IKERGAZTE
NAZIOARTEKO IKERKETA EUSKARAZ

II. IKERGAZTE
NAZIOARTEKO IKERKETA EUSKARAZ

2017ko maiatzaren 10, 11 eta 12
Iruñea, Euskal Herria

ANTOLATZAILEA:
Udako Euskal Unibertsitatea (UEU)

INGENIARITZA ETA ARKITEKTURA

Ahots kantatuaren sintesiaren egokitzapena bertsolaritzarako

Xabier Sarasola, *Eva Navas* eta *Inma Hernaez*

157-164 or.
https://dx.doi.org/10.26876/ikergazte.ii.03.22

Ahots kantatuaren sintesiaren egokitzapena bertsolaritzarako

Xabier Sarasola, Eva Navas eta **Inma Hernaez**

AHOLAB, Euskal Herriko Unibertsitatea (UPV/EHU), Bilbo

Laburpena

Ondorengo paper-ak euskarazko bertsolaritza ahots kantari sintetikoaren sortzeko prozesua deskribatzen du. Bertan, helburu hau lortzeko erabili diren baliabideen eta tresnen deskribapena emango da. Gure sistemak Hidden Markov Model (HMM) bidezko hizketa-sistema estatistikoa oinarri bezala hartzen du ahots kantatua sortzeko. HMM-ak entrenatzeko hizketa datu-base bat erabili beharrean ahots kantatuaren datu-base batez entrenatuko da ezaugarri linguistikoez gain ezaugarri musikalak gehituz audio seinaleei.

Hitz gakoak: Euskarazko ahots-kantatu sintesia, HMM bidezko sintesia, Bertsolaritza

Abstract

This paper describes the process of creation of a singing voice synthesis system for bertsolaritza, the art of singing improvised songs in Basque. The database acquisition process and the tools used to achieve this goal are explained. Our system uses Hidden Markov Model (HMM) based statistical speech synthesis as base to build a singing voice. We use a singing voice database instead of a speech database to train the HMM system with music labels in addition to the usual linguistic labels.

Keywords: Basque singing synthesis, HMM-based singing synthesis, Bertsolaritza

1 Sarrera eta motibazioa

Ahots kantatuaren sintesiak ordenagailu batek edozein kanta kantatzeko aukera ematen du. Nahiz eta 50 hamarkadan ahots kantatua sortzeko aukera erreferentziatuak dauden, 2000 hamarkada arte itxaron behar izan da garrantzi handiko gaia bihurtzeko, batez ere entretenimendu eta aisia eremuetan. Bertsolaritza pisuzko elementu kultural bat izanik, interesgarria da kalitate bertsolaritza sintetikoa lortzea.

Ahots kantatua oso hedatua da bereziki Japonian, non Yamahak (Pompeu i Fabra Unibertsitateko ikerketa taldearekin kolaboratuz) Vocaloid (Kenmochi eta Ohshita, 2007) izeneko sistema komertzial bat garatu zuen 2004 urtean.

Egun bi teknika dira beste teknika guztien gainetik gailentzen direnak ahots-sintesia lortzeko:

- Unitateen selekzioa eta kateamendua (Hunt eta Black, 1996; Raux eta Black, 2003): Teknika honetan difonemak, fonema erdiak edo hainbat fonemako unitateetan segmentatutako datu-base bat erabiltzen da. Sintesia lortzeko, beharrezko ezaugarriak betetzen dituzten unitateak kateatzen dira kateatze-kostua kontuan hartuz. Kateatze-kostuak, originalki kateatuak ez dauden unitateak kateatzeak sintesiaren kalitatean duen eragina baloratzen du. Sortuko den sintesiak kateatze-kostu txikiena duen konbinazioa erabiliko du. Kalitate oneko ahots oso bat lortzeko, hiztun konkretu baten grabazio kopuru handia izan behar da komenigarriki etiketatua (fonema maila, silabak, hitzak, esaldiak, etab.).
- Sintesi estatistiko parametrikoa (Zen *et al.*, 2009; Tokuda *et al.*, 2013): Audioa parametro akustikoen bektoreetan bihurtzen da eta, aurrez egindako entrenamendu batean, sistemak testutik atera daitezkeen ezaugarri linguistiko eta hauei dagokien errealizazio akustikoaren erlazioa 'ikasten' du. Sistema

mota hauek biltegiratze-baldintza minimoak dituzte eta oso malguak dira, oinarritzen diren eredu matematikoak ahots berri edo hizketa-estilo desberdinetara egoki edo manipulatu daitezkeelako (Yamagishi *et al.*, 2009).

