

C O D
Corpus Oral Dialectal

Departament de Filologia Catalana



Reconeixement de veu

Maria Pilar Perea (coordinació)

Desenvolupat per

grubit

Finançat pels projectes "Explotació d'un corpus oral dialectal" [HUM2004-01504/FILO] i "Explotació d'un corpus oral dialectal: anàlisi de la variació lingüística i desenvolupament d'aplicacions informàtiques per a la transcripció automatitzada (ECOD)" [HUM2007 65531/FILO].

Conceptes previs: estudi fonètic

- El programa WavFonemes emmagatzema les síl·labes (normalment difonemes) d'un conjunt de paraules (al voltant de 200) emeses per diferents locutors, les quals serveixen com a base per a l'estudi de les diverses característiques fonètiques:
 - Magnitud
 - Energia
 - Duració
 - Freqüències amb el seu pes, representat pels valors Fourier

Conceptes previs: estudi fonètic

- L'estudi fonètic es pot dividir en 2 fases:
 - Fase d'investigació: ús de mètodes estadístics amb l'objectiu d'estudiar el comportament de les dades recollides d'acord amb les seves diferents característiques fonètiques.
 - Fase d'aplicació: desenvolupament de les xarxes neuronals per millorar el disseny dels patrons, com el també el reconeixement.

Conceptes previs: estudi fonètic

Fase d'investigació

- Els mètodes usats són els següents:
 - **Distribució normal:** indica que el 90% dels valors més grans del fonema que es vol reconèixer equival amb els del patró del fonema.
 - **Interpolació de Lagrange i interpolació trigonomètrica:** indica que els valors correlatius d'una funció no són dispars.
 - **Regressions:** s'han aplicat diversos tipus, com ara la **lineal, logarítmica, exponencial i potencial.**

Conceptes previs: estudi fonètic

Fase d'aplicació

- WavFonemes v1: el disseny del patró s'ha fet prenent com a base les freqüències més representatives de cada síl·laba (valors Fourier més elevats), independentment del locutor, i aplicant una mitjana d'aquests valors per a un mateix fonema, cosa que pot generar un patró ineficaç en el cas de trobar valors molt dispars entre ells
- WavFonemes v2: s'empren les xarxes neuronals (algorisme de Back Propagation) per resoldre relativament el problema anterior, ja que durant el procés d'aprenentatge s'usa una funció que aporta més flexibilitat, i sempre condicionat amb un error mínim establert com a paràmetre

Funcionalitats

- **WavFonemes** és un programa que serveix per:
 - En primer lloc, capturar i dissenyar una base de dades de sons en l'àmbit de llengua catalana
 - En segon lloc, normalitzar la base de dades (disseny de patrons)
 - Com a objectiu final, desenvolupar la capacitat per reconèixer els sons a partir d'aquesta base.

Funcionalitats

- El programa WavFonemes permet:
 - Retallar parts d'un arxiu wav, assignar-hi la sí·laba fonètica corresponent, i guardar la informació a la base de dades
 - Compta amb 2 funcionalitats:
 - **Normalitzar** les sí·labes fonètiques que ja es troben emmagatzemades a la base de dades
 - **Reconèixer** les sí·labes fonètiques després d'haver fet el procés de normalització i a partir de l'aplicació dels diferents mètodes estadístics

Funcionalitats: normalització

- L'objectiu és construir un patró fonètic vàlid per al reconeixement, en el qual només es prenen les freqüències més representatives de cada síl·laba independentment de l'informant, les quals són les que tenen uns valors més elevats
- Es tenen en compte, però, els següents paràmetres per obtenir una bona normalització:
 - Freqüències dins el llindar de la capacitat de l'oïda humana, que oscil·la entre 20 i 10000 Hz
 - Percentatge de recollida de valors, que s'estableix entre el 70 % i el 80 %
 - Es pretén generalitzar; no diversificar

Opcions de Normalització i Reconeixement del fonema

NORMALITZACIÓ DELS FONEMES

Freqüències

Mínima: 20 Hz Màxima: 10000 Hz

Valors (pesos de la freqüència)

