HumanACGAN: conditional generative adversarial network with human-based auxiliary classifier and its evaluation in phoneme perception

Yota Ueda[1], Kazuki Fujii[2], Yuki Saito[1], Shinnosuke Takamichi[1], Yukino Baba[3] and Hiroshi Saruwatari[1]
1. The University of Tokyo, Japan, 2. National institute of Technology, Tokuyama College, Japan, 3. University of Tsukuba, Japan

ICASSP 2021
paper ID: 1681

1. SYNOPSIS

1. Proposes a generative adversarial network (GAN) that represents conditional human acceptable distributions.
2. Demonstrates that the proposed algorithm successfully trains a generator.

2. RESEARCH BACKGROUND

Human-acceptable distribution

Humans can accept outside of real data

Human-acceptable distribution

Conditional human-acceptable distribution for practical use (text-to-speech etc.)

Modeling method

	Real-data distribution	Human-acceptable distribution
Conditional	GAN [1]	HumanGAN [2]
Unconditional	ACGAN [3]	HumanACGAN (ours)

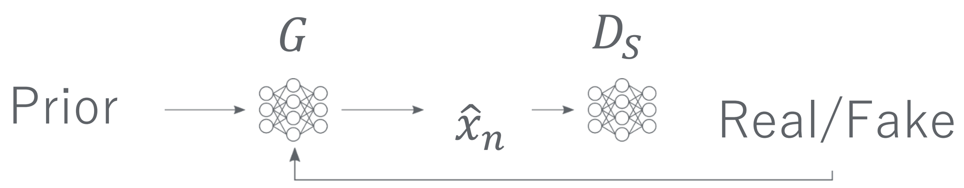
3. RELATED WORKS

	Generator G	Discriminator D_S	Classifier D_C
GAN [1]	DNN	DNN	-
HumanGAN [2]	DNN	Humans	-
ACGAN [3]	DNN	DNN	DNN

GAN [1]

Minimizing $L_S = \sum \log(1 - D_S(\hat{x}_n))$ during generator training

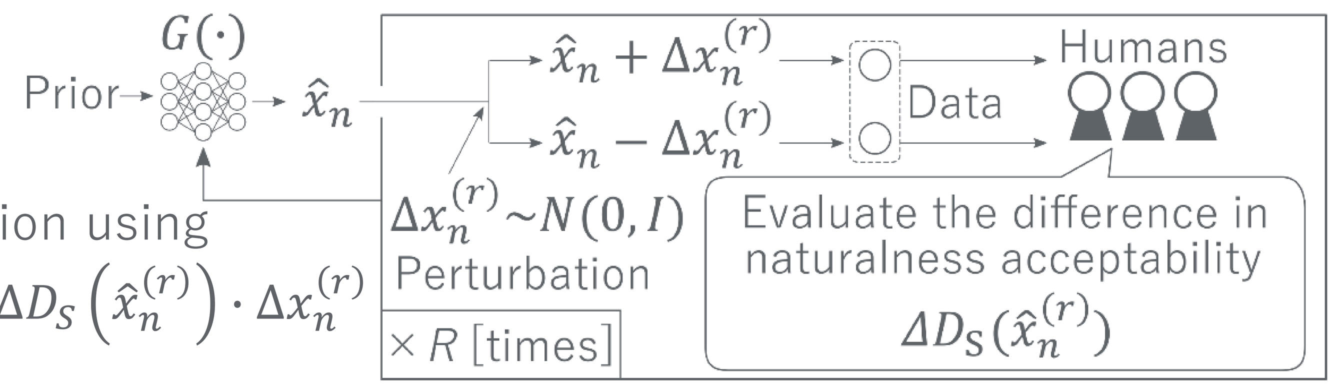
Gradient is calculated analytically.



