How to draft the perfect invention disclosure report for an AI invention

BARDEHLE PAGENBERG

Bastian Best

Dipl.-Inf., German & European Patent Attorney, Partner

January 28, 2019

WHY IS THIS IMPORTANT?

#1: CONFLICTING INTERESTS

OWNER WANTS BROAD PROTECTION

PUBLIC WANTS TO KNOW HOW IT WORKS

Photo by rawpixel on Unsplash

#2: THE RACE TO THE PATENT OFFICE

nton

Cal

HE WILL GET THE PATENT

Comparison of the second contract of the seco

#3: POINT OF NO RETURN

NO CHANGES ALLOWED AFTER FILING

Photo by <u>Sydny Riters</u> on <u>Unsplas</u>

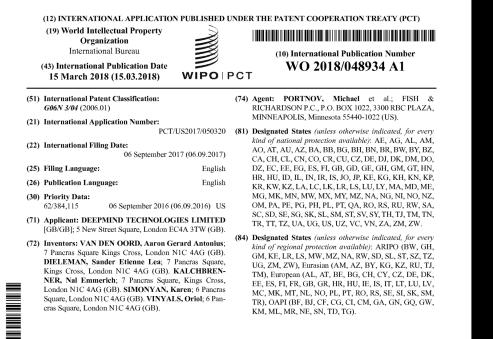
8H00

34520 LE CAYLAR

A0P8L3

PATENTS A HAVE TO BE FILED QUICKLY CANNOT BE EXPANDED LATER

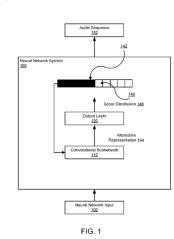
THE ANATOMY OF A PATENT APPLICATION



(54) Title: GENERATING AUDIO USING NEURAL NETWORKS

A1

WO 2018/048934



(57) Abstract: Methods, systems, and apparatus, including computer programs encoded on computer storage media, for generating an output sequence of audio data that comprises a respective audio sample at each of a plurality of time steps. One of the methods includes, for each of the time steps: providing a current sequence of audio data as input to a convolutional subnetwork, wherein the current sequence comprises the respective audio sample at each time step that precedes the time step in the output sequence, and wherein the convolutional subnetwork is configured to process the current sequence of audio data to generate an alternative representation for the time step; and providing the alternative representation for the time step as input to an output layer, wherein the output layer is configured to: process the alternative representation to generate an output that defines a score distribution over a plurality of possible audio samples for the time step.

KM, ML, MR, NE, SN, TD, TG).

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization International Bureau

(43) International Publication Date 15 March 2018 (15.03.2018)



(10) International Publication Number WO 2018/048934 A1

- (51) International Patent Classification: G06N 3/04 (2006.01)
- (21) International Application Number:

PCT/US2017/050320

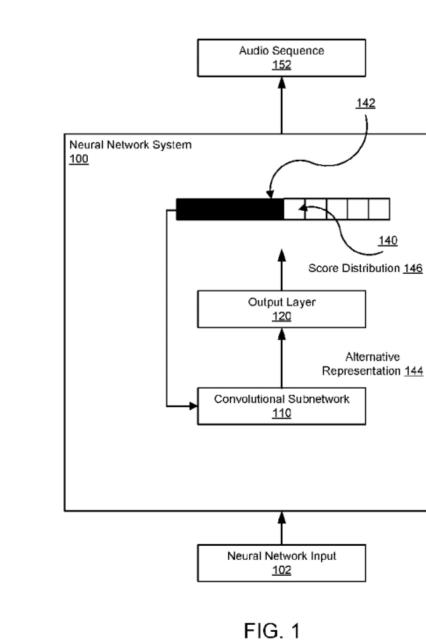
(22) International Filing Date:

06 September 2017 (06.09.2017)

- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 62/384,115 06 September 2016 (06.09.2016) US
- (71) Applicant: DEEPMIND TECHNOLOGIES LIMITED [GB/GB]; 5 New Street Square, London EC4A 3TW (GB).
- (72) Inventors: VAN DEN OORD, Aaron Gerard Antonius;
 7 Pancras Square Kings Cross, London N1C 4AG (GB).
 DIELEMAN, Sander Etienne Lea; 7 Pancras Square,
 Kings Cross, London N1C 4AG (GB). KALCHBRENNER, Nal Emmerich; 7 Pancras Square, Kings Cross,
 London N1C 4AG (GP) SIMONYAN Karan; 6 Paneros

- (74) Agent: PORTNOV, Michael et al.; FISH & RICHARDSON P.C., P.O. BOX 1022, 3300 RBC PLAZA, MINNEAPOLIS, Minnesota 55440-1022 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

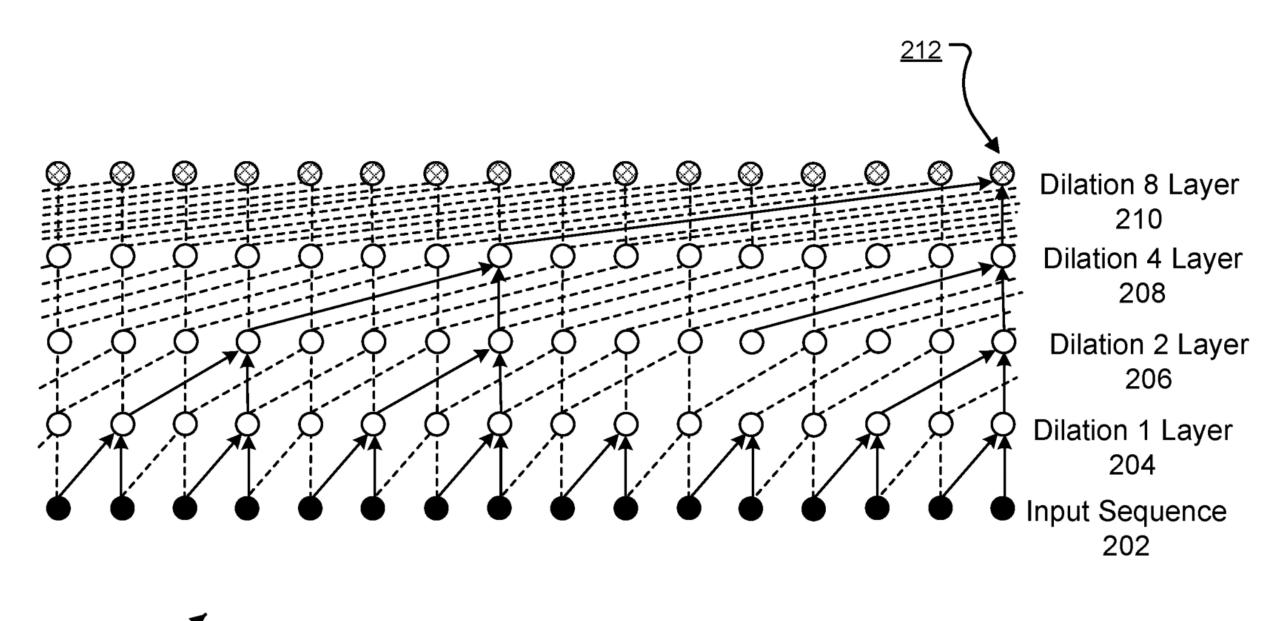
(54) Title: GENERATING AUDIO USING NEURAL NETWORKS

