

Nafar-Lapurtar euskalkiarentzako euskal TTS bat garatzea

E. Navas¹, I. Hernaez¹, D. Erro^{1,2}, J. Salaberria³, B. Oyharçabal³, M. Padilla³

¹Aholab (UPV/EHU)

²IKERBASQUE

³IKER UMR 5478

{eva,inma,derro}@aholab.ehu.es

Abstract

The paper presents a new TTS system for the Navarro-Laburdian dialect based on a standard Basque TTS. A phonetically balanced recording corpus of 4.000 sentences has been designed and two speakers have recorded it. The voice has been built using a high quality speech coder in the context of HMM based speech synthesis. The new dialectal TTS system has been compared in a subjective evaluation with the existing TTS system for standard Basque and with a mixed system that applies the phonetic transcription rules of the dialect, but uses the speech generation module of the standard Basque system. The adaptation of the front-end module with the inclusion of new phonetic transcription rules and new sounds is not enough to get a system that works better than the standard Basque system. The results with the dialectal new voice indicate that users prefer the new dialectal system to the standard Basque one.

Laburpena

Artikulu honek euskarazko TTS estandar batean oinarritutako Nafar-Lapurtar euskalkiarentzat egindako TTS sistema bat aurkezten du. Fonetikoki orekatutako eta 4000 esaldiko corpus grabatua egokitu zaio, bi hizlarik grabatua. Ahotsa kalitate handiko hizketa-kodetzaile baten bidez eraiki da, HMMan oinarritutako hizketa-sintesiaren kontestuan. Euskalkiaren TTS berria era subjektiboan ebaluatu da oraingo euskarazko TTS estandararekin eta baita euskalkiaren transkripzio fonetikoaren arauak aplikatzen dituen sistema-nahasketa batekin, bigarren horrek euskarazko sistema estandarren modulu-sortzailea erabiltzen duelarik. Transkripzio fonetiko eta soinu berriez Front-end moduluarri egindako egokitzapena ez da nahikoa euskarazko sistema estandarra baino hobeto funtzionatzen duen sistema bat lortzeko. Euskalkiaren ahots berriekin lortutako emaitzek erakusten dute erabiltzaileek nahiago dutela euskalki sistema berria euskarazko estandarra baino.

Keywords: TTS, Euskara, Nafarroa

Gako hitzak: TTS, basque, Nafarroa

1. Sarrera

Hizketa-sintesia oso ezaguna den teknologia da, hizkuntza gehienetarako kalitate ona lortu duena. Teknologia hori hizkuntza gutxitu batzuetarako ere garatu da, eta azkenaldian dialektoetarako ere garatzen hasi dira suediera (Beskow eta Gustafson, 2009), Austriako alemana (Pucher et al, 2010), Tianjin-era (Hu et al, 2011) eta iparraldeko sothoera (Langa et al, 2012). Lan hauek guztiak hizketa-sintesi parametrikoki aplikatzen dute Markoven Eredu Ezkutueta (HMM) (Yoshimura et al, 1999) oinarrituz. Sistema estatistikoki parametrikotik batez besteko ereduak entrenatzen dira antzekoak diren unitate natural akustikoak erabilia eta ezaugarri linguistikoak dituzten erabaki-zuhaitzak sortuta. HMMetan oinarritutako hizketa-sintesi teknikek egonkortasun eta malgutasun handia eskaintzen dute ahots berriak sortzeko momentuan egokitzapen eta interpolazio teknikak erabiliz (Zen et al, 2009). Hizketa-sintesi teknika hau hizketa naturalaren datu-baseetan oinarritzen da eta nahiz eta datu-baseen tamaina ez izan kritikoa, fonetikoki aberatsa den corpus handi batez baliatzen da kalitate handiko hizketa sintetikoaren sorreran. Hizketa corpus hauen garapena bereziki garrantzitsua da baliabide mugatuko eta dialekto askoko hizkuntzentzat.

