US011416742B2

(12) **United States Patent**
Sung et al.

(10) **Patent No.:** **US 11,416,742 B2**
(45) **Date of Patent:** **Aug. 16, 2022**

(54) **AUDIO SIGNAL ENCODING METHOD AND APPARATUS AND AUDIO SIGNAL DECODING METHOD AND APPARATUS USING PSYCHOACOUSTIC-BASED WEIGHTED ERROR FUNCTION**

(71) Applicants: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **THE TRUSTEES OF INDIANA UNIVERSITY**, Bloomington, IN (US)

(72) Inventors: **Jongmo Sung**, Daejeon (KR); **Minje Kim**, Bloomington, IN (US); **Aswin Sivaraman**, Bloomington, IN (US); **Kai Zhen**, Bloomington, IN (US)

(73) Assignees: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **THE TRUSTEES OF INDIANA UNIVERSITY**, Bloomington, IN (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1016 days.

(21) Appl. No.: **16/122,708**

(22) Filed: **Sep. 5, 2018**

(65) **Prior Publication Data**

US 2019/0164052 A1 May 30, 2019

Related U.S. Application Data

(60) Provisional application No. 62/590,488, filed on Nov. 24, 2017.

(30) **Foreign Application Priority Data**

Dec. 15, 2017 (KR) 10-2017-0173405

(51) **Int. Cl.**
G06N 3/08 (2006.01)
G10L 19/008 (2013.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06N 3/08*** (2013.01); ***G10L 19/008*** (2013.01); ***G10L 19/032*** (2013.01); *G10L 25/30* (2013.01); *G10L 25/69* (2013.01)

(58) **Field of Classification Search**
CPC G06N 3/08
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,035,177 A * 3/2000 Moses H04H 20/31 380/252
7,921,007 B2 * 4/2011 Van de Par G10L 21/038 704/229
(Continued)

OTHER PUBLICATIONS

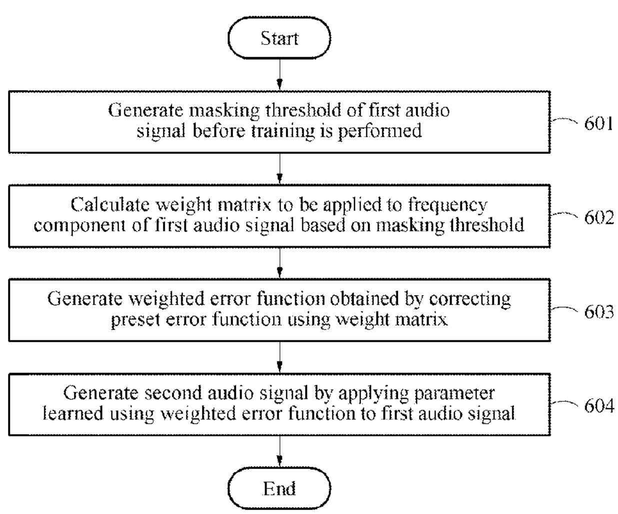
Liu et al., A Perceptually-Weighted Deep Neural Network for Monaural Speech Enhancement in Various Background Noise Conditions, 2017 25th European Signal Processing Conference (EUSIPCO), Aug. 28, 2017, pp. 1-5, IEEE, Kos, Greece.

Primary Examiner — Olisa Anwah
(74) *Attorney, Agent, or Firm* — William Park & Associates Ltd.

(57) **ABSTRACT**

Provided is a training method of a neural network that is applied to an audio signal encoding method using an audio signal encoding apparatus, the training method including generating a masking threshold of a first audio signal before training is performed, calculating a weight matrix to be applied to a frequency component of the first audio signal based on the masking threshold, generating a weighted error function obtained by correcting a preset error function using the weight matrix, and generating a second audio signal by applying a parameter learned using the weighted error function to the first audio signal.

7 Claims, 11 Drawing Sheets



(51) **Int. Cl.**

G10L 19/032	(2013.01)
G10L 25/30	(2013.01)
G10L 25/69	(2013.01)

