

파란학기제 · 기말 발표

ControlV

reference-based audio FX style transfer의 alpha 조절 단조성 분석

DeepAFx-ST 기반 구조에서 단조성 문제 정의와 MonoLoss regularization 검증

MOTIVATION

스타일 변환의 품질이 아니라, 조절 가능성에 집중한다

목표 사용자가 변환 강도를 α 로 연속 조절할 수 있어야 한다

$\alpha = 0$

원본 (input)



$\alpha = 1$

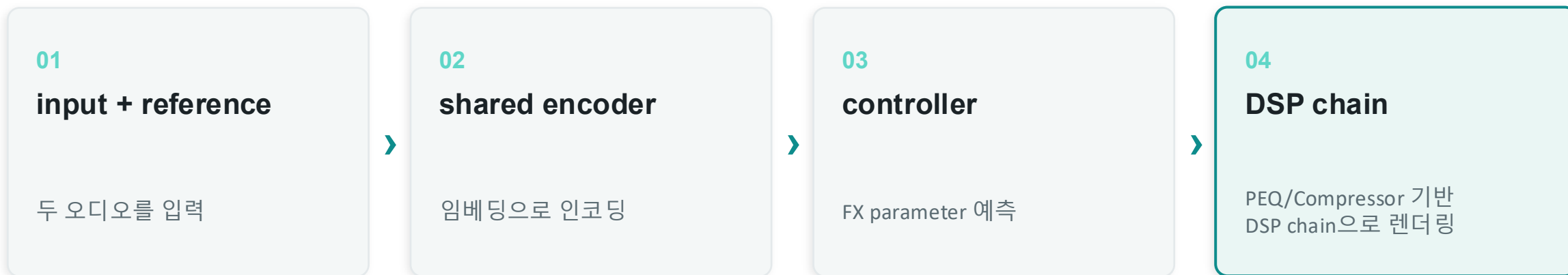
목표 스타일 (reference)

문제

α 를 올려도 소리의 변화가 항상 매끄럽게 이어지지 않는다.

BACKGROUND

DeepAFx-ST 기반 구조



DeepAFx-ST style architecture — DSP 기반 구조라 출력을 FX parameter와 연결해 해석할 수 있다.

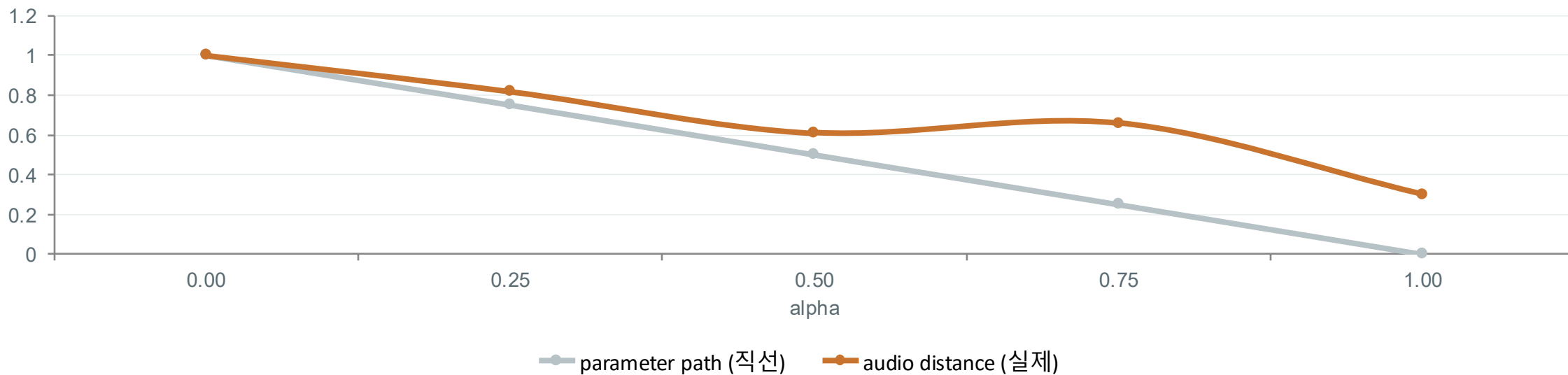
PROBLEM

파라미터를 선형 보간해도 소리는 단조롭게 변하지 않는다

$$p_{\alpha} = p_{\text{default}} + \alpha \cdot (p_{\text{target}} - p_{\text{default}})$$

파라미터가 선형으로 변해도, 실제 audio distance는 단조롭게 변한다고 보장할 수 없다.

