Intonazioa eta Emozioak: Ezagutze automatiko baterantz.

Jon Sanchez, Iker Luengo, Amaia Castelruiz, Xabier Zalbide, Eva Navas, Inma Hernaez

Euskal Herriko Unibertsitatea
Ion, ikerl,amaia,xabier,eva,inma@bips.bi.ehu.es

Abstract

Regarding the developments that have been taking place in the recent years in the field of man-machine communication, it has been noticed the need of going beyond a neutral communication, that is, to reach an emotional kind of communication in order for the machine to be able to recognize human emotions. In this article the authors discuss several research projects undertaken in the last few years that deal with automatic emotion recognition. Furthermore, an experiment on the subject is proposed, and a detailed description is offered of the database created to that purpose.

Laburpena

Gizakia eta makinen arteko komunikazioak azken urteotan jasan duen garapenaren harira, komunikazio neutroaz harago joateko beharra ikusi izan da, hots, emoziozko komunikazio batera heltzeko beharra, makina giza-emozioak ezagutzeko gai izan dadin. Lan honetan azken urteotan emozio ezagutze automatikoaren inguruan egin den hainbat ikerketaren berri emateaz gain, esperimentu bat proposatzen da.

Hitz gakoak: Intonazioa, Ezagutze automatikoa, emozioak.

1. Sarrera

Hizketaren teknologien industriak bultzada handia jaso du azken urteotan. Arlo honen arduretariko bat, makina eta gizakiaren arteko komunikazioa dugu; komunikazio hau ikusizko gertakarien bidez gauzatu ohi da, pantaila bat ikustea esate baterako, eta baita ukituzkoen bidez ere, hala nola, botoi bat sakatu edota teklatu baten tekla sekuentzia. Egun bilatzen den xedea komunikazioa soinuzko gertakarietara bideratzea da, hau da, gizakiok ahotsa eta entzumena erabiliz makinekin komunikatu ahal izatea.

Aurrera pausu hori erdiesteko, sistema bi beharrezkoak ditugu, komunikazioaren noranzko bakoitzerako bana: makina batek gizaki bati hitz egin diezaion, nola edo hala lortzen dituen datuak komunikatuz, ahots-sintesiko sistemak behar dira, testu-ahots bihurketa sistemak hain zuzen ere. Eta, alderantziz, gizaki batek makina bati hitz egin diezaion, eta jakina, makinak ulertzeko moduan, hizketaren ezagutze automatikorako sistemak beharrezkoak dira.

Printzipioz, pentsatzekoa da gizakia eta makinaren arteko komunikazioak aseptikoa, neutroa izan beharko lukeela, makinak emozioak sentitzeko programatuta ez dauden izakiak direlako. Hala eta guztiz ere, komunikazioaren beste aldean pertsona bat dago, emozioak sentitu eta adierazteko gai den pertsona bat, eta, hain zuzen, modu askotan erakusten ditu emoziook, bai irudiaren bidez, keinuekin, edota hitzez eta jokaeraz. Hortaz, makinak bere emozio ez-gizatarrak sentitzeko gai ez direla onartuta, giza emozioak interpretatu eta irudikatzeko gauza izan behar dute [2], gizakiekiko elkarreragina osoa eta aberatsagoa gerta dadin.

Gizakia eta makinen arteko emoziozko elkar eragina, luzaroan, zientzia fikzioaren arloko kontua izan bada ere, hizketaren teknologiek hain heldutasun nabarmena lortzeko bidean daude, ezen makina batek emozio bat imitatu edo gizaki batengan antzemango duen unea oso hurbil baitago. Esate baterako, ahots-testu bihurgailua erabiliz dabilen ipuinak kontatzeko sistema bat askozaz onargarriagoa izango da pasarte triste bat irakurtzean, ahotsari tristura kutsua emateko gai bada, alaitasuna sortu berri onak direnean edo haserrea irudikatu pertsonaietako batek haserre dagoenean. Eta, komunikazioaren beste noranzkoan ere, giza ahotsak oso emotiboetan aldaketa nabarmenak jasaten dituenez, hizketa emozionala identifikatuko duen ahots ezagutze automatikoko sistema bat, eta beraz, ezagutzea egoera berrira birkonfiguratuko duena, egoera guztiei konfigurazio berarekin aurre egingo dien beste sistema bat baino arrakastatsuagoa izango da [20].

Honen guztiaren kasu praktiko eta hurbilagoa Call Centerretan topa dezakegu: erabiltzailearentzako telefono-informazio zerbitzu hauetan egun den joera, ahal den neurrian sistema automatikoak baliatzea da. Gisa honetan, ahotsaren ezagutze sistema automatikoak eta ahots sintesiko sistemak erabiliz, makina bera solasean aritzen da erabiltzailearekin, eta aukera ezberdinak eskainiz ahotsaren bidez ibil daiteke harik eta nahi duen zerbitzua erdietsi arte. Tamalez, sistema hauetan gauzak ez dira horren modu idilikoan gertatzen. Ezagutza ez da beti perfektua, erabiltzaileak makinak itxaroten ez dituen gauzak esaten ditu, hizketa erregistroa ez da berdina... hau dela eta, sarri, erabiltzailea haserretzen ohi da, eta nahi duena lortu ezean telefonoa eskegi lezake eta komunikazioa amaitutzat jo (pozik geratzen ez den erabiltzailea gisa honetako edozein sistemarentzat negatiboa da) [3]. Jakina, arazook ez dira jazotzen telefonoz beste aldean pertsona bat dagoenean. Call Center-a ustiatzen duen enpresarentzat ostera, askoz garestiagoa da komunikazioa gizaki baten bidez sistema automatiko

baten bidez baino. Hortaz, konpromiso bat bilatu beharra dago: erabiltzaileak ahalik eta luzaroen egon ahal izan beharko du, baina, huts eginez gero, sistemak giza-operatzaile batengana bidaltzeko gai izan beharko da. Hala gertatuz gero egungo zerbitzu baten barruan emozioak detektatzeko sistemaren erabilgarritasuna zeharo ondo ikusiko da.

Emozioak detektatzeko sistema bat erdiestea ez da nolanahiko kontua, ordea. Gizaki bat, oro har, beste gizaki batekin berbetan dagoela gai da emozioak zuzen identifikatzeko. Hala ere, ez du beti oso argi ikusten emozio horrek non datzan, ze keinuen bidez ezagutu dezakeen. Hortaz, emozioak ezagutzeko sistema bat lortzearren eman behar den aurreneko urratsa emozio bat zehazki zer den jakitea da, eta giza harremanen zein aldetan errotze den emozio horien komunikazioaren funtsa.

Lan xume honetan ez diogu emozioen giza komunikazioaren gaiari bete-betean helduko, izan ere arlo anitzetan zeresana ematen duen gaia da (hizkuntza hutsa, ahotsaren ezaugarriak, gorputz-lengoaia, jarrerak, keinuak...). Ahots bidezko komunikazioaz baino ez dugu jardungo, eta zehazki intonazio hutsaren bidez igortzeko gai garen emozioaz. Zenbait egileren arabera, ezaugarri honek bakarrik ezin lezake emoziorik igorri [6], baina hala eta guztiz ere, garrantzizko ezaugarri bakarra ez den arren (beraz, beste eragingarri asko kontuan hartu ezean ez dugu igorpen osoa antzemango), eta sarritan beste datuak erabiltzen ez badira antzematen den emozioa zalantzakoa gertatzen den arren, ikerketa askok egiaztatu dute zeregin honetan onartzeko moduko arrakasta lor daitekeela.

Emozioak automatikoki antzemateko gai den sistema bat lortzeko, lehenengo eta behin, emozioek eurek dakartzaten arazoei aurre egin behar zaie, lan honen bigarren zatian egingo dena, hain zuzen: beti ere, emozio bat zer den definitu, igortzeko ze datu behar diren, eta ze modutan sortu eta erregistratu daitezkeen emozio horiek gero eurekin lan egin ahal izateko.

Hirugarren atalean emozioen ezagutzaren arloan garapen egoeraren berri ematen da. Mota honetako sistema batera heltzeko jarraitu beharreko prozedura azalduz, ezagutze-sistemetako pausu bakoitza aztertzen da, asko guk lortu nahi dugunaren parekoak eta beste batzuk oso bestelakoak badira ere (esate baterako, giza antzematea erabiltzen dutenak), garatu nahi den sisteman aplikatzeko baliozko ezagupenak emango dizkigute.

Laugarren atalean, aurreko puntuan gogora ekarri dugun ezagupen guztia erabiliz esperimentu bat planteatuko dugu. Erabiltzeko prozedura azaltzen da (eskuarki aztertutako sistemetan agertzen dena, nahiz eta oso ezberdinak izan), eta hortik abiatuz ezagutze sistemaren atal bakoitza nolakoa izan daitekeen zehazten da: Ezagutu nahi diren emozioak, erabil daitezkeen datu-baseak, datuetatik ondorioetara iristeko eraiki behar diren parametroak, eta azkenik, patroiak sailkatzeko sistemaren diseinua.