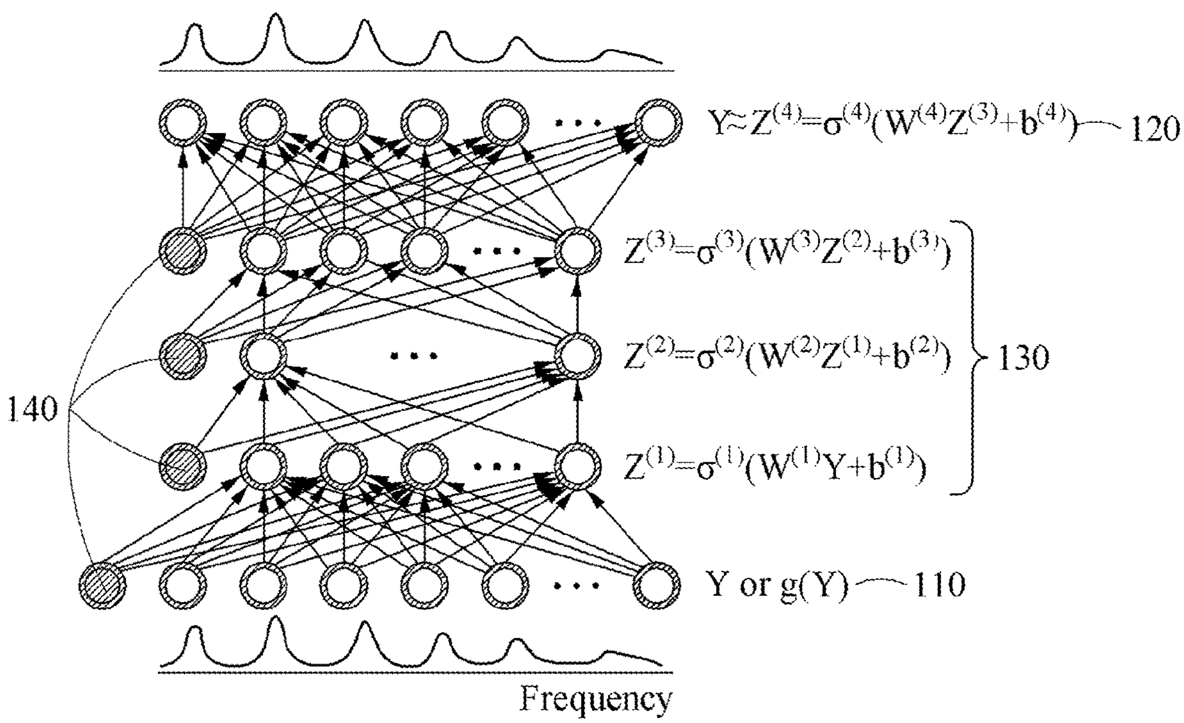
(56) **References Cited**

U.S. PATENT DOCUMENTS

8,588,427 B2 * 11/2013 Uhle H04R 5/04 381/17
10,043,527 B1 * 8/2018 Gurijala G10L 19/018
2003/0115042 A1 * 6/2003 Chen G10L 25/69 704/E19.002
2004/0002859 A1 * 1/2004 Liu G10L 19/035 704/229
2008/0082328 A1 4/2008 Lee
2009/0089049 A1 * 4/2009 Moon G10L 19/035 704/E19.001
2010/0076769 A1 3/2010 Yu
2011/0026724 A1 * 2/2011 Doclo G10K 11/17854 381/71.8
2011/0119039 A1 * 5/2011 Grancharov G10L 25/69 703/2
2012/0020415 A1 * 1/2012 Yang H04N 17/004 375/E7.279
2012/0023051 A1 * 1/2012 Pishehvar G06N 3/049 706/21
2012/0093341 A1 4/2012 Kim et al.
2015/0003625 A1 1/2015 Uhle et al.
2016/0078880 A1 3/2016 Avendano et al.
2016/0247502 A1 8/2016 Park et al.
2020/0234720 A1 * 7/2020 Sung G10L 19/0017
2020/0329330 A1 * 10/2020 Mitchell H04R 29/007

* cited by examiner

FIG. 1



$Y \approx Z^{(4)} = \sigma^{(4)}(W^{(4)}Z^{(3)} + b^{(4)})$
120
$Z^{(3)} = \sigma^{(3)}(W^{(3)}Z^{(2)} + b^{(3)})$
$Z^{(2)} = \sigma^{(2)}(W^{(2)}Z^{(1)} + b^{(2)})$
130
$Z^{(1)} = \sigma^{(1)}(W^{(1)}Y + b^{(1)})$
140
Y or g(Y)
110
Frequency

US011276413B2

(12) **United States Patent**
Lee et al.

(10) **Patent No.: US 11,276,413 B2**
(45) **Date of Patent: Mar. 15, 2022**

(54) **AUDIO SIGNAL ENCODING METHOD AND AUDIO SIGNAL DECODING METHOD, AND ENCODER AND DECODER PERFORMING THE SAME**

(71) Applicants: **Electronics and Telecommunications Research Institute**, Daejeon (KR); **THE TRUSTEES OF INDIANA UNIVERSITY**, Indianapolis, IN (US)

(72) Inventors: **Mi Suk Lee**, Daejeon (KR); **Jongmo Sung**, 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)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 41 days.

(21) Appl. No.: **16/543,095**

(22) Filed: **Aug. 16, 2019**

(65) **Prior Publication Data**

US 2020/0135220 A1 Apr. 30, 2020

Related U.S. Application Data

(60) Provisional application No. 62/751,105, filed on Oct. 26, 2018.

(30) **Foreign Application Priority Data**

Feb. 26, 2019 (KR) 10-2019-0022612

(51) **Int. Cl.**
G10L 15/00 (2013.01)
G10L 19/16 (2013.01)

(52) **U.S. Cl.**
CPC *G10L 19/167* (2013.01)

(58) **Field of Classification Search**
CPC G10L 19/02; G10L 19/0018; G10L 15/00
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

8,484,022	B1 *	7/2013	Vanhoucke	G10L 19/0018 704/240
8,959,015	B2	2/2015	Lee et al.	
9,830,920	B2	11/2017	Rose et al.	
10,397,725	B1 *	8/2019	Bharitkar	H04R 5/02
10,706,856	B1 *	7/2020	Korjani	G10L 19/02

(Continued)

OTHER PUBLICATIONS

Deng et al., "Deep Convex Net: A Scalable Architecture for Speech Pattern Classification"; Aug. 28-31, 2011, Microsoft Reasearch, pp. 2285-2288 (Year: 2011).*

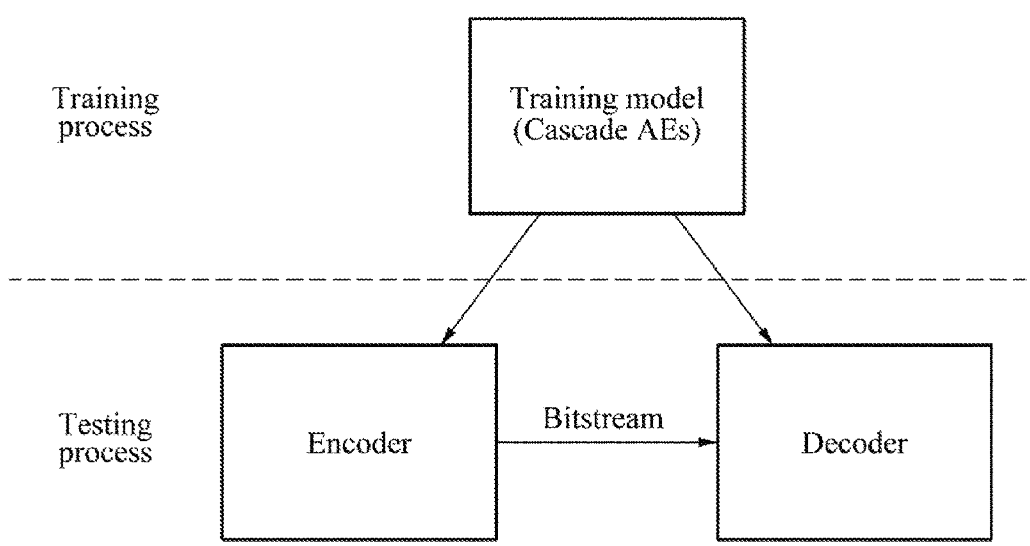
(Continued)

Primary Examiner — Shreyans A Patel
(74) *Attorney, Agent, or Firm* — William Park & Associates Ltd.