Sintesi terminoetan, ahots kantatuak diferentzia garrantzitsuak aurkezten ditu hizketa-ahotsarekin (Garnier *et al.*, 2007). Lehenik, kantuaren intonazioa melodiak eta erritmoak markatzen dute eta ez hainbeste lengoaiaren berezitasunek. Ikuspuntu fonetikotik, seinalearen ehuneko altu batek fonema bokalikoak ditu. Beste alde batetik, ikuspuntu akustikotik ahots kantatuak intentsitate altuagoa erakusten du eta baita zenbait fenomeno espezifiko, adibidez vibratoa (Alemán eta Carlosena, 2004). Beste ezaugarri garrantzitsu bat kantuaren onset bokalikoarekiko erritmoaren sinkronizazioa da, hizketa ahotsean kontsideratzen ez dena (Saino *et al.*, 2006).

Nahiz eta diferentzia hauek izan, aipatutako bi sintesi-teknikak ahots kantatuan aplika daitezke. Aukeraketa-teknikak zenbait fintze eskatzen ditu seinale-prozesamenduan aukeraketa beratik haratago doazenak (Bonada eta Serra, 2007), baita grabazioko datu-baseen diseinu espezifikoa (Martı Umbert eta Blaauw, 2013). Arrakasta komertzial ukaezina izan duten sistemak daude, Vocaloid ezaguna adibidez, teknika hauetan oinarritzen direnak. Sintesi parametrikoak ere, aukeraketa baina askoz berriagoa dena, inpaktu handia izan du ahots kantatuaren munduan. Kasu honetan, eredu estatistikoak ezaugarri linguistikoen elkarrekikotasunaz gain, elkarrekikotasun melodikoak eta kantatutako seinalearen parametro akustikoak atzituko ditu (Oura *et al.*, 2012). Teknika hauen esponente handiena Sinsy sistema da (Oura *et al.*, 2010).

2 Ikerketaren helburuak

Bertsolaritza ahots kantatuko genero oso espezifiko bat denez, hizketa ahotsetik gertuago dagoena beste estilo batzuek baino eta aldi berean karga espresibo garrantzitsu bat duena, interesgarria da ikerketa espezifiko bat egitea oraingo sintesi tekniken egokitasuna aztertzeko kalitate handiko bertsolari sintetiko bat diseinatzeko momentuan. Dakigunaren arabera, gai honetan ez da existitzen ikerketarik IXA eta Aholab taldeen arteko kolaborazioan egindako esperimentu pilotuez gain, BERTSOBOT (Astigarraga *et al.*, 2013) izenez ezaguna dena sortu duen kolaborazioa. BERTSOBOTen demostrazioek interes sozial handia piztu dute eta aurrera ateratzeko hizketa ahotsei dagokien ahots parametrizatuak erabili dira ahots kantatuak sortzeko. Hala ere, hizketa ahotsa eta ahots kantatuaren arteko diferentzia handiek kalitate oso baxuko ahots kantatu bat eman dute emaitza bezala. Gainera, onset eta vibratoaren modelatu gehigarriak ez ziren kontuan hartu integrazio honetan. Aukera bat Vocaloid-erako ahots bat sortzeko sistema honek markatzen dituen etapak jarraitzea izango zen. Nahiz eta kostu handikoa izan, irteera sintetiko on bat eskaini lezake. Hala ere, bertsolari ahots sintetiko bat lortzea ez da lan honen helburu bakarra. Lan honetan teknologia propioa garatu nahi da, ohiz kanpoko bi zirkunstantzien arteko sinergia ustiatzen duena: ahots sintetikoaren sorkuntzan eskuragarri dagoen esperientzia handia eta jakintza sakona, bertsolaritza estilo zaila lantzen den munduko leku bakarrean egotearekin konbinatua. Bertsolaritza generoa erronka partikularki zaila da ahots kantatuaren kalitatea esperimentatzeko, izan ere ahots kalitate baxua instrumentuen atzean ezkutatzea posible delako, baina bertsolaritzak ez dauka instrumenturik.