Amaitzeko, bosgarren atalean zenbait ondorio plazaratzen da.

2. Emozioen arazoa

Pertsona gisa besteek igortzen dizkiguten emozioak deskodifikatu eta geureganatzeko ahalmena garatu dugun arren, makinak gauza bera egin ahal izatea lortu nahi badugu, geure emozioak zehatz-mehatz aztertu beharra daukagu, prozesua ahal den neurrian automatizatzearren.

2.1. Zer da emozio bat?

Egile askok hainbat ikuspuntu eta iritzi eman dituzte galdera honen inguruan [4]. Psikologiaren tradizio akademikoak esaten du guganaino heldu diren azturen adierazpenak edo ekintzak direla, hala nola aurpegia edo gorputzaren mugimenduak. [6]-ren arabera, ekintzaren bat eginarazten digun egoera psikologikoa da. Organismoak egoera zehatz baten aurrean jokatzeko duen gaitasuntzat ere definitu izan da, bai eta jasotzen den hainbat informazio mota prozesatzeko estrategiatzat ere, ondorioren batera heltzearren (ikuspegi hau hain zuzen oso bateragarria da emoziozko sistema automatiko baten ideiarekin).

Oso agerikoa da emozio kontzientea eta inkontzientearen arteko bereizketa. Ohartu barik beti izaten dugu egoera emozionalen bat, kontzienteki zeinuen adierazten ez badugu ere. Badira kontzientzia eta inkontzientzia maila asko, intentsitate eta iraupenarekin zerikusia dutenak.

Hurrengo sailkapena, ordea, ez da horren agerikoa; emozioak unibertso diskretua ala etengabekoa dira? Emozio egoera kopuruan mugatua ote da, eta beraz egoera bakar batean egon gaitezke? Ala gure egoera emozionala identifikatzen duten aldagai etengabe bat edo gehiago ote dira? Psikologoak ez dira gaurdaino ados jarri, nahiz eta teoria zehatz batek nagusitasun hartu duen azkenaldian [4]: Bada egoera emozionalaren unibertso etengabe bat "noranzko emozionala" eta "emozioaren intentsitatea" aldagaiek definitzen dutena, guk identifikatzen ohi ditugun emozioak leku mugatua eta zehatzak dauzkaten etengabeko distribuzio honen barruan.

Laburbilduz, sistema automatiko batez ari garela, jakin beharra dago emozio kontzienteak eta inkontzienteak daudena, eta erabaki, lortu nahi dugun sistemak mota bi ala bakarra bereiztuko duen (kontuan hartuz emozio inkontzienteak gizakiontzako ere detektagaitzak gertatzen direla, eta zer esanik ez makina batentzat, besterik ez bada sistemak behar dituen datuak sortzeko zailtasunagatik). Beste alde batetik, emozioak modu diskretu edo etengabean identifika daitezkeela jakinda, erabaki behar da bietarik zein ezagutuko duen diseinatuko den sistema automatikoak.

2.2. Zeintzuk dira emozio bat komunikatzeko behar diren datuak?

Emozioak modu anitzetan komunikatzen dira. Ikusmenaz, entzumenaz ala ukimenaz adieraz daiteke. Entzumenaz, esate baterako ahotsaren bitartez (lan honen helburua hain zuzen) edo musikaren bitartez (normalean doinu maiorra emozio tristeekin lotzen da). Ikusmenari dagokionez, aurpegiak, keinuek eta gorputzaren mugimenduek komunikazio emozionalaren berri ematen digute. Ukimenari dagokionez, gisa honetako kontaktu fisikoek ere esangura handia dute.

Ikerketek ez datoz bat emozio bat komunikatzeko faktore guztiak batera ageri behar diren ala bakarrarekin nahikoa den esateko orduan. Frijdak [6] egoera emozionalen komunikazioa osotasuna gisa definitzen duen bitartean, beste ikerketa askotan pertsonek emozioak datu multimodal guztiek jakin gabe ezagutzea lortu izan da [16] [12] [7], ezagutze mailak %100ekoak ez diren arren.

2.3. Zer egin behar da datu horiek pilatzeko?

Sistema batek ikasi ahal izateko oinarri-datu batzuk behar direnez gero, komunikazio emozionaleko datu base bat bildu behar da. Gaur egun hala ere, datuak jasotzeko aukerak mugatuak dira. Adibidez, aurpegiaren adierazpena edo entzunezko estimuluak hartzea nahiko erraza da, baina gorputz-lengoaia zaila da kamera baten bidez jasotzeko, eta ukimenezko kontaktu emotiboa kodetzea ere oso lan nekeza da. Beraz, hasieratik bertatik mugapenak ditugu: sistemari ezin dizkiogu komunikazio emozionalaren elementu erakusgarri guztiak eskaini.

Baina hala eta guztiz ere datu-base emozionalak diseinatzean agertzen den arazorik larriena ez da hori. Galdera da: nola sentiarazi ahal zaio pertsona bati emozio zehatz bat hain zuzen sistema automatikoaren diseinugilea erregistratu behar duen momentuan? Hiru irtenbide plazaratu dira: behar den emozioa antzeztuko duten aktoreak erabili, berriemaileek emozio hori sentiaraziko dion egoeraren bat gogoraraztea, edo arazoari beste alderdi batetik heldu: emozioa, jaso nahi den momentuan eragin beharrean, gertatzen den unean jaso.

2.3.1. Aktoreak

Printzipioz, gaitasun dramatikoa duen aktore bat bada ikusleari emozioak komunikatzeko gauza, baita benetakoak izango ez diren emozioak ere. Euren karga dramatiko propioa ezkutatu eta beste bat (ezberdina ala berdina) igortzeko gai diren profesionalak dira.

Hau dela eta, karga emozionala duen datu-base bat erregistratu nahi bada, lehen hurbilketa batean helburua lortzeko nahikoa da eskatutako emozioak antzeztuko dituen aktore bat kontratatzea.

Non dago arazoa, bada? Zenbait egileren arabera antzeztutako emozio horiek ez daukate benetako emozioen sakontasuna eta konplexutasuna, eta ez dira, beraz, emozio-ezagutze sistemetarako baliagarriak. Are gehiago, baliteke antzeztutako emozioa neurriz gainekoa izatea, aktoreak ulergarri bihurtzeko egiten duen ahalegina dela eta. Beraz, beren beregi antzeztutako emozioen bidez garatzen den sistema batek baliozkotasuna gal dezake emozioak berez sentitzen dute jendearekin erabiliz gero.

2.3.2. Oroitzapena

Taktika honetan berriemailea egoeran kokatu egiten da, eduki emotibo handia duten esaldiak aurretiaz irakurriz [16], edota berriemaileak lortu nahi den egoera emozionalera erakarriz, iraganeko egoerak gogoraraziz [11]. Taktika honen aldaera bat honako hau dugu: datu basean sartuko diren aktoreen antzezpenak erabili, berriemaileak imitatu ditzan, edota hortik abiatuz bere bizipen propioak gogoratu [13] [15] [17].

Taktika honen bi arazorik larrienak pertsonek gogoratzeko duten gaitasun ezberdinek sortuak dira: gogora ekartzeko zailtasuna handia dutenek ezin dute emozio sinesgarri edo benetakoak sortu, eta beste muturrean ordea, gogora ekartzeko gaitasun handia dutenak baliteke larregi hunkitzea, eta datu baliozkoak emateko gauza ez izatea [11].

2.3.3. Benetako grabaketak

Benetako emozio komunikazio baten lagin bat lortzeko aukera dagoenean, argi dago beste edozein antzezpen baino hobea dela, edo benetakoagoa behintzat. Arazoa da gaitza dela emoziozko erantzunak noiz gertatuko diren jakitea, eta, oro har, muturreko prozedurak baliatu behar dira horretarako, oso etikoak izaten ez direnak (intimitatean sartzea, berriemaileak egoera arriskutsuetan ezarri...) [20].

3. Ebazpideak: Artearen egoera

Hirugarren atal honetan artearen egoeraren berri emango da, emozioen ezagutzak dituen arazoei egile ezberdinek ematen dizkieten irtenbideak oinarri harturik. Aztertutako bibliografian antzeko metodologia bat biltzen da emozioak detektatzearen auzirako. Honako lau arlo hauek definitu behar dira emaitzak lortzeko:

- Ezagutzarako garrantzizkoak diren emozioak definitzea.
- Erabiliko den datu-basea zehaztea.
- Datu-basearen garrantzizko parametroen hautaketa.
- Detektatze metodoaren definizioa.

Jarraian fase hauek banan-banan komentatuko dira bibliografian agertzen diren ikerketa bakoitzerako.

Kontuan hartuko diren ikerketen artean, gizakiek emozioak ezagutzeko gaitasunaren inguruko batzuk ere izango dira. Honekin, besteak beste, helburu izan dezakegun ezagutzarik gorena zein den zehaztu gura dugu.

