

# 人間GAN : 人間による知覚的識別に基づく敵対的生成ネットワーク

藤井 一貴<sup>1,2</sup>, 齋藤 佑樹<sup>1</sup>, 高道 慎之介<sup>1</sup>, 馬場 雪乃<sup>3</sup>, 猿渡 洋<sup>1</sup>

1, 東京大学 2, 徳山工業高等専門学校 3, 筑波大学

## 人間による知覚評価を用いて生成モデルを学習する敵対的生成ネットワーク

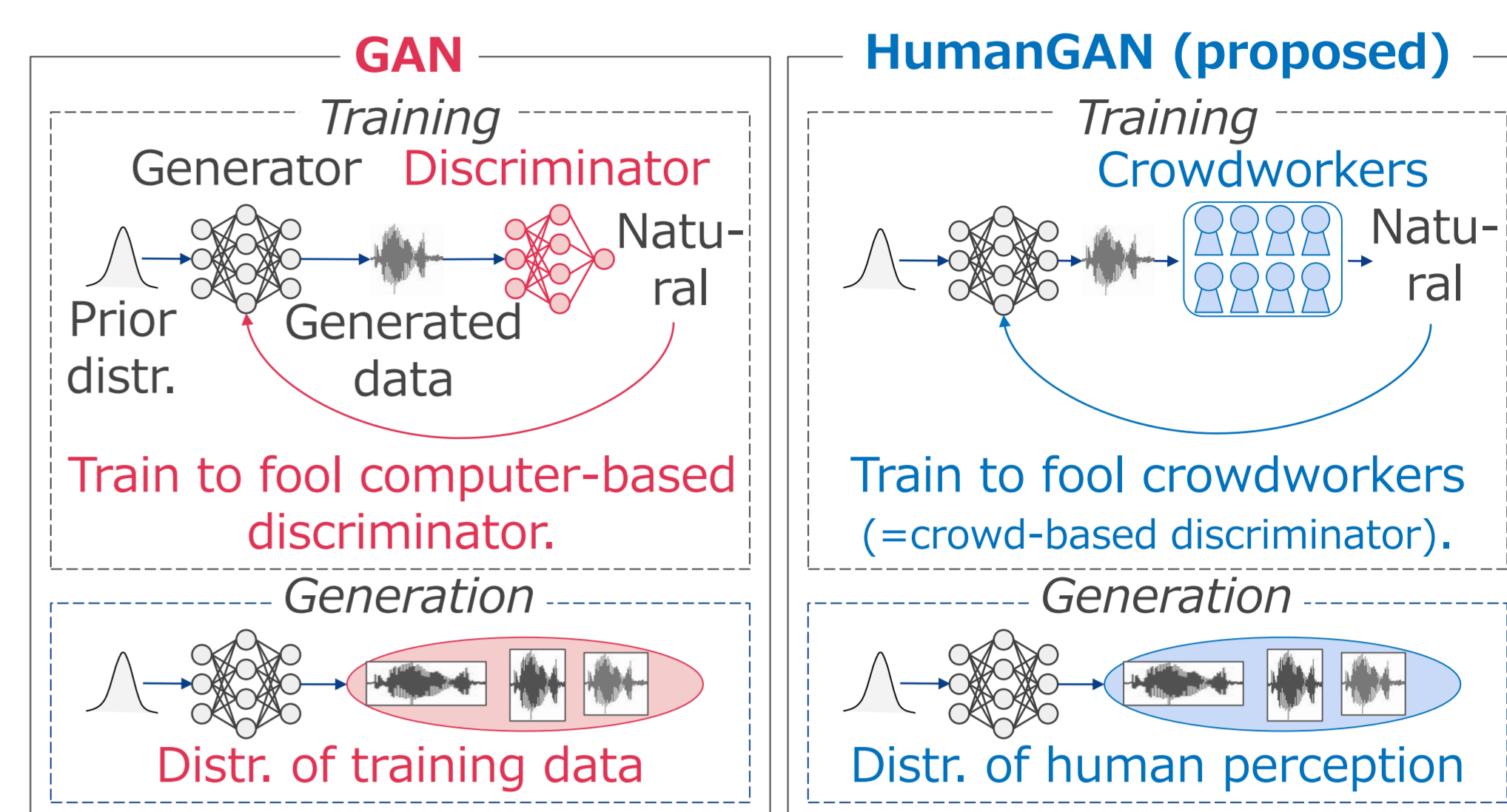
### ■ 敵対的生成ネットワーク (generative adversarial network; GAN) [1]

