方向統計 DNN に基づく振幅スペクトログラムからの位相復元 *

○高道 慎之介, 齋藤 佑樹, 高宗 典玄 (東大院・情報理工),
北村 大地 (香川高専), 猿渡 洋 (東大院・情報理工)

1 はじめに

統計的音声音響信号処理では，振幅スペクトログラムに対する処理がしばしば行われる．そのため，最終的な音声波形を生成するためには，振幅スペクトログラムから位相情報を復元する必要がある．これに対し我々は，von Mises 分布 deep neural network (DNN) に基づく位相復元法 [1, 2] を提案している．この DNN は，入力パラメータ（例えば，振幅）が与えられたときの出力パラメータ（例えば，位相）の条件付き確率分布として von Mises 分布 [3] を仮定する深層生成モデルであり，位相のような周期変数の確率分布のモデル化に適している．更に，同論文で我々は，この DNN が位相そのものよりも群遅延の分布を精度よくモデル化できることを明らかにした．しかしながら，群遅延は，位相に含まれる線形位相成分の影響を受けるため，そのヒストグラムは，正か負のいずれかに偏る．故に，円周上の対称分布である von Mises 分布，並びに，von Mises 分布 DNN は，群遅延モデリングに適さないことが予想される．そこで本稿では，正弦関数摂動一般化ハート分布 DNN を提案し，これを用いた群遅延モデリングを行う．正弦関数摂動一般化ハート分布 [4] とは，von Mises 分布の一般化であり，かつ，円周上の非対称分布を取りうる確率分布である．正弦関数摂動一般化ハート分布 DNN は，この分布を条件付き確率分布として扱う深層生成モデルである．実験結果より，非対称分布の導入により，群遅延の分布を更に精度よくモデル化できることを示す．

2 振幅スペクトログラムからの位相復元

本節では，振幅スペクトログラムからの位相復元について概説する．ここで，$\boldsymbol{x} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_t, \cdots, \boldsymbol{x}_T]$ と $\boldsymbol{y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_t, \cdots, \boldsymbol{y}_T]$ をそれぞれ，振幅・位相スペクトログラムとする．$\boldsymbol{x}_t = [x_{t,0}, \cdots, x_{t,f}, \cdots, x_{t,F}]$ と $\boldsymbol{y}_t = [y_{t,0}, \cdots, y_{t,f}, \cdots, y_{t,F}]$ はそれぞれ，フレーム t における振幅及び位相である．f は周波数ビンのインデックスであり，F はナイキスト周波数に対応する．$x_{t,f}$ と $y_{t,f}$ は実数値であり，$y_{t,f}$ は，2π の周期をもつ周期変数である．本稿では，位相スペクトログラム $\boldsymbol{y}$ から解析的に導出される群遅延を，振幅スペクトログラム $\boldsymbol{x}$ から推定する方法を扱う．

3 正弦関数摂動一般化ハート分布

本節では，正弦関数摂動一般化ハート分布について述べる．Fig. 1 に関係図を示す．von Mises 分布を一般化した対称分布を一般化ハート分布 (generalized cardioid distribution) [5] と呼び，その分布を更に，正弦関数により摂動させた非対称分布を正弦関数摂動一般化ハート分布 (sine-skewed generalized cardioid distribution) [4] と呼ぶ．以降では，これらの分布について述べる．

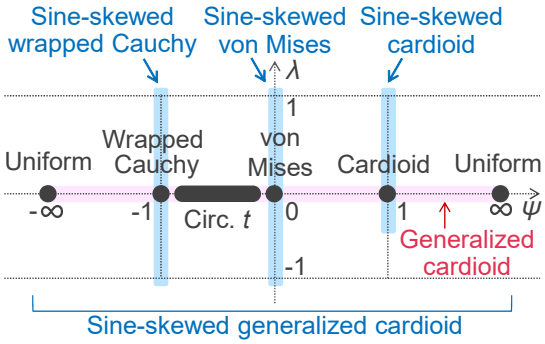

Fig. 1 正弦関数摂動一般化ハート分布の関係図．横軸の ψ と縦軸の λ はそれぞれ，一般化ハート分布と正弦関数摂動のモデルパラメータである．Circ. t は円周 t 分布を表す．