Behin egoera azaldutako lau atal garrantzitsutan azterturik, lanetan erdietsitako emaitzak komentatzen dira, eta baita atal bakoitzaren puntu sendoak komentatu ere.

3.1. Ezagutzarako erabiliko diren emozioen definizioa

Zenbait ikerketa taldek emozioak detektatzera begira dauden lanak egin izan dituzte, nahiz eta guztiek bat etorri ez diren emandako ikuspuntuan. Ondoren, bibliografiako lanen helburuak azaltzen dira, hasi errazenetik eta landuenetara heldu arte.

Badira asmo handirik gabeko ikerketa batzuk. Hauetan aztertutako ahots seinalea emoziozkoa ala neutrala den baino ez da argitu nahi, [3]-an esate baterako modu neutrala ala emozionala den jakin nahi da, edo [20]an hizlaria presiopean dagoen ala ez, hau da, estresa eragiten dion egoeran batean dagoen.

Hurrengo pausua ere detekzio bitarra da, baina emozio gehiago agertzen da. Alegia, emozio ezberdin batzuei dagozkien datuak hartzen dira oinarri, baina ikerketaren xedea emozio bakarra ala emozio multzoa gertatzen den argitzea baino ez da. [1]-ren kasua da: 6 modu ezberdinetatik abiatuz (Neutral, Annoyed, Frustrated, Tired, Amused, Other; Neutrala, Mindua, Frustratua, Nekatua, Jostalaria, Bestelakoak), baina azken xedea hizlaria mindua ala haserre egoteko arrazoirik dagoen argitzea da berez. [8], [9] eta [10]-aren lanetan, beste aldetik, antzera egiten da, eta emozio sorta handi batetik karga "negatiboa" dutenak aurkitu nahi da.

Jarraian, ahots emozionaleko seinaleak "giza-dekodifikatzaileak"[1] erabiliz sailkatzen dituzten sistemak hartuko ditugu aintzat. [16], [12] eta [7]-aren kasuak dira. Oro har, emozio lagin zabalagoak erabiltzen dira, eta gizaki batek egindako sailkatzea dela medio, emozioa eragiten duten parametro-tamainaz bilatzen dira. [7]-aren kasua bereziki garrantzizkoa da, izan ere, parametroetariko batzuk artifizialki moldatzen dira, dekodifikatzaileei birsintetizatutako seinaleak eskaintzeko. Seinale hauek beste benetako seinaleetatik datoz baina parametroetakoren bat aldaturik, igarritako emozioan eragiten den efektua identifikatzeko.

Azkenik, emozio anitzak detektatzeko sistema automatikoak ditugu. Sistema hauetan ezagutu beharreko emozio kopuru mugatua planteatzen da, baina banan-banan ezagutu nahi izaten da: kontua ez da, bada, emozio zehatz bat beste guztien aurrean identifikatzea, baizik eta edozein ahots laginetan aztergai den edozein emozio ezagutuko duen sistema automatikoa garatzea. [11]-ren kasua da, esate baterako. Lan honetan bost emozio modu hartzen dira kontuan (Afraid, Happy, Neutral, Sad, Angry; Beldurra, Poza, Neutrala, Tristura, Haserrea), edota R. Nakatsu, J. Nicholson eta N. Tosaren ikerketa lanak [13] [15] [17]. Lanotan 8 emozio hartzen dira aintzat ezagupenerako: Joy, Teasing, Fear, Sadness, Disgust, Anger, Surprise, Neutral; Alaitasuna, Iseka, Beldurra, Tristura, Nazka, Haserrea, Harridura, Neutrala.

3.2. Erabiliko den datu-basearen definizioa

Azaldu den bezalaxe, kontuan harturiko lan guztiek datu base bat daukate oinarrian. Hala ere, ikerlariek erabilitako datu-baseak oso anitzak izan dira: esperimentu ezberdinak izan dira bai helburuei dagokienez eta eskainitako emaitzei dagokienez ere, eta hortaz, datu-baseak euren artean ez dute zerikusirik izan. Egileak oro har jasotako emaitzekin pozik agertu diren arren, ematen du datu-base hobeak, edo askotan zabalagoak, erabiliz emaitzek ere hobera egingo dutela.

Jarraian egindako ikerketetan erabilitako datu-baseak azaltzen dira:

- [11]-n, 40 hiztuneko datu-basea erabiltzen da, 20 gizonezko eta 20 emakumezko. Hiztunok betebehar batzuk zeuzkaten: ezelako prestakuntza dramatikorik ez izatea, ahots edo entzumen arazorik ez izatea (ezta gripea edo katarroa bezalako gaixotasun arruntak ere), ezta arazo mentalak edo ikasteko zaitasunik ere. Honetaz gain adin-muga batzuk ere ezartzen ziren (18 eta 69 urte bitartean). Berriemaileek emozioak modu sinesgarrian sorzitzaten lortzeko, datu-basera pasatuko zen pasartea irakurri baino lehen, eurengan desiratutako egoera emozionala eragiten zuten egoerak ekartzen ziren gogora[2]. Grabazioak mikrofono profesionala erabiliz egin ziren. Guztira, 7-8 esaldiko 197 paragrafo erdietsi ziren.
- [13], [15] eta [17] ikerlanetan Japoniako hiztunak erabili ziren datu-basea egiteko, 50 gizonezko eta 50 emakumezko hain zuzen. Kasu honetan ere berriemaileek ez dute ezelako prestakuntza dramatikorik. Grabazioak berbak baino ez dira, ez dago esaldirik. Honako hau daukagu, beraz:

[1] Adierazpena [16]tik dator.

[2] Sistema hau hizketa emozionala grabatzeko oso teknika hedatua da, baina, nahiz eta lortutako ahots-laginak ontzat eman arren, badu zenbait arazo. Ikerlari honen kasuan, saiakera batzuetan gogora ekartze hori zela eta ezin izaten zen ahots segmentuaren grabazioa egin: familia egoera triste bat ekartzen zen gogora, eduki tristeko ahots emozionala lortzeko, hala nola senide bat hiltzea. Berriemaile batzuk ezin izan zioten egoerari buru egin, eta ez zuten ezer grabatu nahi izan.

100 hiztun x 100 berba x 8 emozio([3]) = 80000 erregistro

Hizketa emozionala lortzeko, lan honetan irratiko aktore profesional batek 100 berbak grabatu zituen. Bigarren fasean, berriemaileak berba emozio zehatz batekin grabatzera doazenean, aktorearen grabazioa entzun dezakete, hitz zehatz baten intonazio emozionalaren erreferentzia bat izan dezaten

- [8], [9] eta [10]-ean ez da antzeztutako emozioekin lan egiten, benetako emozioekin baino. Call Center batera egindako deiak grabatu dira (8KHz-era egindako grabazioak dira, beraz, μ legearen pekoak). Kasu honetan ez da izan emozioa nola lortu, benetako baita, emozio nola etiketatu baizik. Horretarako hiru etiketatzaile erabili dira, gizakiak, eta grabazioak arakatu dituzte karga emozional negatiboa duten zatiak aurkitzeko[4]. Antzemate emozionala gizakiongan subjektiboa denez gero, hiru etiketatzaileak ez dira beti bat etorri. Hau dela eta, datu fidagarriagoak edukitzearren, etiketatzaile guztiak ados jarri diren grabazioka baino ez dira kontuan hartu datu-baserako. Emakume ahotseko 665 grabazio lortu dira, 133 karga emozional negatibodunak, eta gizon ahotseko 514 grabazio, 122k karga emozional negatiboa dutela.
- [1]-an, berriz ere, datu-basea Call Centerretera egindako deiek osatzen dute (8KHz-ra egindakoak, beraz). Kasu honetan deiak ez dira zeharo benetakoak, ez baitziren zerbitzu bila zebiltzan benetako erabiltzaileak, baizik eta oporrak erreserbatzen ari zirela antzezten zuten boluntarioak (Egileak berak aitortzen du hau dela eta haserretasun maila benetako zerbitzu batean lortuko zirenak baino apalagoak direla). Modu honetan 21899 grabazio lortu dira. Kasu honetan deiak egin ahala etiketatuz joan behar izan dira, aztertu beharreko emozio motak aurkitzearren. 5 giza-etiketatzaileek 49553 anotazio egin zituzten. Honetan ere, etiketatzaileak ez dira ahots zati guztiak juzkatzean bat etorri, eta beraz kontrolatzeko mekanismoa beharrezkoa da. Etiketatzaileak zenbateraino etorri diren bat zehazten duen adostasun parametro bat baliatu da.
- [3]-n, *Wizard of Oz*[5] sistemaren bidez grabatutako 20 telefono-elkarrizketa daude datu-basean, , 2395 zati guztira. Egileak proiektua garatzen ari dira, eta 75 elkarrizketa grabatuta eta anotatuta izateko asmoa daukate. Hizketa emozionala erdiesteko, sistemak, erabiltzailea gisa honetan erreakzionarazten dituen egoeretara eramaten du.
- [16] Sistemako datu-basea grabatzeko, 18 eta 35 urte bitarteko 35 emakume hartu ziren (egia esan 31 izan ziren, 4 alde batera utzi behar izan zirelako arazo fisiologiko edo psikologikoengatik). Desiratutako emozioa duen hizketa emozionala lortzeko honako prozesu hau garatu zen: bukaera berdina baina eduki eta testuinguru emozionala zuten istorioak idatzi ziren, eta emakumeei irakurrarazi zieten. Guztira, datu-basean karga emozionala zuten 620 esaldi eta beste 155 neutroak lortu eskuratu ziren, kromo audio analogikoko zintetan bilduak.
- Mozziconaccik bere ikerketa lanean erabiltzen duen datu-basean [13] 936 esaldi daude: 3 holandar hiztunek (2 gizonezko eta emakume bat) 8 esaldi irakurri zituzten 3 aldiz, 13 emozio modu ezberdin erabiliz. Pilatzea digitalki egin zen, 10KHz-ko laginketa frekuentzia bat erabiliz eta 16 biteko laginak. Emozioa lortzeko teknika gogora ekartzea izan zen: esaldi zehatzak ahoskatu behar zituzten, emozio zehatz horretakoak, eta gero, behin emozio horretaz jantzi ondoren, semantikoki neutroak diren esaldiak irakurri.
- Banse and Shearer-en [7] esperimentuetarako aktoreak erabili ziren. Eduki emozionaleko esaldiak sortzen zituzten, 12 emozio bakoitzeko, guztira 244 adierazpide.
- [20] ikerketa lanean SUSAS (Speech Under Simulated and Actual Stress) datu-basea erabili zen. Datu-base honetan hizketa naturaleko grabazioak daude, erakarpen parkeetan, hegazkinetako kabinetan edo psikiatria kontsultetan egindakoak, eta baita antzeztutako emozioka ere. Guztira 43 berriemaile erregistratu dira datu-basean, eta 6 berba ezberdin erabiliko dira bakoitzeko, 3 estilotan.