Recollida dels valors: 80 %

RECOONEIXEMENT DELS FONEMES

Interval respecte al fonema normalitzat

Marge d'error per sota: 2 % Marge d'error per sobre: 2 %

Reconeix els fonemes al moment d'obrir l'arxiu de so

Acceptar Cancel·lar

Funcionalitats: normalització

Patrons fonètics

- El sistema de reconeixement utilitzats és el que es basa en el de **patrons de veu**
- La base de dades de patrons s'obté mitjançant la selecció de grups significatius de sons prèviament enregistrats a partir d'enquestes a un conjunt d'informants
- Els grups significatius són les **síl·labes**, les quals normalment es poden considerar difonemes (entitats formades per dos sons)

Funcionalitats: normalització

Patrons fonètics

- A l'hora d'incorporar patrons a la base de dades només es tenen en compte les seves característiques espectrals.
- El pas previ consisteix en la transformació del senyal de veu enregistrat directament amb el micròfon (domini en el temps) a freqüències (domini de la freqüència), ja que la informació identificativa i invariable de les vocals i consonants independentment del locutor es codifica en el domini de la freqüència

Funcionalitats: normalització

Patrons fonètics

- S'ha d'aplicar un mecanisme que traspassi a freqüències la informació existent en les mostres de veu que s'han recollit en el domini del temps. Per fer aquesta transformació s'usa la fórmula "Transformada de Fourier" (FFT).
- La "Transformada de Fourier" es basa en el fet que tot senyal sonor, per complex que sigui, es pot descompondre en una suma de funcions periòdiques (sinoïdals i cosinoïdals) simples de diferent freqüència.
- El que s'obté després d'aplicar la fórmula són uns valors que determinen la importància de les freqüències. Les més significatives són les que componen i modelen el patró que es vol crear.

Funcionalitats: normalització

Patrons fonètics

- **L'obtenció de patrons de qualitat** consisteix bàsicament en 2 passos:
 - **Ensinistrament de patrons** (disseny dels patrons): s'intenta d'establir representacions consistents de patrons de veu mínimament fiables.
 - **Comparació de patrons** (reconeixement a partir de patrons dissenyats): es fa una comparació directa entre el senyal desconegut (que cal reconèixer) i tots els possibles patrons adquirits en l'etapa d'entrenament, i es determina quin és el patró que millor s'hi ajusta.

Funcionalitats: normalització

Patrons fonètics

- Inconvenients en el mètode de comparació de patrons:
 - Dificultat per crear una base de dades de patrons de la parla que resulti completa, correcta i, sobretot, significativa (en el nostre cas, de moment ens serveix per a paraules aïllades i per a una base tancada).
 - La informació espectral pot tenir un significat relatiu, a causa de les variabilitats en:
 - Mantenir una pronunciació consistent i constant.
 - Diferències de longitud del tractament de les vocals i consonants, accents que depenen del dialecte, etc.
 - Diferents micròfons quan s'enregistra la veu, la intensitat, etc.
 - L'ambient, com ara sorolls de fons, etc.

Funcionalitats: reconeixement

Mètodes estadístics

- L'objectiu és que, a partir d'un arxiu wav es mostri en la pantalla la divisió de síl·labes fonètiques i es pugui identificar a quina síl·laba fonètica pertany a partir dels patrons emmagatzemats en la base de dades
- S'empren diferents mètodes estadístics per estudiar la dependència entre el valor Fourier i la seva freqüència

Funcionalitats: reconeixement

Mètodes estadístics

- S'ha experimentat la relació de dependència a partir de les següents paraules pronunciades per un mateix locutor masculí:

aigua	alls	alt	alts	ametller	ametlles			
arbres		arc	arcs	avui	becaina	bou		
bouet		buida	cadena	cadira	calor	carn	carrer	
casa		cases	caseta	casetes	cassola	cell	ciutat	
coix		collar	corona	coroneta		cuina	cuinar	
cuixa		despert	despertador	deu	dinar	dissabte		
diumenge		dur	encara	ensenyar		eriçó	és	escola
espiga		esternudar	estrella	estret	ets	examen	faixa	
feina		ferida	ferradura		ferro	flor	foradar	
formatge		fuig	fulla	fusta	fusteta	gelat	gent	germà
gerret		gerro	herba	hort	horts	jove	llaurador	llengua
llet		llibre	llibret	lluna	madeixa	metge		
mobles		mocador	molc	nedar	nou	oberta	ou	
ouet		paper	part	passejar	pebrots		peix	
pèl		pell	pelut	perla	platja	regla	roda	
rodeta		rodó	roja	rosa	sabater		sant	sants
serp		set	sóc	solc	són	talc	taronger	
taronja		tela	teler	tots	tovallola	triple		
tronc		vaca	vesprada	vi	vides	viure	xiquet	