(57) **ABSTRACT**

Disclosed are an audio signal encoding method and audio signal decoding method, and an encoder and decoder performing the same. The audio signal encoding method includes applying an audio signal to a training model including N autoencoders provided in a cascade structure, encoding an output result derived through the training model, and generating a bitstream with respect to the audio signal based on the encoded output result.

12 Claims, 6 Drawing Sheets



(56) **References Cited**

U.S. PATENT DOCUMENTS

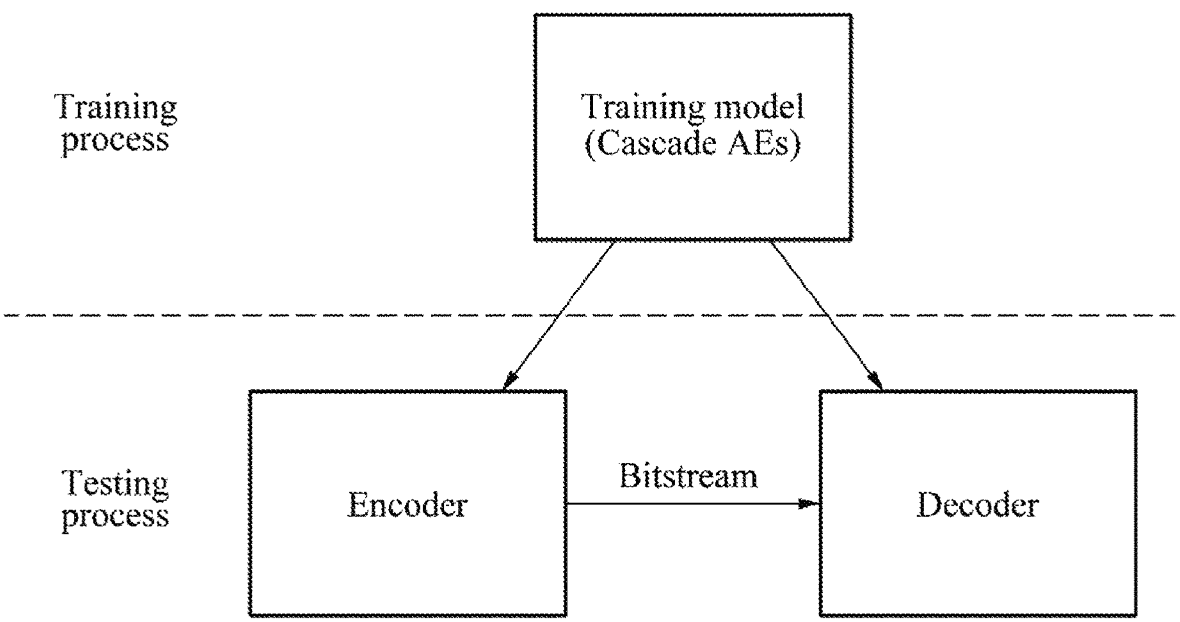
2012/0275609	A1	11/2012	Beack et al.	
2016/0189730	A1 *	6/2016	Du	G10L 21/0272 704/233
2017/0076224	A1 *	3/2017	Munawar	G06N 3/084
2019/0164052	A1 *	5/2019	Sung	G06N 3/088
2019/0198036	A1 *	6/2019	Osako	G06N 3/0454
2020/0090029	A1 *	3/2020	Suzuki	G06N 3/04
2020/0168208	A1 *	5/2020	Mitra	G10L 15/075

OTHER PUBLICATIONS

Kaiming He et al., "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 27-30, 2016, 770-778, IEEE, Las Vegas, NV, USA.

* cited by examiner

FIG. 1


Training process
Training model (Cascade AEs)
Testing process
Encoder
Bitstream
Decoder

US 20210350796A1

(19) United States
(12) Patent Application Publication
(10) Pub. No.: US 2021/0350796 A1
KIM et al.
(43) Pub. Date: Nov. 11, 2021

(54) APPARATUS AND METHOD FOR SPEECH PROCESSING USING A DENSELY CONNECTED HYBRID NEURAL NETWORK

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

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

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

(21) Appl. No.: **17/308,800**

(22) Filed: **May 5, 2021**

(30) **Foreign Application Priority Data**

May 7, 2020 (KR) 10-2020-0054733

Publication Classification

(51) **Int. Cl.**
G10L 15/16 (2006.01)
G06N 3/04 (2006.01)
G06F 17/15 (2006.01)

(52) **U.S. Cl.**
CPC ***G10L 15/16*** (2013.01); ***G06F 17/15*** (2013.01); ***G06N 3/0454*** (2013.01)

(57) **ABSTRACT**

Disclosed is a speech processing apparatus and method using a densely connected hybrid neural network. The speech processing method includes inputting a time domain sample of N*1 dimension for an input speech into a densely connected hybrid network; passing the time domain sample through a plurality of dense blocks in a densely connected hybrid network; reshaping the time domain samples into M subframes by passing the time domain samples through the plurality of dense blocks, inputting the M subframes into gated recurrent unit (GRU) components of N/M-dimension; outputting clean speech from which noise is removed from the input speech by passing the M subframes through GRU components.