3.3. Datu-baseko parametro garrantzitsuen hautaketa

Erabiliko den datu-basea definitu eta landu ostean, esperimentuaren planteamenduan bilatzen zen sailkapena egitea da hurrengo pausua. Baina horretarako, sailkapen sistema datuez hornitu beharra dago, ahots-seinalekoan hain zuzen ere. Giza dekodifikazioa erabiltzen duten sistemetan ez da aurretiaz sarrera parametroen multzoa definitu behar, izan ere, dekodifikatzaileek grabazioka entzun behar dituzte eta.

[3] 8 emozio ezberdinak aurreko atalean azaldu dira.

[4] Aurreko atalean azaldu den bezala, egile hauek ahotsaren ahots karga negatiboa baino ez dute aurkitu nahi.

[5] *Wizard of Oz* (WOZ) sistema batean gizakiak betetako partea bat dago, eta beste parte automatiko bat (normalean sistemaren jokaera gizakiak erabakitzen du eta aurkezpena automatikoa da). Erabiltzaileak ezin ditu gizakiaren parteak ikusi, sistema automatiko hutsaren aurrean dagoela somatuz.

Jarraian sailkapen automatikoko ikerketetan sailkatzaileka hornitzeko erabili diren parametroak azaltzen dira:

- Parametroak [11] ASSESS (Automatic Statistical Summary of Elementary Speech Structures) sistemaren bitartez erauzten dira. Sistema honek ahots seinalearen neurketa-gama automatikoki sortzen du. Guztira 375 neurketa erabiltzeko aukera ematen du, baina horietatik 32 baino ez dira lan honetarako baliatu:
    - Tonuei dagozkien neurketak: tonuaren iraupena, funtzio koadratiko batekin tonuak duen harremana, tonuko pitch-kurbak duen inflexio kopurua.
    - Espektro-neurketak: 250Hz beherako energia.
    - Intentsitate kurbaren neurketak[6]: 3 datu.
    - Intentsitate-kurbaren mutur lokalen neurketak: 4 datu.
    - Intentsitate-kurbaren handitze eta erortzeen magnitudea: 2 datu.
    - Pitch-kurbaren eskuratutako puntuak: puntu kopurua, batezbesteko pitch-a, 'inter quartile range'.
    - Pitch-kurbaren mutur lokalen gaineko neurketak: 3 datu.
    - Pitch-kurbaren handitzeen magnitudea: 2 datu.
    - Intentsitate-kurbaren handitze eta erorketen iraupena: 2 datu.
    - Intentsitate-kurbaren tarte lauen iraupena: 4 datu.
    - Pitch-kurbaren ezaugarrien iraupena: 5 datu.

  Datu bilduma honen gainean bariantzaren analisia egin zen, eta lar ezberdinak ziren datuak ez ziren ezaugarritzat jo, eta hortaz, azken analisitik kanpo utzi ziren.

  Egileak akats bat egiten duela nabarmendu behar da: erabiltzen zuen ASSESS sistemaren bertsioak sarrera datuetan zarata sartzen zuen bug bat zeukan, ondoko bertsioetan arazo hori konpondu den arren.

- R. Nakatsu, J. Nicholson eta N. Tosaren ikerketa lanetan [13] [15] [17], erauzten diren ahots seinaleen ezaugarriak ez dira soilik prosodikoak: parametro fonetikoak ere hartzen dira kontuan. Fase bi dago datuak biltzeko prozesuan: lehenengo eta behin landu beharreko esaldiaren hasiera eta amaiera puntuak detektatzen dira, energia-muga erabiliz (hau da, esaldia energia-muga gainditzen denean hasten dela suposatuko da, eta azpira itzultzen denean amaitzen dela). Orduan, detektatutako esaldi hori 20 bitarte distantziakideetan banatu eta puntu bakoitzeko datu guztiak kalkulatzen dira. Datuok honako hauek dira: alde batetik prosodikoak, pitch eta energiaren balioak, eta bestetik fonetikoak, LPC 12 parametroak eta delta LPC parametroa hain zuzen ere. 20 sarrera puntuetarako hamabosna parametro sortzen dira beraz. Baliook, 300 kokapena bektore baten gisara antolaturik, sailkatze-sistemaren sarrera moduan erabiltzen dira.

- [8], [9] eta [10]-ean, erabilitako ahots-seinaleari dagozkien datuak pitch eta energiaren gainekoak dira: batezbestekoa, mediana, maximoa, minimoa eta heina. Pitch-kurbaren neurketak dauzkaten datu guztiek, kurbaren gune ahostunak baino ez dituzte kontuan izaten. Emozio negatiboak ezagutzeko ez dira soilik ahots-seinalearen datuak kontuan hartzen, informazio semantikoa ere erabiltzen da, izan ere, karga emozional zehatz baterako joera handiagoa duten berbak badaude. Hau dela eta ahots-seinaleaz baino egiten ez den ikerketa hobetu egiten da.

- Sarrera-daturik garrantzitsuenak [1] prosodikoak dira:
    - Bokaleen iraupena (gehienezkoa eta batezbestekoa)[7].
    - Ahoskatze erritmoa: segmentuaren iraupenaren arteko bokale kopurua.
    - Etenei buruzko datuak: eten kopurua, gehienezo iraupena, eten denbora eta hizketa denboraren arteko harremana.
    - Pitch: minimoa, maximoa, burutu gabekoak (neurketak estilizatutako kurben gainean egiten dira, eta osoa informazio erabilgarria ematen dute).
    - Energia: RMS, zati ahostunetan.
    - Espektruak: Lehen Cepstrum koefizientearen batezbestekoa, maiztasun altu eta baxuen arteko batezbestekoa aldea.

  Honetaz gain, beste datu ez prosodikoak erabiltzen dira: esaldiak elkarrizketan duen

[6] Intentsitate-kurbari dagozkien neurketa guztiak etenak kontuan hartu barik egin dira.

[7] Sistema hau hizketaren ezagutze automatikoko beste batekin batera erabiltzen da. Horretatik, bokalen hasiera eta amaierako datuak erauzten dira bere iraupena bertan erabili ahala izateko.

kokapena, eta birsaiakerak eta zuzenketen inguruko datuak.

• [3]-an, sailkatzailerako sarrera-datu gisa 27 ezaugarri prosodiko dituen bektore bat erabiltzen da, bai pitch, energia eta iraupenen ingurukoak. Datu hauek guztiak zuzenean lortzen dira ahots-seinaletik, ez da ezelako informazio semantikorik erabiltzen.