3 Ikerketaren muina

Ahots kantatu sintetikoa bertsolaritzara aplikatzeko ikerketa prozesua hiru zati nagusitan bana daiteke:

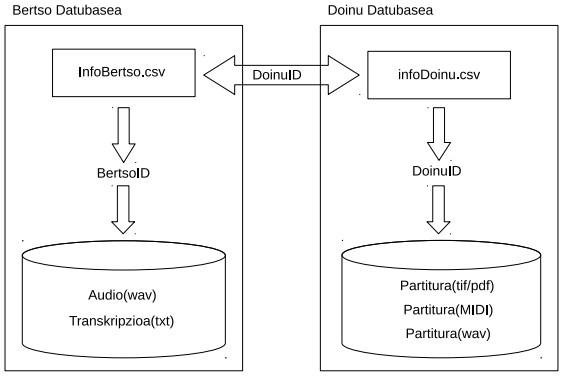
- **Datuak**: Ahots kantatuaren sintesia lortzeko beharrezkoa da ahots naturalaren datu-base bat izatea eredu egokiak eskura izan eta sistema entrenatzeko.
- **Sintesi-sistema**: Sintesi sistemaren aukeraketa garrantzitsua da eskura dauden datuak erabiliz ahalik eta kalitate handieneko ahotsa lortzeko. Sistemak definituko ditu datu-basean egin behar diren etiketatze lanak eta sintesirako erabiliko diren parametroak.
- **Sistemaren egokitzapena**: Aukeratutako sintesi-sistema bertsolaritzarako egokitu behar da. HMM bidezko sistema batean bertsolaritzako kantatzeko estiloa karakterizatzen duten eredu egokiak lortu behar ditu espektro, frekuentzia nagusi eta fonemen iraupenerako. Eredu hauek sortzeko datu-baseko seinale-zatiei testuinguru-etiketa egokiak jarri beharko zaizkie. Unitate selekzio sistema

batean unitate-aukeraketa findu behar da kateatze-kostua eta sintesiaren kalitatea era efizienteenean erlazionatzeko.

3.1 Datuak

Bertsolaritza datu-basea bi zatitan banatzen da 1 Irudian erakusten den bezala: Alde batetik audio grabazioak transkripzio fonetikoekin batera eta grabazio bakoitzaren informazio gehigarria; beste alde batetik musika partiturak daude bertsoetan erabiltzen diren doinuekin. Datu-basearen zati bakoitzak CSV fitxategi bat dauka bertso edo doinu bakoitzaren informazioa daukana. Bertsoen kasuan bertsolaria, doinua, data, etab. eta doinuaren kasuan neurria, bis mota, etab. BertsoID identifikagailuaz bertsoaren audio eta transkripzio fitxategiak atzitu daitezke eta DoinuID identifikagailuak eramango gaitu bertsoaren doinuaren informazio eta fitxategietara.

1 Irudia: Bertsolaritza datu-basearen egitura



Bertso-grabazioak mp3 formatuan lortuak daude bertsolaritza lantzen duten Bertsozale Elkartearen online datu-basetik [1]. Grabazioak 1979tik 2014ra bitarteko lehiaketa eta saio desberdinetako grabazioak dira. Lortutako materiala 1 Taulan erakusten da. Bakarrik bertsolari bakarreko eta puntukako saioak hartu dira kontuan.

1 Taula: Grabazioen informazioa

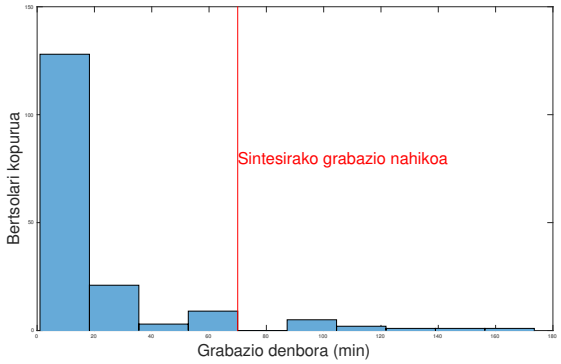
Grabazioen iraupena	6105 min
Grabazio kopurua	2094
Gizonezkoen grabazio kopurua	1860
Emakumezkoen grabazio kopurua	234
Bertsolari kopurua	189
Gizonezko bertsolari kopurua	158
Emakumezko bertsolari kopurua	34

Grabazioen MP3 fitxategiak Windows PCM formatura pasa dira, 44.100 Hz-eko laginketa maiztasunarekin eta 16 bit-ekin lagin bakoitzeko. Bertsolari bakoitzeko grabazioen kopurua 93tik grabazio bakarrera doa. Bertsolari bakoitzeko grabazio-denbora minutu 1etik 314ra doa. Baina grabazioek isiluneak, gaijartzailea eta txaloak dituzte eta beraz, grabazioen erdia besterik ez da ahots kantatua.