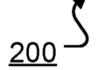


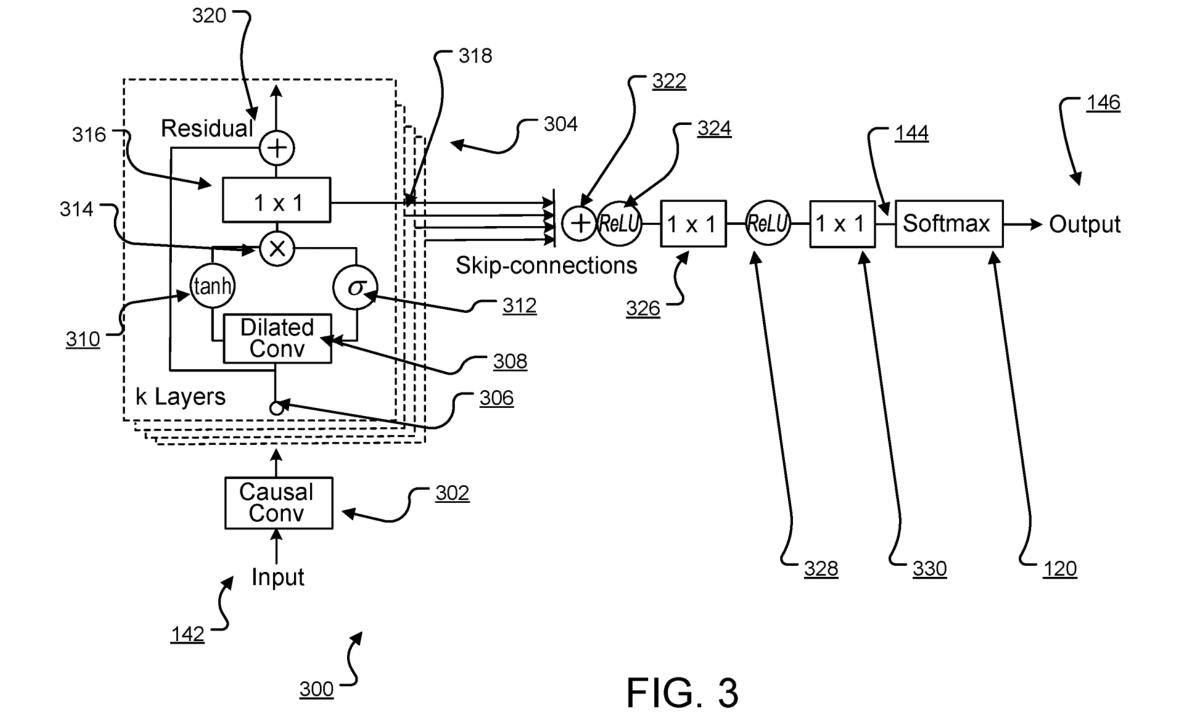
/048934

(57) Abstract: Methods, systems, and apparatus, including computer programs encoded on computer storage media, for generating an output sequence of audio data that comprises a respective audio sample at each of a plurality of time steps. One of the methods includes, for each of the time steps: providing a current sequence of audio data as input to a convolutional subnetwork, wherein the current sequence comprises the respective audio sample at each time step that precedes the time step in the output sequence, and wherein the convolutional subnetwork is configured to process the current sequence of audio data to generate an alternative representation for the time step; and providing the alternative representation for the time step as input to an output layer, wherein the output layer is configured to: process the alternative representation to generate an output that defines a score distribution over a plurality of possible audio samples for the time step.