Euskarak 1.050.000 hizlari inguru ditu eta 5 euskalki nagusi. Euskalki hauetako bat nafar-lapurtar euskalkia da, Lapurdi eta Nafarroa Beherean hitz egiten den euskalki nagusia, Ipar Euskal Herriko bi probintzia historikoak. Hori dela eta, euskalki honek frantsesaren eragina izan du. Ipar Euskal Herrian 73.000 euskal hiztun daude (populazioaren % 30,5) (Basque Government, 2014).

Euskararako testutik hizketa-sintesisirako (TTS) sistema estandarrak, AhoTTS (Hernáez et al, 2001) eta Nuance VocalizerTM, eginda daude, baina ez dago euskalkietarako ezer eginda. Lan honen helburua AhoTTSan oinarritzen den TTS sistema egitea zen, nafar-lapurtar euskalkiarentzat. TTS baten egitura tipikoak bi modulu desberdin ditu: *front-end* edo eredu linguistikoa, sarrerako testua normalizatzen duena eta transkripzio fonetiko esleitzen diona hitz bakoitzari, eta *back-end* edo lehen moduluak emandako errepresentazio linguistikotik hizketa sintetiko sortzen duen sintetizagailua.

Artikuluaren hurrengo sekzioan linguistika-moduluan egindako aldaketak azaltzen dira. Ondoren, 3. sekzioan TTS sistema sortzeko beharrezkoa den hizketa naturalaren datu-basearen diseinua eta grabaketa deskribatzen da. 4. sekzioak ahots sintetiko lortzeko

emandako pausuak aurkezten ditu eta 5. sekzioak sistemaren ebaluazioa zehazten du. Bukatzeko, 6. sekzioan zenbait ondorio aurkezten dira.

2. Hizkuntza-moduluaren moldaketa

AhoTTS eleaniztuna den TTS sistemak euskara baturako barneratzen duen hizkuntza-modulu nafar-lapurtarrera egokitu da soinu eta transkripzio arau berriak sortuz eta orain dituenak birpasatuz. Modulu linguistikoak testu normalizazioa egiten du. Horretarako, grafemetatik fonemetarako konbertsioa arau batzuen arabera gauzatzen da, alegia, lexikoi espezifikoa eta POS etiketaziorako desanbiguazio arau sinpleen arabera.

Lapurdin eta Nafarroa Beherean hitz egiten den euskalkiak desberdintasun fonetiko batzuk ditu euskara batuarekin konparatuz. Adibidez, euskara batuan /n/ eta /l/-ren palatalizazioa arrunta da /i/-ren aurretik doanean, nafar-lapurtar euskalkian ordea ez¹. Soinu berriak ere aurki daitezke, adibidez /h/, euskara batuan ahoskatzen ez dena baina euskalkian laringeko frikari ahoskabe bezala erabiltzen dena. Hala ere, kontsonante aspiratuak galtzen hasi dira euskalki nafar-lapurtarrean ere. Aldaketa handienetako bat /rr/ grafemari dagokio. Ahoskera batuan /rr/ hortz-hobietako dardarkari anizkun ahostuna ahoskatzen da, baina kontsonante honek artikulazio ubularra du (/R/) gaur egungo nafar-lapurtarrean.

Sistemaren garapenaren alderdi problematikoena nafar-lapurtar euskalkian izen berezi frantsesak erabiltzea arrunta izatea da. Izan ere, frantsesaren ahoskera arauak jarraituta esaten direnez, ondo ahoskatzeko beharrezkoa da frantsesaren soinu-sistema gehitzea euskararenari. Nahiz eta garrantzitsua izan TTSaren aplikazio errealerako, gure helburua lehen prototipo honetan euskararako garapenean zentratuta dago. Bitarteko soluzio bezala, frantsesaren fonema garrantzitsuenak hartu dira kontuan, /a~/, /e~/, /o~/, /y/, /2/, /v/ eta /Z/. Ez da frantsesaren transkripzio fonetiko araurik garatu, baina izen berezi frantsesak era egokian transkribatzeko horien transkripzioak dituen hiztegi bat prestatu da eskuz. Aukeraturako fonema frantsesak hiztegiaren bidez egindako transkripzio hauetan azaltzen dira bakarrik.