3.1 一般化ハート分布

一般化ハート分布 [5] は，円周上の確率分布，すなわち 2π の周期をもつ周期変数のための確率分布である．その確率分布 $P^{(\mathrm{gc})}(y; \mu, \kappa, \psi)$ は次式で与えられる．

$$P^{(\mathrm{gc})}(y; \mu, \kappa, \psi) = \frac{(\cosh(\kappa\psi))^{1/\psi} (1 + \tanh(\kappa\psi)\cos(y-\mu))^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}, \tag{1}$$

ここで，y は周期変数，μ は平均または位置パラメータ (location parameter)，κ は集中度パラメータ (concentration parameter)，ψ は実数値をとるモデルパラメータである．$P_{1/\psi}(\cdot)$ は 0 次のルジャンドル陪関数である．一般化ハート分布は，symmetric Jones-Pewsey distribution とも呼ばれ，平均 μ に対して，円周上で対称かつ単峰となる分布である．ψ が特定の値をとるときに，この分布は特殊形となる．その特殊形として，以降では，巻込み Cauchy 分布 (wrapped Cauchy distribution)，von Mises 分布 (von Mises distribution)，ハート型分布 (cardioid distribution) を定義する．ただし，本稿では，ψ が一定の値を持ち，かつ，有限の κ を持つ特殊形のみに限定するため，円周 t 分布 (circular t-distribution) [6] $(0 < \psi < 1)$ と，Cartwright's Power-of-Cosine 分布 [7] $(\psi > 0, \kappa \to \infty)$ は扱わない．

3.1.1 巻込み Cauchy 分布 ($\psi = -1$)

巻込み Cauchy 分布は，Cauchy 分布を巻込み法により円周上の分布に変形したものである．ここでは，κ を ρ に置き換え，巻込み Cauchy 分布 $P^{(\mathrm{wc})}(y; \mu, \rho)$

---

*Phase reconstruction from amplitude spectrograms based on directional-statistics DNNs, by TAKAMICHI, Shinnosuke, SAITO, Yuki, TAMAMUNE, Norihiro (The University of Tokyo), KITAMURA, Daichi (National Institute of Technology, "Kagawa College), and SARUWATARI, Hiroshi (The University of Tokyo)

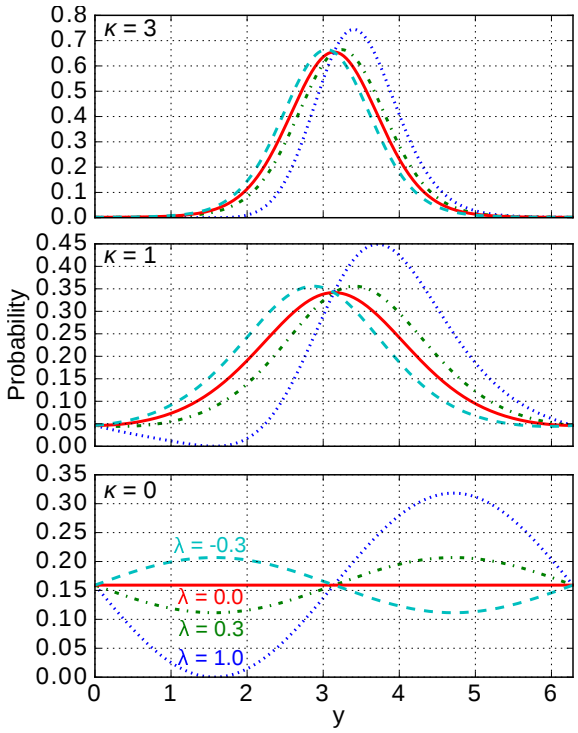

Fig. 2　正弦関数摂動 von Mises 分布の例．平均 μ は π としている．$\lambda = 0.0$ は，通常の von Mises 分布に等しい．

