US 20210233547A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2021/0233547 A1

LEE et al. (43) Pub. Date: Jul. 29, 2021

(54) METHOD AND APPARATUS FOR PROCESSING AUDIO SIGNAL

(71) Applicants: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **The Trustees of Indiana University**, Indianapolis, IN (US)

(72) Inventors: **Mi Suk LEE**, Daejeon (KR); **Seung Kwon BEACK**, Daejeon (KR); **Jongmo SUNG**, Daejeon (KR); **Tae Jin LEE**, Daejeon (KR); **Jin Soo CHOI**, Daejeon (KR); **Minje KIM**, Bloomington, IN (US); **Kai ZHEN**, Bloomington, IN (US)

(73) Assignees: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **The Trustees of Indiana University**, Indianapolis, IN (US)

(21) Appl. No.: **17/156,006**

(22) Filed: **Jan. 22, 2021**

Related U.S. Application Data

(60) Provisional application No. 62/966,917, filed on Jan. 28, 2020.

(30) **Foreign Application Priority Data**

May 12, 2020 (KR) 10-2020-0056492

Publication Classification

(51) **Int. Cl.**

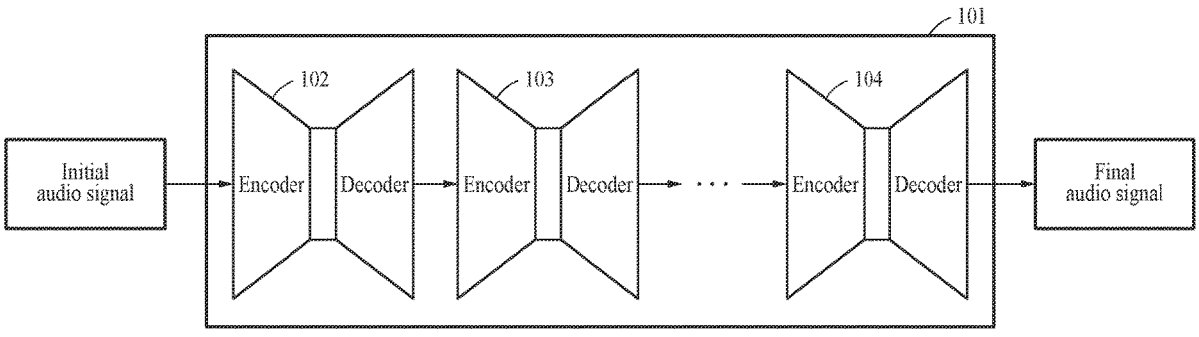
G10L 19/038	(2006.01)
G10L 25/18	(2006.01)
G10L 25/30	(2006.01)
G10L 25/21	(2006.01)
G10L 19/028	(2006.01)
G06N 3/08	(2006.01)

(52) **U.S. Cl.**

CPC ***G10L 19/038*** (2013.01); ***G10L 25/18*** (2013.01); ***G06N 3/08*** (2013.01); ***G10L 25/21*** (2013.01); ***G10L 19/028*** (2013.01); ***G10L 25/30*** (2013.01)

(57) **ABSTRACT**

A method and apparatus for processing an audio signal are disclosed. According to an example embodiment, a method of processing an audio signal may include acquiring a final audio signal for an initial audio signal using a plurality of neural network models generating output audio signals by encoding and decoding input audio signals, calculating a difference between the initial audio signal and the final audio signal in a time domain, converting the initial audio signal and the final audio signal into Mel-spectra, calculating a difference between the Mel-spectra of the initial audio signal and the final audio signal in a frequency domain, training the plurality of neural network models based on results calculated in the time domain and the frequency domain, and generating a new final audio signal distinguished from the final audio signal from the initial audio signal using the trained neural network models.



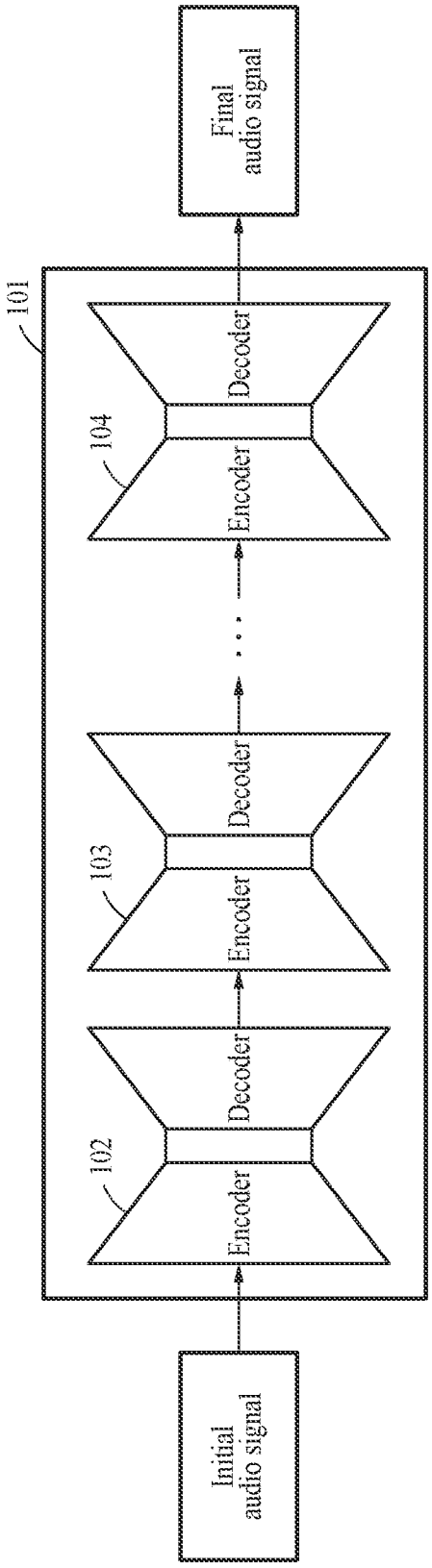

Initial audio signal
101
102
Encoder
Decoder
103
Encoder
Decoder
104
Encoder
Decoder
Final audio signal