3. Hizketa datu-basea grabatzen

3.1. Corpora grabatzeko prestatzen

Euskara batua hizkuntza gutxitua da eta testu gutxi du eskura corpora sortzeko (Hernández et al, 2012). Baliabide gabezia hau oraindik handiagoa da euskalkia tratatzeko momentuan. Sortu nahi dugun TTSak domeinu murrizketarik ez duenez, testuak ahalik eta

iturri gehienetatik jasotzen saiatu gara. Testu soileko 90 MB baino gehiago lortu dira domeinu desberdinetatik: 45 MB Herria aldizkaritik, 9 MB XIX. mendeko liburuetatik eta 42 MB XX. mendeko liburuetatik. Liburu zaharrenak corpusaren parte izateko aukera izan da testuak ortografia modernoarekin ere eskura zeudelako.

Hasierako testu-corpora automatikoki ezabatu zen esaldi problematikoak ekiditeko (adibidez, haien transkripzio anomaloez analisi fonetiko distorsionatu dezaketena hitz kanpotarrak). Ondoren, transkripzio fonetikoak lortu ziren 2. atalean definitutako AhoTTS sistemaren hizkuntza-modulu egokituarekin.

Bukaerako grabaketa-corpora 4.000 esaldi izateko diseinatu zen. Hasierako testu-corpora erraldoi batetik esaldien azpimultzo bat aukeratzeko algoritmo *greedy* (Sesma eta Moreno, 2000) erabili da. Bukaerako esaldiak aukeratzeko erabilitako baldintzak difonemen estaldura maximizatzea zen horiek jasotako datuetan zuten frekuentziaren arabera, beti ere esaldi bakoitzaren hitz kopurua 15era mugatuz, corpora irakurterraza izateko. Soinu frantsesak corpusean egotea ziurtatzeko, bukaerako corpuseko 200 esaldi berezita aukeratu ziren izen berezi frantsesak zituen hasierako testu-corporusetik. Balio ez zuten esaldiak baztertu (adib. Errore gramatikaldun esaldiak) eta ortografia errore batzuk zuzendu ondoren, erauzitako esaldi guztiak iragazki batetik pasa ziren. Aukeraketa- eta zuzenketa-prozesu hau bost aldiz errepikatu zen. 1. taulan erakusten den bukaerako grabaketa-corpora lortuz. Konparazioak egin ahal izateko, 1 taulak antzekoa den euskara baturako egindako corpora [11] ere erakusten du. Euskara batuarekin konparatuz, euskara nafar-lapurtarraren corpusak fonema gehiago hartzen ditu kontuan, baina garrantzitsua da esatea horietatik 35 bakarrik direla euskalkiarenak. Gainontzeko 7ak fonema frantsesak dira, izen berezi frantsesak hobeto ahoskatzeko gehituak. Fonema hauetatik 5 bokalak direnez, difonema kopurua asko hazten da. 2. taulak erakusten du garaturako corpusean difonema ohikoenak zeintzuk diren eta euskara batuko difonema ohikoenekin konparatzen ditu. Ikus daiteke bi corpusek difonema ohikoenak konpartitzen dituztela.

| Kopurua | Navarro-Laburdian Corpus | Standard Basque Corpus |
|----------------------|--------------------------|------------------------|
| Esaldiak | 3999 | 3799 |
| Berbak | 38776 | 38544 |
| Fonema desberdinak | 42 (35+7) | 35 (34+1) |
| Difonema desberdinak | 1079 | 583 |

1. taula: Grabaketa-corporaren ezaugarri nagusiak, euskara batuarekin konparatuta (Sainz et al, 2012)

¹ Soinu kode guztiak SAMPA-n definituak. Euskara SAMPA http://aholab.ehu.es/sampa_basque.htm Frantsesa SAMPA <http://www.phon.ucl.ac.uk/home/sampa/french.htm>.