• [20]-an, TEO (*Teager Energy Operator*) operatzailearen gainean egindako neurri multzo bat erabiltzen da sarrera-datu gisa. Operatzaile hau ahots-seinalea eta bere eratorrietatik definitzen da, eta seinalea maiztasunean modelatutako zatian eta anplitudean modelatutako zatian banatzen du. Zati birekin kalkulatzen dira datuak, horretarako *TEO_Var_FM* (maiztasunean modulatutako zatiari dagokiona) eta *TEO_Auto_Env* eta *TEO_CB_Auto_Env* (anplitudean modulatutako zatiari dagozkionak) balioak datu modura erabiliz.

3.4. Sailkapen metodoaren definizioa

Ahots-seinalea besterik erabiltzen ez duen emozioak ezagutzeko sistema bat, intonazio patroiak sailkatzeko sistema izango da, besterik ez. Aurreko ataletan ze patroi motak eta ze irizpiderekin deskribatu ondoren, sistemak eurak deskribatzeko unea dugu.

- [11]-an sailkapen estatistikoaren aldeko hautua egiten da, 5 metodo ezberdinen bitartez. Gero bakoitzaren emaitzak erkatu egingo dira:
  - Lineala: patroi espazioak sortzen dira eta gero plan linealak erabiliz banatzen dira.
  - SVM[8] antzekotasun linealarekin.
  - SVM Gauss antzekotasunarekin.
  - GVQ[9] ezaugarri batean oinarriturik.
  - GVQ ezaugarri bitan oinarriturik.
- [13] [15] eta [17]-ren sisteman, ostera, sare neuronaletan oinarritutako sailkapen sistema nahiago izan da. OCON sistema erabiltzen da (One Class in One Network); honek esan nahi du 8 azpisare erabili behar dira, eta bakoitzak emozio bana detektatuko du. Azpisareak independenteki molda daiteke funtzionamendu hobea edo zehatzagoa lortzearren. dauka Azpisare bakoitzak honako hau dauka: 300 neuronako sarrera-geruza bat (gogora dezagun sarrera-datuak 300 elementuko bektoreetan daudela), neurona bakarreko irteera-geruza bat (honek esango digu sarrera-patroia azpisareak detektatu behar duen emozioari dagokion ala ez), eta tarteko beste geruza bi, tipologia ezberdin bi izan dezaketenak: 16 neurona lehenengo tarteko geruzak eta 4 bigarrenak, edo 32 neurona lehen tarteko geruzan eta 8 bigarrenean. Azpisarea egokitzeko prozesuan erabakiko da bietako zein moldatzen den ondoen, zatizko emaitzen arabera.
- Ze emozio dagokion emandako esaldi bati azkenik erabakitzeko, 8 azpisareen irteerekin ezartzen da erabakitze-logika bat.
- Leek, Narayananek, eta Pieraccinik, [8], [9] eta [10]-ean, sailkapen estatistikoko hiru sistema ezberdin erabiltzen dituzte. LDC (Gauss-en probabilitate banaketa onartzen duen sailkapen lineala), k-NN (zein patroi dituen hurbileko kide gehien gogoratzen duen sailkapen bat) eta SVM-ak.
- [1]-ean erabiltzen den sistema ere estatistikoa da, baina oraingoan erabakitze-zuhaitzetan oinarriturik, parametro beharrezkoak aurkitu ziren indarreko sistema errepikatzaile baten bitartez.
- [3] ikerketa lanean, emaitza positiboak izan zituzten zenbait metodo azaltzen da: horietariko bat sare neuronaletan oinarriturik (MLP motako sare bat erabiltzen dute, hain zuzen ere, Multi Layer Perceptron), eta beste bi mota estatistikokoak, bata lineala eta bestea CART[10] eta erregresio-zuhaitzak erabiltzen dituena.
- [20]-an sailkapena 5 egoerako Markov-en ezkutuko modeloen bidez egiten da.

Nabarmendu behar da egindako deskribapenean giza-ezagutzan oinarritutako sistemak alde batera utzi direla, sailkapen metodoa beti ere berdina delako: pertsona batzuek grabaketak entzun eta horiei buruzko erabakiak hartzen dituzte.

3.5. Lortutako emaitzak

Jarraian egileek euren ikerketa lanetan nabarmentzen dituzten emaitzak azaltzen dira.

- [11]-an deskribatutako sistemaren lorpenik garrantzitsuenak honako hauek dira: ezagutzarako garrantzizkoak diren parametroen identifikatzea, eta ezagutza zuzenaren tasa 50% izatea. Emaitzok lortu ahala izateko bere datu-basearen %90 erabili dute sistema entrenatzeko, eta %10 probak egiteko. Ez da alde handirik antzematen

[8] *Support Vector Machine*, Bektore Euskarriko Makinak. Ikasten duten makinak dira; sare neuronalen aldaera bat.

[9] *Generative Vector Quantization*, KUNek garatutako sistema, patroiak sailkatzeko beste esperimentu batzuetan oso emaitza onak lortu ditu, hala nola eskuz idatzitako digitoen ezagutza.

[10] *Classification and Regression Trees:* Sailkapen eta Erregresio Zuhaitzak.

sailkatze metodoen artean. Besteak beste, ezagutze maila hobetzeko, datu-base handiagoa erabiltzea proposatzen dute.

- [13], [15] eta [17]-ren lanetarako aurkeztu diren emaitzetan esperimentu mota bi dago: irekia eta itxia izenez ezagutzen direnak. Esperimentu itxiari dagokionez, eskura dauden datu guztiak erabiltzen dira sailkatze-sistema entrenatzeko, eta testa bera ere datu horiekin egiten da. Lokutore gutxi dagoenean emaitzak ematen ditu esperimentu honek, baina ezagutza zuzenen mailak behera egiten du kontuan hartzen den lokutore kopuruak gora egiten duen heinean. Esperimentu irekiari dagokionez, hiztun ezberdinen datuak erabiltzen dira esperimentu zein testerako. Esperimentu honetan, lokutore gutxirekin oso ezagutza tasa apala lortzen da, %20 gutxi gora behera, baina hobera egiten du nabarmen kontuan hartutako lokutore kopuruak gora egiten duen heinean, eta hortaz, 30 lokutore baino gehiagorekin esperimentu ireki eta itxien emaitzek bat egiten dute %50-reko tasa batean.
- [8], [9] eta [10]-ean deskribatutako esperimentuek, emozio negatiboen ezagutzan %30 inguruko tasa bat lortzen dute.
- Berariaz haserrea eta frustrazioa gainontzekoetatik bereizteko lanean, [1]-ean %70 inguruko arrakasta tasa erdietsi da.
- [3]-an deskribatutako lana sistema zabalago baten zatia da. Soilik prosodia erabiliz hizketa emozionalaren ezagutzarako ematen duten emaitza %56koa da.
- [20]-an, estresa hizketan antzemateko arrakasta tasa %80koa izan da. Ingelesaren hizketa ezagutzarako sistema oso baten barnean dago lan hau.
- Aparteko aipamena behar dute giza dekodifikazioaren bidez lortutako emaitzek, gizakiek egindako detektatzea automatikoarentzat erreferentzia izango baita. [12]-an %63ko arrakasta tasa lortzen da.

3.6. Artearen egoera: ondorioak

Emozioen ezagutza automatikorako eman diren irtenbideen inguruko ikerketa honetan honako hau aurkitu dugu:

- Hainbat lan eremu eta arlotan dauden hainbat ikerketa daude, oro har, gehien agertzen dena emozio anitzen ezagutza generikoa da.

- Lan guztietan antzeko metodologia erabiltzen da: ezagutu beharreko emozioen definizioa, datu-basearen diseinua, patroiak sailkatzeko sistemaren garapena.

- Emozioak modu orokorrean ezagutzeko egindako sistemek (hau da, ez bitarra) ez dute balentziaren edo intentsitatearen etengabeko detektatzea egiten, ez bada emozio gutxi batzuen ezagutza diskretua.

- Sistema guztietako datu-baseetan ehunka eta milaka elementu arteko kopurua dago. Oro har, ikerlariek onartzen dute datu-base handiagoekin emaitza hobeak lortzen direla.

- Erabili beharreko datuak ez dira beti prosodiko hutsak. Prosodikoen artean, energia eta pitch-kurba dira parametrorik erabilienak.

- Sailkatze metodorik erabilienak mota ezberdinetako estatistikoak dira, eta ondoren sare neuronalak. Behin baino ez da antzeman HMM erabili izana.

- Emozio orokorren ezagutzarako asmatze-tasa %50 ingurukoa da, eta emozio ezagutza bitarrerako %70 ingurukoa, ordea. Giza-dekodifikatzailea erabiliz emozio orokorrekin lortutako emaitza batzuek %70ko arrakasta tasa erakusten dute. Aipatzeko da giza-dekodifikatzaileek informazio asko ezaugarri supresegmentaleetatik jasotzen dutela.

4. Proposatutako esperimentua

Bibliografian aztertutako lanetan hartutako esperientzia abiapuntutzat harturik, soilik parametro prosodikoetan (intonazioa bereziki) oinarritutako emozioen ezagutzako esperimentu bat planteatu nahi da. Horretarako, eta artearen egoeran antzeman den lau parteko banaketa argia kontuan hartuz, lau kontu horietarako esperimentu proposamen bana zehazten da ondoren.