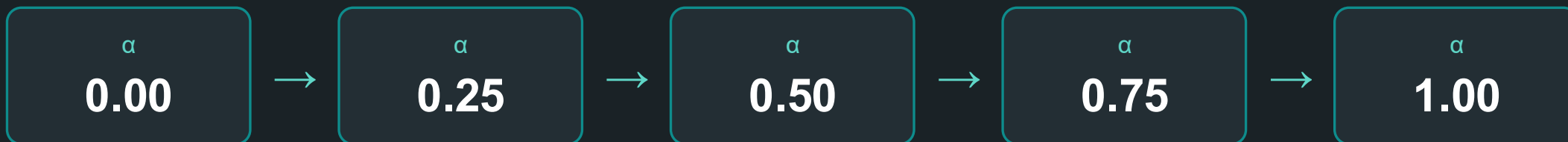
parameter path는 직선, audio distance는 비선형 곡선



METHOD

단조성 위반을 줄이기 위한 MonoLoss

인접한 alpha 구간을 비교해, 높은 alpha의 출력이 낮은 alpha보다 reference에 더 가깝도록 학습한다.



$$L_{mono} = \max(0, d(out_{high}, target) - d(out_{low}, target) + margin)$$

단조성을 위반하는 경우에만 penalty가 발생한다.

Monotonicity rate는 단조성이 **깨진 횟수**를, violation magnitude는 깨졌을 때 **얼마나 크게 깨졌는지**를 본다.

EXPERIMENTAL SETUP

실험 구성

5

model variants

2

eval protocols

500

samples / eval

200

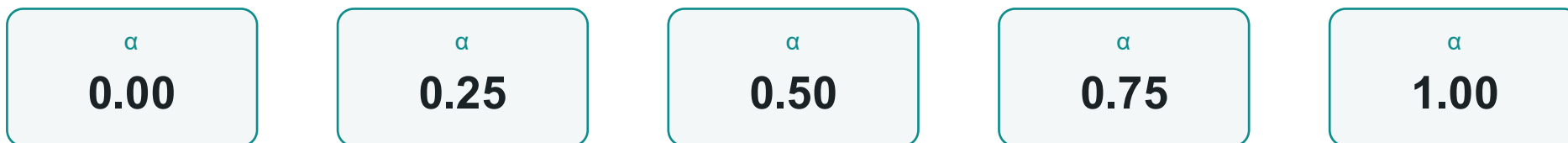
epochs · SageMaker

구분	내용
Models	Baseline · MonoLoss L10 · MonoLoss L30 · Residual L10 · Residual L30
Eval	Random eval · Styled eval (각 500 samples)
Metrics	Endpoint MRSTFT · MRSTFT monotonicity · violation magnitude · transition-wise monotonicity

MonoLoss L10을 중심 비교 대상으로 분석한다.