3.2. Hizlariak aukeratzen

Hizlarien aukeraketa casting arduratsu baten ondoren egin da, hiru gizon eta hiru emakume aukeratu dira, denak irrati-esatari profesionalak. Partaideei eskatu zitzairen testu handi bateko zenbait esaldi irakurtzea. Hizlariak nafa-lapurtar euskalkiaren transkripzio erregelak betetzeko abileziaren arabera aukeratu ziren, baita beraien hizketa seinalea erabiliz lortutako transkodedutako ahotsaren kalitatearengatik ere. Kalitate hau laborategian eginiko ebaluazio subjektibo eta informal batean definitu zen.

| Navarro-Laburdian Corpus | | Standard Basque Corpus | |
|--------------------------|---------|------------------------|---------|
| Difonema | Kopurua | Difonema | Kopurua |
| e-n | 5196 | e-n | 4666 |
| a-n | 4754 | a-n | 3645 |
| t-a | 4635 | t-a | 3224 |
| t-e | 3461 | t-e | 3138 |
| a-k | 3316 | a-k | 3956 |
| k-o | 3040 | k-o | 2571 |

2. taula: Grabaketa-corpusean gehien azaltzen diren diseinuak, euskara batuarekin konparatuta (Sainz et al, 2012)

3.3. Grabatze prozesua

Grabazioak *Gure Irratiaren* irrati estudio profesionalean egin ziren. *Gure Irratia* nafar-lapurtar lurraldean 1981etik finkatutako euskarazko irrati da. Pop iragazki bat jarri zen hizlaria eta mikrofono nagusiaren artean aire-jarioaren presioa gutxitzeko. Saio bakoitza kanpotik monitorizatu zen entzungailuekin.

Grabazioak egiteko 6 saio egin behar izan ziren emakumeentzat eta 5 gizonaentzat. Sesio-arteak aldakuntzak ekiditeko (adib. Ahotsaren kalitatean desberdintasunak, erritmoa, tonua, etab.) zenbait prebentzio-neurri jarraitu ziren. Mikrofonoaren kokapena eta hizlariaren eta mikrofonoaren arteko distantzia bera izan zen grabazio prozesu osoan. Hizlariak zenbait instrukzio eman zitzaizkien sesio luze eta jarraituetan irakurtzearen ondoriozko ahots-nekea ekiditeko. Esfortzurik gabe eta denbora luzez mantentzeko bolumentan hitz egiteko eskatu zitzairen. Sesio bakoitzaren hasieran seinalearen batez besteko anplitudea egokitzen zen azken grabaketa saioaren antzeko maila lortzeko. Hizlariak aurreko saioetako pare bat esaldi entzunarazten zitzairen erritmo eta tonua mantentzeko. Grabazioaren erdian hizlaria erreferentziatik asko urruntzen bazen, instrukzio berriak ematen zitzaizkien. Estiloari dagokionez, hizlariak eskatzen zitzairen estilo naturala erabiltzeko irakurtzean. Euskarazko eta frantseseko hitz batzuen ahoskatzeari buruzko instrukzio gehigarri batzuk ere eman zitzaizkien, transkripzio kanonikoak eta ahoskatze errealak ahalik eta antzekotasun handiena izan zezaten.

4. Ahots sintetikoaren sorrera

Kalitate handiko ahots sintetikoak lortzeko, grabatutako korporaren analisi eta anotazio zehatza pauso erabakigarria da. Hala ere, normalean etiketatzea zehaztasun handikoa izateak lan handia eta denbora asko eskatzen duen eskuzko oharren prozesua ekartzen du. Nafar-lapurtar euskalkiaren ahots sintetikoaren sorreran, etiketatze automatikoa aplikatzea erabaki zen, eskuzko interbentzio gutxi edo bat ere ez eskatzen duena. Ahotsa sortzeko pausua eman baino lehen, grabatutako seinaleen normalizazioa eta segmentazio-prozesua egin ziren ondorengo testuan azaltzen den eran.

