FIG. 1

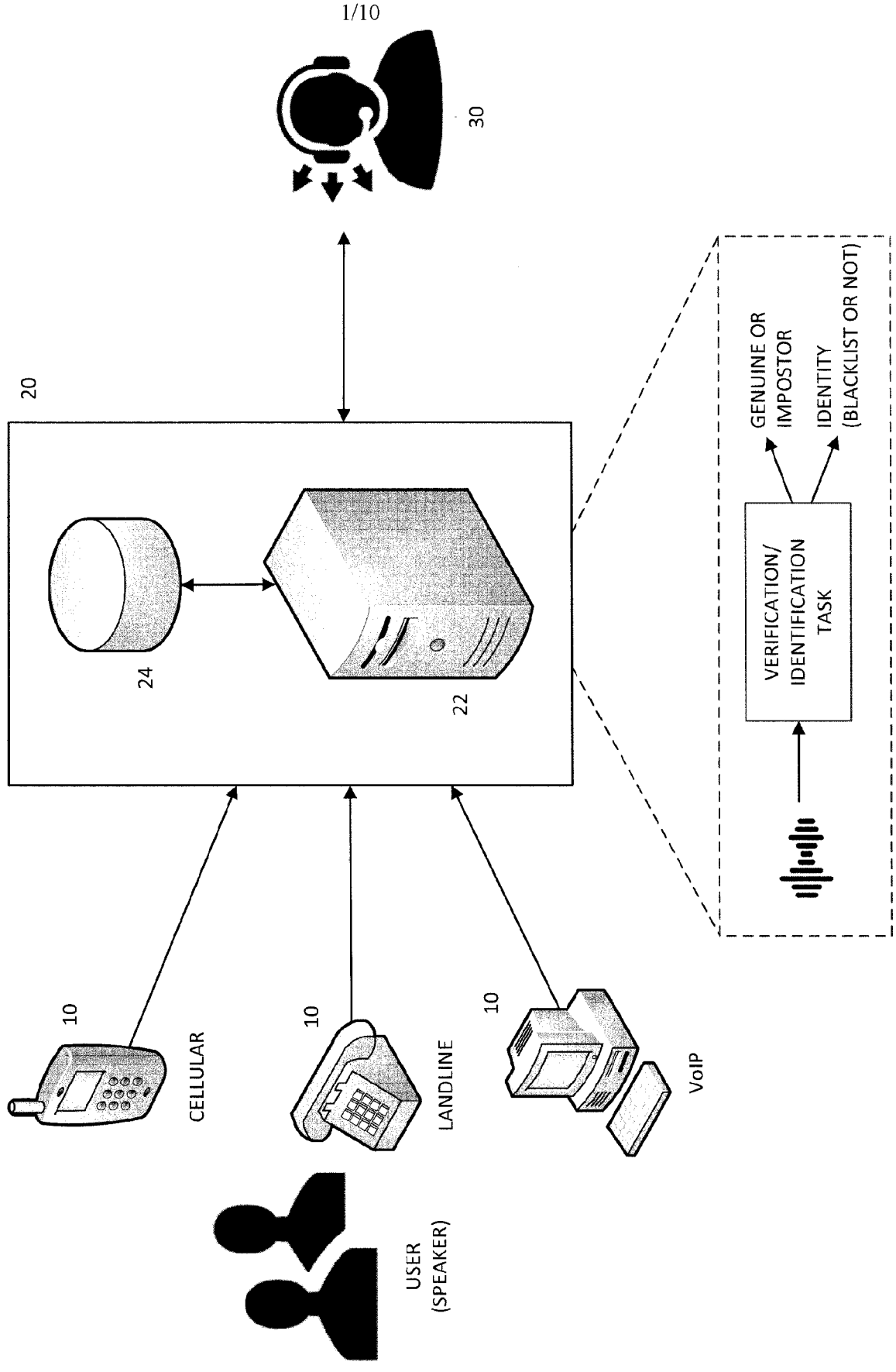

30
20
24
22
GENUINE OR IMPOSTOR
IDENTITY (BLACKLIST OR NOT)
VERIFICATION/ IDENTIFICATION TASK
10
CELLULAR
10
LANDLINE
10
VoIP
USER (SPEAKER)