Funcionalitats: reconeixement Mètodes estadístics

- Resultats per a la paraula 'aigua':

Resultats del reconeixement

Arxiu

Fonemes candidats per a cada sil·laba de l'arxiu aigua_dcr_bellloc.wav

M0-Distr.normal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'a' (Val.:100%)	rə (Val.:82,79)	na (Val.:81,44)	tʃɪw (Val.:78,3%)	'ne (Val.:77,86)
2a sil·laba	wa (Val.:100%)	'dew (Val.:89,19)	'troŋ (Val.:87,55)	dwa (Val.:87,38)	ðwa (Val.:87,37)

M1-Int.Lagrange	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'a' (Val.:100%)	rə (Val.:73,92)	tʃɪ (Val.:72,66)	eg (Val.:70,63)	ma (Val.:69,88)
2a sil·laba	wa (Val.:100%)	'troŋ (Val.:85,87)	'kam (Val.:84,04)	za (Val.:83,76)	'dew (Val.:83,59)

M2-Regr.Lineal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'ju (Val.:84,83)	ro (Val.:84,3%)	'ðeɪ (Val.:79,03)	'pɛr (Val.:78,94)	'ðu (Val.:78,89)
2a sil·laba	'prou (Val.:90,42)	'kuɪ (Val.:89,66)	'kaɪ (Val.:88,22)	bo (Val.:86,98)	eɪ (Val.:86,09)

M3-Regr.Log.	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'kuɪ (Val.:36,42)	ka (Val.:36,13)	'sap (Val.:35,95)	'pɛr (Val.:35,89)	be (Val.:35,72)
2a sil·laba	'sans (Val.:2,93%)	'troŋks (Val.:2,91%)	dɪ (Val.:2,9%)	'molɪ (Val.:2,9%)	'ka (Val.:2,88%)

Guarda
Cancel·la

Funcionalitats: reconeixement Mètodes estadístics

- Resultats per a la paraula 'bou':

Resultats del reconeixement

Arxiu

Fonemes candidats per a cada sil·laba de l'arxiu bou_dcr_bellloc.wav

M0-Distr.normal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'a.ʎ (Val:73,19)	'bou (Val:72,28)	'san (Val:71,37)	'med (Val:71,24)	'gen (Val:67,67)

Guarda
Cancel·la

M1-Int.Lagrange	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'bou (Val:75,87)	're (Val:73,65)	'med (Val:72,69)	'a.ʎ (Val:71,38)	'ɟə (Val:70,48)

M2-Regr.Lineal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'sap (Val:83,07)	'pe (Val:82,91)	'mad (Val:81,05)	'bo (Val:80,5%)	'prow (Val:80,28)

M3-Regr.Log.	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'ret (Val:1,71%)	'ets (Val:1,71%)	'des (Val:1,69%)	'orts (Val:1,67%)	'sok (Val:1,67%)

Funcionalitats: reconeixement Mètodes estadístics

- Resultats per a la paraula 'cadeneta':

Resultats del reconeixement

Arxiu

Fonemes candidats per a cada sil·laba de l'arxiu cadeneta_dcr_bellloc.wav

M0-Distr.normal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	'koj (Val.:78,68)	ʃə (Val.:77,39)	'sans (Val.:76,77)	ro (Val.:76,32)	'troŋks (Val.:76,16)
2a sil·laba	ðe (Val.:100%)	'ro (Val.:68,65)	'ðe (Val.:59,88)	ri (Val.:58,36)	'ɫo (Val.:57,14)
3a sil·laba	'ci (Val.:77,66)	'ðu (Val.:75,45)	βa (Val.:73,21)	'le (Val.:72,76)	'ɫo (Val.:72,67)
4a sil·laba	'aɫ (Val.:78,94)	'med (Val.:78,63)	'gen (Val.:78,6%)	'kam (Val.:77,81)	'dew (Val.:77,25)