- 深層生成モデルの代表例. 生成モデルと識別モデルを用いて互いに学習を行う
- 実在データ (学習データ) 分布に従うデータをランダムにサンプリングすることができる

### ■ 実在データの分布しか表現することができない

- しかし, 人間の知覚は実在データ分布の逸脱に関してある程度の許容範囲を持つ
  - 例: 人間の音声加工して創り出した非実在の音声でも人格性を認める
- 通常のGANでは人間が許容出来る範囲 (知覚分布) は表現不可

### ■ 人間を学習の枠組みに組み込んだ新しいGANを提案する



本発表の主題: 人間の知覚分布を表現可能な生成モデルの学習方法とその実験的評価

### 通常のGAN (従来手法)

#### ■ 従来法の概要

- 生成モデルの表現する分布を **学習データの分布に一致**
- データを生成する **生成モデル** と, 実在データと生成データを識別する **識別モデル** を使用

$$V(G, D) = \sum_{n=1}^N \log D(x_n) + \sum_{n=1}^N \log (1 - D(G(z_n)))$$

G: 生成モデル, D: 識別モデル (DNN), N: 実在/生成データ数,  $x_n$ : 実在データ,  $z_n$ : 入力乱数

#### ■ 生成モデルの学習

- 識別モデルに対して生成データの事後確率を問い合わせ
- 識別モデルが生成データを実在データとして識別するよう学習
- 計算過程は微分可能 → 誤差逆伝搬法による学習が **可能**

#### ■ 識別モデルの学習

- 生成データと実在データを正確に識別するよう学習

#### ■ 問題点

- **人間の知覚分布が実在データ分布よりも広い場合にその領域は表現不可能**

### 人間GAN (提案手法)

#### ■ 提案法の概要

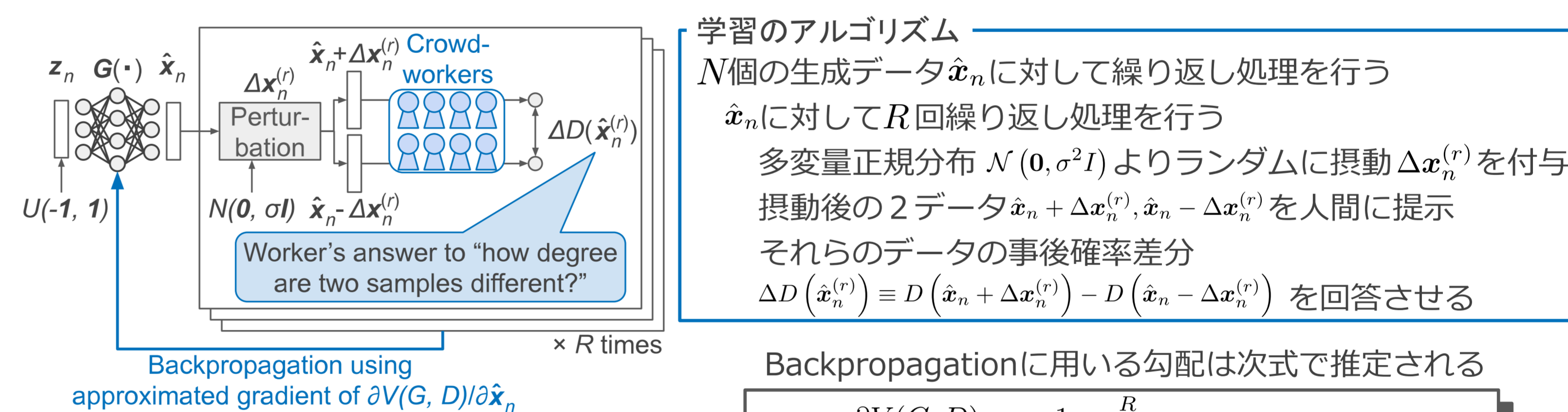
- 生成モデルの表現する分布を **人間の知覚分布に一致**
- **生成モデル** と, 識別モデルの代わりとして **人間の知覚による評価** を使用

$$V(G, D) = \sum_{n=1}^N D(G(z_n))$$

G: 生成モデル, D: 識別モデル (人間), N: 生成データ数,  $z_n$ : 入力乱数

#### ■ 生成モデルの学習

- 人間は微分不可能 → 誤差逆伝搬法による学習が **不可能**
- 人間を「事後確率差分を出力する black-box システム」とみなし勾配推定
  - Natural Evolution Strategies (NES) [2] を用いた学習を提案
  - 摂動と知覚評価に基づく勾配近似により, 誤差逆伝搬法による学習が **可能**