FIG. 2A

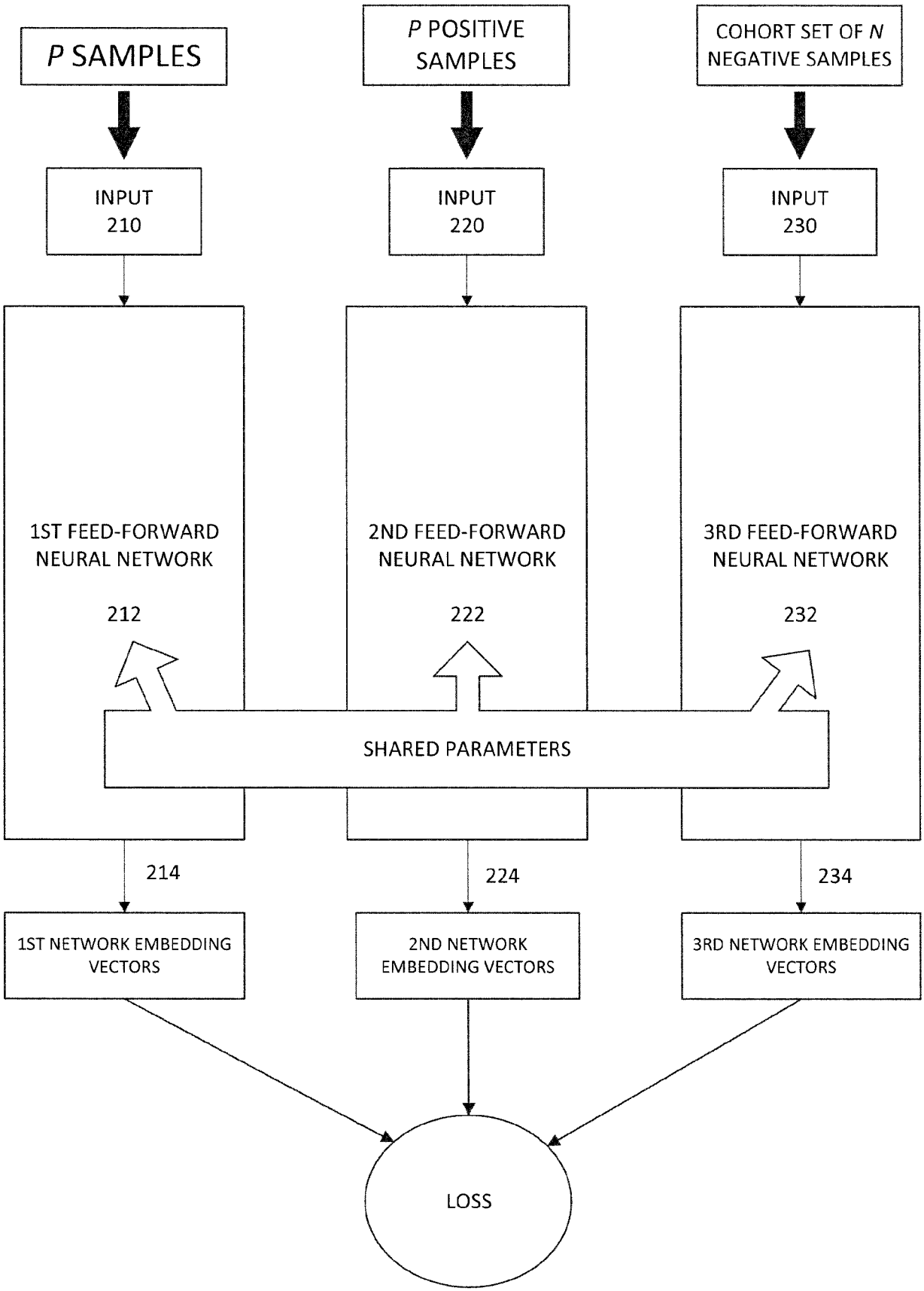

P SAMPLES
P POSITIVE SAMPLES
COHORT SET OF N NEGATIVE SAMPLES
INPUT 210
INPUT 220
INPUT 230
1ST FEED-FORWARD NEURAL NETWORK
212
2ND FEED-FORWARD NEURAL NETWORK
222
3RD FEED-FORWARD NEURAL NETWORK
232
SHARED PARAMETERS
214
224
234
1ST NETWORK EMBEDDING VECTORS
2ND NETWORK EMBEDDING VECTORS
3RD NETWORK EMBEDDING VECTORS
LOSS