4.1. Normalizazioa

Hasteko, uhin-formen fitxategi guztiak 16 KHz-era azpilagindu ziren eta hauen potentzia normalizatu zen. Normalizazio-prozesu hau oso garrantzitsua da bolumen desberdintasun handiegiak ekiditeko ahots bereko grabazioetan. Normalizazioa uhin-forma bakoitzean ondorengo eran egin zen: seinale bakoitzean ahots zatiak automatikoki determinatuak izan dira Praat-en (Boersma and Weenink, 2010) laguntzarekin. Potentzia batez bestekoa aldatu zen -25dB-era ITU-T-P.56-ak (ITU-T, 2011) espezifikatzen duen eran. Normalizazio-prozesuak seinalea saturaziora eramaten bazuen, zati problematikoak automatikoki detektatuak eta egokiro ahulduak ziren leihok zuzen baten barruan. Leihok horren mugak eremu problematikoaren kanpoan zeroa zeharkatzen duen puntu hurbilenean daude. Hurbilketa simple honek adierazpen batzuen hasieran gehiegizko bolumena ekiditen du eta esaldien potentzia inguratzailea mantentzen du.

4.2. Segmentazioa

Grabazio-corpuseko esaldiak eskuz berrikusi eta zuzendu ziren hizlariak esandako esaldiekin bat etor zitezkeen. Etenak ere sartu ziren eta puntuazio zeinua gehitu edo ezabatu ziren hiztunak egindako geldialdiekin bat etor zitezkeen. Azkenik, hiztun bakoitzeko lexiko pertsonalizatua sortu zen kontuan hartutako transkripzio arauak betetzen ez zituzten ahoskatzeak barneratzeko.

2. atalean azaldutako fonetika-transkribatzailea testu-fitxategiz elikatuta eta dagokion lexiko pertsonalizatua erabilia, fonema-sekuentziak eta ortografia-etenaldiak sortu ziren. Fonema sekuentzia horiek, normalizatutako seinaleekin batera, lerrotatze behartuan oinarritutako hizlari automatikoaren menpekora den segmentazioa egiteko erabiltzen ziren. HTK tresna (Young et al, 2006) erabili zen segmentazio prozesuan. Lehenik eta behin, lotutako egoeradun trifonema ereduak entrenatu ziren hasiera arrunt batean, etenaldi laburrak hitzaren mugetan txertatu gabe. Azkenik, etenaldien alboko fonemen segmentazio-mugak birdefinitu ziren potentzia inguratzailea eta iraupen atipikoak erabiltzen dituen algoritmo simple eta eraginkor baten bidez.

4.3. Ahotsaren eraketa

Lehenik, ahots seinaleak Aholab taldean garatutako kalitate handiko AHOCODER bokoderraren bidez (Erro et al, 2011) aztertu ziren. Ondoren, hizkuntza-etiketa egokiak prestatu ziren (Erro et al, 2010) eta HTS sistema (Zen et al, 2006) erabili zen HMM ereduak entrenatzeko. Orain arte emakumezko ahotsa besterik ez da sortu. Sintesi-prozesuan soinu ahostunen arteko ez-jarraitasunak ikusi ziren, nagusiki hitzen mugetan. Arazo honen eragina gutxitzeko, atalase ahoskabeak 0,5etik (balio lehenetsia) 0,2ra murriztu da. Horrela, zati gehiago jotzen dira ahostuntzat.

5. Ebaluazio subjektiboa

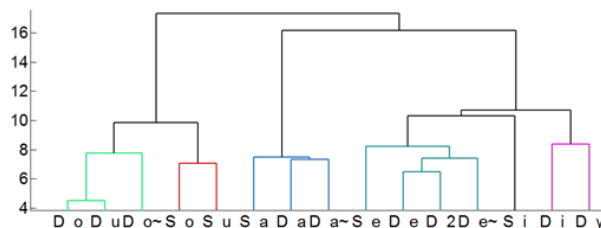
TTS sistemen ebaluazioa proba subjektiboen bidez egin izan ohi da, non benetako erabiltzaileek sistemaren alderdi desberdinei buruz beren iritzia ematen duten. Ebaluazio honen xedea hizkuntza-modulua egokitzearen eta euskalki baten hizketa sintetikoa sortzeko ahotsa grabatzearen egokitasunari buruz informazioa lortzea da. Hori dela eta hiru sistema mota sartu dira ebaluazioan, denak AhoTTSn oinarrituak eta lexikoa konpartitzen dutenak:

- Euskara batua (estandarra bezala definitua 2 eta 3 irudietan)
- Artikulu honetan azaltzen den lanean garatutako sistema dialektala (dialektala)
- Sistema mistoa, 2. atalean eta euskarazko TTS estandarren amaieran deskribatutako modulu linguistiko aldatuak erabiltzen dituenak.