WO 2018/048934 PCT/US2017/050320	WO 2018/042934 PCT/US20/2/05020	WO 2018/148934 PCT/US2417/054320	WO 200848034 PCT/US3473/64230	WO 2018/04/2014 PCT/US36170/6428	W0 2018/04/204 PCT/US/04/20
GENERATING AUDIO USING NEURAL NETWORKS	output that defines a score distribution over a plurality of possible audio samples for the time sters.	a conditioning input. The conditioning input may be global (substantially time- independent) and or local (time-dependent). The conditioning input may comprise, for	second, providing a greater level of granularity than other neural network-based audio generation systems. The neural network system can achieve results that significantly	DETAILED DESCRIPTION FIG. 1 shows an example neural network system 160. The neural network system	The output layer 120 is configured to, at each of the time steps, receive the alternative representation at the time step and generate a score distribution over possible
BACKGROUND	Some of the many advantages of such a system are described later. The system	example, text, image or video data, or audio data, for example an example of a particular	outperform the state of the art on audio generation tasks, e.g., by generating speech from	100 is an example of a system implemented as computer programs on one or more	audio samples for the time step. The score distribution includes a respective score for
This specification relates to processing and generating audio using neural networks	can use the score distribution to select a sample for the current time step, by sampling from the distribution. The output may, but need not necessarily, comprise one score for	speaker or language or music. The neural network input may comprise an embedding of the conditioning input. For example in a text-to-speech system a global conditioning	text that is of higher quality than state of the art techniques. A single trained neural network system can be used to generate different voices by conditioning on the speaker	computers in one or more locations, in which the systems, components, and techniques described below can be implemented.	each of multiple possible audio samples. In some implementations, the output layer 120 is a softmax output layer. For example, the output layer 120 can receive the alternative
Neural networks are machine learning models that employ one or more layers of	each possible audio sample value, for example 256 scores for 256 possible values. In can	input may comprise a speaker embedding and a local conditioning input may comprise	identity. By using convolutional neural network layers, e.g., causal convolutional layers,	The neural network system 100 generates sequences of audio data that each	representation 144 and process the alternative representation 144 to generate a score
nonlinear units to predict an output for a received input. Some neural networks include one or more hidden layers in addition to an output layer. The output of each hidden layer	thus be useful to compress or compand the audio sample values, which may be amplitude	linguistic features. The system may be configured to map the neural network input, or a	initeal of recurrent neural network layers, e.g., instead of long short-term memory (LSTM) layers, the neural network system can achieve these advantaneous results while	include a respective audio sample at each of multiple time steps, e.g., an output sequence	distribution 146.
is used as input to the next layer in the network, i.e., the next hidden layer or the output	values, to reduce the number of model outputs. In some implementations the correlational neural network layers are causal	conditioning input, from a low er sampling frequency to the audio sample generation frequency, for example by repeating the input or upsampling the input using a neural	not needing as many computational resources to train as other systems that do include	of audio data 152. Generally, each time step in a given audio sequence corresponds to a respective	In particular, when the neural network system 100 is configured to generate raw audio data, the score distribution includes a respective score for each of multiple possible
layer. Each layer of the network generates an output from a received input in accordance with current values of a respective set of resumeters.	convolutional neural network layers, as described in more detail later. In particular, the	network. Thus the neural network input may comprise features of a text segment and the	recurrent neural network layers, resulting in a reduced training time. By employing	time in an audio waveform and the audio sample at the time step characterizes the	amplitude values. When the neural network system 100 is configured to generate
with current values of a respective set of parameters.	audio-processing convolutional neural network layers may include one or more dilated causal convolutional neural network layers. Assin as described in more detail later, a	output sequence may represent a verbalization of the text segment; and/or the neural network input may comprise speaker or intonation pattern values; and/or the neural	convolutional layers rather than recurrent layers, the computation of the neural network system can be more easily batched and more easily namifelized a p because the layers	waveform at the corresponding time. In some implementations, the audio sample at each time sten in the sequence is the annihule of the audio waveform at the corresponding	compressed or companded values, the score distribution includes a respective score for each of multiple possible compressed or companded values.
SUMMARY	dilated convolutional neural network layer applies a convolution to non-adjacent values in	network input may include one or more of: speaker identity information, language	of the network do not have to be unrolled for each time step, allowing the computation of	time, i.e., the sequence generated by the neural network system 100 is a taw audio	Once the output lay er 146 has generated the score distribution for a given time
This specification describes how a system implemented as computer programs on one or more computers in one or more locations can generate a sequence of audio data	a sequence, i.e., as defined by the outputs from a previous layer. This can increase the receptive field of the convolutional subnetwork by orders of magnitude whilst preserving	identity information, and speaking style information. Alternatively the output sequence represents a piece of music.	the system to be performed more efficiently. Additionally, by employing dilated causal convolutional layers, the receptive field of the convolutional subnetwork and, therefore,	waveform. In some other implementations, the audio sample at each time step in the sequence is a compressed or companded representation of the waveform at the	step, the neural network system 100 can select an audio sample to be included in the output sequence at the given time step from the multiple possible audio samples in
that includes a respective audio sample at each of multiple time steps. For example, the	the input (time) resolution and maintaining computational efficiency.	The convolutional subnetwork may comprise residual connections, for example	the quality of the audio generated by the system, can be improved without greatly	corresponding time. For example, the audio sample can be a µ-law transformed	accordance with the score distribution for the given time step. For example, the neural
sequence of audio data can represent speech in a particular natural language or a piece of music	In some implementations the convolutional neural network layers include multiple stacked blocks of dilated convolutional neural network layers. Each block may	a connection from an input of a convolutional layer to a summer to sum this with an intermediate output of the layer. This effectively allows the network to be trained to skip	increasing the computational cost of generating the audio. The details of one or more embeddiments of the subject matter described in this	representation of the waveform. More specifically, the neural network system 100 generates audio sequences	network system 100 can select an audio sample by sampling from the score distribution, i.e., sampling from the possible audio samples in accordance with the scores in the score
In one innovative aspect a neural network system implemented by one or more	comprise multiple datased convolutional neural network layers with increasing dilation.	or partially skip a layer, thus speeding up convergence and facilitating training of deeper	specification are set forth in the accompanying drawings and the description below.	autoregressively. That is, for each particular time step in an output audio sequence, the	distribution so that each audio sample is selected with a likelihood that corresponds to the
computers is configured to generate an output sequence of audio data that comprises a respective audio sample at each of a plantility of time steps. The neural network system	For example the dilation may be increased by a factor n for each successive layer up to a limit within each block. This can further increase the recentive field size.	models. The convolutional subnetwork may additionally or alternatively comprise skip connections. For example directly form each of one or many intermediate layers of the	Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.	neural network system 100 generates the audio sample at the time step conditioned on the audio samples that have already been senerated as of the particular time step, i.e. on	score for the audio sample, or can select the possible audio sample having the highest score according to the score distribution.
may comprise a convolutional subnatwork comprising one or more audio-processing	In some implementations one or more of the convolutional neural network	connections, for example directly from each of one of more intermediate tayers of the convolutional subnetwork to one or more operations that directly generate the alternative		audio samples that have areaay ocen generated to of the particular time step, i.e., or audio samples at time steps that are earlier than the particular time step in the audio	score according to the score distintuition. The convolutional subnetwork 110 generally includes multiple audio-processing
convolutional neural network layers; and an output layer. The correlational subnetwork may be configured to, for each of the plurality of time steps: receive a current sequence of	layers may have gated activation units. For example a rectified linear or other unit	representation that is provided to the output layer.	BRIEF DESCRIPTION OF THE DRAWINGS FIG. 1 shows an example neural network system.	sequence.	convolutional neural network layers. More specifically, the audio-processing
audio data that comprises the respective audio sample at each time step that precedes the	following a convolution implemented by a layer may be replaced by a gated activation unit. In a gated activation unit the output may be a combination of two (causal)	In some implementations processing the current sequence of audio data using the corrolational submetwork, to generate an alternative representation for the time step,	FIG. 2 shows an example mean network system. FIG. 2 shows a visualization of an example block of dilated causal convolutional	The neural network system 100 includes a convolutional subnetwork 110 and an output layer 120.	convolutional neural network layors include multiple causal convolutional layors. A causal correlational layor is a correlational layor that operator on an input
(current) time step in the output sequence. The convolutional subnetwork may further be configured to process the current sequence of audio data to generate an alternative	convolutions, a main convolution and a gate convolution. The convolutions may each be	re-uses values computed for provious time steps. The re-used values may comprise	layers. FIG. 3 shows an example architecture for the convolutional subnetwork.	At each time step during the generation of an audio sequence, the convolutional	sequence that has a respective input at each of multiple time steps by, for each time step,
representation for the time (current) step. This alternative representation may thus	applied to some or all of the same outputs from the previous layer. The combination may involve a non-linear activation function applied to the gate convolution, for example an	values derived from application of a convolutional filter to the audio sample data or data derived therefrom. The re-used values may be stored at one time step and retrieved at a	FIG. 3 shows an example architecture for the convolutional submetwork. FIG. 4 is a flow diagram of an example process for generating an audio sample at	subnetwork 110 is configured to receive the current audio sequence, i.e., the audio sequence that has already been generated as of the time step, and to process the current	generating an output that depends only on the inputs at the time step and at the time steps before the time steps in the input sequence, i.e., and not on any inputs at any time steps