FIG. 2B

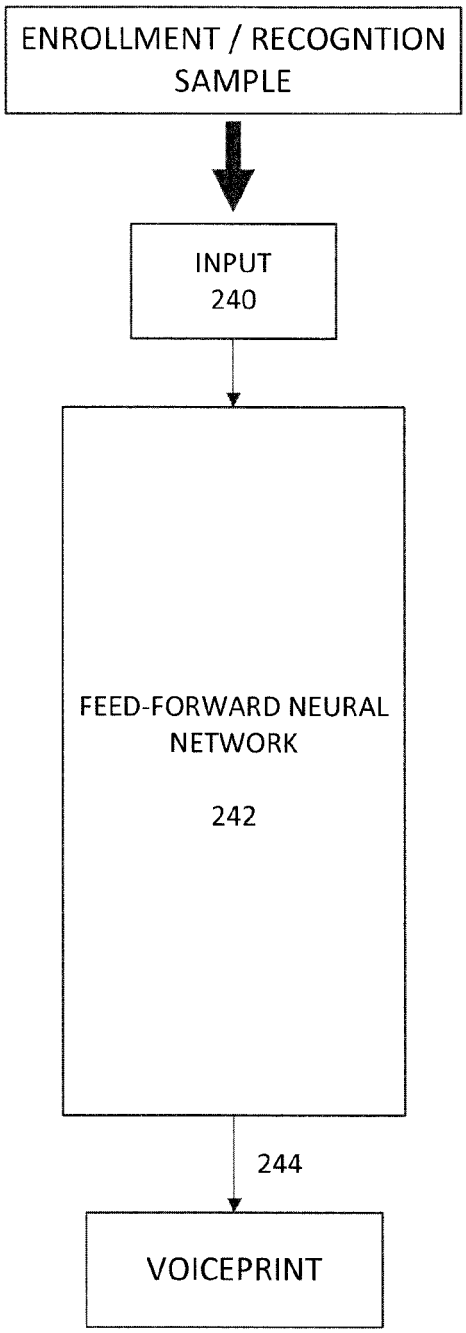

ENROLLMENT / RECOGNTION SAMPLE
INPUT 240
FEED-FORWARD NEURAL NETWORK 242
244
VOICEPRINT

FIG. 3A

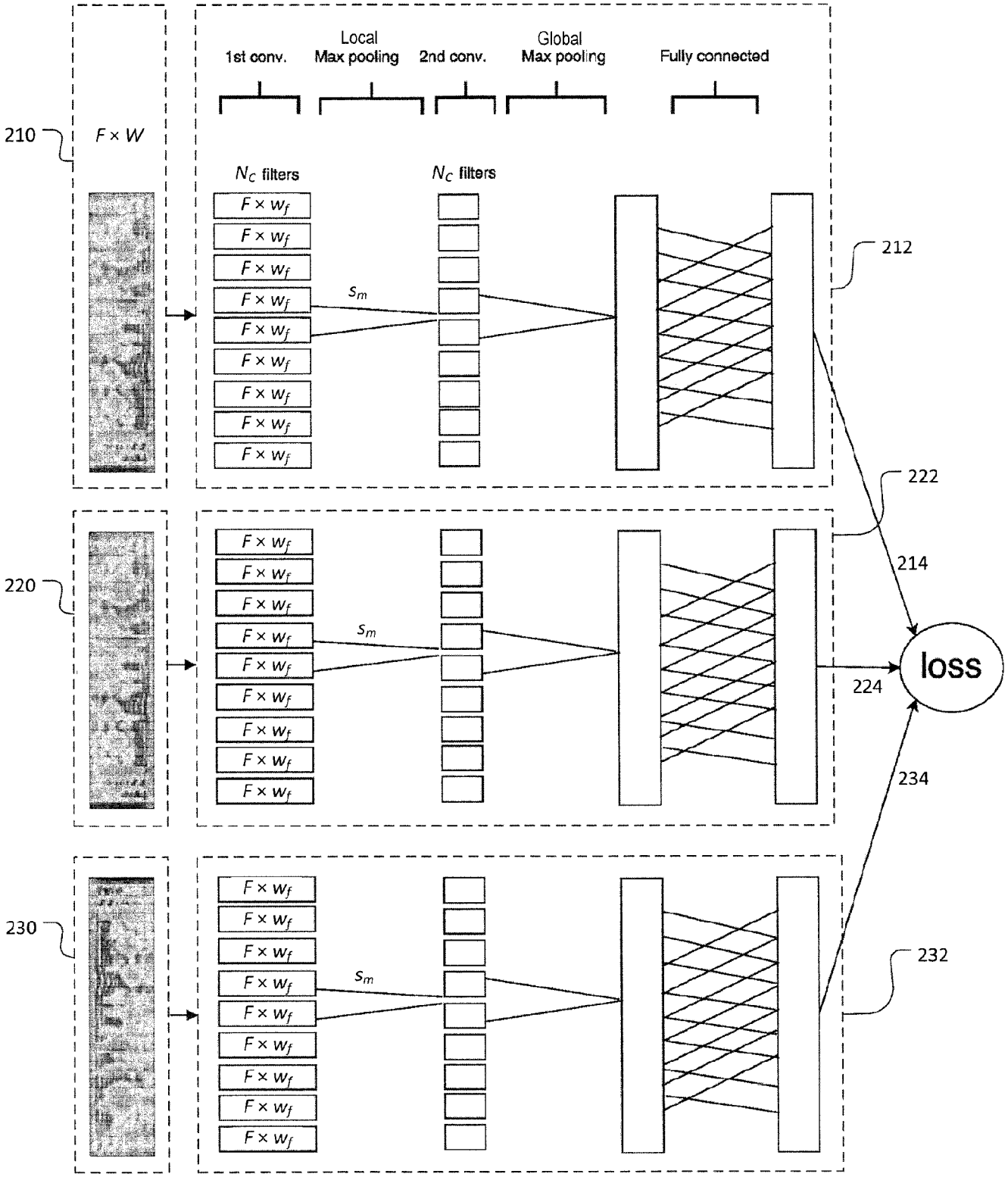

1st conv.
Local Max pooling
2nd conv.
Global Max pooling
Fully connected
210
F × W
N_C filters
N_C filters
F × w_f
s_m
212
222
214
220
loss
224
234
230
232