HumanGAN [2]

Maximizing $L_S = \sum D_S(\hat{x}_n)$ during generator training

Estimating gradient using Natural evolution strategy (NES) [4]



Backpropagation using

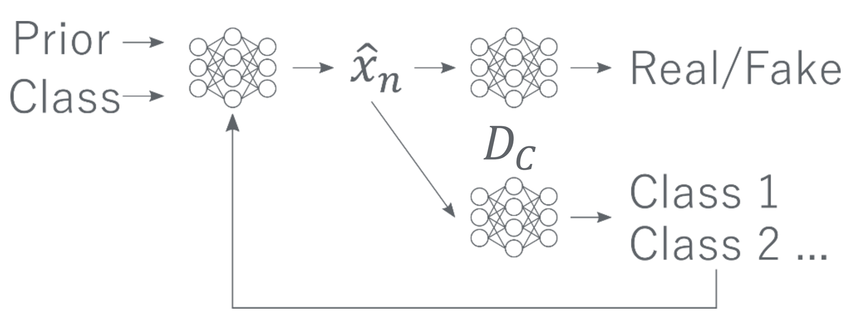
$\partial L_S / \partial \hat{x}_n \approx \frac{1}{2\sigma^2 R} \sum_{r=1}^{R} \Delta D_S\left(\hat{x}_n^{(r)}\right) \cdot \Delta x_n^{(r)}$

ACGAN [3]

Maximizing $-L_S + \lambda L_C$ during generator training

$L_C = \sum \log D_C(x_n, c_n) + \sum \log(D_C(\hat{x}_n), c_n))$

Gradients are calculated analytically.



4. PROPOSED METHOD: HumanACGAN

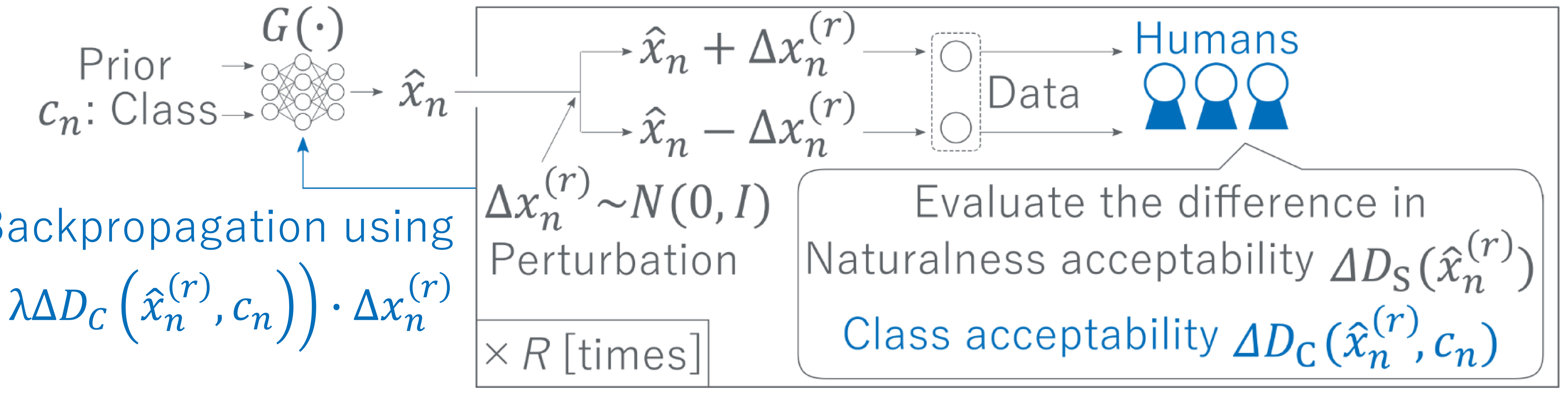
	Generator G	Discriminator D_S	Classifier D_C
HumanACGAN	DNN	Humans	Humans

Humans act as a discriminator and a classifier.