comprise a numeric representation, i.e., an ordered collection of numeric values, in which the current sequence of audio data has been encoded by the convolutional subnetwork, for	activation with a (0,1) range such as a sigmoid. This may then multiply a value from the	later time step when the same filter is applied to the same (or some of the same) audio	a given time step in an audio sequence.	audio sequence to generate an alternative representation for the time step. For example,	after the time step in the input sequence. In some cases, the causal convolutional layers
the current sequence of autoo data has been encoded by the convolutional submerview, for example encoding features of the current sequence. The output layer may be configured	main convolution; a non-linear activation function may, but need not be, applied to the main convolution. Such an approach may, assist in capturing more complex structure	sumple data or data derived therefrom. This can make the system computationally more efficient and hence faster, because there is no need to re-compute the stored values.	Like reference numbers and designations in the various drawings indicate like elements.	when generating an audio sample 140 in the audio sequence 152; the convolutional subnetwork 110 can receive a current audio sequence 142 that includes the audio samples	are implemented by applying a normal convolution and then shifting each output of the normal convolution by a few time steps, i.e., shifting each output forward by (filter length
to, for each of the plurality of time steps: receive the alternative representation for the	within the data.	Particular embodiments of the subject matter described in this specification can be		that precede the audio sample 140 in the audio sequence 152 and process the current	-1) time steps, prior to applying the activation function for the convolutional layer,
time step, and process the alternative representation for the time step to generate an	The alternative representation from the convolutional subnetwork at each time step may be conditioned on a neural network input, for example a latent representation of	implemented so as to realize one or more of the following advantages. The neural network system can generate on the order of tens of thousands of audio samples per		audio sequence 142 to generate an alternative representation 144.	where "filter length" is the length of the filter of the corrolation that is being applied.
	sup my or consistent of a scalar network man, for example a state representation of	a construction of the second o		4	
	*			·	*
WO 2018/049934 PCT/US2017/050220	WO 2018/049934 PCT/US20170/60120	WO 2018/148934 PCT/US2017/056220	WO 2018/049934 PCT/US2017/056320	WO 2018/048934 PCT/US2017/050220	WO 2018/048934 PCT/US2017/05020
To increase the receptive field of the audio-processing correlational layers without requiring an excessive number of layers or filters of excessive length, some or all	on an identity of the speaker, i.e., so that the speech is generated to sound like the voice of the speaker. In this example, the neural network system 100 can obtain a vector encoding	layer 208 has dilation four, the filter of the layer 208 is applied to outputs that are separated by three outputs in the output sequence generated by the layer 206. Because the	In particular, when the activation function is a gated activation function and the output sequence being generated is not conditioned on a neural network input, the layor	$\mathbf{x} = tanh(W_{f,k} * \mathbf{x} + V_{f,k} * \mathbf{y}) \odot \sigma(W_{g,k} * \mathbf{x} + V_{g,k} * \mathbf{y}),$ where $V_{f,k} * \mathbf{y}$ and $V_{g,k} * \mathbf{y}$ are respective 1 x 1 convolutions. In some implementations,	final output 320 generated by the last layer or can refrain from computing a final output, i.e., can refrain from performing the 1 x 1 correlation and the residual sum for the last.
of the audio-processing convolutional lay ers can be dilated causal convolutional lay ers.	the identity of the speaker, e.g., a one-hot encoded vector identifying the speaker, and	dilated causal convolutional layer 210 has dilation eight, the filter of the layer 210 is	304 applies an element-wise non-linear function 310 which, in the example of FIG. 3 is	where $V_{f,k} + y$ and $V_{g,k} + y$ are respective 1 x.1 convolutions. In some implementations, the system 100 directly receives the sequence y as the neural network input, i.e., directly	layer in the stack.
A dilated convolution is a convolution where the filter is applied over an area larger than its length by skipping input values with a certain step that is defined by the dilation value	condition the generated speech on the obtained vector. Generally, the audio sequences are conditioned on the neural network input by	applied to corputs that are separated by seven outputs in the output sequence generated by the layer 208.	the rowh function, to the output of the dilated convolution with the main filter and applies an element-wise gating function which, in the example of FIG. 3, is the signosid function,	receives a sequence that has the same resolution as the output sequence. In other	Once the processing of all of the layers 304 in the stack of dilated convolutional layers has been completed, the convolutional subnetwork 110 sums 322 the skip outputs
its length by skipping input values with a certain step that is defined by the duation value for the dilated convolution. By incorporating dilated causal convolutions, the audio-	Generally, the autos sequences are constituoted on the neural network input by conditioning the activation function of some or all of the convolutional layers in the	FIG. 3 shows an example architecture 300 for the convolutional subnetwork 110	an element-wise gaining function which, in the example of FRG. 3, is the signson function, to the output of the dilated convolution with the gate filter. The layer 304 then performs	implementations, the system 100 receives a sequence having a lower resolution, i.e., with a lower sampling frequency, that the output sequence. In these cases, the system can	generated by the layers 304. The convolutional subnetwork 110 can then apply one or
processing neural network layers effectively operate on their inputs with a coarser scale than with a normal convolution.	convolutional subnetwork. That is, the output of the activation function and, accordingly, the output of the convolutional law or, is dependent not only on the output of the	of FIG. 1. As described above, in the example architecture 300, the dilated causal convolutional layers that are in the convolutional subnetwork have residual connections	an element-wise multiplication 314 between the output of the non-linear function 310 and the output of the gating function 312 to generate the activation function output.	generate the sequence y by processing the lower resolution sequence using a transposed	more non-linear functions, one or more 1 x 1 convolutions, or both to the sum 322 to generate the alternative representation 144. In particular, in the example of FIG. 3, the
inter with a normal convolution. In some implementations, the audio-processing neural network layers include a	the output of the convolutional tayer, it dependent not only on the output of the convolution performed by the layer but also on the neural network input.	and skip connections.	the output of the gaining function 312 to generate the activation function output. More specifically, when the element-wise non-linearity is towh and the element-	(hearned upsampling) convolutional network to generate the sequence y or can repeat values from the lower resolution sequence across time to generate the sequence y.	generate the anomative representation 144. In particular, in the example of PAL 5, the convolutional subnetwork 110 applies an element-wise non-linearity 324, e.g., a ReLU,
stack of multiple blocks of dilated causal correlational layers. Each block in the stack can include multiple dilated correlational neural network layers with increasing dilation.	Conditioning an activation function of a correlational layer on the neural network input will be described in more detail below with reference to FIG. 3.	In particular, in the architecture 300, the convolutional subnetwork 110 includes a	wise gating function is the zigwood function, the output of the activation function z for a	As an example, when the local features are linguistic features for use in text to	followed by a 1x1 convolution 326, followed by another element-wise non-linearity 328,
can include multiple dilated conventional neural network ayers with increasing dilation. For example, within a block, the dilation can double for each layer starting from an initial	FIG. 2 shows a visualization 200 of an example block of dilated causal	causal convolutional layer 302 that processes the current output sequence 142, i.e., by applying a causal convolution to the current output sequence 142.	haver k satisfies: $\mathbf{z} = \tanh(W_{f,K} * \mathbf{x}) \otimes \sigma(W_{g,K} * \mathbf{x}).$	speech generation, the linguistic features can include some or all of phone, syllable, word, phrase, and utterance-level features of the text. Example sets of linguistic features that	and followed by a final 1x1 convolution 330, to generate the alternative representation 144.
dilation, and then return to the initial dilation for the first lay er in the next block. As an	convolutional layers. In particular, the example block includes a dilated causal	The convolutional subnetwork 110 then processes the output of the causal convolutional layer 302 threach a stack of dilated causal convolutional layers.	where W _{FA} is the main filter for the layer k, x is the layer input, * denotes a causal dilated	can be used are described in Zen, Heiga. An example of context-dependent label format	As described above, the output layer 120 then processes the alternative
illustrative example, the dilations of the dilated convolutional layers in a block can be, in order: 1, 2, 4, , 512. A simplified example of a block of dilated causal convolutional	convolutional lay or 204 with dilation one, a dilated causal convolutional lay or 206 with dilation two, a dilated causal convolutional layer 208 with dilation four, and a dilated	convolutional layer 302 through a stack of dilated causal convolutional layers. Each dilated causal convolutional layer 304 in the stack applies a dilated causal	convolution. \otimes denotes element-wise multiplication, and $W_{g,h}$ is the gate filter for the layer k.	for HMB4-based speech spectrasis in English, 2006. URL http://htt.sp.niiech.ac.jp?/Download and Zan, Heiga, Senice, Andrew, and Schuster,	representation 144 to generate the score distribution 146. FIG: 4 is a flow diagram of an example process 400 for generating an radio
layers is described below with reference to FIG. 2.	causal convolutional layer 210 with dilation eight.	convolution 308 to the input 306 to the illated causal convolutional layer 304. As	When the output sequence being generated is conditioned on a neural network	http://htts.sp.nitech.ac.pt/TDownload and Zan, Honga, Sentor, Andrew, and Schueler, Milor. Statistical parametric spruch synthesis using dwp marral networks. In Proc.	sumple at a given time step in an audio sequence. For convenience, the process 400 will
In some implementations, the correlational subnetwork includes residual connections, skip connections, or both. An example architecture of the correlational	In the visualization 200, the block of dilated causal convolutional layers are operating on a current input sequence 202 to generate an output sequence. In particular,	described above, in some implementations, the dilated causal convolutional layers in the stack are arranged in blocks, with the dilation of the dilated causal convolutions applied	input, the layer 304 also conditions the output of the activation function on the neural	ICASSP, pp. 7962-7966, 2013.	be described as being performed by a system of one or more computers located in one or more locations. For example, a neural network system, e.g., the neural network system
subnetwork that includes both residual connections and skip connections is described	the visualization 200 visualizes using hold arrows how the block generates the output 212	by each layer increasing within a given block and then restarting at the initial value for	network input. In particular, the nerv-linear function and the gating function each take as input a combination of the corresponding dilated correlation output and an input	Because the architecture 300 includes skip connections and residual connections for the dilated causal convolutional layers, the layer 304 then performs a 1 x 1	100 of FIG.1, appropriately programmed, can perform the process 400.
