

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
19 May 2023 (19.05.2023)



(10) International Publication Number
WO 2023/086025 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/SG2022/050815

(22) International Filing Date:
09 November 2022 (09.11.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10202112485R 10 November 2021 (10.11.2021) SG

(71) Applicant: NATIONAL UNIVERSITY OF SINGAPORE [SG/SG]; 21 Lower Kent Ridge Road, Singapore 119077 (SG).

(72) Inventors: LIM, Brian, Youliang; c/o National University of Singapore, School of Computing / Department of Computer Science, 21 Lower Kent Ridge Road, Singapore 119077 (SG). ZHANG, Wencan; c/o National University of Singapore, School of Computing / Department of Computer Science, 21 Lower Kent Ridge Road, Singapore 119077 (SG).

(74) Agent: SPRUSON & FERGUSON (ASIA) PTE LTD; P.O. Box 1531, Robinson Road Post Office, Singapore 903031 (SG).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,

(54) Title: METHOD AND SYSTEM FOR GENERATING AN EXPLAINABLE PREDICTION OF AN EMOTION ASSOCIATED WITH A VOCAL SAMPLE

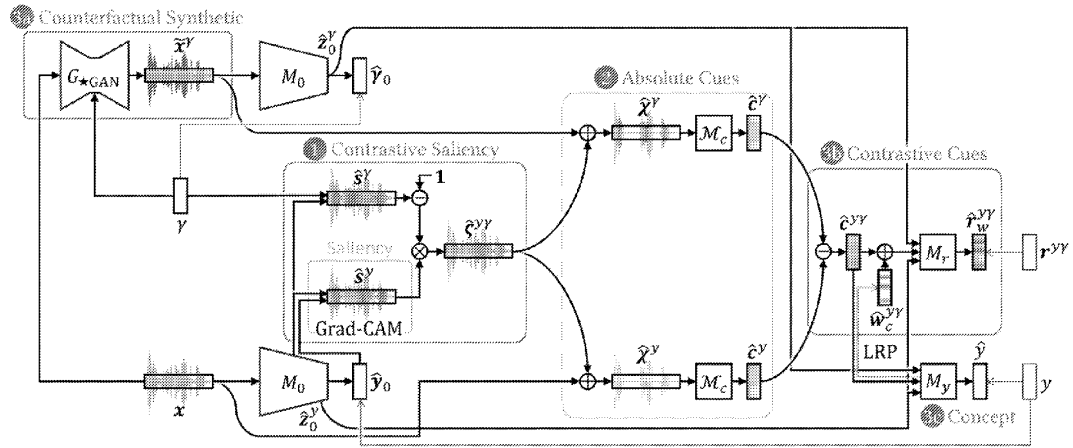


FIG. 2A

(57) Abstract: A method and a system for generating an explainable prediction of an emotion associated with a vocal sample are disclosed. The method includes receiving, by a processing device, a vector representation (z_{x_0}) of an initial prediction (y_0) of the emotion associated with the vocal sample (x), a counterfactual synthetic vocal sample (xY) associated with the vocal sample (x) and an alternate emotion (y) different from the initial prediction (y_0) of the emotion, a vector representation (z_{xY}) of an emotion prediction (y_0) associated with the counterfactual synthetic vocal sample (xY), vocal cue information (cy , cy) associated with the vocal sample (x) and the counterfactual synthetic vocal sample (xY) and attribution explanation information ($iVc7$) associated with relative importance of the vocal cue information (cy , cy) in prediction of the emotion. The method also includes determining, using the processing device, numeric cue differences (cyY) between the vocal cue information (cy) associated with the vocal sample (x) and the vocal cue information (cy) associated with the counterfactual synthetic vocal sample (xY), generating, using the processing device, cue difference relations infor-

[Continued on next page]

LY, MA, MD, 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, CV, 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, MC, ME, 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:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

mation ($r^{\wedge}$) based on the attribution explanation information ($iv^{\wedge 7}$), the numeric cue differences (cyy), the vector representation ($z_{..}$) of the initial prediction ($y0$) and the vector representation ($z_{..}$) of the emotion prediction ($y0$) associated with the counterfactual synthetic vocal sample (xY) using a first neural network (Mr), generating, using the processing device, a final prediction (y) of the emotion based on the numeric cue differences (cyy), the vector representation ($z_{..}$) of the initial prediction ($y0$) and the vector representation ($z_{..}$) of the emotion prediction ($y0$) associated with the counterfactual synthetic vocal sample (xY) using a second neural network (My), and generating, using the processing device, the explainable prediction of the emotion associated with the vocal sample (x) based on at least the counterfactual synthetic vocal sample (xY), the final prediction (y) of the emotion and the cue difference relations information ($r^{\wedge}$).