



- (51) **International Patent Classification:**
G10L 17/00 (2013.01) *G06N 20/00* (2019.01)
G10L 15/00 (2013.01)
- (21) **International Application Number:**
PCT/US2021/020891
- (22) **International Filing Date:**
04 March 2021 (04.03.2021)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/985,757 05 March 2020 (05.03.2020) US
- (71) **Applicant:** **PINDROP SECURITY, INC.** [US/US]; 817 West Peachtree Street, NW, Suite 770, Atlanta, Georgia 30308 (US).
- (72) **Inventors:** **PHATAK, Kedar**; 817 West Peachtree Street, NW, Suite 770, Atlanta, Georgia 30308 (US). **KHOURY,**

Elie; 817 West Peachtree Street, NW, Suite 770, Atlanta, Georgia 30308 (US).

- (74) **Agent:** **SOPHIR, Eric**; Foley & Lardner LLP, 3000 K Street N.W., Washington, District of Columbia 20007 (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, IT, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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,

(54) **Title:** SYSTEMS AND METHODS OF SPEAKER-INDEPENDENT EMBEDDING FOR IDENTIFICATION AND VERIFICATION FROM AUDIO

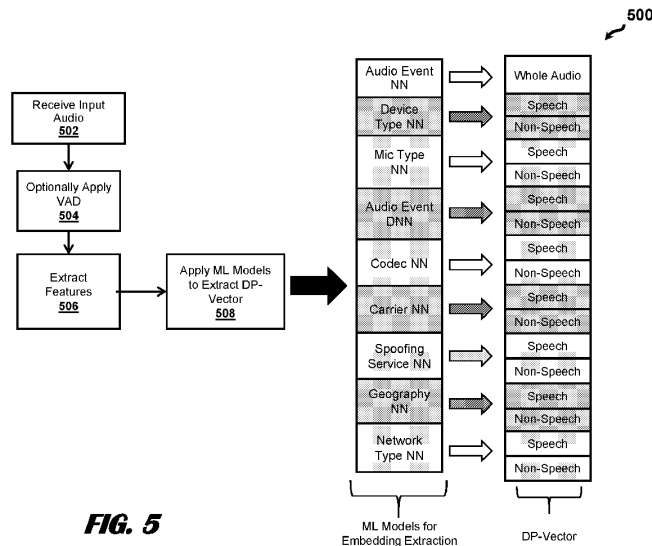


FIG. 5

(57) **Abstract:** Embodiments described herein provide for audio processing operations that evaluate characteristics of audio signals that are independent of the speaker's voice. A neural network architecture trains and applies discriminatory neural networks tasked with modelling and classifying speaker-independent characteristics. The task-specific models generate or extract feature vectors from input audio data based on the trained embedding extraction models. The embeddings from the task-specific models are concatenated to form a deep-phoneprint vector for the input audio signal. The DP vector is a low dimensional representation of the each of the speaker-independent characteristics of the audio signal and applied in various downstream operations.

[Continued on next page]

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, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*