4.1. Ezagutzarako erabiliko diren emozioen definizioa

Mota orokorreko emozioen detektatze automatikoa egin nahi dugu, hau da, ez bitarra, [11], [13], [15] eta [17]-an azaltzen diren modukoa, eta horietan egiten den bezala, emozio kopuru mugatu bat modu diskretuan antzeman. Emozio multzo mugatu hori zein izango den zehaztu egin beharra dago, eta horretarako aurkitutako bibliografian usuen ageri direnak bilatuko ditugu. Horiek guztietatik MPEG4 [21] estandarrak definitzen dituen 6 emozio arketipo hartuko ditugu:

- Anger (Haserrea)
- Happiness (Poza)
- Surprise (Harridura)
- Sadness (Tristura)
- Disgust (Nazka)
- Fear (Beldurra)
- Neutral (Neutrala)

Ikerketan emozio kontzienteak identifikatzea bilatzen da, inkontzienteak, giza dekodifikatzaileentzat ere antzemangaitzak diren neurrian, datuak biltzeko eta anotatzeko orduan arazoak ematen dituztelako.

4.2. Erabiliko den datu-basearen definizioa

Planteatu den esperimenturako, datu-base bat erabili behar da, ondoren sailkapen sistema entrenatu ahal izateko. Datu-base honetan eskatutako emozio bakoitzaren datu kopuru beharrezko bat izan behar da, zertarako? Bada sailkapen sistemak erabakiak hartzeko datu nahikoak izan ditzan, eta hala bada, garrantzizko parametroak zuzen identifikatzeko. Honetaz gain, lokutore anitzeko datu-basea izatea desiragarria da, emozioak ezagutzeko sistema orokor batek emoziook hainbat modutan adierazten dituzten pertsona mota ezberdinekin jokatu behar izango duelako.

Intonazio patroien bidezko detektatze sistemetara lehenengo hurbilketa baterako [19]-an deskribatzen den hizketa emozionaleko datu-basea erabiltzea proposatzen da, Euskal Herriko Unibertsitatearen Elektronika eta Telekomunikazio Sailean burututakoa. Datu-base multimodala da, behar ziren emozioak antzezten dituen aktore batek egina, ahots seinalean zein bideotan. Lan honetarako audioa baino ez da erabili behar izango. Datu-base honetan eduki mota bi dago: emozio guztietarako komunak diren 97 esaldi daude, hau da, eduki semantikoa berbera da baina 7 modu ezberdinetan irakurtzen da, neutroa barne. 97 esaldi hauetaz gain, emozio bakoitzerako, eduki semantikotik hurbilago zeuden beste 75 grabatu dira. Beraz guztira ondorengoa dago:

(97 finko + 75 emozioaren menpe) x 7 emozio = 1204 erregistro.

Datu-basea euskara batuan grabatu da. Laringrogafo baten bidez grabatu zen, eta hortaz, ahots seinalearekin batera eta lerrokaturik, ahots-korden oinarrizko maiztasuna lortzeko aukera ematen digun seinale bat daukagu eskura.

Datu-base honen helburua modelo ez neutroetan sintesia egiteko emozioei buruzko informazioa jasotzea izan zen. Hau dela eta, oso landua dago: fonema mailan anotatuta dago, eta laringografoak emandako intonazio-kurbak hobetu egin dira, berrikusita, araztuta eta parametrizatuta daude.

Datu-base hau lehenengo hurbilketa gisa baliatuko dugu, eskutan dugulako eta lantze prozesua oso aurreraturik dagoelako. Hala ere, ezin ahantzi dezakegu datu-base berria grabatzea beharrezkoa izango dela, honako mugapen hauek direla eta:

- Aipatutako datu-basea berriemaile bakarrarekin grabatu da. Honek esan nahi du, kasurik eta onenean, berau oinarri hartzen duen sistema automatiko bat egindako emozioen errepresentaziora ezin hobeto moldatzeko gai izango dela, baina ez gainontzeko hiztun guztien prosodia emozionalera. Hau garrantzizko mugapena da, izan ere [13] [15] eta [17]-ren egileek lokutore gutxi erabilita egindako esperimentu ireki batean %20 inguruko arrakasta tasa baino ez zuten erdietsi.

- Hizketa emozionala lortzeko, aktore batez baliatu gara. Honek egile batzuen arabera arazo bi dakartza: lehenengo eta behin, sortuko dituen emozioak ez direla “benetakoak” izango [4], eta estereotipoei lotuta egongo direla; beraz, ez dute izango benetako emozioen aberastasuna eta konplexutasuna. [16]. Hortaz, datuokin elikatzen den sistema bat, a priori, ez da antzeztutakoak ez diren emozioak ezagutzeko gai. Bigarrenez, aktoreak igorriko duen emozioa ulergarria gerta dadin nahi duenez gero, gehiegizko antzezpena egin lezake, eta hau dela eta, ezagutza orokorrean baino zehatzagoa izan daiteke, benetako emozioak ezagutzen ahaleginduko den sistema batean lortuko direnak baino emaitza hobeak lortuz [3].

Jatorriz sintesirako diseinatu zen datu-base hau erabilgarria izango zaigu ezagupenak pilatzeko emozioen ezagutza proiektuaren ondorengo faseetan. Arestian esan dugun bezala, ordea, emozioen ezagutzako esperimentu oso bat egin ahal izateko, datu-base berri bat bildu beharko da, aurrekoaren abantailak jaso eta haren akatsak gaindituko dituena.

- Gure sistema jende desberdinen emozioak ezagutzea nahi badugu, datu-basean derrigorrez lokutore anitzak egon behar dira. Lokutore kopuruari dagokionez, kontsultatutako bibliografian oso aldagarri gertatzen da, hasi [13]-ren 3 lokutoreetatik eta [8], [9] eta [10]-ren 665 lokutoreetara arte. Guk planteatu nahi dugun datu-basearekin antz handien duten datu-baseen esperimentuek, hau da, alde batetik [13], [15], [17], eta bestetik [11], 100 eta 40 lokutore erabiltzen dituzte hurrenez hurren. Hortaz, kopuruok xedetzat hartuko ditugu, kontuan hartuz beti ere berriemaileak lortzeko zailtasuna, batez ere grabazio sesioak oso luzeak diren kasuan. Zailtasun hau dela eta baliteke kopuru hori ez gainditzea.

- Lan honen bigarren atalean azaldutako arazoak kontuan hartuta, argi ikusten da hizketa emozionalaren benetako gertaerak jasotzea ez dela batere erraza. Aztertutako bibliografiaren artean, emaitzarik onenak lortu dituzten bi ikerketa lanek honako modu honetan egin dira: [11]-n lokutoreei emozioa gogora erakartzen zaie aurretik sortutako egoera baten bitartez, eta [13], [15] eta [17]-an, aktore baten hasierako grabazio bat oinarri hartuta, lokutoreek emozio hori antzeztu edo imitatu behar dute. Emozio pilaketa mota biek arazoak dauzkate: lehenak, kontu teknikoak alde batera utzita, ezbai etikoa dakarkigu, lokutoreak zenbait egoera emozionaletara erakartzea mingarria gerta daiteke eta. Bigarrenean, arazorik larriena zera da: benetako emozioak ez direla sortzen, antzeztutakoak baizik; printzipioz baztertzekoa ematen duen planteamendu hau gora-behera, erabiltzen duten egileek nahiko emaitza onargarriak lortzen dituzte. Beraz, erabiltzeko modukotzat jo dezakegu. Kasu honetan, erreferentzia

moduan erabiltzeko, aurretiaz aktore batek antzeztutako emozioa grabatuta izatea abantaila handia da.

- Grabatutako datu-basearen gainontzeko alde tekniko guztiez baliatu gaitezke: grabaketa ekipamendua adibidez, laringografoa izan daiteke[11]. Corpusari dagokionez ere, sintesirako datu-basean erabilitakoaz balia gaitezke, nahiz eta jatorrizkotik ateratako corpus murriztu bat erabiliz ere emaitza onargarririk lor daitekeen, azken batean oinarri hartzen ditugun ikerketa lanetakoak askoz txikiagoak dira. Honetaz gain, corpusaren tamaina murriztuz lokutore berriak lortzea errazagoa izango da. Grabaketaren hizkuntza euskara batua izango da.

Euskaraz grabatutako datu-basea erabiltzen dugunez, hizkuntza honetara mugatu beharko dugu ezagutza fasean. Edozein hizkuntza multzorako baliteke prosodiaren bidezko emozioen adierazpena asko ez aldatzea batetik bestera, baina ezin dugu suposatu hizkuntza baterako lortutako emaitzak beste baterako baliagarriak izango direnik.