FIG. 3B

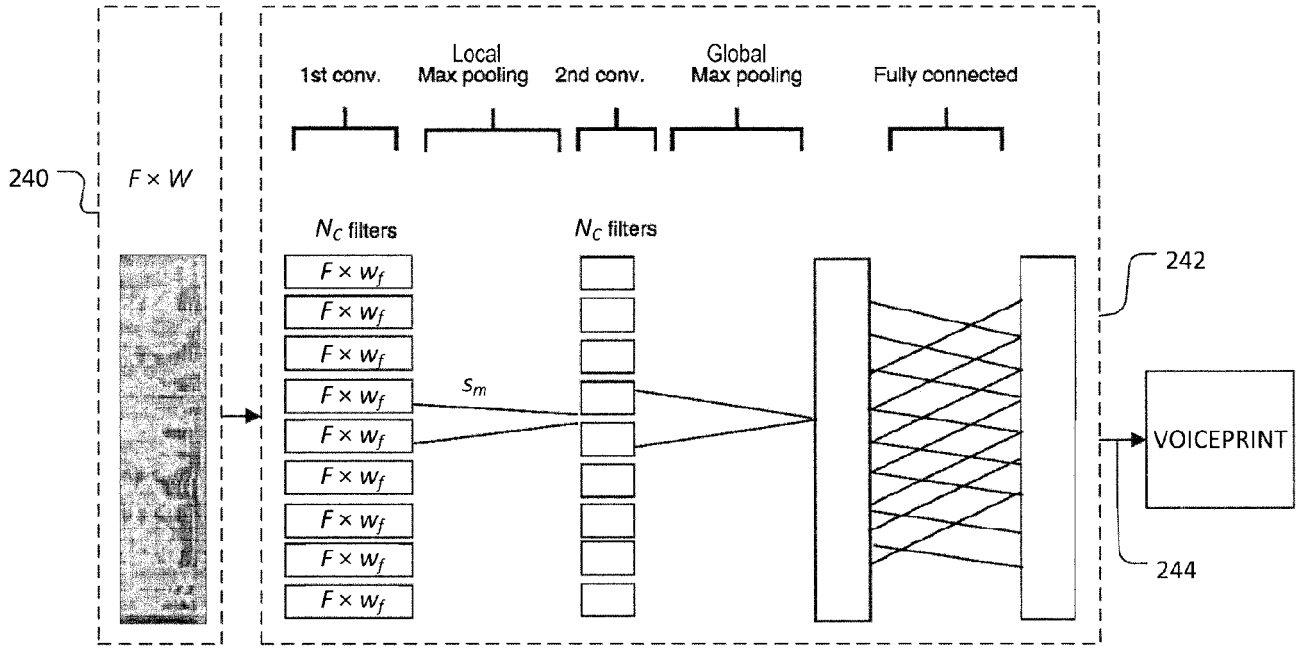

240
F × W
1st conv.
Local Max pooling
2nd conv.
Global Max pooling
Fully connected
N_C filters
N_C filters
F × w_f
s_m
242
VOICEPRINT
244