5.1. Sistema Mistoaren Garapena

Sistema mistoa eraikitzeke, hizkuntza-moduluaren egokitzea bakarrik behar da, ez da grabaziorik egin behar eta ez da testu corpusik prestatu behar. Baina, soinu-multzoa desberdina denez euskara batuan eta Nafar-lapurtar euskalkian, beharrezkoa da dialektoen soinuak gertuen dituzten estandarera mapeatzea. Horretarako, *clustering* fonetiko itsua egin da. Gauss-en nahasketen eredu (GMM) entrenatu da 12 MFCCrekin hizketa sintesiaren datu-base dialektala erabiltzen duen soinu bakoitzarengatik eta baita euskara batuan egindako emakumezkoen hizketaren sintesirako datu-base batez, AhoSyn (Sainz et al, 2012) hain zuzen. Eredu horiek baliagarri dira soinu bakoitzaren ezaugarri fonetikoak era konpaktuan irudikatzeke eta soinuak aplikatzeko erregresio zuhaitzak taldekatzeke erabiltzen dira. 1. irudian erakusten dira bokalak taldekatzerakoan lortutako dendogramak. Euskara estandarren bokalak "S" aurrizkiarekin adierazten dira SAMPA kodearen aurretik eta euskalkiaren sistemako bokalak "D" atzizkiarekin adierazten dira. Adibidez, 1. irudian ikus daiteke /e~/, /2/ eta /e/ euskalkiaren

bokalak /e/ estandarretik gertuago daudela beste edozein baino, beraz, euskalkiaren modulu linguistikotik aterako diren hiru soinu hauek /e/ estandarera mapeatuko dira, sistema nahasiaren bukaera estandarrean sintetizatuak izateko. Era berean, gainontzeko bokal eta kontsonante berriak euskara batuaren soinu-multzoko soinueta mapeatu dira Euskal phoneset estandarrean. Multzokatze itsu honen ondorioz, /R/ soinua /G/-ri mapeatua azaltzen da eta /h/ /x/-ri. Sudur bokalen kasuan, /n/ fonema bat dago sartuta mapeatutako bokal ostean sudurkaria zaintzeko.

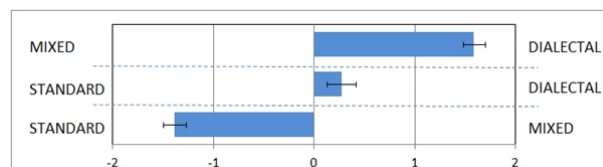


1. irudia: Bokal-soinuen dendograma

5.2. Ebaluazioaren emaitzak

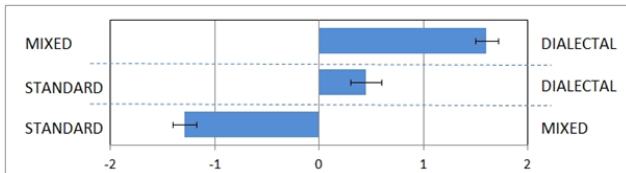
Lan honetan ebaluazio konparatiboak esleitu dira. Ebaluatzaileek web interfaze baten bidez entzun behar dituzte esaldi baten bertsiotik sintetiko desberdinak eta gehien gustatu zaiena aukeratu behar dute. Bost puntuko eskala erabili da, -2tik 2ra graduatzen, non, -2ak esan nahi duen "Oso garbi daukat lehen seinalea nahiago dudala", -1ak "Gutxigatik nahiago dut lehen seinalea", 0ak "Ezin dut aukeratu, uste dut biak kalitate bera dutela", 1 "Gutxigatik nahiago dut bigarren seinalea" eta 2ak "Oso garbi daukat bigarren seinalea nahiago dudala".

Bi esaldi mota sartu dira ebaluazioan: Esaldien erdiak hitz guztiak ditu euskarazkoak eta beste erdiak frantsesezko izen bereziak ditu. Era honetan izen frantsesen presentziak erabiltzaileek sistema bat edo bestea aukeratzeari eragiten dion zehaztu daiteke.