を，次式で定義する．

$$P^{(\mathrm{wc})}(y;\mu,\rho) = \frac{1-\rho^2}{1+\rho^2-2\rho\cos(y-\mu)} \tag{2}$$

ρ は $0 \leq \rho < 1$ の値を取り，Cauchy 分布の尺度パラメータに対応する．

3.1.2　von Mises 分布 ($\psi \to 0$)

von Mises 分布は，2 次元の等方性ガウス分布を単位円上で条件つけることで導出される確率分布である．von Mises 分布 $P^{(\mathrm{vm})}(y;\mu,\kappa)$ を，次式で定義する．

$$P^{(\mathrm{vm})}(y;\mu,\kappa) = \frac{\exp(\kappa\cos(y-\mu))}{2\pi I_0(\kappa)}, \tag{3}$$

ここで，$I_0(\cdot)$ は 0 次の第 1 種変形 Bessel 関数である．このときの κ は，ガウス分布の精度（分散の逆数）に対応するパラメータであり，非負値をとる．

3.1.3　ハート分布 ($\psi = 1$)

ハート型分布の名称は，変数の極座標変換後の分布形状がハート型曲線であることに由来する．ハート型分布 $P^{(\mathrm{ca})}(y;\mu,\rho)$ を次式で定義する．

$$P^{(\mathrm{ca})}(y;\mu,\rho) = \frac{1}{2\pi}(1+2\rho\cos(y-\mu)) \tag{4}$$

ρ の値の範囲は，$0 \leq \rho \leq 1/2$ である．ρ を負値とすることも可能だが，同様の分布形状を正値の ρ でも表現可能であるため，本稿では除外する．

3.2　正弦関数摂動

正弦関数摂動 [4] は，一般化ハート分布のような円周上の対称分布を，非対称分布に変形する手法である．正弦関数摂動一般化ハート分布 $P^{(\mathrm{ssgc})}(y;\mu,\kappa,\psi,\lambda)$

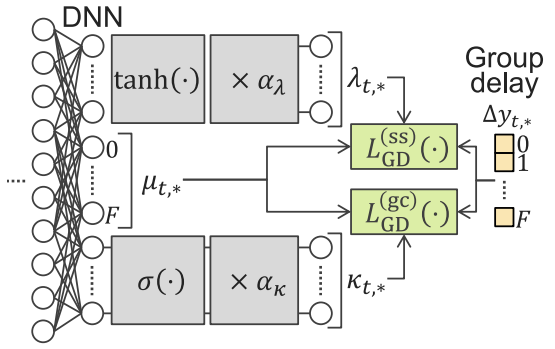

Fig. 3　DNN の出力層から損失関数計算までのアーキテクチャ．$\sigma(\cdot)$ はシグモイド関数である．ハート型分布または巻込み Cauchy 分布を使用する際には，$\kappa_{t,*}$ を $\rho_{t,*}$ に，α_κ を α_ρ に置換することに注意する．α_λ 及び α_κ は，事前に決定されるスカラ値である．

を，次式で定義する[1]．

$$P^{(\mathrm{ssgc})}(y;\mu,\kappa,\psi,\lambda) = P^{(\mathrm{gc})}(y;\mu,\kappa,\psi)\cdot P^{(\mathrm{ss})}(y;\mu,\lambda) \tag{5}$$

$$P^{(\mathrm{ss})}(y;\mu,\lambda) = (1+\lambda\sin(y-\mu)) \tag{6}$$

λ は，摂動のパラメータであり，$-1 \leq \lambda \leq 1$ の値をとる．例として，正弦関数摂動 von Mises 分布，すなわち $P^{(\mathrm{ssgc})}(y;\mu,\kappa,\psi\to 0,\lambda)$ の分布形状を Fig. 2 に示す．$\kappa = 0$ のとき，von Mises 分布は一様分布と等しくなるため，正弦関数摂動 von Mises 分布はスケールされた正弦関数をバイアスした形状となる．一方，κ が大きくなるにつれ λ による形状変化の効果は小さくなるが，最瀕値を μ から動かし，かつ，非対称分布に変形していることが分かる．この傾向は，ハート分布と巻込み Cauchy 分布でも同様である．