M1-Int.Lagrange	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	ro (Val.:70,16)	'le (Val.:66,58)	'ði (Val.:65,93)	en (Val.:65,81)	tʃe (Val.:64,91)
2a sil·laba	ðe (Val.:100%)	'ro (Val.:61,99)	'ðe (Val.:60,04)	ri (Val.:59,41)	'ɫo (Val.:57,18)
3a sil·laba	'ci (Val.:71,65)	'ðu (Val.:70,21)	'le (Val.:70,11)	'ɫo (Val.:70,1%)	βa (Val.:67,9%)
4a sil·laba	'pe (Val.:79,05)	ple (Val.:76,55)	'mad (Val.:75,02)	ko (Val.:72,56)	'sel (Val.:71,92)

M2-Regr.Lineal	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	ðe (Val.:77,31)	ri (Val.:75,82)	'ro (Val.:70,5%)	'a (Val.:64,48)	'ðe (Val.:63,79)
2a sil·laba	ðe (Val.:98,38)	'ro (Val.:85,8%)	ri (Val.:77,91)	'ðe (Val.:75,96)	ro (Val.:72,38)
3a sil·laba	ro (Val.:87,96)	'ðu (Val.:85,52)	'ðe (Val.:82,27)	'ɫo (Val.:82,15)	'ju (Val.:80,75)
4a sil·laba	ʒe (Val.:91,25)	'ɫi (Val.:86%)	o (Val.:85,84)	diw (Val.:85,37)	'pe (Val.:84,28)

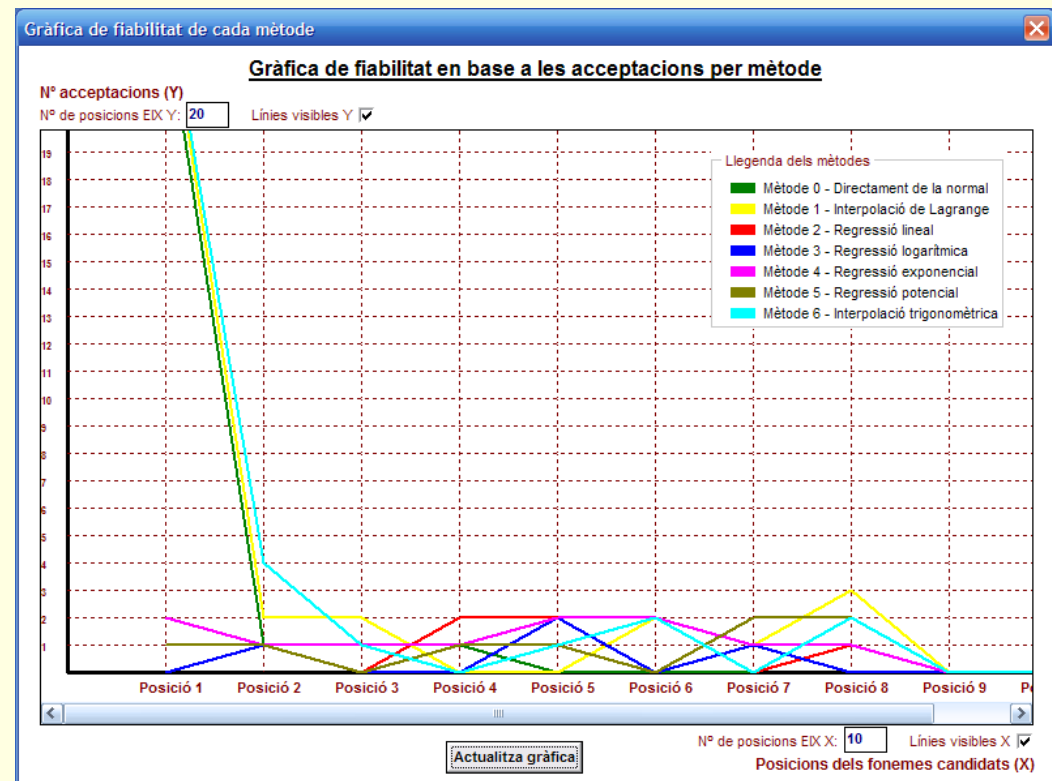
M3-Regr.Log.	Candidat 1	Candidat 2	Candidat 3	Candidat 4	Candidat 5
1a sil·laba	tʃi (Val.:89,26)	ʒet (Val.:73,63)	ra (Val.:70,78)	'mo (Val.:68,64)	'a (Val.:68,35)
2a sil·laba	ro (Val.:86,37)	ʒet (Val.:85,12)	e (Val.:82,07)	'a (Val.:77,55)	tʃi (Val.:74,2%)
3a sil·laba	'le (Val.:81,2%)	'ri (Val.:75,82)	βa (Val.:63,18)	e (Val.:54,82)	ro (Val.:49,71)
4a sil·laba	fè (Val.:2,11%)	'bo (Val.:2,09%)	'sok (Val.:2,08%)	'solks (Val.:2,05%)	ða (Val.:2,05%)