below with reference to FIG. 3. In some intelementations, the neural network system 100 systematics and/o.	that is the output at the time step that is currently the last time step in the current input sequence 202 and the output sequence.	the first layer in the next block. In some implementations, the dilated causal convolutional layers in the stack have	generated from the neural network input.	convolution 316 on the activation function output.	The system provides a current audio sequence as input to the convolutional subnetwork (step 402). The current audio sequence is the audio sequence that has already
sequences conditioned on a neural network input. For example, the neural network	As can be seen from the visualization 200, because each layer in the block is a	a gated activation function in which the output of an element-wise non-linearity, i.e., of a	More specifically, when the neural network input includes global features and is therefore the same for all of the time steps in the sequence, the element-wise non-linearity	The layer 304 provides the output of the 1×1 correlation as the skip output 318 of the layer and adds the residual, i.e., the layer input 306, and the output of the 1×1	been generated as of the given time step, i.e., a sequence that includes the output audio
system 100 can generate the audio sequence 152 conditioned on a neural network input	causal convolutional layer, the output 212 depends only on outputs that are at the last current time step or time steps before the last current time step in the various sequences	conventional activation function, is element-wise multiplied by a gate vector. In some of these implementations, the dilated causal convolution 308 includes two dilated causal	is storb and the element-wise gating function is the stgrootd function, the output of the	convolution to generate the final output 320 of the lay er 304. The convolutional	sumples at time steps before the given time step. As described above, the convolutional subnetwork includes audio-processing convolutional neural network layers, e.g., illated
In some cases, the neural network input includes one or more local features, i.e.,	operated on by the layers in the block.	convolutions on the layer input 302 - a first dilated causal correlation between a main	activation function z for the layer k satisfies: $\mathbf{x} = \tanh(W_{e,\mathbf{x}} * \mathbf{x} + V_{e,\mathbf{x}}^T \mathbf{h}) \odot \sigma(W_{a,\mathbf{x}} * \mathbf{x} + V_{e,\mathbf{x}}^T \mathbf{h}).$	subnetwork 110 then provides the final output 320 as the layer input to the next dilated convolutional layer in the stack.	causal convolutional layers, and is configured to process the current sequence of audio
one or more features that are different for different time steps in the output sequence. For example, the neural network system 100 can obtain as input linguistic features of a text	Additionally, as can be seen from the visualization 200, the layers in the block are arranged in order of increasing dilation, with the first layer in the block, i.e., dilated	filter for the layer 304 and the layer input 306 and another dilated causal convolution	where $V_{i,k}^{T}$ is a main learnable linear projection (of h to the main component of the	In some implementations, the layer 304 performs two 1 x 1 convolutions on the	data to generate an alternative representation for the given time step.
segment and can generate an audio sequence that represents a verbalization of the text		hadronese a serie (Direction the Incom Will and the Incom Intern Will - Incohore of desce		activation function output, one with a residual filter and the other with a skip filter. In	
	causal convolutional layer 204, having dilation one and the last layer in the block, i.e.,	between a gate filter for the layer 304 and the layer input 306. In others of these implementations, dilated causal convolution 308 is a single dilated causal convolution and	activation function) for the layer \hat{v} , h is the neural network input, and $V_{f,\theta}^T$ is a gate		The system provides the alternative representation as input to an output layer, e.g., a softmax output layer (step 404). The output layer is configured to process the
segment, i.e., the neural network system 100 can function as part of a test-to-speech	causal convolutional layer 204, having dilation one and the last layer in the block, i.e., dilated causal convolutional layer 204, having dilation eight. In particular, as is shown by	implementations, dilated causal convolution 308 is a single dilated causal convolution and half of the output of the single convolution is provided as the output of the dilated causal	activation function) for the layer k , h is the neural network input, and $V_{f_{n}}^{2}$ is a gase learnable linear projection (of h to the gate component of the activation function) for the layer k .	these implementations, the layer 304 provides the output of the corrolation with the skip filter as the skip corput 318 of the layer and adds the residual and the output of the $1 \propto 1$	a softmax output layer (step 404). The output layer is configured to process the alternative representation to generate a score distribution over possible audio samples for
segment, i.e., the neural network (system 10% can function in part of a text-to-speech system that converts written text to spoken speech and also includes a component that verbaliens the audio sequence generated by the neural network system 100.	causal convolutional layer 204, having dilation one and the last layer in the block, i.e.,	implementations, diluted causal convolution 398 is a single diluted causal convolution and half of the compare of the single convolution is provided an the compare of the diluted causal convolution between the main filter for the layer 3044 and the layer is paral 3065 and the other half of the sequent of the single convolution is reprovided an the compare of the diluted causal	kamuable linear projection (of h to the gate component of the activation function) for the layer k. Alternatively, when the noural network input includes local features, i.e., features	these implementations, the layer 304 provides the output of the convolutions with the skip filter as the skip output 318 of the layer and table the residual and we output of the 1 x.1 convolution with the metiscalar filter to expensive the fund output 320 of the layer 304	a softmax output layer (step 404). The output layer is configured to process the alternative representation to generate a score distribution over possible and/o samples for the time step. The system selects an and/o sample for inclusion in the and/o sequence at the
system that converts written text to upsken speech and also includes a component that verbalion the audio sequence generated by the neural network system 100. In some other cases, the neural network input includes one cer rece global	canal convolutional lays "Do I, having attractions one and the last lays in the block, i.e., dilated canad convolutional lays 20M having allutions eight. In particular, an is shown by the blod arrows in the trainformation 20M, housance the dilated canad convolutional by are 20M han dilations ones, the filter of the lays 20M is applied to adjusterat impairs in the cannot impt requester 20M. Beams of hald and canad convolutional lays 20M hand dilaton hou.	implementation, of latest canad convolution 39 fits a strught dilated canad convolution and half of the compart of the single encondution is provided an the compart of the dilated eaund convolution between the mann filter for hell per 30 hald the large vin part. Dilated the other half of the compart of the single convolution is reported and not except of the dilated canad convolutions between the part of life for the large 30 hald not large vin part. Since	learnable linear projection (of h to the gata component of the activation function) for the layer k. Alternatively: when the neural network input includes local features; i.e., features that charge from time step to time step, the system 100 obtains a sequence y that includes	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions, 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The stark of the stark,	a softnax orquita layer taiga 440. The extracts layer it configured to parsens the alternative representation to generate a score databation over possible and/o samples for the time stap. The system valotat are and/or sample for industrian in the and/or sequence at the given time step in accordance with the score distribution top 4400. For example, the
system that converts written text to spoken speech and also includes a component that verbalizes the audio sequence generated by the neural network system 100.	causal convolutional layer 204, having dilution one and the last layer in the block, i.e., dilated causal convolutional layer 204, having dilation equit. In sparafordir, as is shown by the bold answers in the visualization 2009, because the dilated causal convolutional layer 204 has dilation ones, the filter of the layer 2044 is applied to adjacent inputs in the current	implementations, diluted causal convolution 398 is a single diluted causal convolution and half of the compare of the single convolution is provided an the compare of the diluted causal convolution between the main filter for the layer 3044 and the layer is paral 3065 and the other half of the sequent of the single convolution is reprovided an the compare of the diluted causal	kamuable linear projection (of h to the gate component of the activation function) for the layer k. Alternatively, when the noural network input includes local features, i.e., features	finese implementations, the layer 304 provides the sutput of the convolution with the skip fibure as the skip output 315 of the layer and slads the resolution and the entry of the 1 x 1 convolutions with the isoskial fibure to generate the final output 320 of the layer 2304. The convolutional subsections, 1110 the providen the final output 320 as the lay or	a softmax output layer (step 404). The output layer is configured to process the alternative representation to generate a score distribution over possible and/o samples for the time step. The system selects an and/o sample for inclusion in the and/o sequence at the
system that converts written text to upoken spaceh and also includes a component that vertilation that and/o separate generated by the neural networker system 100. In some other cases, the neural network input includes one or more global features, i.e., one or more features that are the server howeghout the entity output.	canad convolutional baye 394, Having Blatton server and de hard have of the Mock, i.e., differed canadi convolutional baye 294, horing difference in the hardwards, are as shown hy first-bidl arrows rule for the level 204 horing difference in the hardwards are able on hy first-bidl arrows rule for the for the level 204 horing direct and a convolutional layer 204 has allowner server, the first of the level 204 horing direct anglesen inspirate of the conven- inger sequences 202. Because the dideal canade convolutional layer 204 has different her filter for the layer 204 rule galaction equation has a second server of the second second her filter of the layer 204 rule galaction equation has a second second second by our compare in the second second second second second second second second second second second seco	implementation, dissol canal consolution NPA is a single databat canal consolution and half of the compact of the single consolution in provides at a compact of the distance tanal consolution between the main filters for the lart of the lart base in pice NPA is and the other half of the compact of the single consolution in provides at a compact of the distance tanal consolution between and be gate filter for the lart pice 20% and for the single ATM the filter of the larget consolution in provides at a compact of the distance tanal consolution between one be gate filter for the larget pice at 20% and for the pice target. Not The field and consolutional larget 95% and for the pice target of the	learnable linear projection (off to the gate component of the activation function) for the layer sk.: Alternatively, when the result network input includes local futures, i.e., futures that charge throm time using to time step, the system 100 column is sequence y that includes a set of futures for each investige in the outpace spacement. The outpace of the activation	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The the stark the scark,	a softnax orquita layer taiga 440. The extracts layer it configured to parsens the alternative representation to generate a score databation over possible and/o samples for the time stap. The system valotat are and/or sample for industrian in the and/or sequence at the given time step in accordance with the score distribution top 4400. For example, the