4　正弦関数摂動一般化ハート分布 DNN に基づく群遅延推定

正弦関数摂動一般化ハート分布 DNN を用いて，各フレーム・周波数の群遅延を振幅スペクトログラム $\boldsymbol{x}$ から予測する方法を提案する．

4.1　群遅延の定義

群遅延は，位相の周波数微分の負値として定義される．本項では，群遅延を一次差分で近似する．

$$\Delta y_{t,f} = -(y_{t,f+1} - y_{t,f}) \tag{7}$$

フレーム t・周波数 f における位相 $y_{t,f}$ は周期変数であるが，群遅延 $\Delta y_{t,f}$ もまた周期変数である．以降では，フレーム t における群遅延を $\Delta\boldsymbol{y}_t = [\Delta y_{t,0},\cdots,\Delta y_{t,f},\cdots,\Delta y_{t,F}]$ とする．

4.2　正弦関数摂動一般化ハート分布 DNN の学習

正弦関数摂動一般化ハート分布を条件付き確率分布として有する DNN を学習する．ここで，DNN を $\boldsymbol{G}(\cdot)$ とすると，フレーム t において $\boldsymbol{G}(\cdot)$ は，以下の 3 つのモデルパラメータを生成するように学習される．

$$\boldsymbol{\mu}_t = [\mu_{t,0},\cdots,\mu_{t,f},\cdots,\mu_{t,F}] \tag{8}$$

$$\boldsymbol{\kappa}_t = [\kappa_{t,0},\cdots,\kappa_{t,f},\cdots,\kappa_{t,F}] \tag{9}$$

[1]簡易化のため，摂動項を確率分布のように表記 ($P^{(\mathrm{ss})}(\cdot)$) するが，この項を全区間で積分すると 2π になることに注意する．

$$\boldsymbol{\lambda}_t = [\lambda_{t,0}, \cdots, \lambda_{t,f}, \cdots, \lambda_{t,F}] \tag{10}$$

ここで，$\mu_{t,f}$，$\kappa_{t,f}$，$\lambda_{t,f}$ はそれぞれ，フレーム t，周波数インデックス f におけるモデルパラメータ μ，κ，λ である．

DNN 学習時の損失関数は，次式に示す負の対数尤度であり，これを最小化するように DNN を学習する．

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{ssgc})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t, \boldsymbol{\lambda}_t) \\ &\quad = -\sum_{f=0}^{F} \log P^{(\mathrm{ssgc})}(\Delta y_{t,f}; \mu_{t,f}, \kappa_{t,f}, \psi, \lambda_{t,f}) \\ &\quad = -\sum_{f=0}^{F} \log P^{(\mathrm{gc})}(\Delta y_{t,f}; \mu_{t,f}, \kappa_{t,f}, \psi) \\ &\qquad -\sum_{f=0}^{F} \log P^{(\mathrm{ss})}(y; \mu_{t,f}, \lambda_{t,f}) \\ &\quad = L_{\mathrm{GD}}^{(\mathrm{gc})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t, \psi) + L_{\mathrm{GD}}^{(\mathrm{ss})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t) \end{aligned} \tag{11}$$

ここで，

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{gc})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t, \psi) \\ &\qquad \equiv -\sum_{f=0}^{F} \log P^{(\mathrm{gc})}(\Delta y_{t,f}; \mu_{t,f}, \kappa_{t,f}, \psi) \end{aligned} \tag{12}$$

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{ss})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t) \\ &\qquad \equiv -\sum_{f=0}^{F} \log P^{(\mathrm{ss})}(\Delta y; \mu_{t,f}, \lambda_{t,f}) \end{aligned} \tag{13}$$

とする．以降ではまず，一般化ハート分布 DNN の損失関数 $L_{\mathrm{GD}}^{(\mathrm{gc})}(\cdot)$ の特殊形として，ハート分布 DNN，von Mises 分布 DNN，巻込み Cauchy 分布 DNN の損失関数を定義し，次に正弦関数摂動に由来する損失関数 $L_{\mathrm{GD}}^{(\mathrm{ss})}(\cdot)$ を定義する．Fig. 3 は，本節で述べる損失関数の計算過程を図示したものである．