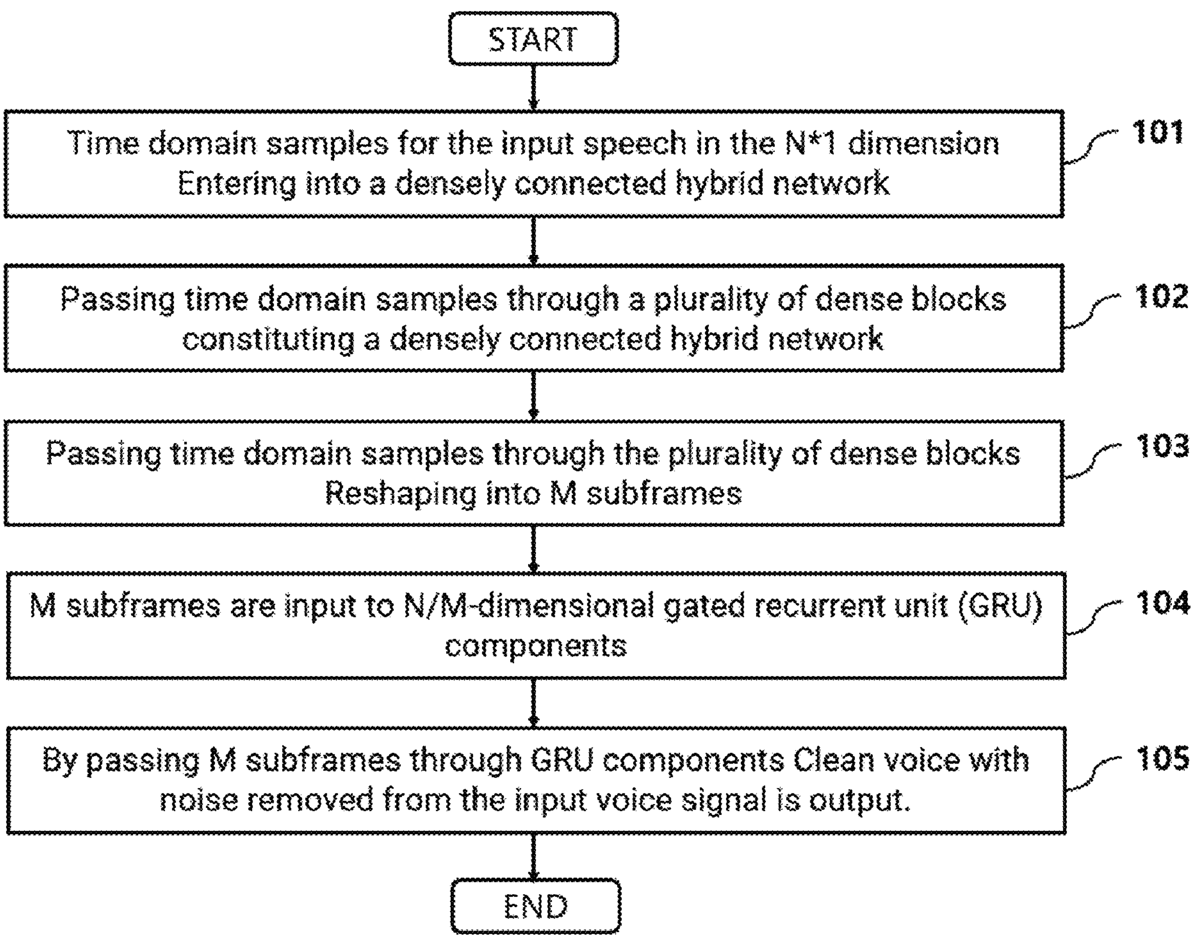


FIG. 1

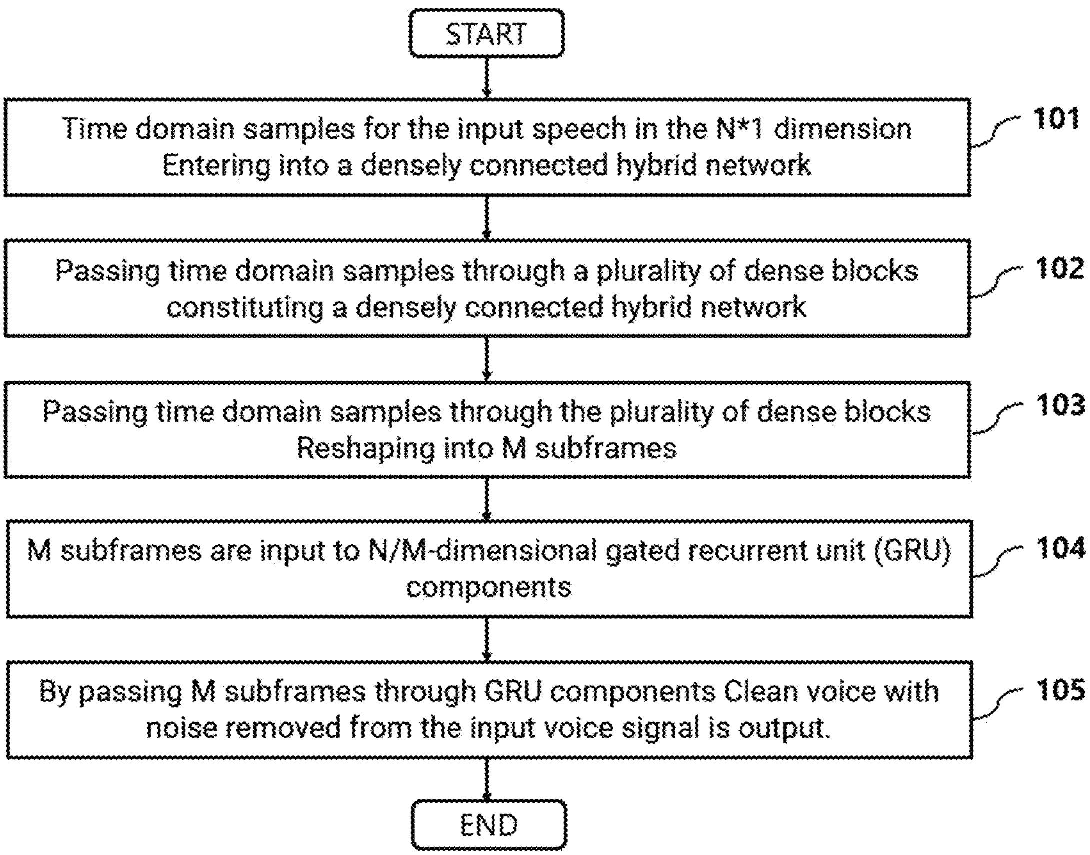
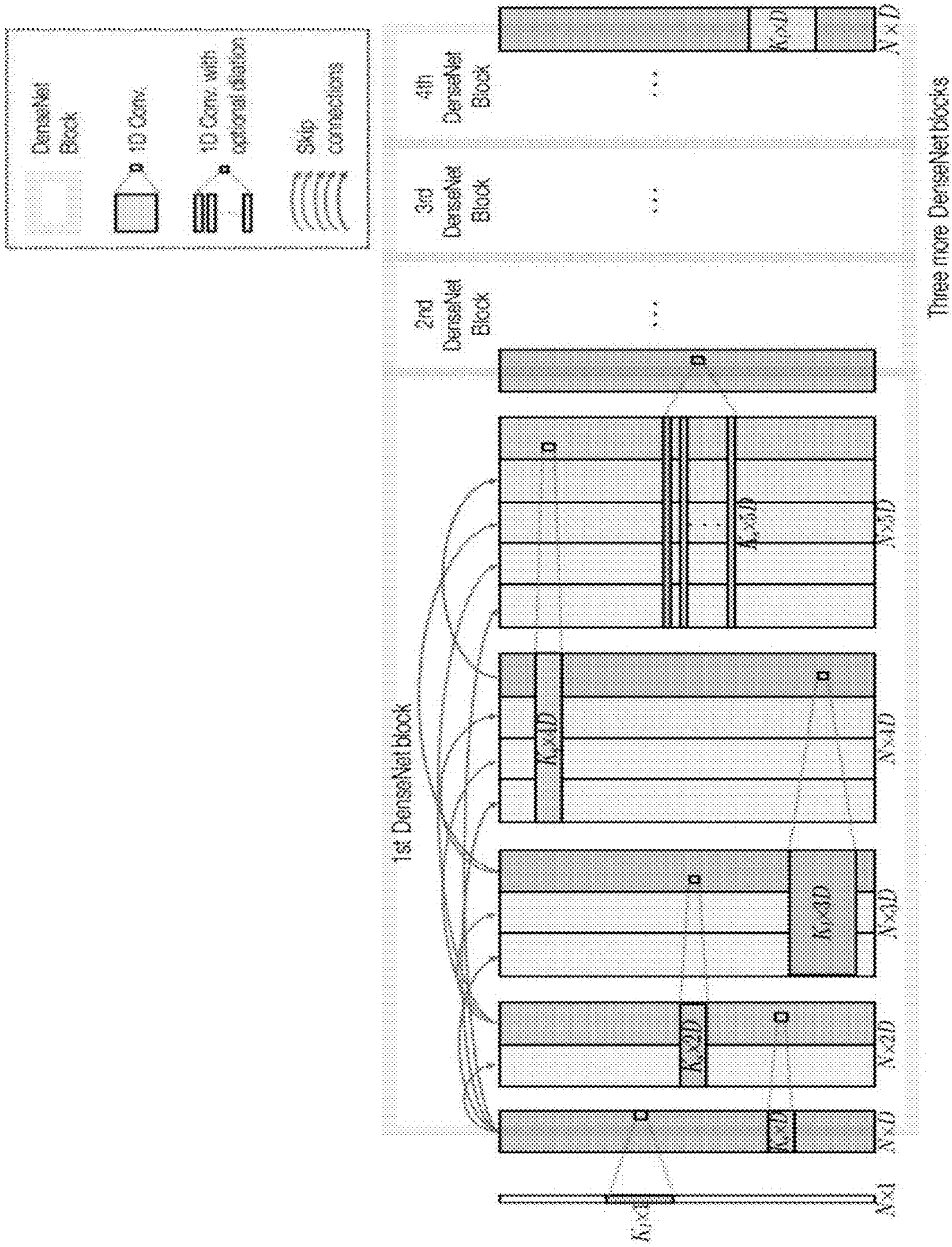

START
Time domain samples for the input speech in the N*1 dimension Entering into a densely connected hybrid network
101
Passing time domain samples through a plurality of dense blocks constituting a densely connected hybrid network
102
Passing time domain samples through the plurality of dense blocks Reshaping into M subframes
103
M subframes are input to N/M-dimensional gated recurrent unit (GRU) components
104
By passing M subframes through GRU components Clean voice with noise removed from the input voice signal is output.