Guarda
Cancel·la

Funcionalitats: reconeixement Mètodes estadístics

■ Resultats dels mètodes estadístics:

- S'obté un alt percentatge d'encert per als mètodes d'interpolació de Lagrange i trigonomètrica
- En general, la distribució normal també és acceptable



Funcionalitats: reconeixement

Mètodes estadístics

- Resultats dels mètode estadístics (continuació):
 - Mètode de regressió lineal:
 - Alt encert de les vocals
 - Alt encert en les síl·labes que contenen /ʎ/, /t/, /k/, /p/, /ɾ/, /tʃ/, /m/
 - Mètode de regressió logarítmica:
 - Alt encert en les vocals, especialment les obertes
 - Alt encert en síl·labes que contenen /ʎ/, /d/, /ð/, /s/, /k/, /t/, /p/, /ts/, /f/
 - Mètode de regressió exponencial:
 - Alt encert en les vocals (especialment /a/ i /e/), baix encert en les consonants
 - Alt encert en síl·labes que contenen els fonemes /ʎ/, /n/, /m/, /β/, /b/, /ð/, /z/, /s/
 - Mètode de regressió potencial:
 - Alt encert en les vocals /e/, /o/, /u/
 - Alt encert en /b/, /ʎ/, /s/, /f/, /t/
 - Baix encert en /n/ i /ð/

Funcionalitats: reconeixement

Mètodes estadístics

- Conclusions dels mètodes estadístics:
 - Com a regla general, s'obté un alt percentatge d'encert per als mètodes d'interpolació de Lagrange i trigonomètrica, la qual cosa significa que, donada una ona amb un cert nombre de valors Fourier, es pot trobar un polinomi que connecti per a la majoria dels seus valors
 - La distribució normal també és acceptable, la qual cosa significa que aprox. el 90 % dels valors tenen el 90 % de la suma total d'aquests valors
 - Els fonemes oclusius (/p/, /t/, /k/) tendeixen a tenir una dependència lineal o logarítmica, com també les vocals obertes
 - Els fonemes de vocals tancades tenen una dependència potencial o exponencial
 - Els fonemes vibrants simples (/r/) tendeixen a un comportament lineal, així com les vocals
 - Els fonemes fricatius (/s/) tenen un comportament logarítmic o potencial

Funcionalitats: reconeixement

Xarxes neuronals

- **Back Propagation:** algorisme d'aprenentatge que s'utilitza per entrenar les xarxes neuronals
- L'algoritme consisteix a minimitzar l'error en un nombre determinat d'iteracions per mitjà del càlcul de derivades parcials
- Com a 'input' s'incorporen arxius wav corresponents als patrons.
- Cada patró està format de 100 valors, corresponents als valors Fourier de l'interval de freqüències entre 100 Hz i 10000 Hz, en trams de 100 en 100 Hz

Funcionalitats: reconeixement Xarxes neuronals

■ Configuració de la xarxa:

- Es determina el nombre de capes que ha de tenir la xarxa neuronal (màxim 3) i es guarda a la base de dades
- Es determina el màxim d'error (1 ó menys) i nombre màxim d'iteracions (en principi a 10000) que ha de tenir l'entrenament
- L'aprenentatge s'estableix en 0.2; d'aquesta manera la velocitat és baixa i evita problemes d'oscil·lacions al voltant del mínim error
- El nombre de capes de sortida ha de ser el mateix que el de patrons amb els que s'ha entrenat la xarxa