FIG. 3C

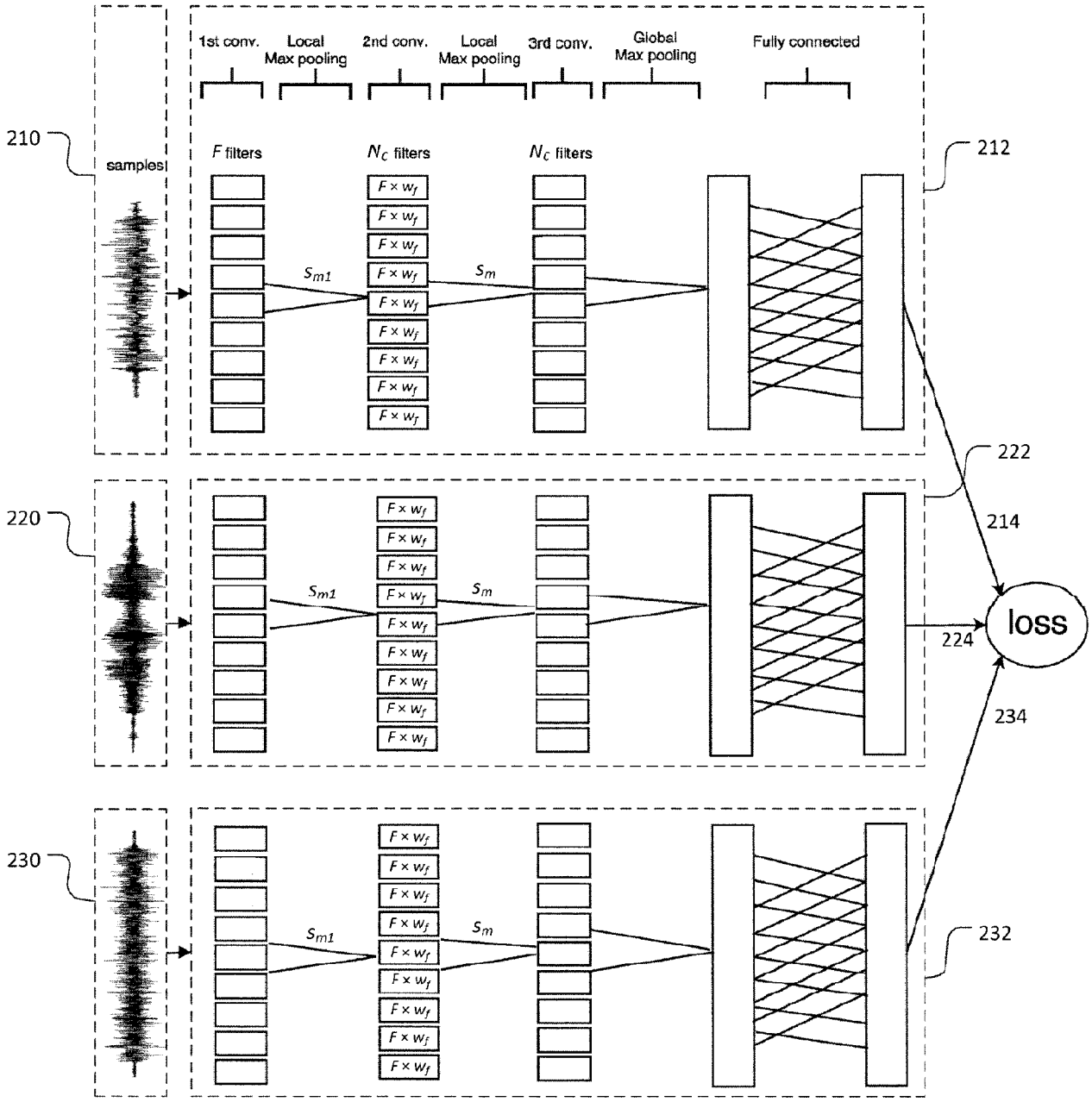

1st conv.
Local Max pooling
2nd conv.
Local Max pooling
3rd conv.
Global Max pooling
Fully connected
210
samples
F filters
N_C filters
N_C filters
212
S_{m1}
S_m
$F \times w_f$
220
222
214
loss
224
230
234
232