Doinuei dagokienez, 3019 doinu daude eskura baina 202 besterik ez dira erabiltzen datu-baseko bertso-grabazioetan.

[1] www.bertsozale.eus

2 Irudia: Grabazio denbora bertsolariko



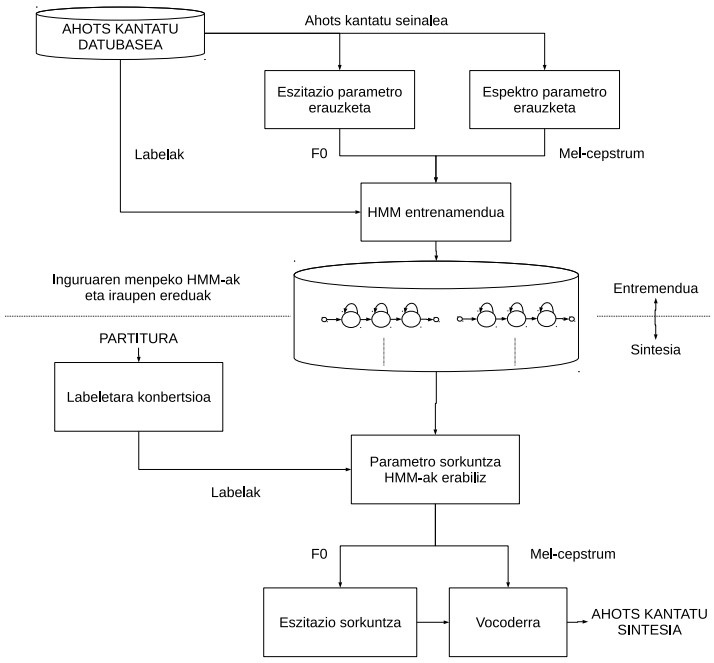
2 Irudian erakusten den bezala, bertsolari gutxik dauka 70 minutu grabazio baino gehiago, ahots bat sortzeko normalean erabiltzen den kopurua HMM sintesi-sistema bat erabiltzen denean (Oura *et al.*, 2010).

Bakarrik 10 bertsolarik betetzen dute baldintza hau, 9 gizonek eta emakume 1ek. Hala ere, material gutxiago daukaten bertsolariak ahots kantatuaren ereduak erabiliz egokitu ahalko dira etorkizunean.

3.2 Sintesi sistemak

Sarreran azaldu den moduan, sintesirako erabiliko den sistemak aukeratzeak eskura dauden datuen araberakoa da. Grabazioen kateatze sistemek kalitate handiagoa lortzen dute baina datu-base handiagoa behar dituzte. Bertsolari bakoitzetik dugun ahots denbora maximoak 70 minutu ingurukoak direnez daukagun aukera bakarra HMM bidezko sintesia egitea da oraingoz. Sistemaren eskema orokorra 3 Irudian ikus daiteke.

3 Irudia: HMM bidezko ahots kantatu sintesi-sistemaren eskema orokorra



Sintesi estatistikoa, entrenamendu fasean lortutako eredu estatistiko multzo batean oinarritua dago, geroago sintesi-parametroak sortzeko balioko dena.

Entrenamendu fasean espektro- eta eszitazio-parametroak erauzten dira bertso grabazioetatik. Ondoren parametro hauen ereduak sortzen dira grabazio beretatik ateratako testuinguru-etiketak erabiliz. Entrenamendu honekin pitch (F0) eta espektro ereduak (Mel-cepstrum), HMM egoeren iraupen ereduak eta nota eta fonemen sinkronizazio ereduak lortzen dira. Azken ezaugarri hau musika-tempoari lotua dago eta bertsolaritza estiloa beste kanta estilo batzuetatik banatuko duena izango da. Ereduen lortzea iterazio asko eskatzen dituen prozesu bat da, non iterazio bakoitzean, grabazioen etiketatze errore posibleak zuzentzen diren, parametro desberdinak doitzen diren, etab.