system the converts writes that to update speech that allow includes a comparement that venduation due tande coapsenge compared by the secant ancorea (years 10.0. In to sense other cases, the name in serves of target includes or set or more global features, i.e., our or runs distance that on the same frequency but be entire angus sequence. As an example, the neural network system 100 can generate speech conditioned	canal convolutional type 2014, being fattions are not the last lase in the block, i.e., dialeted can enviroation the 2014. Noting fattions of the structures, as shown by fitte-bill arrows rule transformation. 2014. Noting fattions of the structure of the point 2014 to dialism sea, the fitter of a structure point structure impacts are presented point and the structure of the point of the structure of the structure of the input areas and the fitter areas of the structure of the structure of the fitter of the layer 2014, single of the structure of the structure of the couper supersized by the layer 2014. However the dialism consolitoriand of the structure presented by the layer 2014. However the dialism consolitoriand	implementations, data durand convolution WH as single direct canad convolution and half of the comparise of the surge convolution in provide also developed in the chards canad convolution between the ensure filter for late (JP 2) and a fut hay ingo. We can the chards half of the property of the surge convolution in provide also developed in the chards cannot increase the specific filter for hat (JP 2) and fut hay ingo. We can also convolution between the specific filter for hat (JP 2) and fut has increase. The filter data can convolution of the JP 39 Abs and thermation because of the services of the hard process convolution of the property of the dilated canad convolution.	learnable linear projection (off to the gate component of the activation function) for the layer sk.: Alternatively, when the result network input includes local futures, i.e., futures that charge throm time using to time step, the system 100 column is sequence y that includes a set of futures for each investige in the outpace spacement. The outpace of the activation	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The the stark the scark,	a softmax andper last er state particular, best or transformed in spreases the alternative representation to partners as some institutions one possible anados samples for for store step The control market and an instrumentation of the control market and analysis of the given time any in accordance with the scores of minimum (in the assessed for my nummer and samples a possible andos sample in accordance with the score dominium of the assessed for the score of minimum (in the assessed for the score dominium).
system the converts writes that to update speech that allow includes a comparement that venduation due tande coapsenge compared by the secant ancorea (years 10.0. In to sense other cases, the name in serves of target includes or set or more global features, i.e., our or runs distance that on the same frequency but be entire angus sequence. As an example, the neural network system 100 can generate speech conditioned	canal convolutional type 2014, being fattions are not the last lase in the block, i.e., dialeted can enviroation the 2014. Noting fattions of the structures, as shown by fitte-bill arrows rule transformation. 2014. Noting fattions of the structure of the point 2014 to dialism sea, the fitter of a structure point structure impacts are presented point and the structure of the point of the structure of the structure of the input areas and the fitter areas of the structure of the structure of the fitter of the layer 2014, single of the structure of the structure of the couper supersized by the layer 2014. However the dialism consolitoriand of the structure presented by the layer 2014. However the dialism consolitoriand	implementations, data durand convolution WH as single direct canad convolution and half of the comparise of the surge convolution in provide also developed in the chards canad convolution between the ensure filter for late (JP 2) and a fut hay ingo. We can the chards half of the property of the surge convolution in provide also developed in the chards cannot increase the specific filter for hat (JP 2) and fut hay ingo. We can also convolution between the specific filter for hat (JP 2) and fut has increase. The filter data can convolution of the JP 39 Abs and thermation because of the services of the hard process convolution of the property of the dilated canad convolution.	learnable linear projection (off to the gate component of the activation function) for the layer sk.: Alternatively, when the result network input includes local futures, i.e., futures that charge throm time using to time step, the system 100 column is sequence y that includes a set of futures for each investige in the outpace spacement. The outpace of the activation	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The the stark the scark,	a softmax andper last er state particular, best or transformed in spreases the alternative representation to partners as some institutions one possible anados samples for for store step The control market and an instrumentation of the control market and analysis of the given time any in accordance with the scores of minimum (in the assessed for my nummer and samples a possible andos sample in accordance with the score dominium of the assessed for the score of minimum (in the assessed for the score dominium).
optime that construct writes the top locality angues that all also that the compared that versifies the state on oppinger again multi-the term and movies do great multi-the Historic allow cannot be meand in order top provide the state of the state finance, it is a strength of the state of the state of the state of the state support. As are example, the stated means of symm. The state parameter speed conditioned supports is due compared, the stated means of symm. The state parameter speed conditioned	canal convolutional type 2014, being fattions are not the last lase in the block, i.e., dialeted can enviroation the 2014. Noting fattions of the structures, as shown by fitte-bill arrows rule transformation. 2014. Noting fattions of the structure of the point 2014 to dialism sea, the fitter of a structure point structure inspiration for enrome imper suppose 2020. Excernite the dialistic parat convolutional type 2024 to dialism to no. for filter of the layer 2014 supplied to structure that are appendent by one compares that coupter suppose 2020. Excernite the dialistic parat convolutional of the object paration of the layer 2014. Distance the dialistic consolitoriand coupter supporting by the layer 2014. Distance the dialistic convolutional	implementations, data durand convolution WH as single direct canad convolution and half of the comparise of the surge convolution in provide also developed in the chards canad convolution between the ensure filter for late (JP 2) and a fut hay ingo. We can the chards half of the property of the surge convolution in provide also developed in the chards cannot increase the specific filter for hat (JP 2) and fut hay ingo. We can also convolution between the specific filter for hat (JP 2) and fut has increase. The filter data can convolution of the JP 39 Abs and thermation because of the services of the hard process convolution of the property of the dilated canad convolution.	kernahu inter upperson of vis ub ag as composed on the activation function) for the legal. Automatively, where the smarth serves is type includes hard function. I, is former that due from timer top to time type in possibility and accesses by that includes as and features for each time step in the argue sequence. The support of the activation functions 2 for the lege i does stantified.	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The the stark the scark,	a submax copel are or of any (40). The entryick ber in conjugated to process the atomic representation to approach constraintion one opposition abounding for the true ray. The format formation and/or margine constraintion (40). For example, the given true are pre-associated with the user a doublewise (40). The example, the system can sample a possible and/or arepli in accordance with the users doublewise.
system the converts writes that to update speech that allow includes a comparement that venduation due tande coapsenge compared by the secant ancorea (years 10.0. In to sense other cases, the name in serves of target includes or set or more global features, i.e., our or runs distance that on the same frequency but be entire angus sequence. As an example, the neural network system 100 can generate speech conditioned	canal convolutional type 2014, being fattions are not the last lase in the block, i.e., dialeted can enviroation the 2014. Noting fattions of the structures, as shown by fitte-bill arrows rule transformation. 2014. Noting fattions of the structure of the point 2014 to dialism sea, the fitter of a structure point structure inspiration for enrome imper suppose 2020. Excernite the dialistic parat convolutional type 2024 to dialism to no. for filter of the layer 2014 supplied to structure that are appendent by one compares that coupter suppose 2020. Excernite the dialistic parat convolutional of the object paration of the layer 2014. Distance the dialistic consolitoriand coupter supporting by the layer 2014. Distance the dialistic convolutional	implementations, data durand convolution WH as single direct canad convolution and half of the comparise of the surge convolution in provide also developed in the chards canad convolution between the ensure filter for late (JP 2) and a fut hay ingo. We can the chards half of the property of the surge convolution in provide also developed in the chards cannot increase the specific filter for hat (JP 2) and fut hay ingo. We can also convolution between the specific filter for hat (JP 2) and fut has increase. The filter data can convolution of the JP 39 Abs and thermation because of the services of the hard process convolution of the property of the dilated canad convolution.	learnable linear projection (off to the gate component of the activation function) for the layer sk.: Alternatively, when the result network input includes local futures, i.e., futures that charge throm time using to time step, the system 100 column as sequences y that includes a set of futures for each investige in the outpace spacement. The outpace of the activation	these implementations, the line of Via (provide the energy of a fiber consolutions with the skip fibure at the skip coupled 31% of the layer and adde the residual and the couplet of the 1 × 1 convolutions with the residual fiber to generate the final coupter 250 of the 10 we 7.0%. The convolutional substretions 11 the providen the final coupter 250 as the layer input to the next differed convolutional by our in the stark. The the stark the scark,	a softmax andper last er state particular, best or transformed in spreases the alternative representation to partners as some institutions one possible anados samples for for store step The control market and an instrumentation of the control market and analysis of the given time any in accordance with the scores of minimum (in the assessed for my nummer and samples a possible andos sample in accordance with the score dominium of the assessed for the score of minimum (in the assessed for the score dominium).