endpoint가 아니라, alpha 경로 전체를 평가했다

① 각 sample에서 alpha 5개 지점을 생성



② 인접 지점 사이 4개 transition을 점검

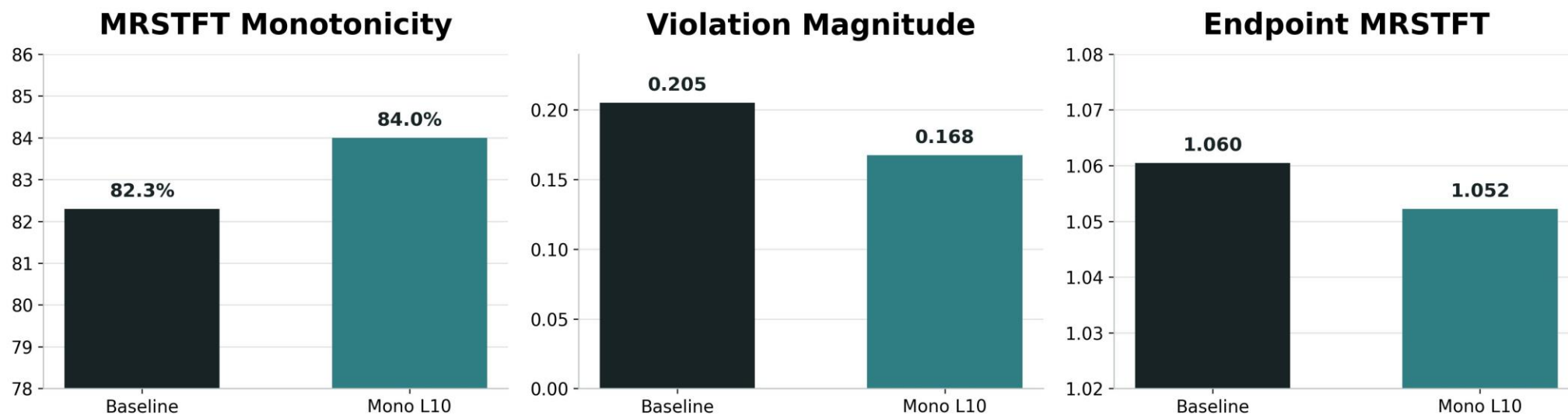


Pass 조건 $d(\text{out_high}, \text{target}) \leq d(\text{out_low}, \text{target})$