Reconeixement de veu utilitzant xarxes neuronals

Configuració de la xarxa

Nº capes: 3 Nº unitats a capes de neurones (la 1ª columna equival a la capa INPUT, la resta a capa HIDDEN):

	Nivell 0	Nivell 1
Nº d'unitats	40	10

Màxim error: 1

Màxim iteracions: 10000 Nº unitats a capa de sortida (corresponent al nº de patrons): 12

Aprenentatge: 0.2

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Pat Patrons
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Pat Esborrar tots
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Pat
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Pat

1º pas) Inicia entrenament Para entrenament 2º pas) Guarda xarxa

Iteració actual: 2424
Error actual de l'entrenament: 1.04867575017054

Reconeixement

Obre arxiu WAV

[Tria la XN] Reconeix!

Resultats

Funcionalitats: reconeixement Xarxes neuronals

■ Reconeixement a través de la XN:

- S'obre un dels patrons per comprovar si l'entrenament ha estat el correcte
- Se selecciona la XN que s'ha guardat amb els valors d'entrenament
- La funció d'activació utilitzada és la sigmoïdal:
$$Y := 1 / (1 + \text{Exp}(-X));$$
- La graella inferior mostra el percentatge d'aproximació entre els patrons 'input' i els valors 'output' de la XN

Reconeixement de veu utilitzant xarxes neuronals

Configuració de la xarxa

Nº capes: 3 Nº unitats a capes de neurones (la 1ª columna equival a la capa INPUT, la resta a capa HIDDEN):
Nº d'unitats: 10 (Nivell 0), 10 (Nivell 1)

Màxim error: 1

Màxim iteracions: 10000 Nº unitats a capa de sortida (corresponent al nº de patrons): 12

Aprenentatge: 0.2

Patrons

Esborrar tots

1º pas) Inicia entrenament Para entrenament 2º pas) Guarda xarxa

Iteració actual: 2424
Error actual de l'entrenament: 1.04867575017054

Reconeixement

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Particulars\UB Obre arxiu WAV

2capes Reconeix!

Resultats

	1º fonema	2º fonema	3º fonema	4º fonema
Resultats 1º:	ble.wav_2 (9.82%)	ble.wav_2 (9.97%)	ble.wav_2 (11.91%)	ble.wav_2 (9.5%)
Resultats 2º:	ble.wav_1 (8.63%)	ble.wav_1 (8.78%)	ble.wav_1 (10.64%)	ble.wav_1 (8.33%)

Funcionalitats: reconeixement

Xarxes neuronals

- Patrons utilitzats per a l'entrenament:

- a
- ai
- ble
- bres
- cel
- co
- dei
- du
- es
- ge
- ja
- jer
- la

Funcionalitats: reconeixement Xarxes neuronals

- Resultats de XN amb 1 capa per al patró 'a':

Reconeixement de veu utilitzant xarxes neuronals

Configuració de la xarxa

Nº capes: 1 Nº unitats a capes de neurones (la 1ª columna equival a la capa INPUT, la resta a capa HIDDEN):

Màxim error: 1

Màxim iteracions: 10000 Nº unitats a capa de sortida (corresponent al nº de patrons): 12

Aprenentatge: 0.2

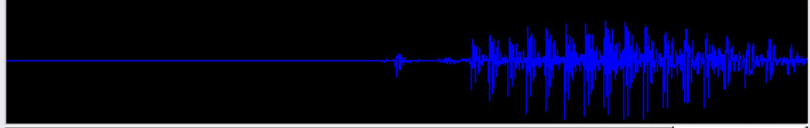
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Patrons
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Esborrar tots
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes

1ª pas) Inicia entrenament Para entrenament 2ª pas) 1capultima Guarda xarxa

Iteració actual: 2189
Error actual de l'entrenament: 4.69491930005654

Reconeixement

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Particulars\UB Obre arxiu WAV



1capultima Reconeix !