4.2.1 巻込み Cauchy 分布 DNN の損失関数

損失関数 $L_{\mathrm{GD}}^{(\mathrm{wc})}$ は，次式で与えられる．

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{wc})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t) = \\ &\quad \sum_{f=0}^{F} \left\{ \log\left(1 + \rho_{t,f}^2 - 2\rho_{t,f}\cos(\Delta y_{t,f} - \mu_{t,f})\right) \right. \\ &\qquad \left. - \log\left(1 - \rho_{t,f}^2\right) \right\} \end{aligned} \tag{14}$$

$\mu_{t,f}$ は実数値をとる．$\rho_{t,f}$ は $0 \leq \rho_{t,f} < 1$ の値をとるため，DNN の出力に対し sigmoid 関数をかけ，更に $\alpha_\rho = 0.99$ をかけたものを $\rho_{t,f}$ とする (Fig. 3 参照)．

4.2.2 von Mises 分布 DNN の損失関数

損失関数 $L_{\mathrm{GD}}^{(\mathrm{vm})}$ は，次式で与えられる．

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{vm})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t) = \\ &\quad \sum_{f=0}^{F} \left\{ \log\left(I_0(\kappa_{t,f})\right) - \kappa_{t,f}\cos(\Delta y_{t,f} - \mu_{t,f}) \right\} \end{aligned} \tag{15}$$

$\kappa_{t,f}$ は非負値をとるが，本稿では，$\log(I_0(\kappa_{t,f}))$ の値の発散を防ぐため，$\kappa_{t,f}$ の範囲を $0 \leq \kappa_{t,f} \leq 10$ とし，Fig. 3 の α_κ を 10 とする．なお，$\partial I_0(\kappa_{t,f}) / \partial \kappa_{t,f} = I_1(\kappa_{t,f})$ で与えられるため，通常の backpropagation により，DNN の学習が可能である．$I_1(\cdot)$ は 1 次の第 1 種変形 Bessel 関数である．

4.2.3 ハート分布 DNN の損失関数

損失関数 $L_{\mathrm{GD}}^{(\mathrm{ca})}$ は，次式で与えられる．

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{ca})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\kappa}_t) = \\ &\quad \sum_{f=0}^{F} -\log\left(1 + 2\rho_{t,f}\cos(\Delta y_{t,f} - \mu_{t,f})\right) \end{aligned} \tag{16}$$

$\rho_{t,f}$ は $0 \leq \rho_{t,f} \leq 1/2$ の値をとるため，Fig. 3 の α_ρ を 0.5 とする．

4.2.4 正弦関数摂動に由来する損失関数

Eq. (11) に示したとおり，正弦関数摂動一般化ハート分布 DNN の損失関数 $L_{\mathrm{GD}}^{(\mathrm{ssgc})}(\cdot)$ は，一般化ハート分布 DNN の損失関数 $L_{\mathrm{GD}}^{(\mathrm{gc})}(\cdot)$ と，正弦関数摂動に由来する損失関数 $L_{\mathrm{GD}}^{(\mathrm{ss})}(\cdot)$ の和で与えられる．後者の $L_{\mathrm{GD}}^{(\mathrm{ss})}(\cdot)$ は次式で与えられる．

$$\begin{aligned} &L_{\mathrm{GD}}^{(\mathrm{ss})}(\Delta\boldsymbol{y}_t, \boldsymbol{\mu}_t, \boldsymbol{\lambda}_t) = \\ &\quad -\sum_{f=0}^{F} \log\left(1 + \lambda_{t,f}\sin(\Delta y_{t,f} - \mu_{t,f})\right) \end{aligned} \tag{17}$$

$\lambda_{t,f}$ は $-1 \leq \lambda_{t,f} \leq 1$ の値をとるが，$\lambda_{t,f} = -1$ かつ $\mu_{t,f} = \Delta y_{t,f} + \pi N/2$，もしくは，$\lambda_{t,f} = 1$ かつ $\mu_{t,f} = \Delta y_{t,f} - \pi N/2$ のとき (N は整数) に，Eq. (17) が発散するため，本稿では，Fig. 3 の α_λ を 0.99 とする．