RESULT 1

MonoLoss는 단조성 위반의 크기를 줄였다

MonoLoss Improves Random-Eval Monotonicity

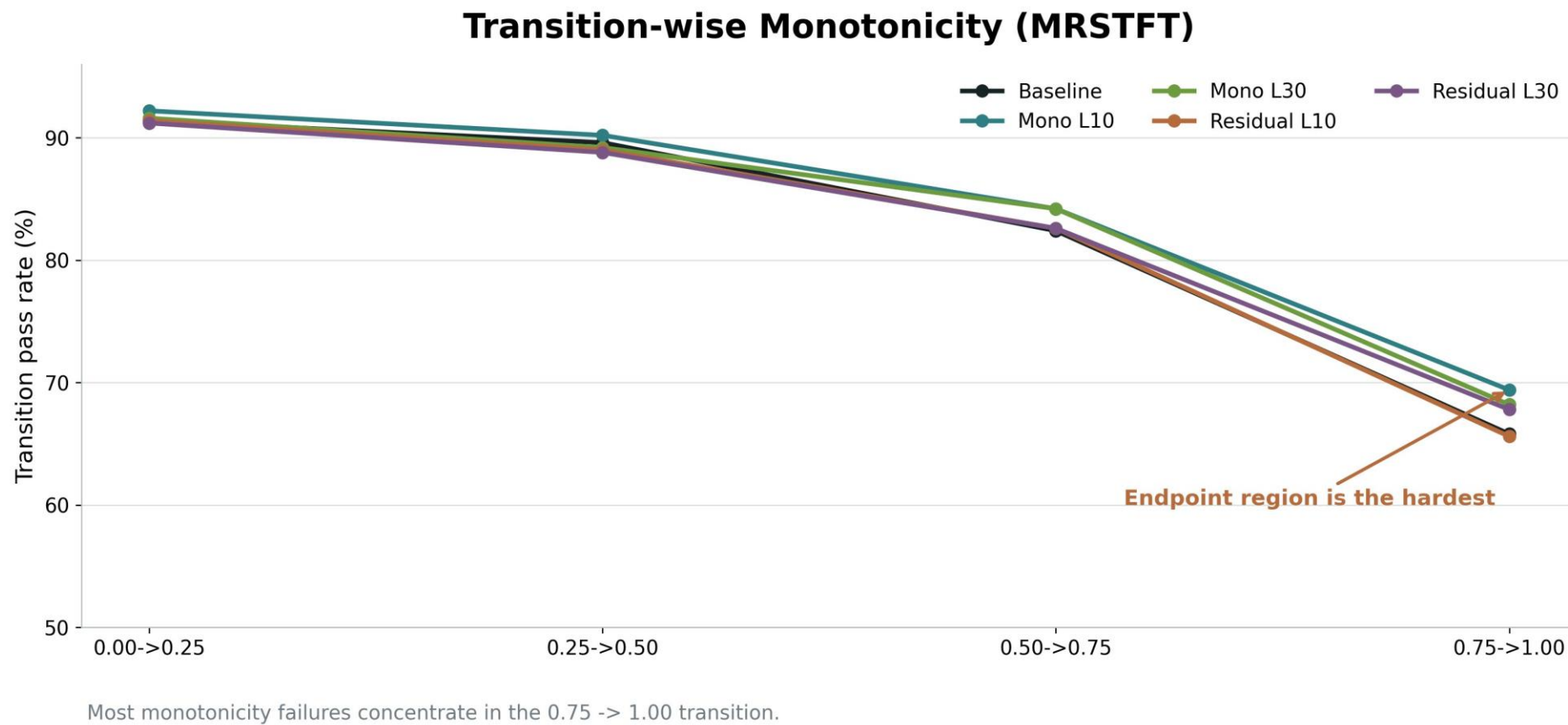


MRSTFT-based metrics only; n=500 random FX samples

핵심 단조성 성공률은 82.3% → 84.0%로 소폭 증가했지만, violation magnitude는 0.205 → 0.168로 약 18% 감소했다.