Sintesi estatistikoek hizketa ahotsari eskaintzen dizkioten aukeretako bat lortutako ahotsa hizketa estilo berrietara moldatzeko malgutasuna da, sorburu grabazioetan agertzen ez diren emozio, azentu eta ahots berriak lortzeko erabil daitekeena. Modu berean, bertsolari berrien ahotsak sortzeko ere aplikagarriak izan daiteke. Beraz, bertsolari ahots berriak sortzeko ahots egokitzapen eta bihurketa teknikak aztertuko dira, ahots kantaturako beharrezkoak diren moldaketak eginez.

3.3 Sistemaren egokitzapena

Aurreko atalean azaldu den bezala, HMM sistema batek seinale etiketatu baten testuinguru-etiketa eta ezaugarri espektralen arteko erlazioa ikasiz lortzen du ahots-sintesia. Hau dela eta, testuinguru-etiketek markatzen dute sortuko den ahotsak zein ezaugarri hartuko dituen entrenamenduan datu-basetik. Etiketa hauek kontuan hartzen dute seinalearen zati bakoitza ez dela seinale zati isolatu bat. Fonema batek nota, silaba edo esaldi batean duen kokapenak eta aurretik eta ondoren dituen fonema edo notek eragingo dute seinalearen forman. Hau kontuan hartuta fonema bakoitzari emango zaizkion etiketak ondorengoak izango dira.

- **Fonema**: aurreko bi fonemak, oraingo fonema eta ondorengo 2 fonemak.
- **Posizioa notan**: aurreko, oraingo eta ondorengo fonemen posizioa notan.
- **Zortziduna**: Aurreko, oraingo eta ondorengo notaren zortziduna.
- **Nota**: aurreko, oraingo eta ondorengo nota.
- **Iraupena**: Aurreko, oraingo eta ondorengo notaren iraupena.
- **Notaren posizioa**: Aurreko, oraingo eta ondorengo notaren posizioa esaldian.

Etiketa hauek erabiltzeak esan nahi du seinalean denbora markez seinalatutako fonema bakoitzak etiketa hauek definiturik izan behar dituela. Fonema bakoitzean zein nota kantatzen ari den era automatikoan etiketatu da, doinuaren partitura oinarritzat hartuz. Ahotsaren frekuentzia nagusia, eskuz etiketatuko fonema labelak eta partiturak erabiliz testuinguru-etiketak sortu dira.

Testuinguru-etiketa hauekin egindako ahots sintetikoaren kalitatea ezegokia izanik, etiketa hauen analisia egin da eta aurkitu da partiturak eta grabazioen arteko koherentzia ez dela egokia.

3.3.1 Noten iraupena

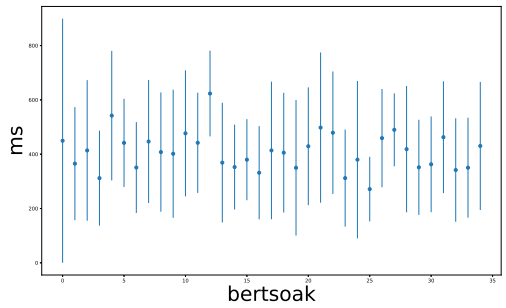
Nota bakoitzaren iraupena nota motaren bidez definitzen da (zuria, beltza, kortxea...). Informazio hau doinuen partiturak grabazioekin lerrokatuz lortzen da eta HMM-ak era egokian entrenatzeko iraupen erreala eta etiketakoak erlazioa izan behar dute baina 4 Irudian ikusten den bezala partiturek markatutako denborak ez dira egokiak bertso grabazio errealentzat. Nota beltzek bertso desberdinetan batezbesteko denbora berak ez izatea arazo bat da entrenamendu fasean, baina gerta daitekeen zerbait da kontuan hartzen badugu bertso desberdinak erritmo desberdinarekin kantatuak izan daitezkeela. Baina hau kontuan hartuta ere, bertso bakoitzean iraupen etiketa bera duten noten artean dagoen desbideratze estandarra handiegia izaten jarraitzen du.