FIG. 4

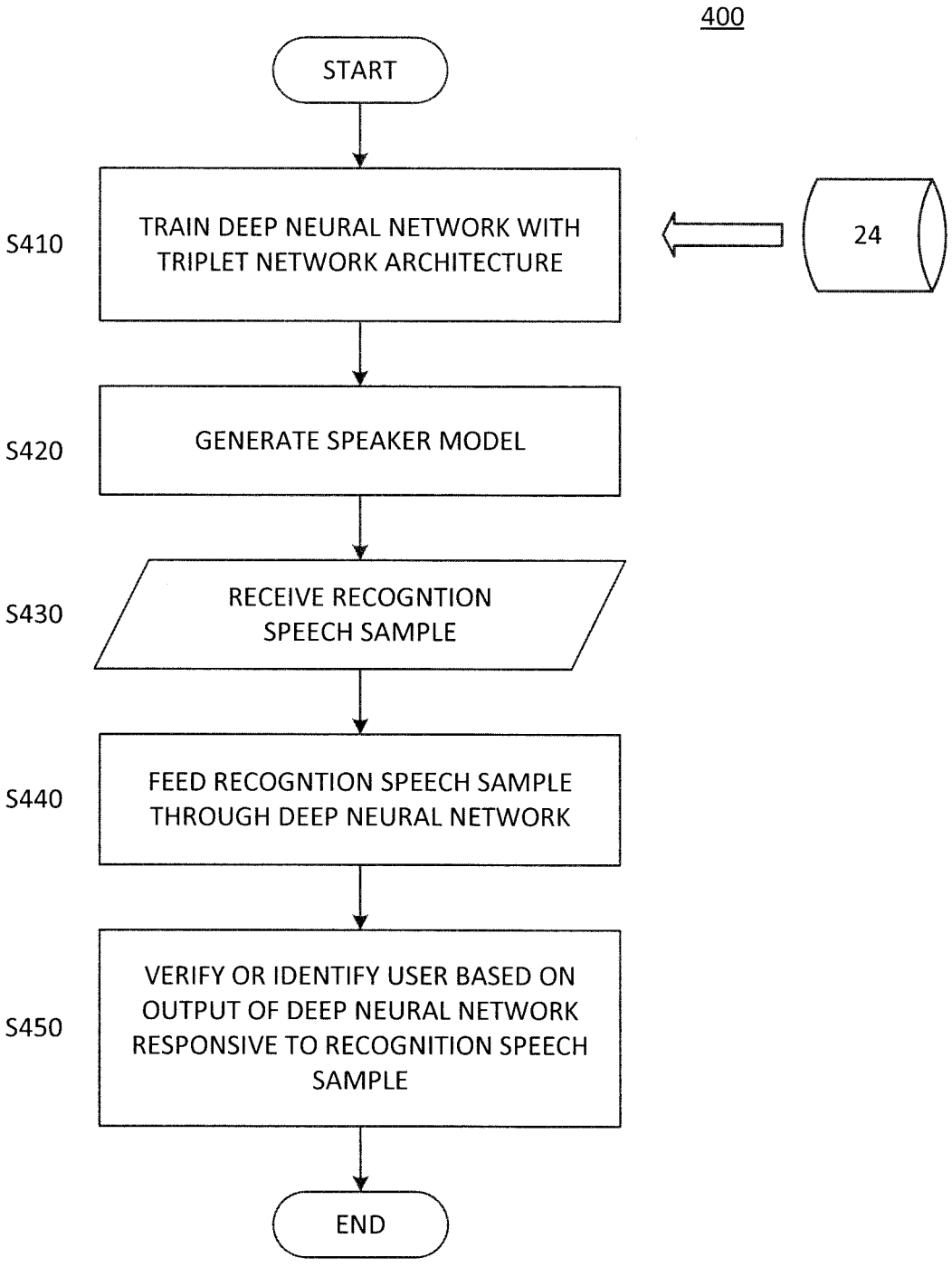

400
START
S410
TRAIN DEEP NEURAL NETWORK WITH TRIPLET NETWORK ARCHITECTURE
24
S420
GENERATE SPEAKER MODEL
S430
RECEIVE RECOGNTION SPEECH SAMPLE
S440
FEED RECOGNTION SPEECH SAMPLE THROUGH DEEP NEURAL NETWORK
S450
VERIFY OR IDENTIFY USER BASED ON OUTPUT OF DEEP NEURAL NETWORK RESPONSIVE TO RECOGNITION SPEECH SAMPLE
END