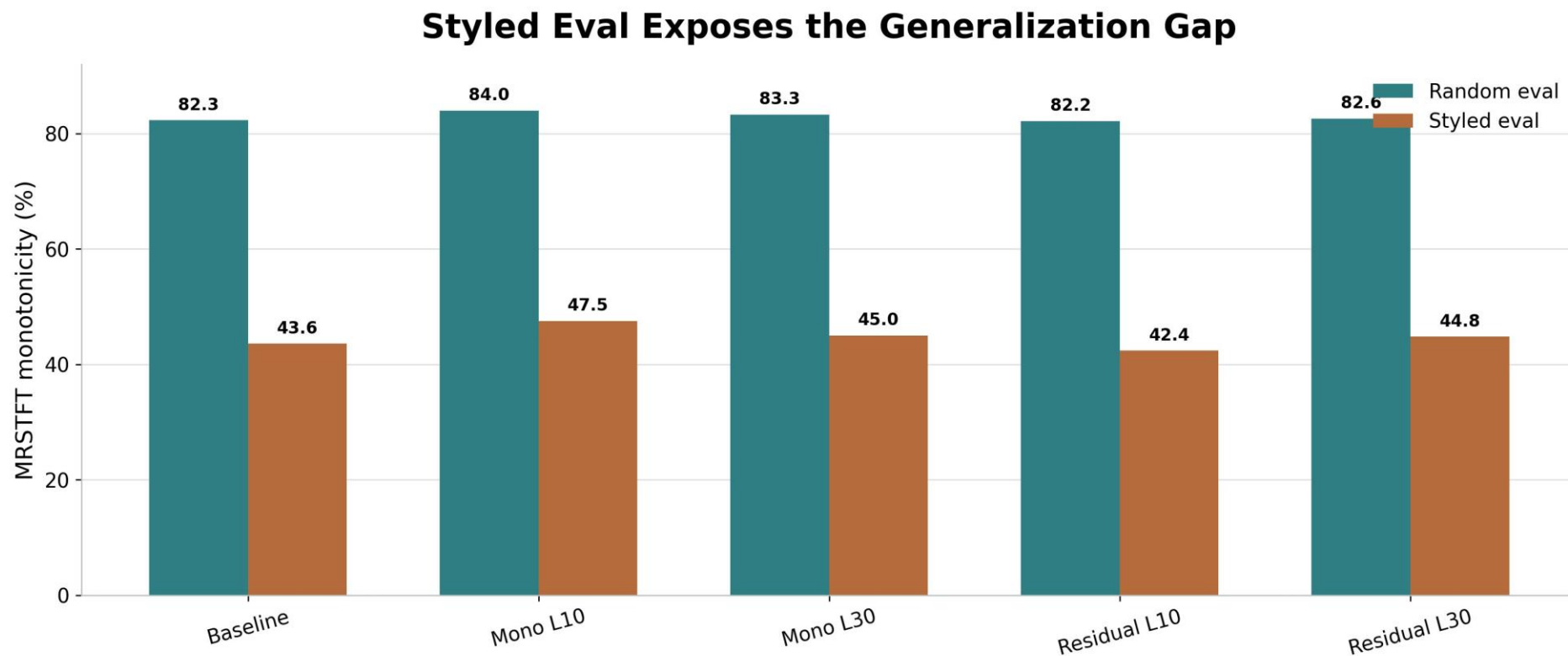
RESULT 2

alpha 후반 구간에서 단조성 위반이 가장 많이 발생했다



핵심 endpoint 품질 실패가 아니라, 0.75 → 1.00 구간에서 target distance가 항상 더 줄어들지는 않았다는 의미다.

Styled eval에서 현실 조건의 어려움이 드러났다



Random FX interpolation is easier; named style transfer is a harder, more realistic test.

결론

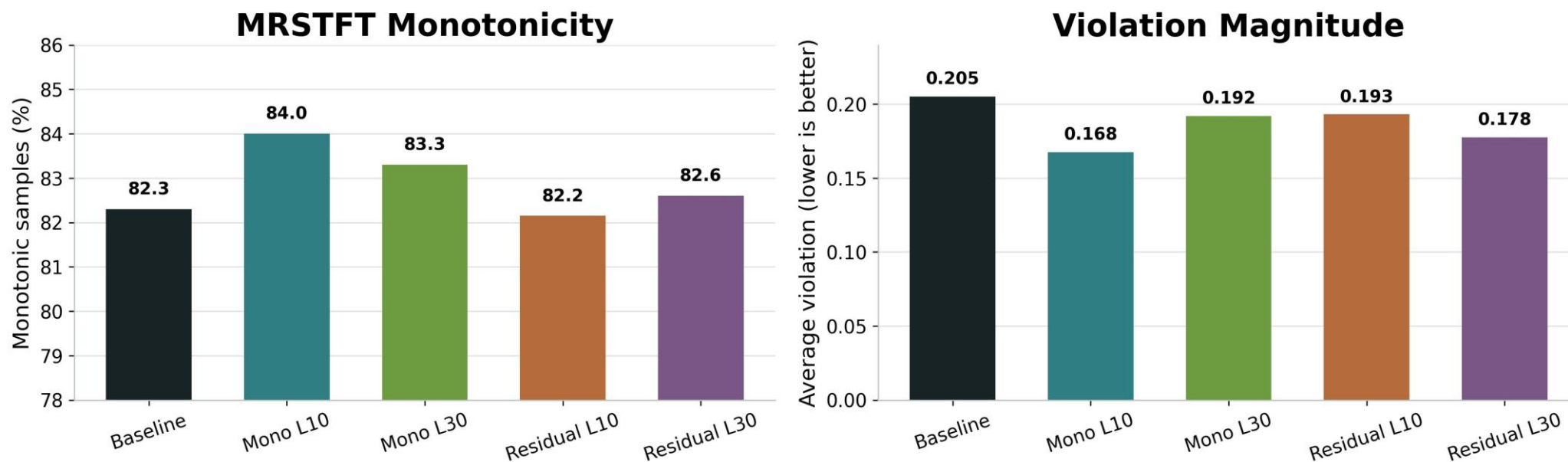
- 1 DeepAFx-ST 기반 audio FX style transfer에서 alpha monotonicity 문제를 정의했다.
- 2 MonoLoss는 단조성 성공률을 크게 높이기보다, violation magnitude를 줄이는 regularization 효과를 보였다.
- 3 Styled eval에서는 reference·dataset·style 정의의 영향이 크게 나타나 realistic style transfer의 어려움을 확인했다.

