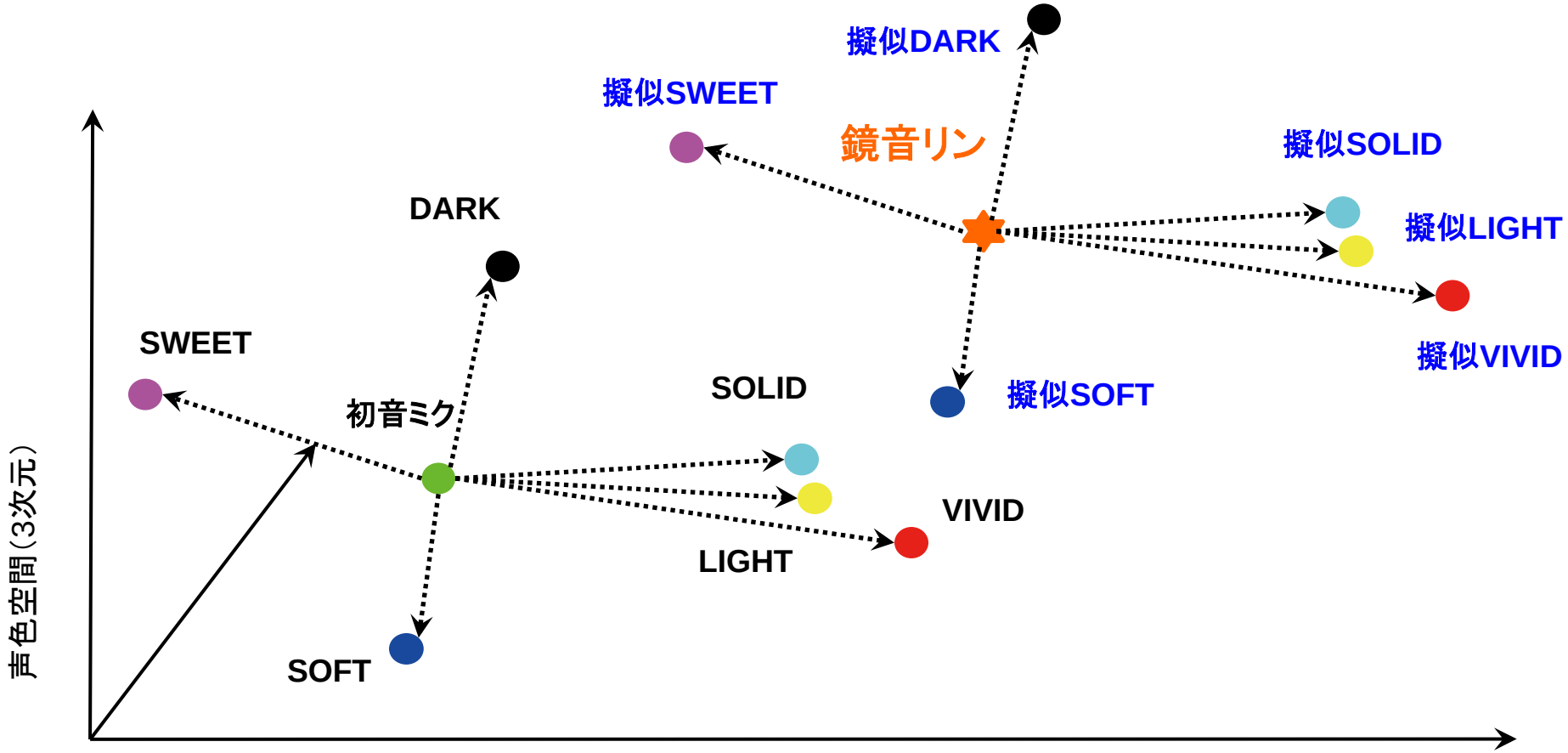
鏡音リン・擬似アペンドで声色変化を真似る

- 声色空間上で鏡音リン・擬似アペンドを生成



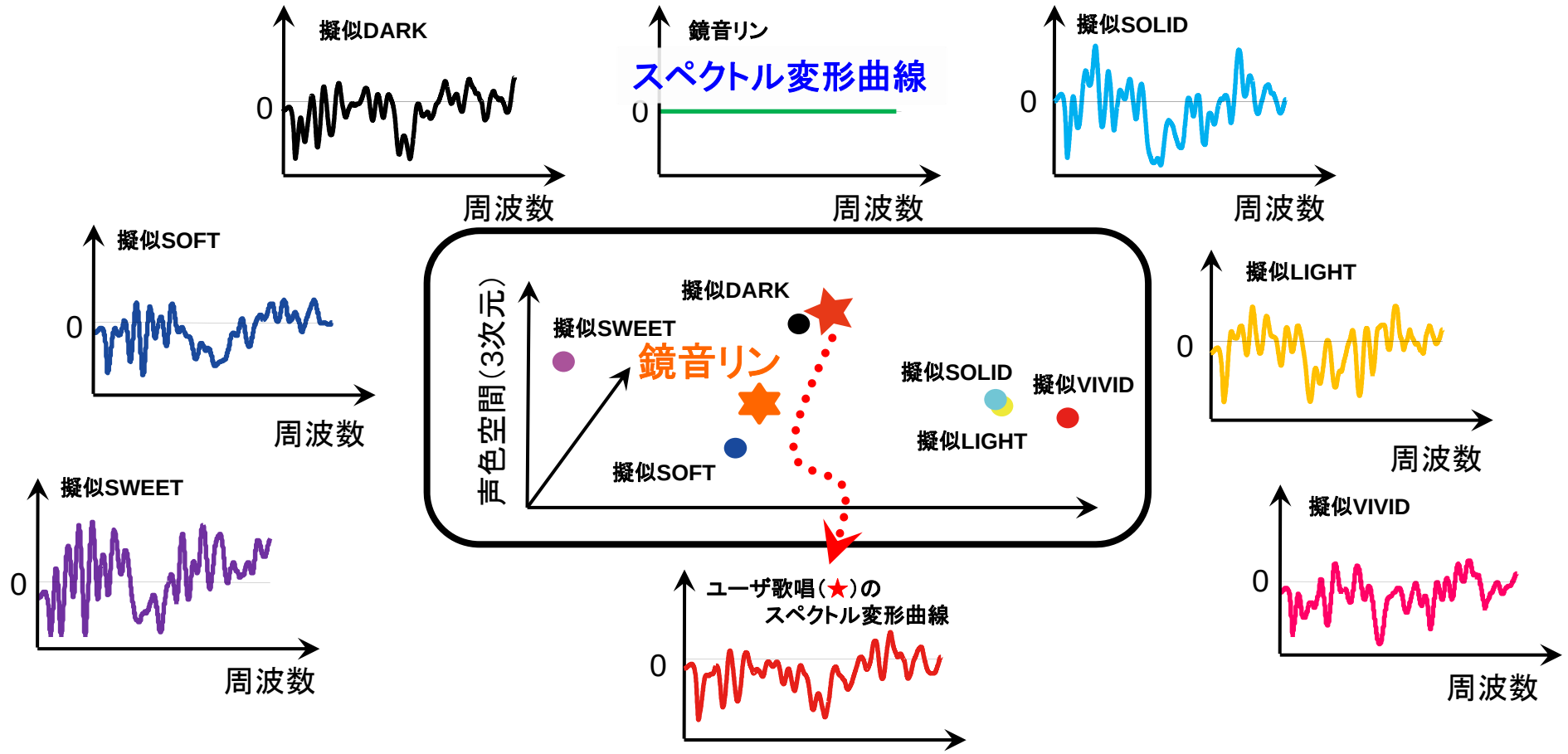
鏡音リン・擬似アペンドで声色変化を真似る

- 声色空間上で鏡音リン・擬似アペンドを生成



鏡音リン・擬似アペンドで声色変化を真似る

- 声色空間上で鏡音リン・擬似アペンドを生成



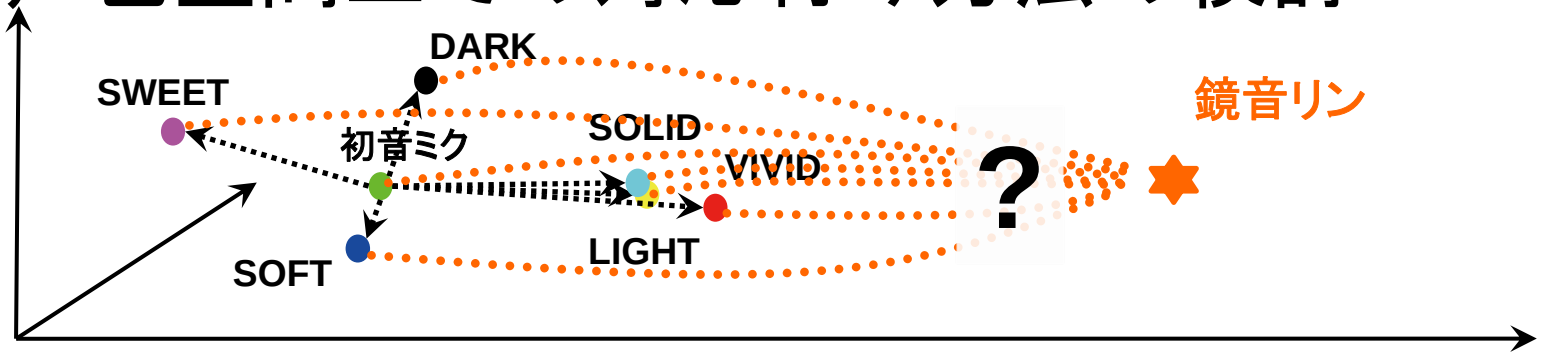
擬似アペンドの合成とぼかりす2の課題

- 擬似アペンドの合成方法の検討
  - **スペクトル変形曲面の最適化（再推定）**が必要
    - 現在の単純な入れ替えでは不十分

擬似アペンドの合成とぼかりす2の課題

- 擬似アペンドの合成方法の検討
  - **スペクトル変形曲面の最適化（再推定）**が必要
    - 現在の単純な入れ替えでは不十分

- 声色空間上での対応付け方法の検討



おわりに

まとめ

- **VocaListener1（ぼかりす1）**を最大限活用し
  - 声色空間の構成
  - ユーザ歌唱の声色変化を反映させた歌声合成

  が行える**VocaListener2（ぼかりす2）**を提案

- 今後
  - 応用で述べた「汎用的な声色転写」等の検討
  - 声色変化の新たな活用法の検討
  - 「人間らしい」歌唱や知覚に関する知見を得て活用

VocaListener2

ユーザ歌唱の音高と音量だけでなく 声色変化も真似る歌声合成システムの提案

中野 倫靖, 後藤 真孝
（産業技術総合研究所）

2010年7月28日
第86回音楽情報科学研究会(SIGMUS)