105
END

FIG. 2


DenseNet Block
1D Conv.
1D Conv. with optional dilation
Skip connections
1st DenseNet block
2nd DenseNet Block
3rd DenseNet Block
4th DenseNet Block
Three more DenseNet blocks (the same architecture)

US 20210233547A1

(19) **United States**

(12) **Patent Application Publication**

LEE et al.

(10) **Pub. No.: US 2021/0233547 A1**

(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.**

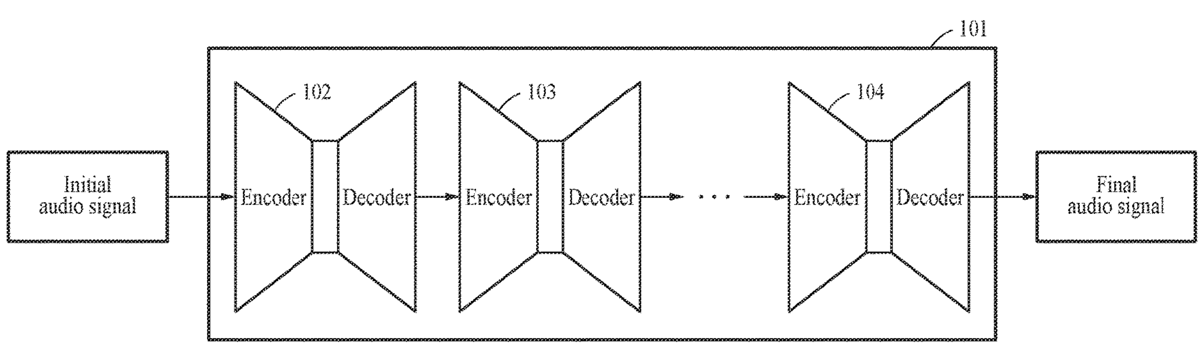
G10L 19/038	(2006.01)
G10L 25/18	(2006.01)
G10L 25/30	(2006.01)
G10L 25/21	(2006.01)
G10L 19/028	(2006.01)
G06N 3/08	(2006.01)

(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.