Backpropagationに用いる勾配は次式で推定される

$$\frac{\partial V(G, D)}{\partial \hat{x}_n} = \frac{1}{2\sigma^2 R} \sum_{r=1}^R \Delta D(\hat{x}_n^{(r)}) \cdot \Delta x_n^{(r)}$$

$\sigma$ : 定数の標準偏差,  $r$ : 摂動の index ( $1 \leq r \leq R$ ),  $I$ : 単位行列

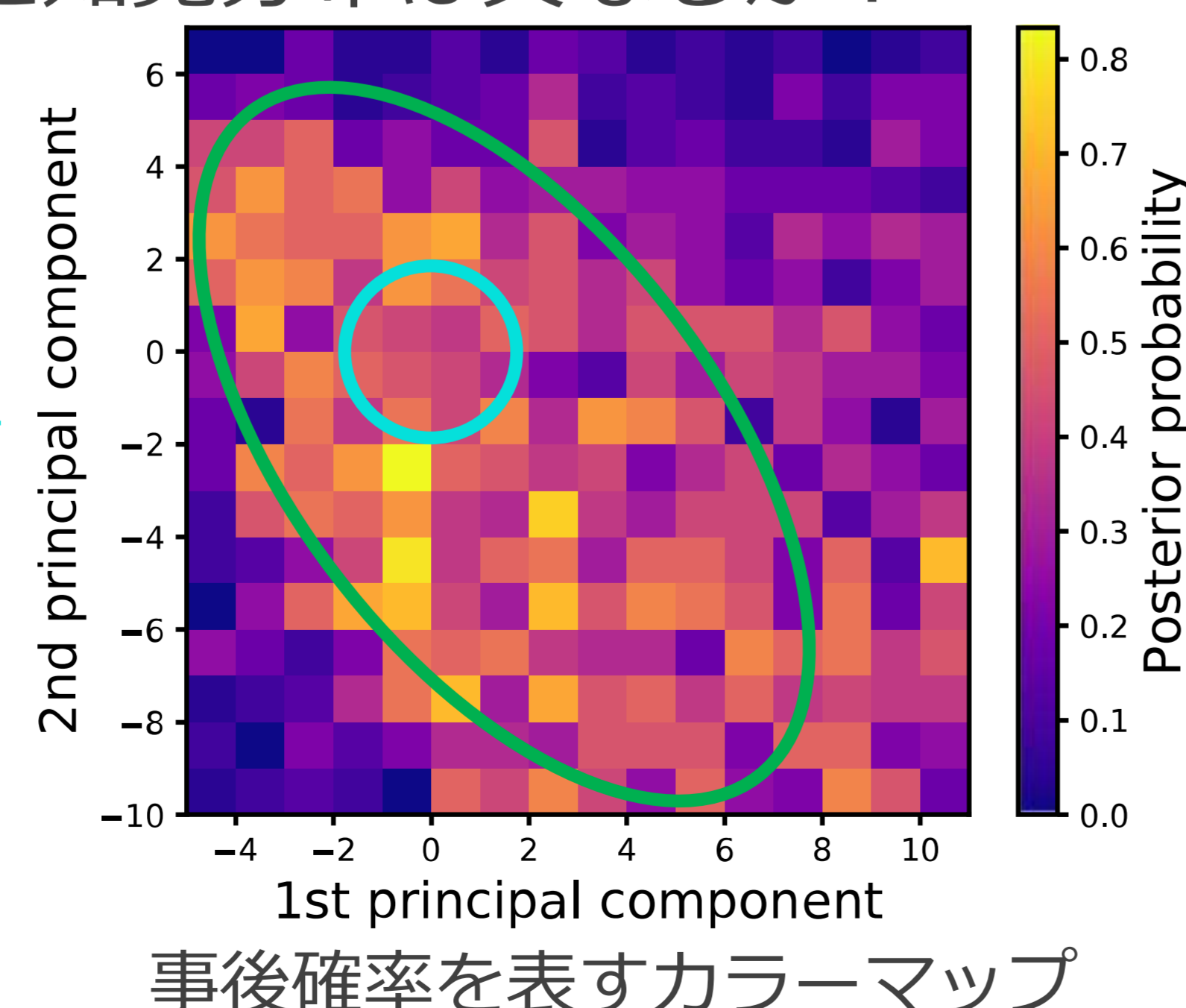
### 音声の自然性に関する実験的評価

#### ■ 実験条件

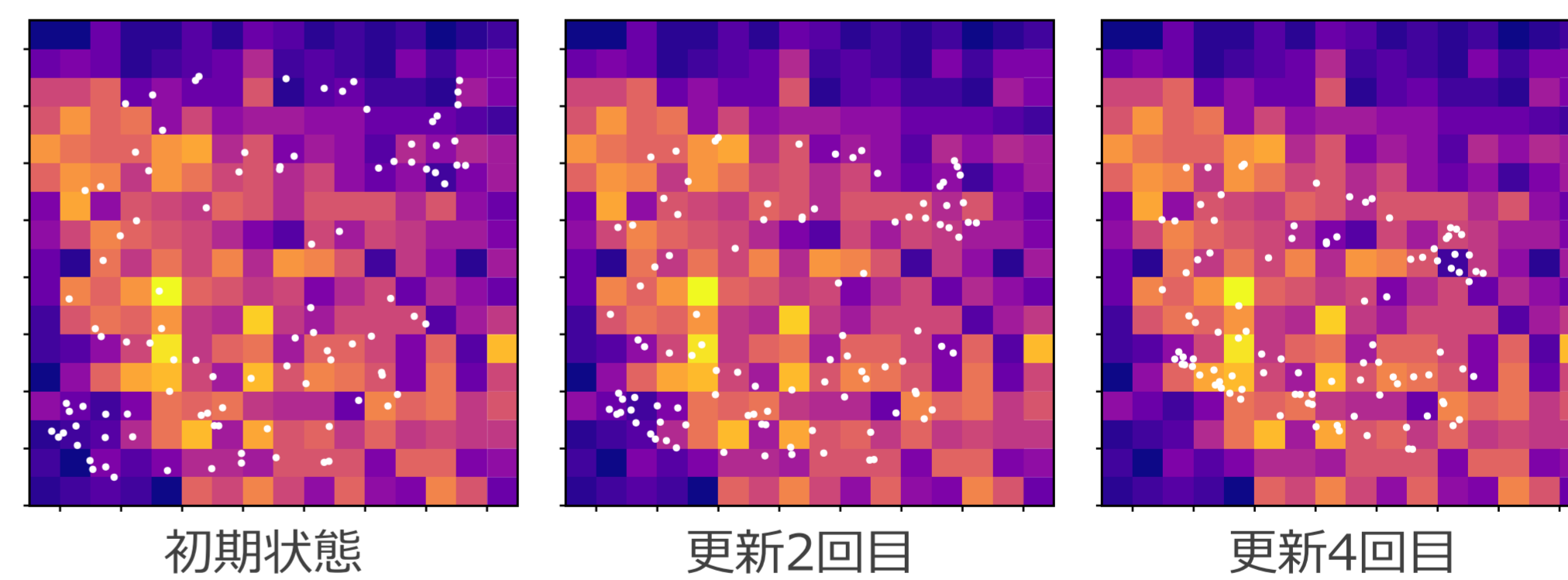
項目	値・設定
使用データ	JVPDコーパス[3] (女性話者199名) の 513次元対数スペクトル包絡の主成分分析により得られる第1/2主成分 (平均0分散1に正規化済)
サンプリングレート	16 kHz
入力乱数 $z_n$	2次元一様乱数 $U(-1, 1)$
DNN アーキテクチャ	Feed-Forward 2 - 4×2 (sigmoid) - 2 (linear)
最適化手法	確率的勾配降下法

#### ■ 実験①: 実在データ分布と知覚分布は異なるか?

- 2次元空間をグリッド分割, 各グリッドにおける音声の自然性の主観評価を実施
- **通常のGANで表現可能な分布と知覚分布は異なる**  
= 人間GANによる枠組みが必要であることが示された