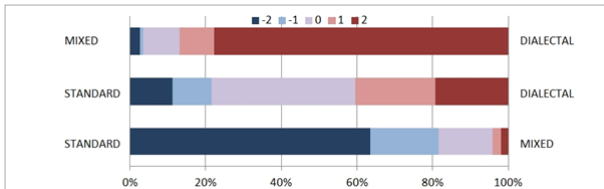
2. irudia: Euskarazko testuak izen frantses gabe erabiltzean lortutako emaitzak, %95eko konfiantza tartearekin.

Hogeita sei pertsonak, denak nafar-lapurtar euskalkiaren natiboak, hartu dute parte ebaluazioan. Izan frantsesik gabeko emaitzak 2. irudian erakusten dira eta izen frantsesdunak 3. irudian.



3. irudia: Euskarazko testuak izen frantses gabe erabiltzean lortutako emaitzak, %95eko konfiantza tartearekin.

Erabiltzaileek era argian nahiago dituzte euskalkiaren edo euskara batuaren sistema nahastua baino. Euskalki sistema ere nahiago dute euskara batuarena baino bi testu motetarako, baina diferentzia ez da sistema nahastuarekin lortu dena bezain handia. Hau izan daiteke euskara batuaren ahotsaren kalitatea altua delako, zeinetarako data-basearen segmentazioa eskuz birpasatua izan den. Beste alde batetik, euskara batuak medioetan duen posizio menderatzailearen ondorioz, Behe Nafarroako eta Lapurdiko hiztunak ohituta daude euskara batua entzuzera eta ondo barneratzen dute, nahiz eta beraien euskalkia ez izan. Nahiz eta konfiantza tartek ez duten 0 balioa hartzen kasuen %95ean, aukera guztiak esanguratsuak dira. 4. irudiak erakusten du TTS sistemen puntuazio-banaketa izen frantsesik gabeko testuak erabili direnean. Euskara batua eta euskalkiaren arteko konparazioan, jendearen %40k euskalkia nahiago izan du eta %20k euskara batua aukeratu du. 2. irudian ikusi den desberdintasun txikia bi sistemen artean aukeratu ezin izan duen % 40-aren ondorioz gertatu da.



4. irudia: Euskarazko testuak izen frantses gabe erabiltzean lortutako emaitzak, %95eko konfiantza tartearekin.

6. Ondorioak

Nafar-lapurtar euskalkiarentzat TTS sistema berri bat garatu da, euskara batuaren TTS sistema den AhoTTS erabiliz oinarri bezala. Nafar-lapurtar euskalkiaren 90 MB testu plano bildu dira eta fonetikoki orekatutako 4.000 esaldi aukeratu dira grabaketa-corpus batetik. Ahotsa sortzeko kalitate handiko hizketa-kodifikatzailea erabili da, HMMan oinarritutako hizketa sintesi testuinguruan. Euskalkiaren TTS berria sistema subjektibo baten bidez ebaluatu da, euskara batukoarekin eta baita euskalkiaren transkripzio arauak aplikatzen dituen eta euskara batuaren hizketarako modulu sintesi-sortzailea erabiltzen duen sistema nahastu batekin konparatuta. Aurre-moduluaren egokitzapena, transkripzio arau fonetiko berriak eta soinu berriekin, ez da nahikoa euskara batuaren sistema baino hobeto dabilen sistema bat lortzeko, gutxienez

soinu-multzokatze itsua aplikaturik lan honetan. Soinuak mapeatzeko, jakintzan oinarritutako ikuspuntu linguistikoa erabiltzeak emaitza hobek eman ditzake. Grabaketarik ezin bada egin, gomendagarriagoa da euskara batuaren sistema erabiltzea. Baina euskalkiaren ahots berriak erakusten du erabiltzaileek nahiago dutela euskalkiaren sistema euskara batuarena baino.

7. Esker onean

Autoreek eskertu nahi dute Iñaki Sainz proiektu honetan egin duen lan handiarengatik, Josu Landa grabaketa-corpusa sortzeko euskalkiaren testuak bildu eta hornitu dituelako eta ebaluazio subjektiboan parte hartu duen jende guztia.