FIG. 5

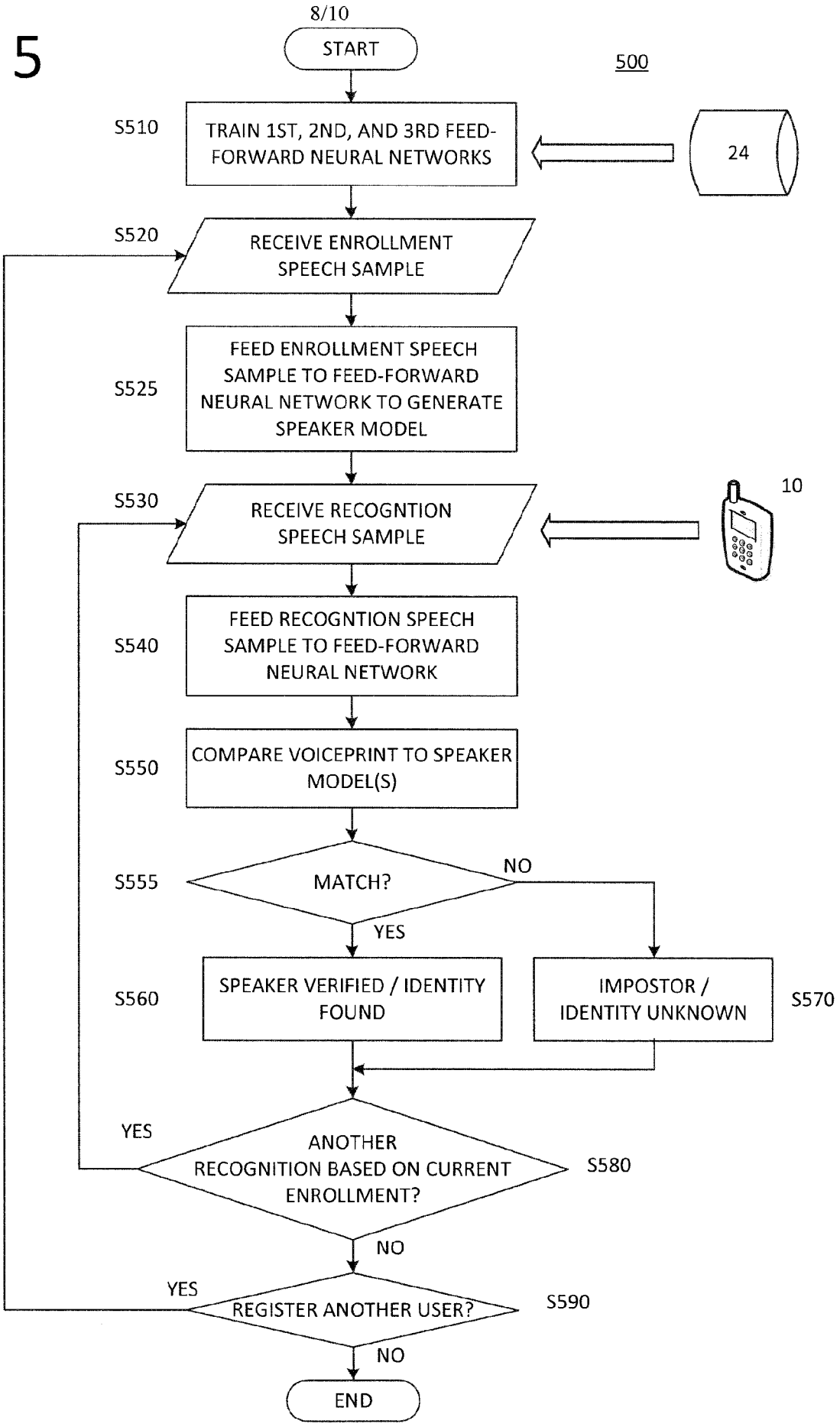
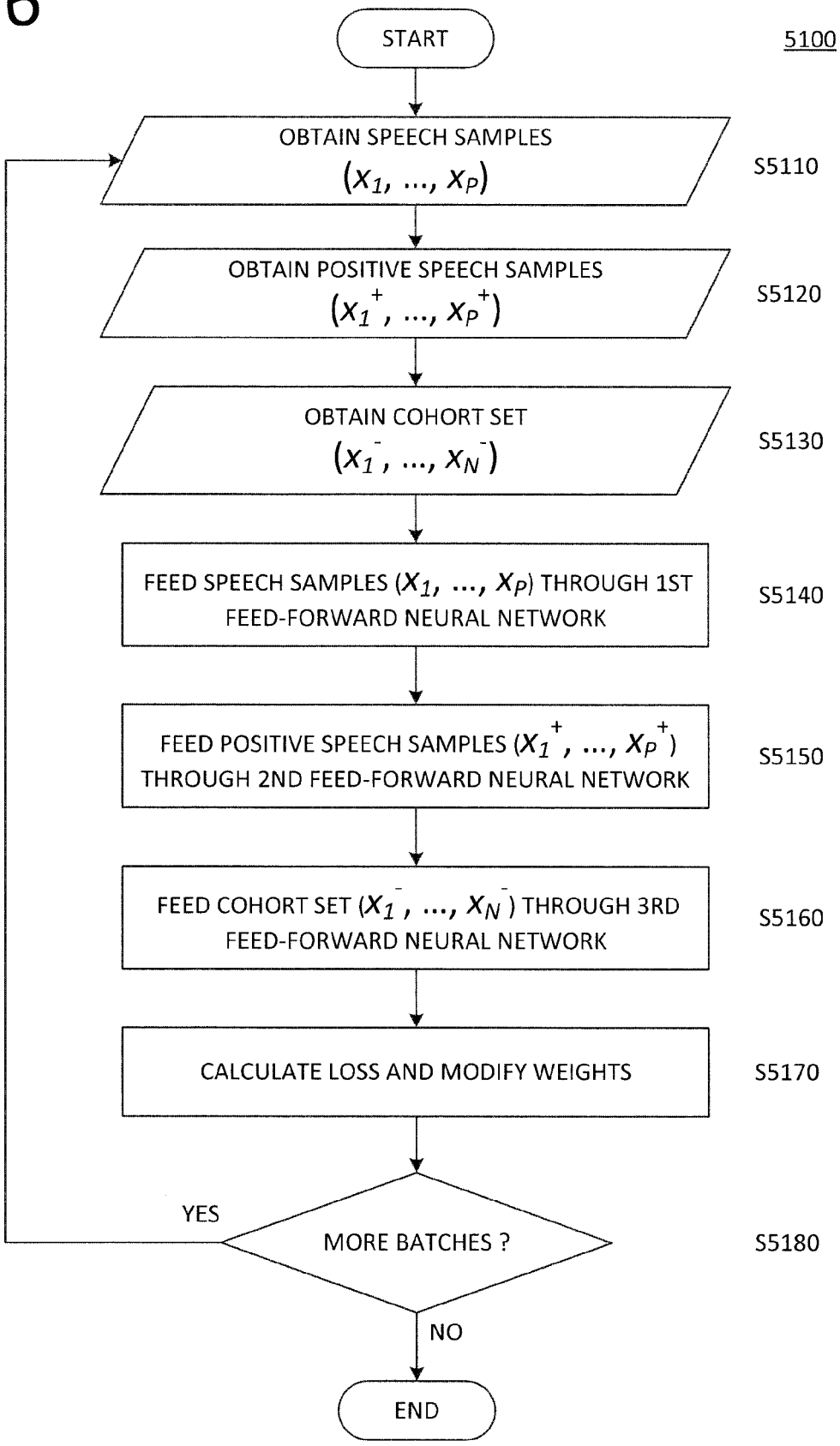