FIG. 1

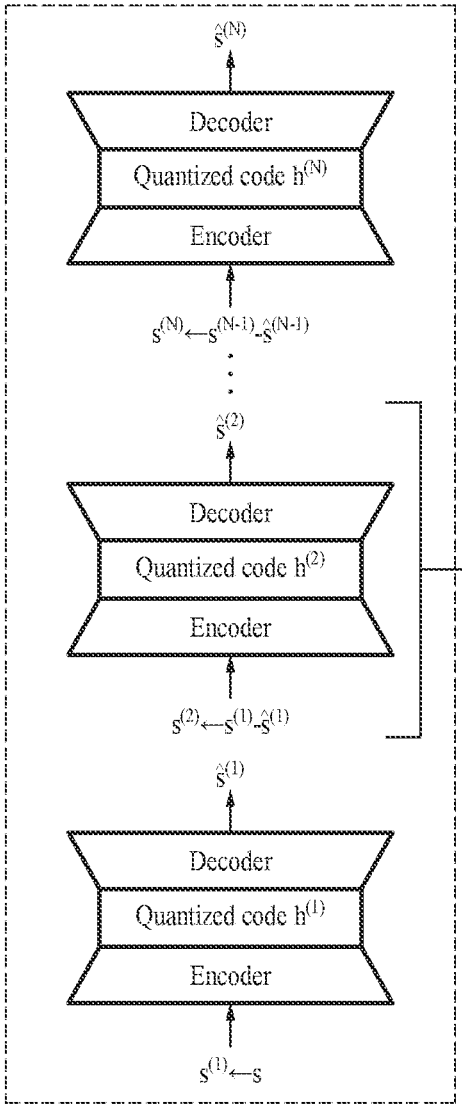

$\hat{s}^{(N)}$
Decoder
Quantized code $h^{(N)}$
Encoder
$s^{(N)} \leftarrow s^{(N-1)} - \hat{s}^{(N-1)}$
$\hat{s}^{(2)}$
Decoder
Quantized code $h^{(2)}$
Encoder
$s^{(2)} \leftarrow s^{(1)} - \hat{s}^{(1)}$
$\hat{s}^{(1)}$
Decoder
Quantized code $h^{(1)}$
Encoder
$s^{(1)} \leftarrow s$

FIG. 2A

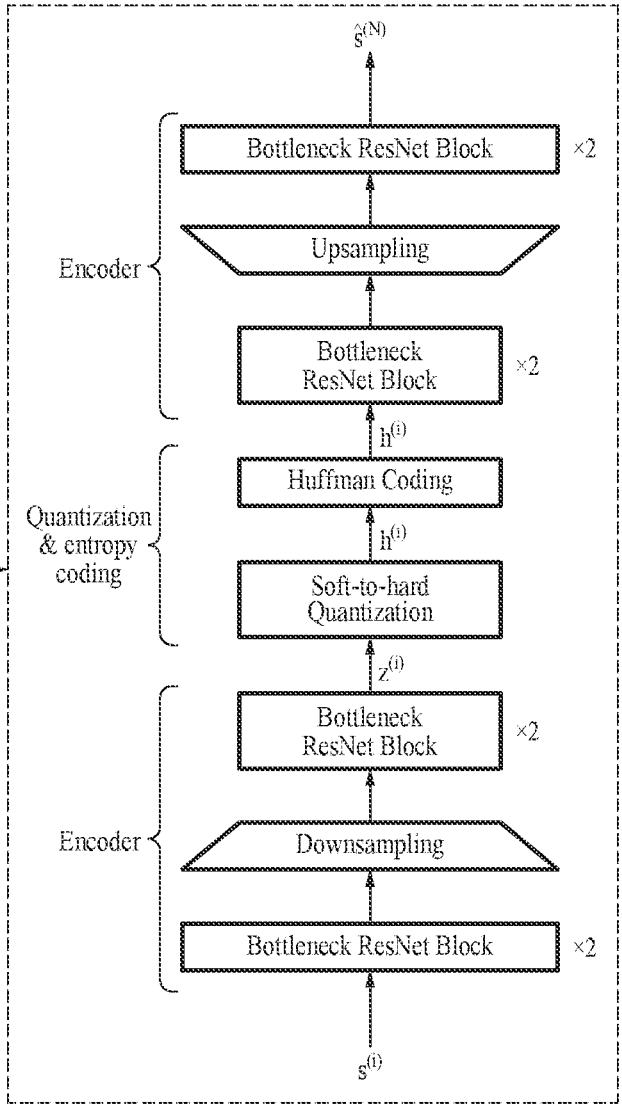

$\hat{s}^{(N)}$
Encoder
Bottleneck ResNet Block
×2
Upsampling
Bottleneck ResNet Block
×2
$h^{(i)}$
Quantization & entropy coding
Huffman Coding
$h^{(i)}$
Soft-to-hard Quantization
$z^{(i)}$
Encoder
Bottleneck ResNet Block
×2
Downsampling
Bottleneck ResNet Block
×2
$s^{(i)}$

FIG. 2B