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.

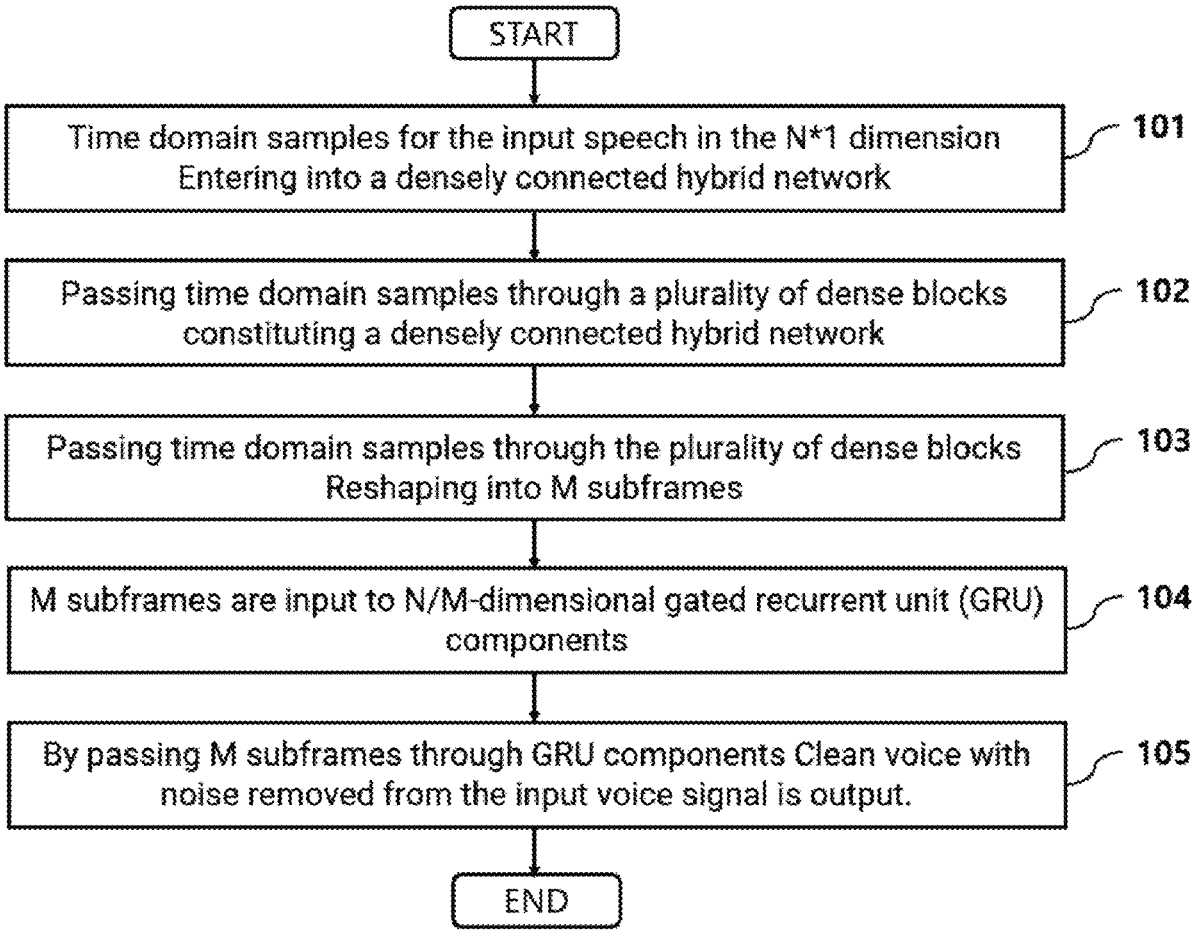


FIG. 1

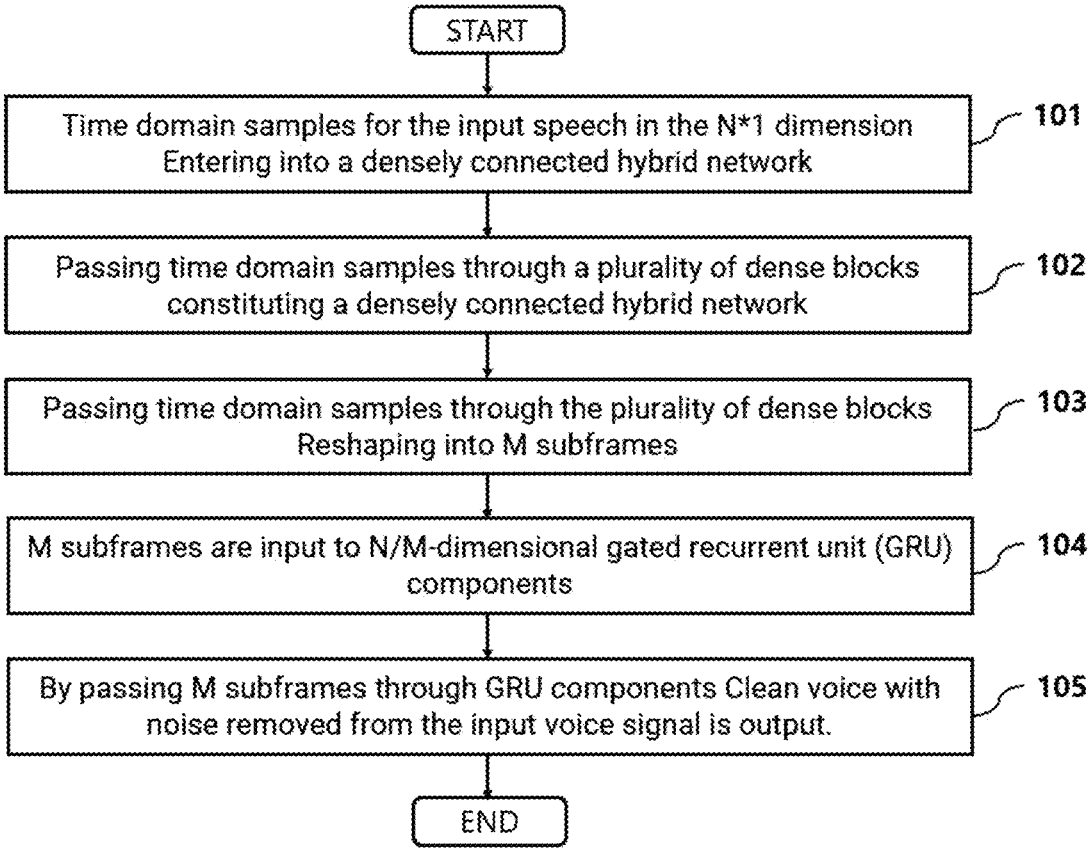
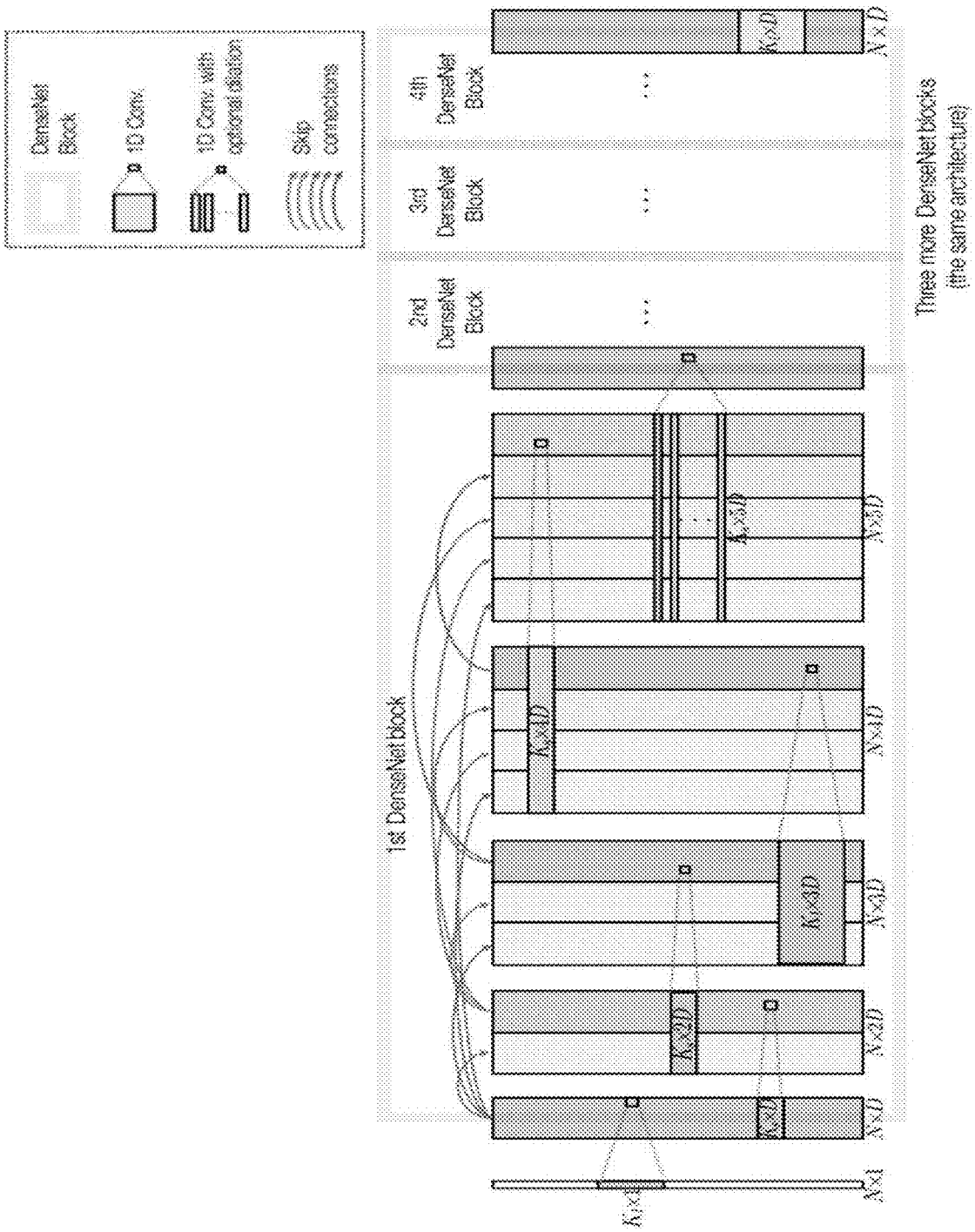

START
Time domain samples for the input speech in the N*1 dimension Entering into a densely connected hybrid network
101
Passing time domain samples through a plurality of dense blocks constituting a densely connected hybrid network
102
Passing time domain samples through the plurality of dense blocks 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)