



Elhuyarren hizketa-sintesi neuronal berria

ARG.: Jackie Niam/Shutterstock.com

2014. urtean zerbitzu teknologiko berri bat plazaratu zuen Elhuyarrek: hizketaren sintesia. Testuak audio bihurtzea ahalbidetzen duen teknologia hori modu askotako zerbitzuak eskaintzeko baliatu izan dute geroztik gure bezeroek. Hizkuntza- eta hizketa-teknologiek biziki egin dute aurrera sare neuronalen teknologiari esker, eta Elhuyarren ere hizketa-sintesiko teknologia neuronal propioa garatu dugu, lehengo baino kalitate hobearekin eta aukera berriekin. Ezagut dezagun hizketa-sintesi neuronalaren web-zerbitzu berria.

[Hizketa-teknologien](#) barruan, bi teknologia dira nagusi: [ASR \(Automatic Speech Recognition\) edo hizketaren ezagutza](#) deritzona, zeina hizketa-audio bat transkribatzean edo testu bihurtzean datzan, eta [TTS \(Text-To-Speech\) edo hizketaren sintesia](#) deitutakoa, zeina testu bat irakurri edo audio bihurtzean datzan. Duela urte batzuetatik, bi-biak lantzen ditugu euskararentzat Elhuyarrek sortutako [Orai adimen artifizialeko](#) zentroan, eta, bi-bietan oinarrituta, euskaraz dabilzan zerbitzuak

gizarteratzen ditu Elhuyarrek. ASRari dagokionez, [Adu.eu](#)s transkripzio-, azpitolazio- eta diktaketa-zerbitzua [2020an jarri genuen martxan](#). TTSari dagokionez, [2014tik dago online testua ahots bihurtzeko zerbitzua eta webgunea](#).

Zerbitzu hori [Euskal Herriko Unibertsitateko Aholab ikertaldea](#)k garatutako [AhoTTS](#) teknologian oinarrituta zegoen; garai horretan zegoen teknologiarik onenarekin garatuta zegoen, eta euska-

Igor Leturia Azkarate
Informatikaria eta ikertzailea



raz zebilen bakarra zen. Eta ordutik hona igarotako urteotan, hainbat toki eta kasutan erabili da teknologia hori: zenbait webguneren edukia irakurri beharrean entzutea ahalbidetzeko ([Elhuyar aldizkaria](#), [Zientzia.eus](#), [EITB.eus](#), [Sarean.eus](#), [ezgaitasunak dituzten pertsonentzako EHUko zerbitzua](#), [irakasleek material eta baliabide pedagogikoak partekatzeko Hezkuntza Sailaren Aamaraua webgunea...](#)), dislexia edo irakurtzeko bestelako zailtasunak dituzten ikasleek webguneak eta dokumentuak hobeto barneratzen laguntzeko Hezkuntza Sailaren [Irakurle Digitalean](#), [Elhuyar hiztegia](#) edo Justizia Sailaren [Justiziagelako](#) euskara-ikastaroetan hitzak nola esaten diren erakusteko, [Mycroft](#) euskarazko bozgorailu adimendunear...

TTS teknologia neuronal berria

Ordutik hona, [hizkuntza- eta hizketa-teknologia](#) guztiak, TTSa barne, [sare neuronal sakonak \(deep neural networks\)](#) edo [ikasketa sakona \(deep learning\)](#) izenez ezagutzen den teknologiaren bidez funtzionatzera pasatu dira, emaitza askoz hobek ematen dituelako. Eta orduan sintetikoki sortzen genuen hizketa nahiko naturala zela bagenioen ere (eta hala zen orduko estandarrentzat), gaur egun sare neuronalekin lortzen dena askoz ere naturalagoa da, ia-ia benetako hizketatik bereizezina izateraino.

Bada, azken urte parean-edo, Orai-n euskarazko hizketa-sintesi neuronalak garatzen aritu gara, eta jada badugu gure sistema propioa. Oso kalitate onekoa da, ahoskeran, intonazioan, prosodian... bene-benetako hizketa dirudiena. Gainera, egungo

teknologiekin lehen gehiago kostatzen zen gauzak egin ditzakegu eta funtzionalitate berriak izan. Adibidez, lehengo sisteman, sortu nahi zen ahots ezberdin bakoitzeko, eredu bat entrenatu eta grabazio-kopuru askotxo behar izaten zen. Gaur egun, aldiz, eredu bakarrean ahots askoz gehiago izan ditzakegu, eta grabazio-denbora askoz gutxiagorekin; beraz, errazago sor ditzakegu ahots sintetiko berriak.