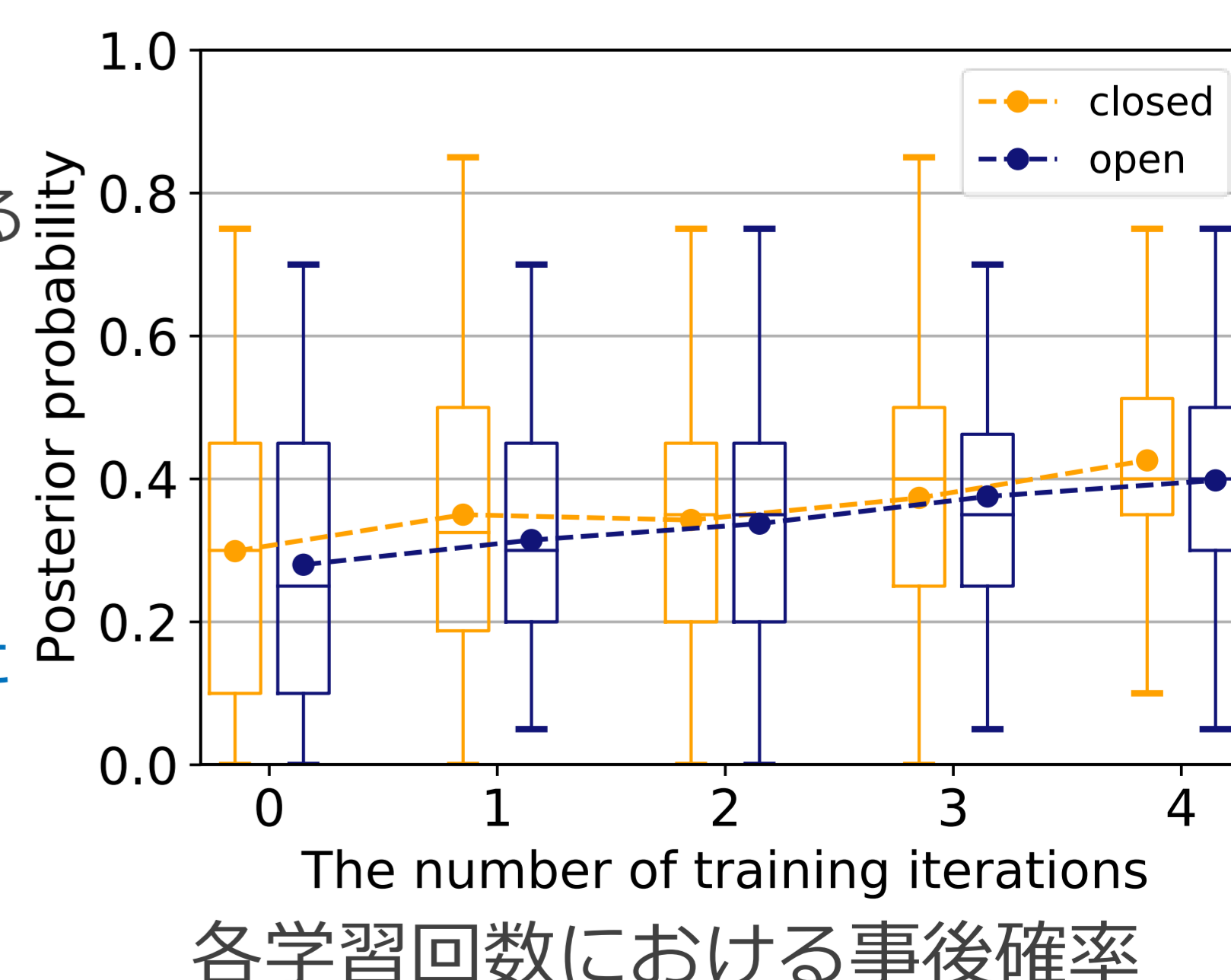
#### ■ 実験② 学習の様子を確認 (白い点は生成データ)



- 学習の反復により, 生成データは事後確率の高い領域に遷移 = 人間GANは知覚分布を表現する生成モデル学習を可能にする

#### ■ 実験③ 学習の反復により事後確率は上昇するか?

- 音声の自然性に関する事後確率の問い合わせ
  - 学習に用いた乱数による特徴量 (closed)
  - 新しい乱数による特徴量 (open)
- 両者とも事後確率は上昇 = 人間GANにより学習された生成モデルが知覚分布を表現し得る事を定量的に確認された



#### 参考文献

[1] Goodfellow et al., Proc. NIPS, 2014. [2] Ilyas et al., Proc. ICML, 2018. [3] <http://research.nii.ac.jp/src/en/JVPD.html>