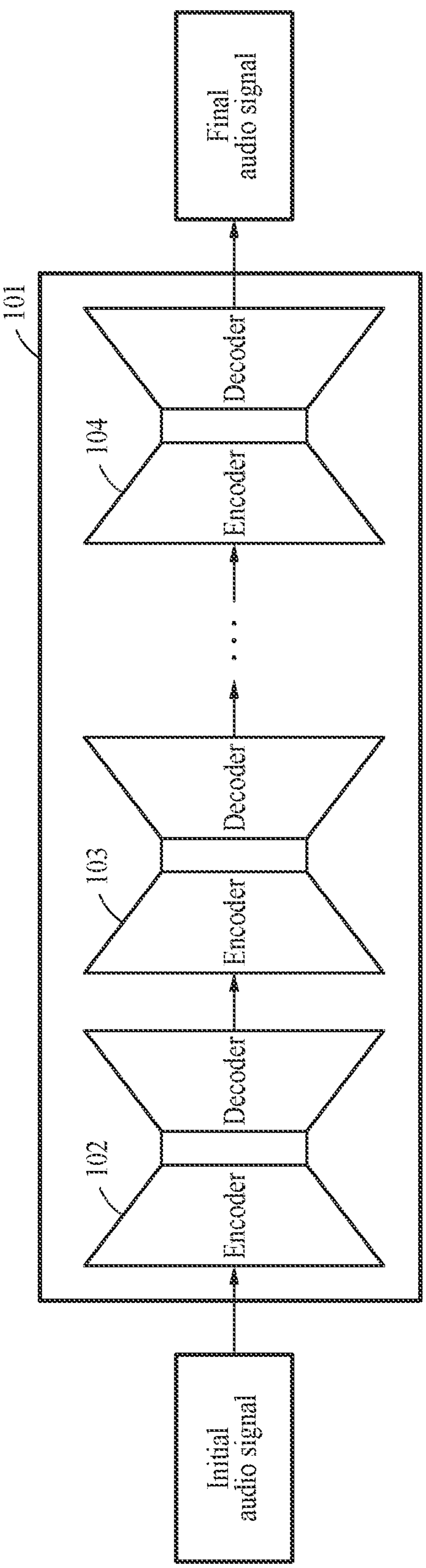
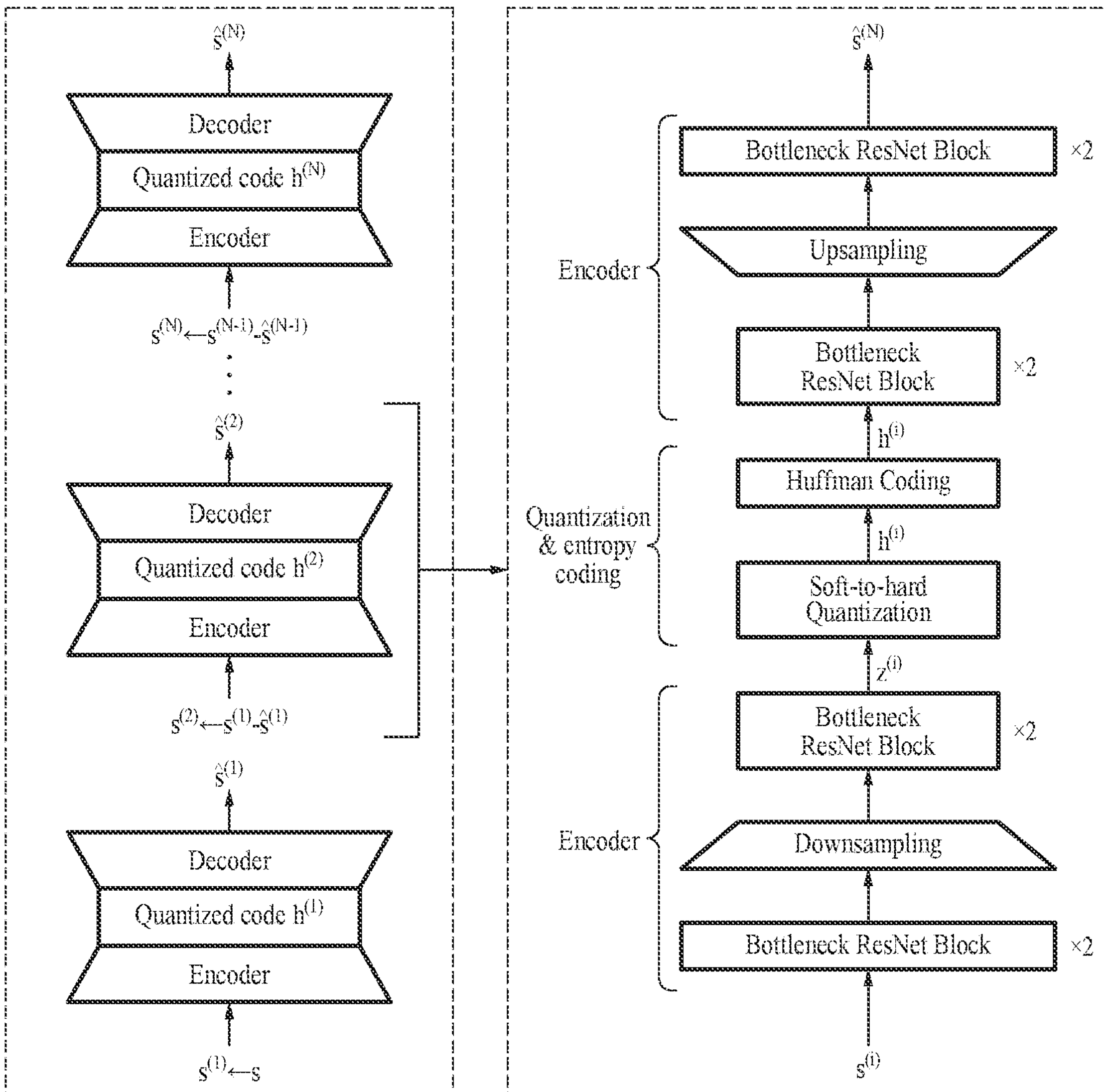