State that control writering take to be partial marked on the p	canal construintend up to 2014, being datatois near and be includes in the Hocks, i.e., distance and consolitation large structure datatois and a structure data to the structur	Implementations, dilato canda cironolatori Met su sulpid altadi canda cironolatori and di conducti che un di conducti con los possi da di conducti con los che di cale a candi constrainto biossano fato galitto di conducti con los conducti con los conducti non di conducti con los conducti con los conducti con los conducti internationa los conducti conducti con los conducti con los conducti non di conducti con los conducti con los conducti con los conducti di conducti con los conducti con los conducti con los conducti di conducti con los conducti con los conducti con los conducti di conducti con los conductivi con los conducti con los conducti di conducti con los conductivi con los conducti con los conducti di conducti conductivi con los conducti los conducti con los conducti di conducti con los conductivi con los conducti con los conducti di conducti con los conductivi con los conducti con los conducti di conductivi con los conductivi con los conductivi conducti con los conductivi di conductivi con esta conductivi con los conductivi con los conductivi di conductivi con esta conducti los conductivi conductivi conductivi di conductivi con esta conducti los conductivi con los conductivi di conductivi con esta conducti los conductivi con los conductivi di conductivi conductivi con esta conducti los conductivi di conductivi conductivi con esta conducti los conductivi di conductivi conductivi con esta conductivi conductivi conductivi di conductivi conductivi con esta conductivi conductivi conductivi di conductivi con esta conductivi conductivi conductivi conductivi di conductivi conductivi conductivi conductivi conductivi conductivi conductivi conductivi conductivi conductivi di conductivi conductivi conductivi conductivi conductivi conductivi di c	kennalis later proposition of %n to large compose and the activation function of the large set. A dependent of the set of the sector of the large set of the set of the set of the set of the sector functions for the layer i from set of the set of the sector of the sector functions for the layer i from set of the sector of the set of the set of the sector of the sector of the set of the sector of the sector of the set of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sector of the sect	Bere implementations, the logs 3/3 provides the experiment of the consultation with the start fitting and one should be consultated one should be consultated one should be consultated one should be consultated on showing and the start provides the entry of the entry of the entry of	a submax copical are of used (6). The ending to give to compare a distribution or oppication are supported as a support or constraintion or oppication and supports of the interime. The interime is a support of the interiment of the interiment of the distribution of
cytater that correct writering to be update and a bench and a bench or update and upd	canal constructional by e20 M, basing dilations steered the task layer in the folds, i.e., dilated cancel consolitations layer 3. More particularly a structure of the fold and consolitation layer the balances and the fold and canad consolitation layer solves and another the fold and canad consolitation layer and the fold and canad consolitation layer and the fold and canad compares 20. Research with and canad consolitation layer 20 for definition to many anyones 20. Research with and canad consolitation of the fold and canad compares anyones and the fold of the layer 20. Because the definition of the fold and canad consolitation of the sequences and the fold of the layer 20. Research and and consolitation of the sequences and the fold of the layer 20. Research and the fold and canad consolitation and the sequences and the fold of the layer 20. Research and the fold and canad consolitation and the sequences and the fold of the layer 20. Research and the fold and canad consolitation and the sequences and the fold of the layer and the sequences and compares presented by the layer and the sequences and the sequences and the sequences and the fold of the layer and the sequences and the sequences and compares presented by the layer and the sequences and the sequences and the sequences and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares and the sequences and the sequences and the sequences and compares	Implementations, dilata canadi convolution 34 htt and a studyed idead is canadi convolution and and convolution barries given convolutions in provide and is convolution in the classical canadi convolutions between the mean filter for far large 19 states for the large 19 states of the black for canadi convolutions between the gas filter first large 19 states and the large trape. 31 means convolutions between the gas filter first large 19 states and the large trape 35 means at the large trape and the large trape 35 means and the large trape 35 means at the large trape and the large provide trape and the large trape 35 means a transmission function of the large 35 means and the large trape 35 means the large trape and the large providence of the large state and the large states and the large provide states are stated as the large state and the large states and the large provide states are stated as the large states and the large states and the large states and endower providence are an anothed composed states, are often states.	kenatabi kara proposition of \$% ub aga composed of the activation function of the the second of t	Bere implementations, the logs 3/3 provides the energy of the convolution with the stary time show process 1/1 of the logs of a log dotted show provides the energy 4/2 of the log or 3/4 convolution with the model failure target stars the first and any 0/2 of the log or 3/4 the transmission of the logs of the log dotted show the log dotted show the transmission of the logs of the log dotted show the log dotted show the transmission of the log dotted show the log dotted show the MOMENTATION OF THE SHOW THE AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND AND	a submax cope La et or (a) 40. The explor Lay is no constrained on opposite the advance of the form of the source of advance of the source of th
starts the control writering to be paired and paired being to set of the pair of the paired being to set of the paired being t	cand considuring by 2014, being dations one of the tark pair in the Model, i.e., different and considering large structure structures, as a shows by field alreases nets restardants? Not because the field and considered the tark pair of the Model and the	Implementation, and an and a conclusion. We have a single details can add convolution and and conception of the single convolution is notestical de constrained the details can add constrained the single particular bare for the large bar and the single single bar add constrained the single particular bar add constrained the single particular bare methods and constrained and the single particular bar add constrained the details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar add details are an ordinations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or gas and deviations or gas and deviations or gas and single gas are deviations or gas and deviations or gas and deviations or gas and single gas are deviations or gas and deviations or gas and deviations of the deviation	tearable later properties of the totage compose and the activation function of the later sectors and the sector of the sector of the sector of the sector of the sectors of the later Automatic sectors of the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the secto	The implementation, the large N3 inprovides the expansion of the convolutions with the starting the observation of the 11 of the top with the convolution of the top of the 11 of the convolution with the expansion of the convolution of the 11 of the top with the convolution of the top with the convolution of the 11 of the 11 of the convolution of the 11 of the CONVOLUTION OF the 11 of the CONVOLUTION OF the 11 of the CONVOLUTION OF the 11 of the CONVOLUTION OF the 11 of th	a subma could be only only. The entropy to pix to any only only only only only only only on
citatistic carrenti verittisti kuo junki anglo kui dala suki kuo junki panga kuo junki junki panga kuo junk	canal considuring of pp 201, being dations are not be in the pix in the tota, i.e., disting and acconsistent pix pre-total canadian sets and the total pix	Implementations, difficult caused aroundance 30 He is analyed alread to caused aroundance and analyzed of the strain filter for large to 30 star for high tegs to 30 and 30 star excitations between the many filter for large to 30 and 40 star in the strain of caused aroundance between the strain filter for large to 30 and 40 strain of the strain of caused aroundance 30 strain of the strain of the strain of the strain of caused aroundance 30 strain of the strain of the strain of the strain of caused aroundance 30 strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the strain of the stra	branches lotter projection of \$\$ to the gate composed of the activation function() of the lotter	<text><text><text><text></text></text></text></text>	a subma copie die origi (e)). The equipite lyer is conjugated partners the distribution or possible and and the submatrix of the possible and supplies for the sure ray.
optime that construit writen to top update angle and a land series of your of the structure	cand considuring by 2014, being dations one of the tark pair in the Model, i.e., different and considering large structure structures, as a shows by field alreases nets restardants? Not because the field and considered the tark pair of the Model and the	Implementation, and an and a conclusion. We have a single details can add convolution and and conception of the single convolution is notestical de constrained the details can add constrained the single particular bare for the large bar and the single single bar add constrained the single particular bar add constrained the single particular bare methods and constrained and the single particular bar add constrained the details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar details and the single particular bar add constrained the single particular bar add details are an ordinations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or proceeding languages, and in can be deplecied as any firm, single gas are deviations or gas and deviations or gas and deviations or gas and single gas are deviations or gas and deviations or gas and deviations or gas and single gas are deviations or gas and deviations or gas and deviations of the deviation	tearable later properties of the totage compose and the activation function of the later sectors and the sector of the sector of the sector of the sector of the sectors of the later Automatic sectors of the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the sector of the sector of the sectors of the sectors because the sector of the secto	The implementation, the large N-10 provides the energy of the convolution with the strip time of the operation of the other stress and the formation and the stress of the stress of the energy of the other stress of the other stress of the other stress of the stress of the stress of the stress of the stress of the stress of the convolution of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stress of the stre	<text><text><text></text></text></text>
startistic control treation topolar indication control treation in the startistic control treatic control treation in the startistic control treation in	canal constrainting the 2014, being dilation strend by last layer in the the SA, i.e., different and consolitating the strength strength and strength streng	Implementations, did cauda cau	branches lotter propies of the tot gate composed of the activation function of the distribution. The formation of the distribution of the distr	<text><text><text><text></text></text></text></text>	<text></text>
starting the control tracting to update appeads and its method to compare promoting the tracting starting st	canal considuring of pp 201, being failures in each test large in the Host, i.e., distant consolution large to 2014, being failures site, being failures site, being failure site, being failures site, being failures site, being failure site, being failures site, being fa	Implementations, did cauda cireo violation 3 Mit a study di badia cauda cireo violation and di construit di sung di construit di sung di sung di sung di sung di sung di sung di sung di di sung di construit di sung di sung di sung di sung di sung di sung di sung di di sung di sung di sung di di sung di sung di sung di sung di sung di sung di sung di sung di di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung di sung	branches lorge represents of the totage composed of the activation function of the totage composed of the activation function of the totage composed of the activation of t	<text><text><text><text></text></text></text></text>	<text><text><text></text></text></text>
<text><text><text><text></text></text></text></text>	erand considuring the 2014, being dilation seem of the inclusion is a shown by the big dilation seem of the inclusion is an above the big dilation seem of the inclusion is a shown by the big dilation seem of the inclusion is a shown by the big dilation seem of the inclusion is a shown by the big dilation seem of the inclusion is a shown by the big dilation seem of the inclusion	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	Independencies of the origin of the origi	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>
<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text></text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text><text></text></text></text></text></text></text></text></text>	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>	<text><text><text><text></text></text></text></text>