Hizkuntza anitzeko ezagutza sendoa egin nahi izanez gero, hizkuntza anitzetako datu-basea izan beharko genuke oinarrian, bibliografian aurkitutako ikerketa lanen artean halakorik ez badago ere. Hortaz, arlo honetan aurrerapausorik ematea ez da oraindik planteatzen.

4.3. Datu-baseko parametro garrantzitsuen hautaketa

Hasieratik bertatik, lan honek ahots-seinalean oinarrituriko emozioen ezagutza duela helburu onartu da, ahots-seinalearen elementu prosodikoak baino ez erabiliz. Beraz, hauek izango dira sailkapen sisteman sartuko direnak. Bibliografian ikusi dugun bezala, prosodiaren barruan lau datu garrantzitsu daude: energia, abiadura, etenak eta, batez ere, oinarrizko maiztasunaren kurba.

4.3.1. Energia

Ezagutzari begira, energiaren kurbak garrantzi mugatua dauka. Zenbait emoziotan garrantzitsu gerta daiteke, baina hautatutako lan gehienek oso modu orokorrean baino ez dute aintzat hartzen: bataz besteko energia, desbideratze estandarra, heinak. 4 datu hauek sartuko dira, isiltasuna ez diren ahots zatiekin soilik kalkulatu behar direla kontuan hartuz, hau da, energiarako gutxieneko mugatik gora daudenak. Energia honela kalkulatzeak arazoak ekar ditzake, izan ere energia hiztun, ekipamendu edo sesio aldaketekin ere alda daitekeelako. Aurreneko hurbilketa bat egiteko erabiliko den datu-basean lokutore, sesio edo ekipamendu aldaketarik ez dagoen arren, grabatu nahi den lokutore anitzeko datu-basean:

- Badira lokutore aldaketak, eta energian eragina izan dezakete. Hala ere, honen eragina murrizteko metodoak ere badaude.

- Grabaketa guztiak ekipamendu berbera erabiliz egiteko asmoa dago: mikrofonoa eta laringografoa.

- Lokutore ezberdinen grabaketak sesio ezberdinetan gauzatuko dira, baina lokutore bakoitzeko sesio bakarra egingo da, eta beraz, sesioen arteko aldea lokutoreen artekoarekin parekatu daiteke, lokutore bakoitzarengan ez baitago sesio aldaketarik.

Lokutore anitzeko datu-base batekin ari garenez lanean, sistemak lokutoreen arteko ezberdintasunak sortzeko gai izan dadin da helburua; izan ere, emaitzarik onenak lortu dituzten ikerketa lanek lokutore anitzak erabiltzen zituzten, eta orokortzea lortu dute.

4.3.2. Abiadura

Erabili den lanean, esate-abiadura segmentuaren iraupena zati bokale kopuru moduan neurtzen zen. Aurreneko hurbilketa bat egiteko erabiltzen dugun sintesirako datu-basea fonema mailan etiketaturik dagoenez gero, hauetariko zeintzuk dagozkion bokale bati igarri daiteke, [1]-an erabiliko parametro berbera erabiliz. Etiketatzen hau lanabes automatikoak erabiliz egin zen. Lokutore anitzeko datu-base osoa grabatu ostean, arestian aipatutako lanabesak erabil daitezke, eta hortaz, datu berbera ere erabili. Hala ere, abiadura neurtzeko beste parametro egokiagoak ere kontuan artuko dira.

4.3.3. Etenaldiak

Aztertutako lanen batean ere [1], etenen iraupena eta kopurua ere sarrera-datu gisa erabiltzen dute. Lehenengo hurbilketarako erabiliko den datu-basean etenak ez daude markaturik. Lanabes automatikorik ere ez da sortu. Hortaz, datu hau eskura eduki ahal izateko, etiketatze automatikoko prozesu neketsu bat burutu beharko litzateke, edota horretarako lanabesak sortu.

Parametro honetaz baliatzen den ikerketa lana emaitzarik onenak lortzen dituena da; hala ere, lortzen dituztenek ez dute erabiltzen, eta pentsatzekoa da parametroa sartzeak anotatze lan neketsu hori justifikatzen duen garrantzizko datu gehigarririk ez duela emango

4.3.4. Pitch kurbak

Emozioak detektatzeko orduan prosodiaren elementurik garrantzitsuena dira. Arazoa da etengabeko kurba delako eta datu-baseko seinalea bakoitzaren denbora une ezberdinetan balio ezberdina hartzen duelako. Gainera, ahots-seinale guztiek ez dute iraupen berbera izango. Hau dela eta, sailkapen sistema batean

[11] Bideoa grabatzea ez da ezinbestekoa gure esperimenturako, baina sintesirako grabatutako datu-basearen kasuaren moduan, datuak bideoan edukitzea onuragarria suerta daiteke gerora begira beste proiektu batzuetarako.

sarrera-datu gisa sartu ahal izateko moduan adieraztea lortu beharko da, ahalik eta informazio gehien emanez. Kuantifikazio estatistiko eta parametrizazioak kontuan hartuko ditugu.

Kuantifikazio estatistikoa

Metodo honen asmoa datu estatistikoak pitch-kurbatik lortzea da (batezbestekoa, desbideratze estandarra, balio maximoa, balio minimoa,...) eta eurekin sailkatze sistema elikatu. [11]-an edo [1]-an erabilitako sistema da, datu garrantzizkoak zeintzuk diren jakinez gero nahiko emaitza onargarriak lor ditzake. Hala ere, a priori, ez dakigu emozioaren analisirako zeresanik handiena izango duten datuak zeintzuk izango diren, eta beraz, honetaz ezagupen handiago ez dagoen bitartean ez zaio metodo honi helduko.

Laginketa

Metodo honetan seinalearen denbora-puntu kopuru finkoa aukeratzen da eta horien gainean sistema elikatzeko datu garrantzizkoa kalkulatu. [11]-an seinalearen definizio denbora-eremuan banatutako 20 puntu erabiltzen dituzte. Modu hau oso eragingarri gerta daiteke seinale labur eta antzeko iraupenekin, hala ere, proposatutako esperimentuan erabiliko den datu-basearen egoera ez da halakoa, iraupen oso ezberdineko seinaleak egon daitezke ordea, eta sarritan lar luzeak ere. Beraz, zenbait ikerketa lanetan emaitza onak eman dituen metodo honek ez du ematen oraingoan gai honetarako egokiena izango denik.

Parametrizazioa

Metodo honetan pitch-kurba erabat irudikatu nahi da, edo gutxienez hurbilketa nahiko adierazgarri bat, parametro kopuru mugatua bat erabiliz[12]. Erabil daitezkeen parametrizazio ezberdinak daude, [5] eta [15]-an dokumentatzen den bezala; ez ditugu horiek guztiak sakonean ikusiko.

Sintesirako grabatu den datu-basea, azaldu den bezalaxe, oso landua dago, baita pitch-kurbaren aldetik ere. Pitch-kurba markatze automatikoko lanabesak erabiliz sortu da, eta bere parametrizazioa ere hala burutu da. Lanabes automatiko horiek grabatuko dugun lokutore anitzeko datu-basearen pitch-kurben eredua lortzeko ere balia daitezke.

Sintesirako datu-basearen parametrizazioa (eta beraz, grabatuko den datu-base berriarena) metodo ezberdin biren bitartez burutu da: bietako bat, Fujisakiren metodoa, euskararako izan duen luzapena [14]-an zehazten dena, eta bestea, pitch-kurbak estilizatzeko PSP bektoreetan oinarrituta. [18].

Sailkapen sistemaren sarrera gisa metodo bi hauetako bat erabiltzera begira, kontuan hartu behar da estilizazioan oinarritutakoaren bitartez fonema ahostun bakoitzerako bektore bana sortuz anotatu dela datu-basea. Honek esan nahi du parametro kopurua oso aldakorra dela (izan ere, esan bezala, datu-baseko erregistroen luzera oso aldakorra baita), eta gainera, kasu askotan oso altua da. Beraz, metodo hau erabiltzea, kasu askotan, parametro hauen laginketak edo kuantifikazio estatistikoak baliatuz baino ez litzateke zentzuzkoa izango, eta honek ez du abantaila handirik ekartzen pitch-kurbaren kuantifikazio estatistiko zuzenaren aurrean. Honen bitartez, delta eta pultsuei dagozkien datuak sartzeaz gain, sailkapen sistemari Fujisakiren modeloa funtzionatzen ari den egoeraren berri emango dioten parametro sorta bat ere sartu behar da. Esate baterako, modelatzea ezberdina da esaldi laburretarako edo eten anizdun luzeetarako, hortaz, esaldi bakoitzaren eten kopurua sailkatzaileak erabil dezakeen datua da.[13].

Egindako hautaketa eskura dauden lanabesetan oinarritzen da, baina gerora begira beste parametrizazio mota bat egitea ez da baztertzen, hala nola Bézierren funtzioetan oinarritzen dena [5], parametrizazio benetakoagoa eta emaitza hobeak izan ditzagun.