Maximizing $L_S + \lambda L_C$ during generator training

$L_S + \lambda L_C \; (L_C = \sum D_C(\hat{x}_n, c_n))$

Estimating gradient using NES



Backpropagation using

$$\frac{\partial(L_S + \lambda L_C)}{\partial \hat{x}_n} \approx \frac{1}{2\sigma^2 R} \sum_{r=1}^{R} \left(\Delta D_S\left(\hat{x}_n^{(r)}\right) + \lambda \Delta D_C\left(\hat{x}_n^{(r)}, c_n\right) \right) \cdot \Delta x_n^{(r)}$$

5. EXPERIMENTAL EVALUATIONS

Experimental conditions

Corpus	JVPD [5]
Analysis/synthesis	WORLD [6]
Optimization	Stochastic gradient descent
Data space	Applying PCA (principal component analysis) to log spectral envelope and using 1st and 2nd PC as data space

Experiment 1: estimation of gradient

Method

Preparing perturbated data $\hat{x} + \Delta x$ and $\hat{x} - \Delta x$

Synthesizing speech waveform from $\hat{x} + \Delta x$ and $\hat{x} - \Delta x$

Humans' tasks

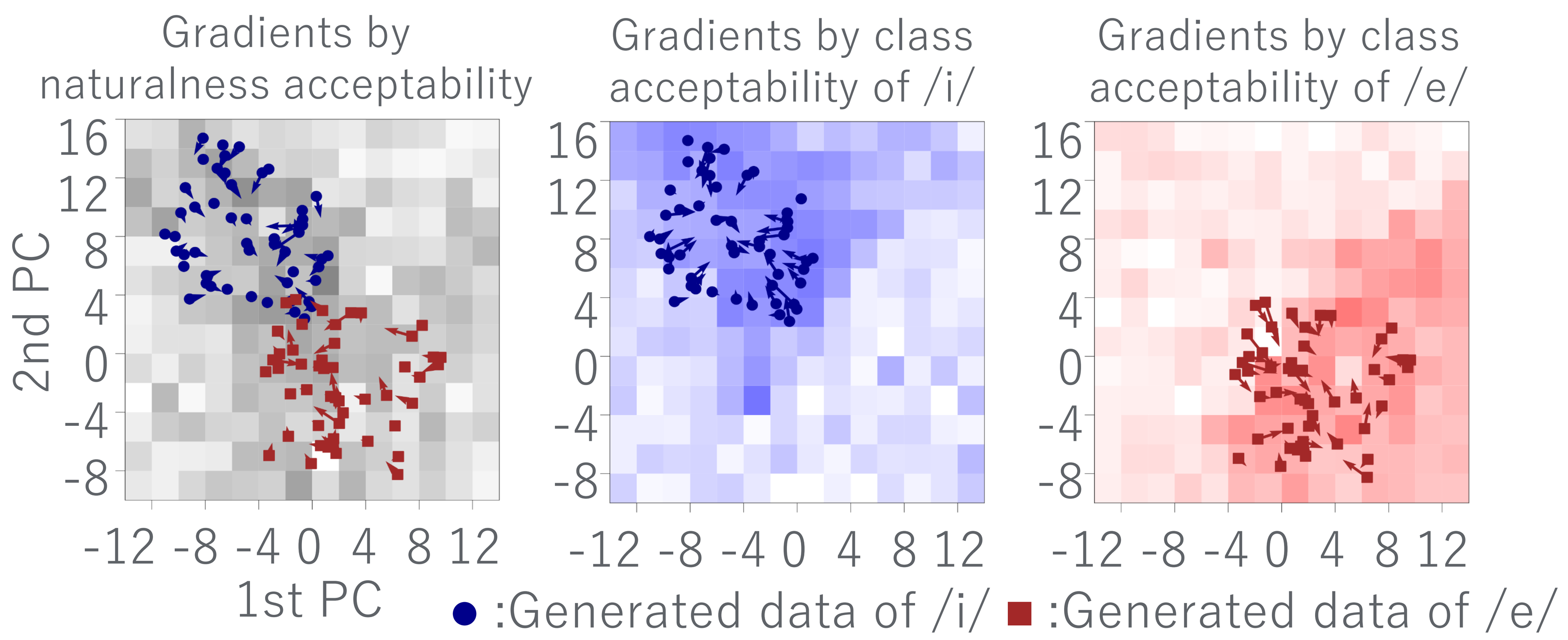
As discriminator: Evaluates difference in naturalness acceptability D_S

As classifier: Evaluates difference in class acceptability D_C

Results

Gradients pointed to darker (i.e., higher posterior) zones.