WHAT IS CLAIMED IS:

 A neural network system implemented by one or more computers, wherein the neural network system is configured to generate an output sequence of audio data that comprises a respective audio sample at each of a plurality of time steps, and

wherein the neural network system comprises:

a convolutional subnetwork comprising one or more audio-processing convolutional neural network layers, wherein the convolutional subnetwork is configured to, for each of the plurality of time steps:

receive a current sequence of audio data that comprises the respective audio sample at each time step that precedes the time step in the output sequence, and

process the current sequence of audio data to generate an alternative representation for the time step; and

an output layer, wherein the output layer is configured to, for each of the plurality of time steps:

receive the alternative representation for the time step, and process the alternative representation for the time step to generate an output that defines a score distribution over a plurality of possible audio samples for the time step.

BASIC PRINCIPLE

Broad scope of protection, but probably not patentable

FALLBACK POSITIONS!

CONRETE REALIZATION

Likely patentable, but too narrow to be enforced

Photo by <u>Jeremy Bishop</u> on <u>Unspla</u>

INGREDIENTS OF A PERFECT INVENTION DISCLOSURE REPORT

Photo by <u>Dose Juice</u> on <u>Unspla</u>

WHAT'S THE INVENTION ACTUALLY MADE OF?

POSSIBLE WORK-AROUNDS?



Photo by Dose Juice on Unsplash

WHAT'S THE UNIQUE CONCEPT OF THE INVENTION*?

Photo by Dose Juice on Unsplash

* THE ONE THING THE CUSTOMER CANNOT LIVE WITHOUT WHICH IS NOT STRAIGHT-FORWARD

caler kale spin jeme ging

PRESSE

WHICH PRIOR ART IS THERE?

upouls Ban strai et liux

trawbe lax. 3anane

TEIN PR

Photo by <u>Dose Juice</u> on <u>Unsplash</u>

WHAT ARE THE DIFFERENCES <u>REALLY</u>?

Photo by Jeremy Bishop on Unsplash

SPECIFIC REALIZATION OF THE INVENTION (AND WORK-AROUNDS!)

CONCEPTUAL INVENTIVE CONTRIBUTION

FALLBACK POSITIONS TO GET FROM "CONCEPTUAL" TO "SPECIFIC"



GET THE FULL GUIDE HERE:

20 pages of step-by-step instructions

Includes examples and a handy template

BARDEHLE PAGENBERG

How to draft the perfect invention disclosure report

Prinzregentenplatz 7 81675 München T +49.(0)89.928 05-0 F +49.(0)89.928 05-444 info@bardehle.de www.bardehle.com

Ask me for a copy: <u>bastian.best@bardehle.de</u>

BARDEHLE PAGENBERG Partnerschaft mbB Patentanwälte Rechtsanwälte Amtsgericht München Partnerschaftsregister 1152 ISO 9001 certified In cooperation with:

BARDEHLE.COM



BARDEHLE PAGENBERG

BARDEHLE PAGENBERG Partnerschaft mbB Prinzregentenplatz 7 81675 München

in X y G+ □ Fi

"Law Firm of the Year" 2016 for Intellectual Property Law – named by *Best Lawyers*[®] and *Handelsblatt* "TOP-KANZLEI Patentrecht 2017" – awarded by *WirtschaftsWoche* "Germany Trade Mark Firm of the Year" 2018 – honored by *Managing IP*