Initial audio signal
101
102
Encoder
Decoder
103
Encoder
Decoder
104
Encoder
Decoder
Final audio signal

FIG. 1



ŝ(N)
Decoder
Quantized code h(N)
Encoder
s(N)←s(N-1)-ŝ(N-1)
ŝ(2)
Decoder
Quantized code h(2)
Encoder
s(2)←s(1)-ŝ(1)
ŝ(1)
Decoder
Quantized code h(1)
Encoder
s(1)←s
ŝ(N)
Bottleneck ResNet Block ×2
Upsampling
Encoder
Bottleneck ResNet Block ×2
h(i)
Huffman Coding
Quantization & entropy coding
h(i)
Soft-to-hard Quantization
z(i)
Bottleneck ResNet Block ×2
Encoder
Downsampling
Bottleneck ResNet Block ×2
s(i)

FIG. 2A

FIG. 2B

US 20210142812A1

(19) **United States**

(12) **Patent Application Publication**

KIM et al.

(10) **Pub. No.: US 2021/0142812 A1**

(43) **Pub. Date: May 13, 2021**

(54) **RESIDUAL CODING METHOD OF LINEAR PREDICTION CODING COEFFICIENT BASED ON COLLABORATIVE QUANTIZATION, AND COMPUTING DEVICE FOR PERFORMING THE METHOD**

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

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

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

(21) Appl. No.: **17/098,090**

(22) Filed: **Nov. 13, 2020**

Related U.S. Application Data

(60) Provisional application No. 62/934,868, filed on Nov. 13, 2019.

(30) **Foreign Application Priority Data**

Nov. 13, 2020 (KR) 10-2020-0152071

Publication Classification

(51) **Int. Cl.**

G10L 19/08	(2006.01)
G10L 19/032	(2006.01)
G10L 19/26	(2006.01)
G10L 21/0208	(2006.01)
G10L 25/30	(2006.01)
G10L 13/02	(2006.01)
G06N 3/08	(2006.01)

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

CPC ***G10L 19/08*** (2013.01); ***G10L 19/032*** (2013.01); ***G10L 19/265*** (2013.01); ***G06N 3/08*** (2013.01); ***G10L 25/30*** (2013.01); ***G10L 13/02*** (2013.01); ***G10L 21/0208*** (2013.01)

(57) **ABSTRACT**

Disclosed are a method for coding a residual signal of LPC coefficients based on collaborative quantization and a computing device for performing the method. The residual signal coding method includes: generating encoded LPC coefficients and LPC residual signals by performing LPC analysis and quantization on an input speech; Determining a predicted LPC residual signal by applying the LPC residual signal to cross module residual learning; Performing LPC synthesis using the coded LPC coefficients and the predicted LPC residual signal; It may include the step of determining an output speech that is a synthesized output according to a result of performing the LPC synthesis.