Lan hau L'Euroregion Aquitaine-Euskadi-k IPARRAHOTSA proiektuaren barruan, Espainiako Ekonomia eta Kompetentzien Ministerioak (SpeechTech4All project, TEC2012-38939-C03-03) eta Eusko Jaurlaritzak (Ber2tek project, IE12-333) finantzatu dute.

8. Aipamenak

- Beskow, J., Gustafson, J. (2009). Experiments with Synthesis of Swedish Dialects. *FONETIK 2009*, 28-29. (Stockholm.)
- Pucher, M., Schabus, D., Yamagishi, J., Neubarth, F., Strom, V. (2010) Modeling and interpolation of Austrian German and Viennese dialect in HMM-based speech synthesis. *Speech Communication*, 52(2), 164-179.
- Hu, Q., Tao, J., Zhao, C.(2011). HMM-based Tianjin Dialect speech synthesis using bilateral question Set. *2011 IEEE International Workshop on Machine Learning for Signal Processing 2011*, 1-4. Beijing.
- Langa, R., Manamela, J., Gasela, N. (2012). Synthesis of dialect speech for an underresourced language. *SATNAC 2012*.
- Yoshimura, T., Tokuda, K., Masuko, T., Kobayashi, T., Kitamura, T. (1999). Simultaneous modeling of spectrum, pitch and duration in HMMbased speech synthesis. *EUROSPEECH 1999*, 2374-2350.
- Zen, H., Tokuda, K., Black, A.W. (2009). Statistical parametric speech synthesis. *Speech Communication*, 51, 1039-1064.
- Basque Government: *Fifth Sociolinguistic Survey: Basque Autonomous Community, Navarre and Iparralde* http://www.euskara.euskadi.net/r59-738/en/contenidos/informacion/sociolinguistic_research2011/en_2011/2011.html, 2014
- Hernaiz, I., Navas, E., Murugarren, J. L., Etxebarria, B (2001). Description of the AhoTTS Conversion System for the Basque Language. *4th ISCA Tutorial and Research Workshop (ITRW) on Speech Synthesis*, paper 202.
- Hernaiz, I., Navas, E., Odriozola, I., Sarasola, K., Diaz de Ilarraza, A., Leturia, I., Diaz de Lezana, A., Oyharcabal, B., Salaberria, J.(2012). The Basque language in the digital age/Euskara aro digitalean. *METANET White Paper Series*. Springer

- Sesma, A. Moreno, A.(2000): CorpusCrt 1.0: Diseño de corpus orales equilibrados (in Spanish). *Technical report*, UPC
- Sainz, I., Erro, D., Navas, E., Hernaez, I., Sanchez, J., Saratxaga, Odriozola, I. (2012). Versatile Speech Databases for High Quality Synthesis for Basque. *LREC 2012*, 3308-3312, Istanbul.
- Boersma, P., Weenink, D. (2010): *Praat: doing phonetics by computer [Computer program]*. Version 5.1.38, retrieved 2 June 2010 from <http://www.praat.org/>
- International Telecommunication Union (ITU-T) (2011). Recommendation ITU-T P.56, Objective measurement of active speech level [On-line] <https://www.itu.int/rec/T-REC-P.56-201112-I/en>
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P. (2006). *The HTK Book*, version 3.4
- Erro, D., Sainz, I., Navas, E., Hernández, I. (2011). HNM-Based MFCC+f0 Extractor Applied to Statistical Speech Synthesis. *ICASSP 2011*, 4728-4731 Florence.
- Erro, D., Sainz, I., Luengo, I., Odriozola, I., Sanchez, J., Saratxaga, I., Navas, E., Hernaez, I. (2010). HMM-based Speech Synthesis in Basque Language using HTS. *FALA 2010*, 67-70, Vigo
- Zen, H., Nose, T., Yamagishi, J., Sako, S., Black, A. W., Masuko, T., Tokuda, K.(2006). The HMM-based speech synthesis system (HTS) version 2.0. *SSW6*, 294-299, Bonn.