START
500
S510
TRAIN 1ST, 2ND, AND 3RD FEED-FORWARD NEURAL NETWORKS
24
S520
RECEIVE ENROLLMENT SPEECH SAMPLE
S525
FEED ENROLLMENT SPEECH SAMPLE TO FEED-FORWARD NEURAL NETWORK TO GENERATE SPEAKER MODEL
S530
RECEIVE RECOGNTION SPEECH SAMPLE
10
S540
FEED RECOGNTION SPEECH SAMPLE TO FEED-FORWARD NEURAL NETWORK
S550
COMPARE VOICEPRINT TO SPEAKER MODEL(S)
S555
MATCH?
NO
YES
S560
SPEAKER VERIFIED / IDENTITY FOUND
IMPOSTOR / IDENTITY UNKNOWN
S570
YES
ANOTHER RECOGNITION BASED ON CURRENT ENROLLMENT?
S580
NO
YES
REGISTER ANOTHER USER?
S590
NO
END

FIG. 6


START
5100
OBTAIN SPEECH SAMPLES
$(X_1, \ldots, X_P)$
S5110
OBTAIN POSITIVE SPEECH SAMPLES
$(X_1^+, \ldots, X_P^+)$
S5120
OBTAIN COHORT SET
$(X_1^-, \ldots, X_N^-)$
S5130
FEED SPEECH SAMPLES $(X_1, \ldots, X_P)$ THROUGH 1ST FEED-FORWARD NEURAL NETWORK
S5140
FEED POSITIVE SPEECH SAMPLES $(X_1^+, \ldots, X_P^+)$ THROUGH 2ND FEED-FORWARD NEURAL NETWORK
S5150
FEED COHORT SET $(X_1^-, \ldots, X_N^-)$ THROUGH 3RD FEED-FORWARD NEURAL NETWORK
S5160
CALCULATE LOSS AND MODIFY WEIGHTS
S5170
YES
MORE BATCHES ?
S5180
NO
END

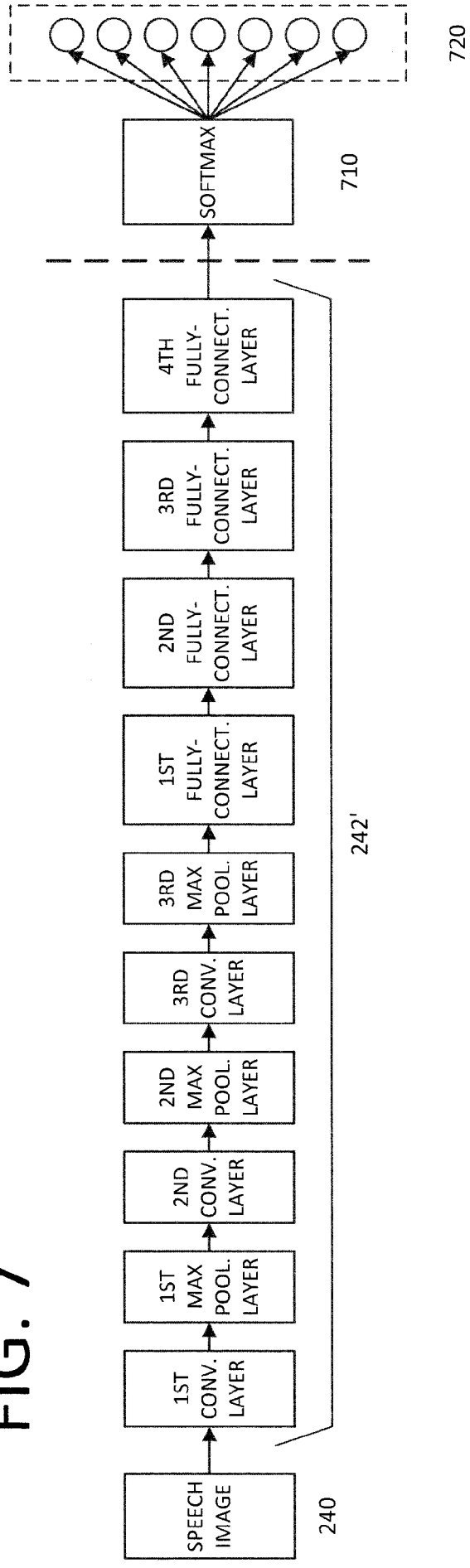

SPEECH IMAGE
240
1ST CONV. LAYER
1ST MAX POOL. LAYER
2ND CONV. LAYER
2ND MAX POOL. LAYER
3RD CONV. LAYER
3RD MAX POOL. LAYER
1ST FULLY-CONNECT. LAYER
2ND FULLY-CONNECT. LAYER
3RD FULLY-CONNECT. LAYER
4TH FULLY-CONNECT. LAYER
242'
SOFTMAX
710
720

FIG. 7