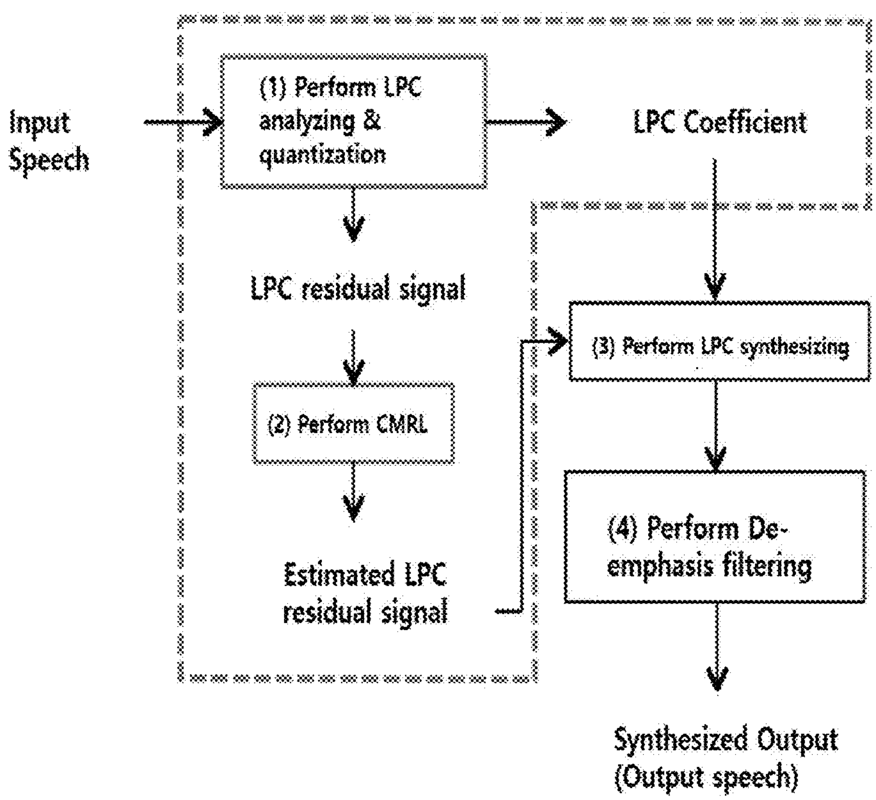


FIG. 1

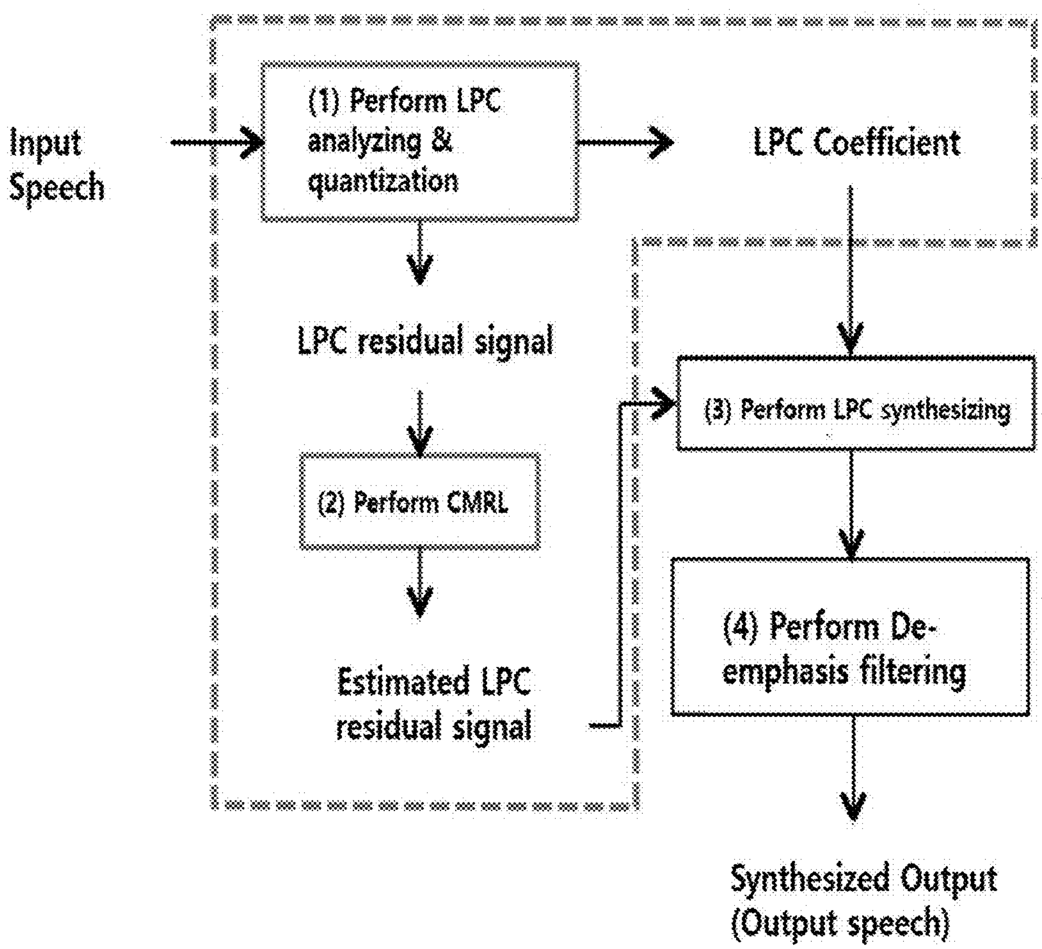
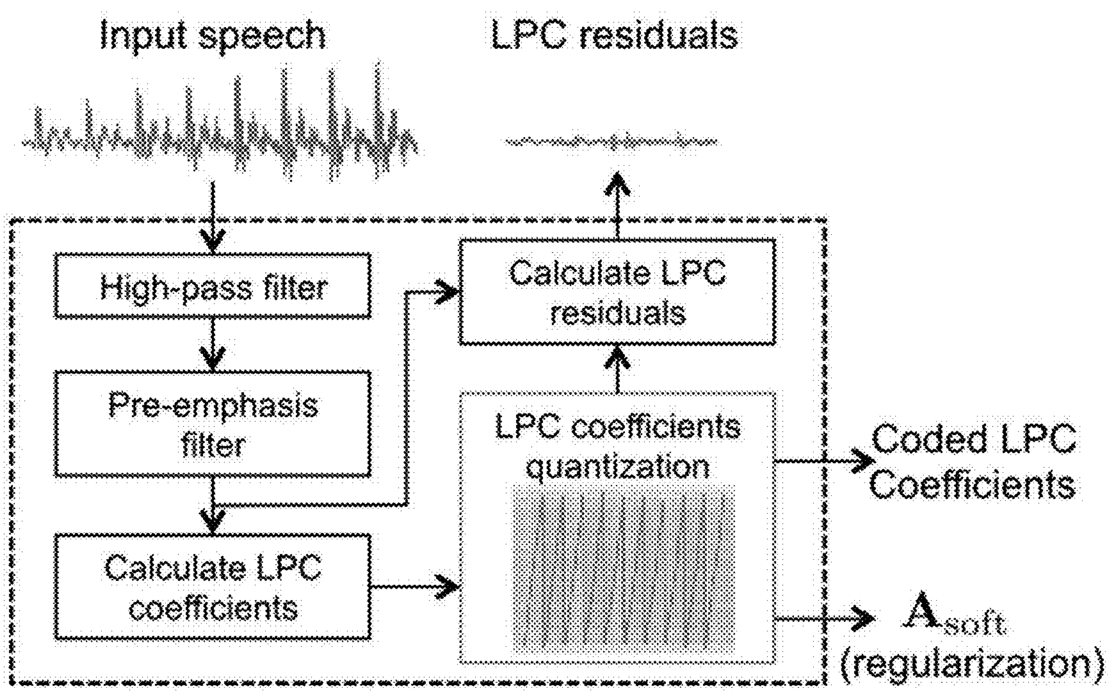

Input
Speech
(1) Perform LPC analyzing & quantization
LPC Coefficient
LPC residual signal
(2) Perform CMRL
Estimated LPC residual signal
(3) Perform LPC synthesizing
(4) Perform De-emphasis filtering
Synthesized Output
(Output speech)

FIG. 2


Input speech
LPC residuals
High-pass filter
Pre-emphasis filter
Calculate LPC coefficients
Calculate LPC residuals
LPC coefficients quantization
Coded LPC Coefficients
$\mathbf{A}_{\mathrm{soft}}$
(regularization)