Laburbilduz, pitch-kurbak sistemak kontuan hartu ahal izateko erarik hoberena deskribatutako hiruen artean (kuantifikazio estatistikoa, laginketa, parametrizazioa) hautatu behar da, ikerketa sakon baten ondoren.

4.4. Sailkapena metodoaren definizioa

Bibliografian sailkapen metodo ezberdin asko agertzen da, baina begirada hemen planteatzen den lanarekin zerikusirik handiena dutenetan jarri behar da, [13] [15] [17] eta [11]. [13] [15] [17]-an OCON sare neuronalen sistema bat erabiltzen da. Sistema honetan, detektatu beharreko emozio bakoitzerako tarteko geruza bi zehatz-mehatz moldatzen dira, eta gero, erabakitze-logika batek irteera egokia ematen du. [11]-an metodo estatistiko batzuk proposatzen dira, (lineal, SVM, GVQ) horiekin guztiekin lortutako emaitzak oso antzekoak dira. [1]-an, guri dagokigunaren pare-pareko esperimentua ez bada ere, erregresio-zuhaitzen bidezko sailkapena ere proposatzen dute. Zuhaitz hauek, sarrera-parametro bakoitzari buruzko informazio argia ematen dute, eta hori abantaila handia da.

Zein metodo izango den egokiena zehaztea ezinezkoa denez gero, esperimentuaren atal honetarako

[12] Ikerketa honetan parametrizazio kuantitatiboak baino ez dira landuko, kualitatiboek pitch-kurbaren deskribapenak besterik ez baitituzte eskaintzen, eta oro har subjektiboak izaten dira. Kurbaren balioak ez dira modu numerikoan azaltzen.

[13] Aurreko atalean azaldu den bezalaxe, etenak ez daude datu-basean anotaturik. Honek esan nahi du ezin dezakegula izan kokapena edo iraupenaren berri. Esaldiko eten kopurua, ordea, badakigu zein den.

sailkapenerako 3 metodo ezberdin erabiltzea proposatzen da.:

- 7 azpisareetan oinarritutako sare neuronalen sistema bat, detektatu beharreko emozioetarako bana, estilo neutroa barne eta irteerak emango duen logika batek elkarturik.

- Diskriminatzaile lineal bidezko sailkapen estatistikoa.

- Erregresio-zuhaitzetan oinarritutako sistema.

4.5. Laburpena

Proposatzen den esperimentuak MPEG estandarrak oinarrizkotzat jotzen dituen 6 emozioak detektatzea du helburu, eta baita tonu neutroa ere. Horretarako, lehenengo hurbilketa batean, Euskal Herriko Unibertsitateko Ahots Seinalearen Tratamenduko Taldeak garatutako emozioen datu-basea erabiliko da. Honetaz gain, eta, ezin bazter dezakegu emaitzak hobetu eta orokor bihurtzeko beste datu-base bat grabatzea. Printzipioz sarrera-datuak honako hauek izango dira:

- Energiaren gaineko estatistikoak.

- Abiadura, segmentuaren iraupenaren arteko bokale kopuru gisa neurturik.

- Intonazioa, kuantifikazio estatistikoa, laginketa edo parametrizazioa erabiliz.

Eza da baztergarria, honetaz gain, etenei buruzko informazioa ematea.

Sailkapen metodoei dagokienez, hiru aldaera ezberdin planteatzen dira: bat OCON sare neuronaletan oinarriturik, beste bat diskriminatzaile linealetan, eta azkena erregresio-zuhaitzetan oinarriturik.

5. Ondorioak

Gizakiok emozioak ulertzeko dugun gaitasuna esparru automatikora ekartzea ez da batere lan erraza, nahiz eta harantz goazen. Lan honetan artearen egoerari buruzko ikerketa bat egin da, metodologia finko bat azalean utziz, jorratu beharreko atal ezberdinetarako taktikak konprobatuz. Aztertu diren emaitzak honako modu honetan adieraz daitezke, laburki:

Hiru esperimentu mota hartu dira kontuan; giza-dekodifikazioko esperimentuak, detektatze automatiko bitarreko esperimentuak eta detektatze automatiko orokorreko esperimentuak. Hiru esperimentu mota hartu dira kontuan: giza-kodifikazio esperimentuak, detekzio automatiko bitarreko esperimentuak, detekzio automatiko orokorreko esperimentuak. Giza-dekodifikazioaren bitartez burututako esperimentuekin lortutako arrakasta mailak %80 ingurukoak dira, detekzio automatiko bitarraren bidez egindakoetan maila %70ra jaisten da, eta detekzio automatiko orokorreko esperimentuetan, %50ren inguruko arrakasta maila lortu ohi da.

Azkenik, artearen egoeraren ikerketan jasotako ezagupenez baliaturik, euskararako emozioen ezagutza automatikorako esperimentu bat izan beharko lukeena planteatu da, bai eta ezagutza egokia erdiesteko erabil daitekeen hainbat aldaera.

6. Aipamenak

[1] J. Ang, R. Dhillon, A. Krupski, E. Shriberg, A. Stolke: “Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog”. 2001: A Speaker Odyssey. The Speaker Recognition Workshop, ISCA 2001.

[2] C. Bartneck: “How convincing is Mr. Data’s smile: Affective expressions of machines”. User Modeling and User Adapted Interaction 11: 279-295. Kluwer Academic Publishers, 2001.

[3] A. Batliner, K. Fischer, R. Huber, J. Spilker, E. Nöth: “Desperately seeking emotions or: Actors, Wizards, and Human Beings”. ITRW on Speech and Emotion, ISCA 2000.

[4] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Vortsis, S. Kollins, W. Fellenz, J.G. Taylor: “Emotion Recognition in human-computer interaction”. IEEE Signal Processing Magazine, Jan 2001.

[5] D. Escudero: “Modelado Estadístico de Entonación con Funciones de Bézier: Aplicaciones a la Conversión de Texto a Voz en Español”. Tesis Doctoral. Universidad de Valladolid, 2002.

[6] N.H. Fridja: “The Emotions”. Cambridge University Press, 1986.

[7] I.T. Johnstone, R. Banse, K.R. Scherer: “Acoustic Profiles from Prototypical Vocal Expressions of Emotion”. Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm, Sweden, 4, 2-5, 1995.

[8] C.M. Lee, S.S. Narayanan, R. Pieraccini: “Combining Acoustic and Language Information for Emotion Recognition”. Proc. ICSLP’02, 2002.

[9] C.M. Lee, S.S. Narayanan, R. Pieraccini: "Classifying Emotions in Human Machine Spoken Dialogs". Proc. ICME'02, 2002.

[10] C.M. Lee, S.S. Narayanan, R. Pieraccini: "Recognition of Negative Emotions from the Speech Signal". Proc. 'Automatic Speech Recognition and Understanding', 2002.

[11] S. McGilloway, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, S. Stroeve: "Approaching Automatic Recognition of Emotion From Voice: A Rough Benchmark". ITRW on Speech and Emotion, ISCA 2000.

[12] S.J.L. Mozziconacci: "Modeling Emotion and Attitude in Speech by Means of Perceptually Based Parameter Values". User Modeling and User Adapted Interaction 11: 297-326. Kluwer Academic Publishers, 2001.

[13] R. Nakatsu, J. Nicholson, N. Tosa: "Emotion recognition and its application to computer agents with spontaneous interactive capabilities". Knowledge-Based Systems, 13, 497-504. Elsevier Science B.V., 2000.

[14] E. Navas: "Modelado Prosódico del Euskera Batúa para Conversión de Texto a Habla". Tesis Doctoral. Euskal Herriko Unibertsitatea, 2003.

[15] J. Nicholson, K. Takahashi, R. Nakatsu: "Emotion Recognition in Speech using neural networks". Neural Computing and Applications, 9: 290-296. Springer-Verlang London Limited, 2000.

[16] C. Sobin, M. Alpert: "Emotion in Speech: The acoustic attributes of Fear, Anger, Sadness and Joy". Journal of Psycholinguistics Research, Vol. 28, N.4. Plenum Publishing Corporation, 1999.

[17] N. Tosa, R. Nakatsu: "Emotion-based, Multi-person interactive Theater -Romeo & Juliet in Hades-" Workshop on Emotion-Based Agent Architectures (EBAA '99)

[18] M. Vivanco: "Modelado de la Entonación en Euskera para la Conversión de Texto a Voz Usando Métodos Basados en Corpus". Proyecto de Fin de Carrera. Euskal Herriko Unibertsitatea, 2002.

[19] S. Zaldumide: "Modelado de las emociones para conversion de texto a voz en euskera". Proyecto de Fin de Carrera. Euskal Herriko Unibertsitatea, 2003.

[20] G. Zhou, J.H.L. Hansen, J.F. Kaiser: "Nonlinear Feature Based Classification of Speech under Stress". IEEE Transactions on Speech and Audio Processing, Vol. 9, n.3, March 2001.

[21] Standard MPEG4: http://mpeg.telecomitalialab.com/