3.3.2 Frekuentzia nagusia

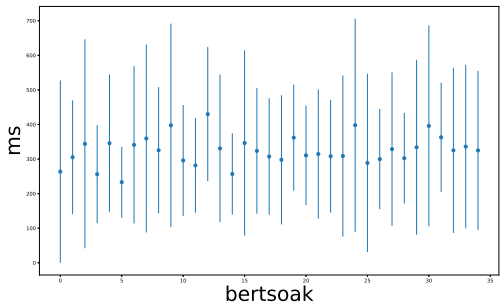
Fonema bakoitzean kantatzen den nota zein den markatu behar da etiketetan ere bai, baina 5 Irudian erakusten du bertso desberdinetan nota berdin bezala markatutako fonemen frekuentzia nagusia asko aldatzen dela. Irudian nota beraren batezbesteko balioa eta desbideratze estandarra irudikatzen da. Frekuentzia cent-etan irudikatuta dago, non tonuerdien arteko desberdintasunak 100 cent diren.

4 Irudia: Eskala desberdinetako Re notaren frekuentzia nagusia

(a) Beltzen denbora

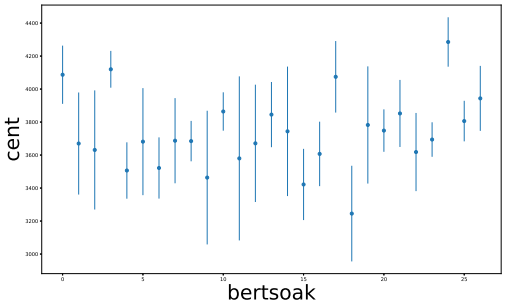


(b) Kortxeen denbora

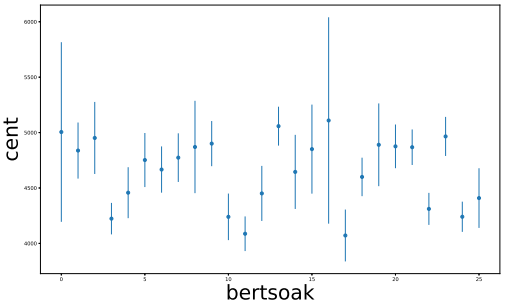


5 Irudia: Eskala desberdinetako Re notaren frekuentzia nagusia

(a) 4. zortziduneko Re



(b) 5. zortziduneko Re



4 Ondorioak

Bertso sintesia sortzeko sistemaren egitura sortu da baina oraindik ez da lortu kalitate nahikoa duen ahots kantatuko sintesia. Arrazoi nagusia automatikoki sortutako testuinguru-etiketen koherentzia falta da. Etiketa hauek lortzeko erabilitako partiturak eta grabazioen arteko ezberdintasunek etiketa ezegokiak sortzen dituzte eta honek erlazio ezegokiak ikastera eramaten du sistema. Hau konpontzeko bi bide nagusi azpimarra daitezke etorkizunerako:

- **Etiketatze automatikoa hobetu**: Koherentzia falta hauek existitzen direla jakinda, etiketatzean arazo hauek detektatu eta etiketa egokiak sortzea.
- **Parametroen behartzea sintesian**: Ahots kantatua sintetizatzen denean erabiltzen diren espektro, frekuentzia nagusi eta iraupen ereduak HMM-en ikasketatik datoz guztiz. Zenbait ereduren kalitatea ez bada hobetzea lortzen testuinguru etiketak hobetuz, arau bidezko ereduen bidez ordezka daitezke nahi diren ezaugarriak lortzeko ahots sintetikoak.

5 Etorkizuna

Etorkizunerako lana hiru puntu nagusitan bana daiteke:

- Entrenamenduaren hobekuntza
- Sistema ebaluatzeko era berriak

- Ahots kantatu berriak

Lortu den ahots kantatuaren kalitatea oraindik ez da behar bezalakoa, hau hobetzeko HMM-en entrenamendu fasean erabiltzen diren labelak aldatu daitezke. Label egokiak gehituz, inguru informazio gehiago eman daiteke HMM-ei ikasketa egokiagoa izateko.

Sintesiaren kalitatea egiaztatzeko ez dira oraindik froga estandarrik egin. Fase honetara ez gara iritsi sintesiaren kalitatea ez dela horrelako froga egiten planteatzeko bezain handia. Maila horretara iristerakoan egiteke egongo da bi motatako frogen artean aukeratzea:

- **Froga subjektiboak**: Hainbat entzulek ezberdin konfiguratutako sintesi-sistemak ebaluatzea.
- **Froga objektiboak**: Lortutako ahotsean objektiboki neur daiteken ezaugarriak ebaluatzea. Musikan tempoa eta tonua hain garrantzitsuak izanik, fonema eta silaben iraupenak eta frekuentzia nagusia (F0) desiratuak diren balora daiteke.