5 実験的評価

5.1 実験条件

実験的評価は，単一話者による読み上げ音声コーパス JSUT [8] を用いて実施した．学習データは，サブセット BASIC5000 に含まれる 5,000 文（約 6 時間），評価データは，サブセット ONOMATOPEE300 に含まれる 300 文である．サンプリング周波数は 16 kHz である．フレーム分析における窓長，シフト長，フーリエ変換長はそれぞれ，400 サンプル (25 ms)，80 サンプル (5 ms)，及び 512 サンプルとする．使用した窓関数は，ハミング窓である．DNN への入力特徴量は，当該フレーム及びその前後 2 フレームの対数振幅スペクトルを連結したベクトルである．入力特徴量は，学習時に平均 0・分散 1 に正規化する．DNN のアーキテクチャは，Feed-Forward 型であり，3 層・512 ユニットの gated linear hidden unit [9] を持つ．DNN のモデルパラメータは乱数により初期化する．最適化アルゴリズムには，AdaGrad [10] を利用する．

本稿では，以下の確率分布を持つ DNN を比較する．カッコ内は，DNN 学習時の損失関数である．

- 巻込み Cauchy (Eq. (14))
- 正弦関数摂動巻込み Cauchy (Eq. (14)+Eq. (17))
- von Mises (Eq. (15))
- 正弦関数摂動 von Mises (Eq. (15)+Eq. (17))
- ハート (Eq. (16))
- 正弦関数摂動ハート (Eq. (16) + Eq. (17))

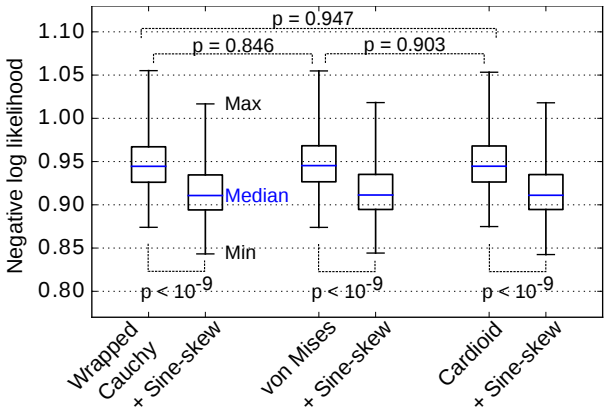

Fig. 4　テストデータに対する，各確率分布 DNN の負の対数尤度の箱ひげ図．箱は，第 1・第 3 四分位点を表す．p は p 値である．比較として，一様分布の負の対数尤度は，1.83 である．”+sine-skew”は，正弦関数摂動を加えた分布を表す．

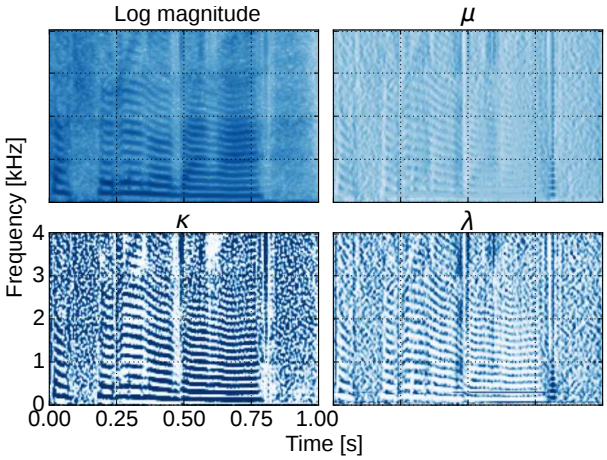

Fig. 5　対数振幅スペクトログラム，及び DNN から生成された正弦関数摂動 von Mises 分布の $\mu_{t,f}$，$\kappa_{t,f}$，$\lambda_{t,f}$．濃い色ほど大きい値であることを表す．$\kappa_{t,f}$ の範囲は 0 から 10，$\lambda_{t,f}$ の範囲は-0.99 から 0.99 である．本研究で扱う周波数帯域は 0〜8 kHz であるが，ここでは，0〜4 kHz のみを示す．