Gradients for generator training were estimated correctly.



Experiment 2: increase of objective function

Method

Training generator for 4 iterations using estimated gradient

Perceptual test for generated data from initialized and trained generator

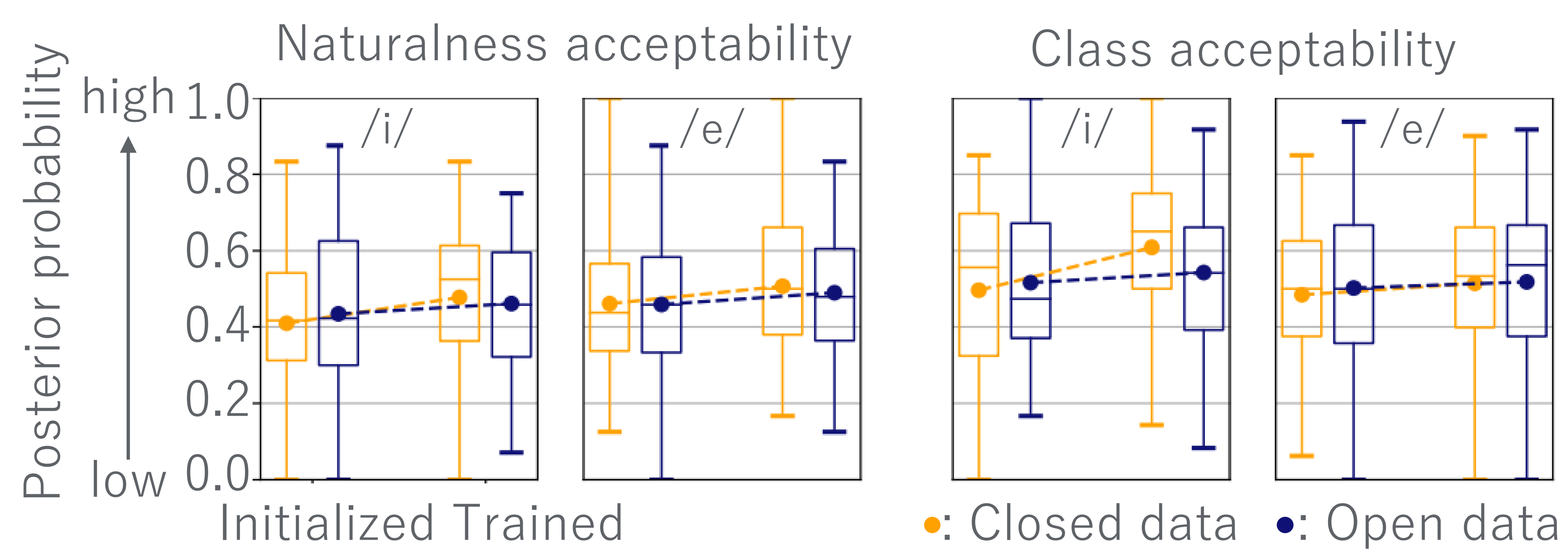
Evaluated sample

Closed data (used in training) and open data

Results

Prior probabilities of naturalness and class acceptability increased.

Generator was trained successfully.



References [1] Goodfellow et al., *Proc. NIPS*, 2014. [2] Fujii et al., *Proc. ICASSP*, 2020. [3] Mirza et al., arXiv:1411.1784, 2014. [4] Ilyas et al., *Proc. ICML*, 2018. [5] " http:// research.nii.ac.jp/src/en/JVPD.html [6] Morise et al., *IEICE*, 2016.