Discussion

- 파라미터 보간이 곧 지각적 보간을 의미하지는 않는다
- endpoint 품질과 alpha 경로의 단조성은 다른 문제다
- MonoLoss는 control behavior를 개선할 수 있지만, style quality 자체를 보장하지는 않는다
- Residual path는 이번 설정에서 MonoLoss보다 뚜렷하게 우수하지 않았다

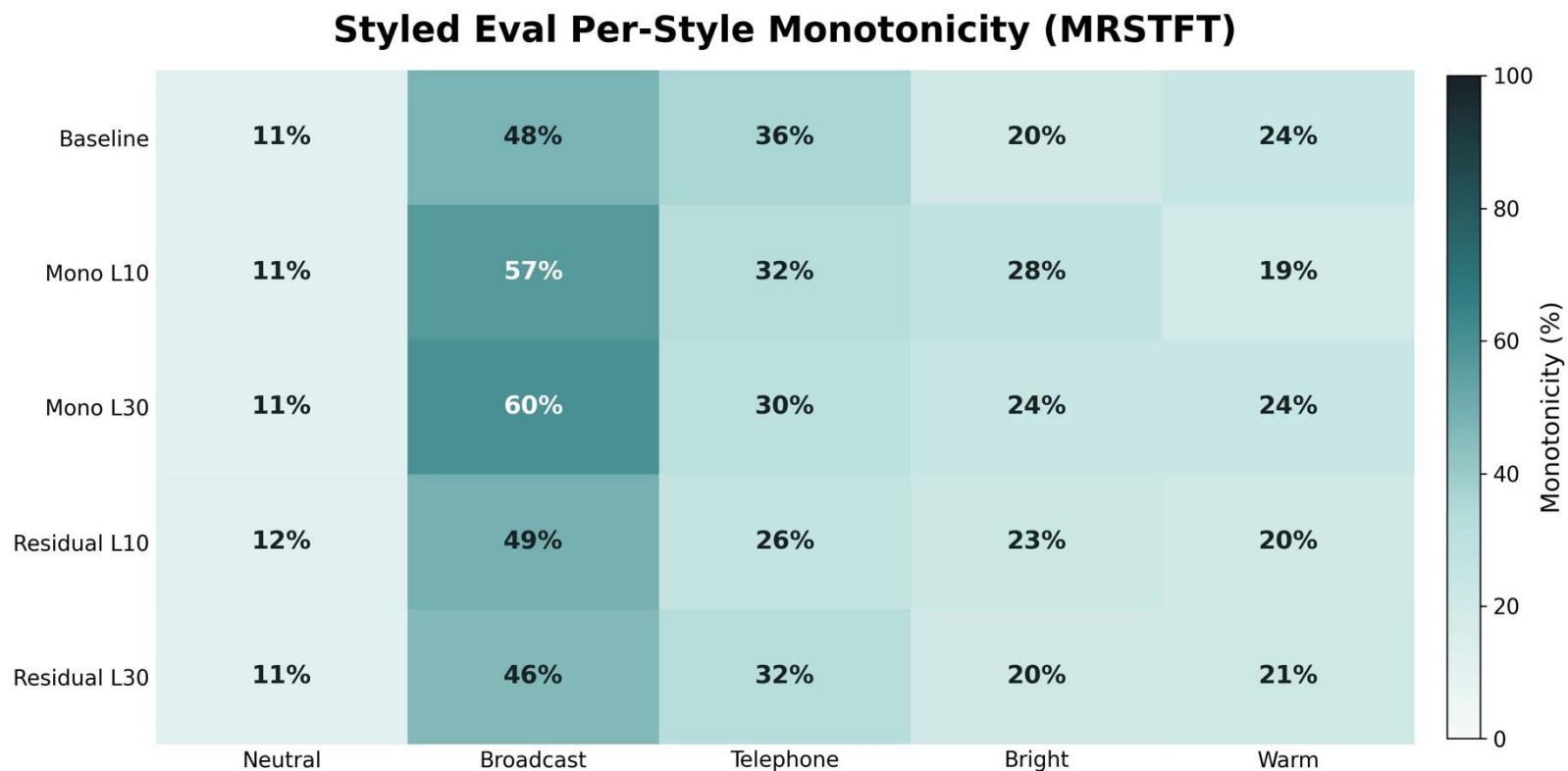
Random eval — 전체 모델 비교

Random Eval: Monotonicity vs. Violation Trade-off



Mono L10 gives the cleanest main improvement; residual variants did not clearly dominate.

Styled eval — 스타일별 단조성



Neutral/Warm are weak cases, while Broadcast/Telephone are comparatively easier in this setup.