5.2　実験結果

テストデータに対し，各分布 DNN の負の対数尤度を計算した．負の対数尤度は，1 発話毎に全フレーム・周波数で平均し，それをテストデータの全発話で平均した．計算結果を Fig. 4 に示す．

巻込み Cauchy，von Mises，ハート分布 DNN 間では，尤度に大きな差は見られない．一方で，正弦関数摂動を加えることにより，負の対数尤度は，有意に向上することが分かる．以上より，群遅延の分布の非対称性を，正弦関数摂動により捉えられることが明らかになった．

DNN から生成された，正弦関数摂動 von Mises 分布のモデルパラメータの例を Fig. 5 に示す．図より，有声音のスペクトルにおいて調波成分の現れる周波数（すなわち，F_0 の定数倍の周波数）において，$\kappa_{t,f}$ は大きい値をとるため，Fig. 2 上のような，比較的尖った分布となることが分かる．一方，調波成分以外の周波数（すなわち，F_0 の半分の奇数倍の周波数）においては，$\kappa_{t,f}$ は 0 に近く，$\lambda_{t,f}$ は，比較的大きい正値をとることが分かる．すなわち，Fig. 2 下のような，バイアスかつスケールされた正弦関数のような分布となる．これらについては，更に調査する必要がある．

6　まとめ

本稿では，振幅スペクトログラムから群遅延を推定する際の，群遅延の分布の非対称性を捉えるために，正弦関数摂動一般化ハート分布を条件付き分布として有する DNN を提案した．実験結果から，正弦関数摂動により実現される円周上の非対称分布を導入することで，従来の対称分布よりも高精度に群遅延の分布をモデル化できることを示した．今後は，skewness と音声的特徴の関係性を分析し，主観的音声品質への影響を調査する．

謝辞：本研究の一部は，セコム科学技術支援財団，JSPS 科研費 18K18100 の助成を受け実施した．また，本研究の一部は，東京大学 計数工学科 近藤 樹氏の協力を得て実施した．

参考文献

[1] 高道慎之介, 齋藤佑樹, 高宗典玄, 北村大地, and 猿渡洋, “von Mises 分布 DNN に基づく振幅スペクトログラムからの位相復元,” in **情報処理学会研究報告**, 東京, Jun. 2018, pp. 1–6.

[2] S. Takamichi, Y. Saito, N. Takamune, D. Kitamura, and H. Saruwatari, “Phase reconstruction from amplitude spectrograms based on von-Mises-distribution deep neural network,” in *Proc. IWAENC*, Tokyo, Japan, Sep. 2018 (to appear).

[3] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Sons Ltd., 1999.

[4] T. Abe and A. Pewsey, “Sine-skewed circular distributions,” *Statistical Papers*, vol. 52, no. 3, pp. 683–707, Aug. 2011.

[5] M. C. Jones and A. Pewsey, “A family of symmetric distributions on the circle,” *Journal of the American Statistical Association*, vol. 100, no. 472, pp. 1422–1428, Dec. 2005.

[6] K. Shimizu and K. Iida, “Person type vii distributions on spheres,” *Communications in Statistics - Theory and Methods*, vol. 31, no. 4, pp. 513–526, 2002.

[7] D. E. Cartwright, “The use of directional spectra in studying the output of a wave recorder on a moving ship,” *Ocean Wave Spectra*, pp. 203–218, 1963.

[8] R. Sonobe, S. Takamichi, and H. Saruwatari, “JSUT corpus: free large-scale japanese speech corpus for end-to-end speech synthesis,” vol. abs/1711.00354, 2017.

[9] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, “Language modeling with gated convolutional networks,” vol. abs/1612.08083, 2016.

[10] J. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” *Journal of Machine Learning Research*, vol. 12, pp. 2121–2159, 2011.