Azkenik lehen ahotsaren kalitatea egokia dela ikustean, beste kantarien ahotsak sintetizatzera iritsi nahi da entrenamenduan horien grabazioak erabiliz.

6 Esker onak

Autoreek eskerrak eman nahi lizkioke Bertsozale Elkarteari, bertso-ekitaldien digitalizazioan egin duten lanarengatik. Lan honek asko erraztu du HMM bidezko ahots kantatu sistema erabiltzeko datu-basea sortzea.

Lan hau UPV/EHU (Ayudas para la Formación de Personal Investigador), Eusko Jaurlaritza (ELKAROLA project, KK2016/00087) eta Espainiako Ekonomia, Industria eta Leihakortasun Ministerioak FEDER laguntzarekin(RESTORE project,TEC2015-67163-C2-1-R) finantzatu dute.

Erreferentziak

Alemán, Ixone Arrobarren, eta Alfonso Carlosena. 2004. *Signal Processing Techniques for Singing and Vibrato Modeling*.

Astigarraga, Aitzol, Manex Agirrezabal, Elena Lazkano, Ekaitz Jauregi, eta Basilio Sierra. 2013. Bertsobot: the first minstrel robot. In *Human System Interaction (HSI), 2013 The 6th International Conference on*, 129–136. IEEE.

Bonada, Jordi, eta Xavier Serra. 2007. Synthesis of the singing voice by performance sampling and spectral models. *IEEE signal processing magazine* 24.67–79.

Garnier, Maëva, Nathalie Henrich, Michele Castellengo, David Sotiropoulos, eta Daniele Dubois. 2007. Characterisation of voice quality in western lyrical singing: from teachers' judgements to acoustic descriptions. *Journal of interdisciplinary music studies* 1.62–91.

Hunt, Andrew J, eta Alan W Black. 1996. Unit selection in a concatenative speech synthesis system using a large speech database. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 1, 373–376. IEEE.

Kenmochi, Hideki, eta Hayato Ohshita. 2007. Vocaloid-commercial singing synthesizer based on sample concatenation. In *Interspeech*, volume 2007, 4009–4010. Citeseer.

Marti Umbert, Jordi Bonada, eta Merlijn Blaauw. 2013. Systematic database creation for expressive singing voice synthesis control.

Oura, Keiichiro, Ayami Mase, Yoshihiko Nankaku, eta Keiichi Tokuda. 2012. Pitch adaptive training for HMM-based singing voice synthesis. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 5377–5380. IEEE.

Oura, Keiichiro, Ayami Mase, Tomohiko Yamada, Satoru Muto, Yoshihiko Nankaku, eta Keiichi Tokuda. 2010. Recent development of the HMM-based singing voice synthesis system-Sinsy. In *Seventh ISCA Workshop on Speech Synthesis*.

Raux, Antoine, eta Alan W Black. 2003. A unit selection approach to f0 modeling and its application to emphasis. In *Automatic Speech Recognition and Understanding, 2003. ASRU'03. 2003 IEEE Workshop on*, 700–705. IEEE.

Saino, Keijiro, Heiga Zen, Yoshihiko Nankaku, Akinobu Lee, eta Keiichi Tokuda. 2006. An hmm-based singing voice synthesis system. In *INTERSPEECH*.

Tokuda, Keiichi, Yoshihiko Nankaku, Tomoki Toda, Heiga Zen, Junichi Yamagishi, eta Keiichiro Oura. 2013. Speech synthesis based on hidden markov models. *Proceedings of the IEEE* 101.1234–1252.

Yamagishi, Junichi, Takashi Nose, Heiga Zen, Zhen-Hua Ling, Tomoki Toda, Keiichi Tokuda, Simon King, eta Steve Renals. 2009. Robust speaker-adaptive hmm-based text-to-speech synthesis. *IEEE Transactions on Audio, Speech, and Language Processing* 17.1208–1230.

Zen, Heiga, Keiichi Tokuda, eta Alan W Black. 2009. Statistical parametric speech synthesis. *Speech Communication* 51.1039–1064.