“Pertsona bat beste hizkuntza batean ‘hitz egiten’ jar dezakezu, hizkuntza horren ideiarik ere izan gabe”

Horrez gain, eredu eleaniztunak ere sor daitezke, eta sei hizkuntzatan sortu ditugu: euskara, gaztelania, frantsesa, ingelesa, katalana eta galegoa. Haien bidez lor daiteke pertsona batekin grabazioak hizkuntza jakin batean (demagun, euskaraz) egitea, baina gero grabazio horiekin entrenatutako ereduak gai izatea beste hizkuntza batean (demagun, ingelesa, edo frantsesa, edo katalana) sintesia egiteko pertsona horren ahotsarekin! Hau da, pertsona bat beste hizkuntza batean “hitz egiten” jar dezakezu, hizkuntza horren ideiarik ere izan gabe!

Web-zerbitzua, aukera eta erabilera anitzekoa

Aurten, Elhuyarrek martxan jarri du teknologia neuronal berrian oinarritutako web-zerbitzua, <https://ttsneuronal.elhuyar.eus/> helbidean.

“Esaldi batzuk irakurtzen hamar minutu grabatu besterik ez da behar, gure ahotsarekin hizketa-sintesia egin ahal izateko”

Aipatutako sei hizkuntzen artean eta haietako bakoitzean bi edo lau ahots ezberdinen artean aukeratu dezakegu, testu bat eman, eta berak hizketa bihurtuko du. Ahotsen kalitatea webguneko testu-kutxaren bidez egiazta dezakegu.

Gainera, nahi badugu, gure ahots pertsonalizatua ere sor dezakegu, guk bakarrik erabili ahal izango duguna. Horretarako, esaldi batzuk irakurtzen hamar minutu grabatu besterik ez da behar, eta ondoren gure ahotsarekin hizketa-sintesia egin ahal izango dugu, grabatutako hizkuntzan edo beste edozeinetan. Horrela sortutako ahots pertsonalizatuen adibideak [Elhuyaraldizkariaren](#) edota [Goienaren](#) webguneetan entzun ditzakegu (Elhuyarren kasuan, beste hizkuntzetara pasatuta ere bai).

Teknologia baliatzeko, modu bat baino gehiago dago. Sinpleena eta errazena testu-kutxaren bidezkoa da: hartan, nahi dugun testua itsatsi, eta audioa sortuko dugu. APlA ere eskaintzen dugu, eta horrela gure aplikazioan edo zerbitzuan txerta daiteke. Eta webgune bat irakurri beharrean entzuteko aukera eman nahi bada, barra erreproduzitzaila baten kodea ere ematen dugu, webgunean erraz-erraz txertatuta aukera hori eskaintzeko.

TTS zerbitzuaren aurretiko bezeroak zerbitzu eta ahots berrietara migratu ditugu jada, eta hainbat enpresa berri ere ari dira baliatzen ([Tokikom](#), [Skura](#), [Batura](#), [Ulma](#), [Ibil](#), [Naiz...](#)), haietako asko ahots pertsonalizatuekin.

Eta zer erabilera du benetan TTS teknologiak, zertarako ari dira erabiltzen bezero horiek? Bada,

erabilera posible anitz ditu. Webguneak barra erreproduzitzailaren bidez irisgarriago edota eskuragarriago egitea da ohikoenetako bat (mugikorrean oinez edo garraio publikoan goazela entzuteko, adibidez). APlaren bidez, eta ASRarekin konbinatuta, makinekin edo app-ekin elkarrekintza hizketa bidez egitea ere ahalbidetzen da. Testu-kutxa erabilia, podcast entzungarri bat sor dezakegu zuzenean testutik grabatu beharrik gabe, edo gure ikus-entzunezkoarentzat off-eko ahotsak eta bestelakoak sortu. Etorkizun hurbilean, bikoizketa (erdi)automatiko egin ahal izatea ere posible izango da, [Aditu gure azpitulazio eta itzulpen automatiko zerbitzuari](#) TTSa gehituz.

Oraingoz, gure TTS teknologiak hizketa neutroa sortzen du, eta hori nahikoa da komunikabide edo webgune bateko edukia irakurtzeko, makina batek hitz egiteko edo off-eko ahotsetarako. Baina Orain norabide askotan ikertzen jarraitzen dugu, ahots emoziodunak ere edukitzeko, sortzen den hizketa parametrizatu eta nahieran moldatu ahal izateko (tarte bakoitzaren abiadura, intonazioa, bolumena...), grabaketak eginda eredu propioak entrenatu beharrik gabe lagin txiki batekin ahots bat imitatuz sintesia egin ahal izateko... Gauza guztiz beharrezkoak, horrelako tresnak gero eta presenteago dauden mundu gero eta teknologizatuagoan eus-kara ere presente egon dadin. ●