Resultats

	1º fonema	2º fonema	3º fonema	4º fonema
Resultats 1º:	a.wav_0 (89.24%)	a.wav_1 (97.76%)	a.wav_2 (97.07%)	ble.wav_2 (0.96%)
Resultats 2º:	ge.wav_0 (2.03%)	ge.wav_0 (2.06%)	a.wav_0 (2.34%)	ge.wav_3 (0.5%)

Funcionalitats: reconeixement Xarxes neuronals

- Resultats de XN amb 1 capa per al patró 'ai':

Reconeixement de veu utilitzant xarxes neuronals

Configuració de la xarxa

Nº capes: 1 Nº unitats a capes de neurones (la 1ª columna equival a la capa INPUT, la resta a capa HIDDEN):

Màxim error: 1

Màxim iteracions: 10000 Nº unitats a capa de sortida (corresponent al nº de patrons): 12

Aprentatge: 0.2

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes
D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes

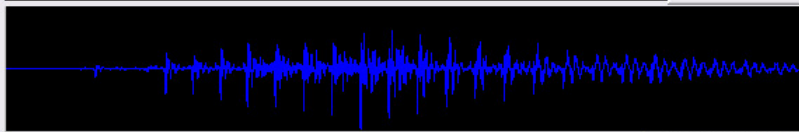
Patrons
Esborrar tots

1ª pas) Inicia entrenament Para entrenament 2ª pas) 1capultima Guarda xarxa

Iteració actual: 2189
Error actual de l'entrenament: 4.69491930005654

Reconeixement

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documentos\Els Meus Projectes Particulars\UB Obre arxiu WAV



1capultima Reconeix !

Resultats

	1ª fonema	2ª fonema	3ª fonema
Resultats 1ª:	aj.wav_0 (97.89%)	aj.wav_1 (95.77%)	ge.wav_3 (1.53%)
Resultats 2ª:	ge.wav_0 (1.56%)	a.wav_2 (1.03%)	ge.wav_0 (0.09%)

Funcionalitats: reconeixement Xarxes neuronals

- Resultats de XN amb 1 capa per al patró 'du':

Reconeixement de veu utilitzant xarxes neuronals

Configuració de la xarxa

Nº capes: 1 Nº unitats a capes de neurones (la 1ª columna equival a la capa INPUT, la resta a capa HIDDEN):

Màxim error: 1

Màxim iteracions: 10000 Nº unitats a capa de sortida (corresponent al nº de patrons): 12

Aprenentatge: 0.2

Patrons

Esborrar tots

1ª pas) Inicia entrenament Para entrenament 2ª pas) 1capultima Guarda xarxa

Iteració actual: 2189
Error actual de l'entrenament: 4.69491930005654

Reconeixement

D:\DOCUMENTS I CONFIGURACIONS\Francesc\Mis documents\Els Meus Projectes Particulars\UB Obre arxiu WAV

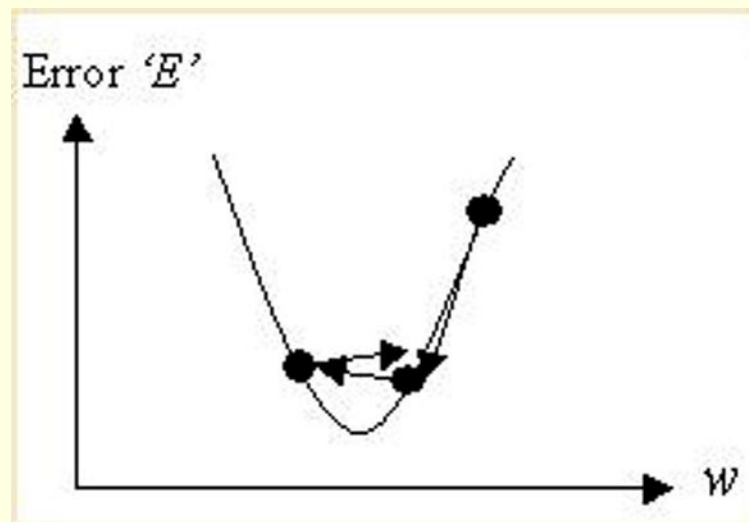
1capultima Reconeix!

Resultats

	1ª fonema	2ª fonema	3ª fonema	4ª fonema
Resultats 1ª:	du.wav_0 (91.36%)	du.wav_1 (93.13%)	du.wav_2 (92.67%)	du.wav_3 (96.02%)
Resultats 2ª:	ge.wav_0 (2.9%)	bres.wav_6 (4.73%)	dej.wav_0 (7.47%)	ja.wav_0 (2.72%)

Funcionalitats: reconeixement Xarxes neuronals

- A falta de proves empíriques evidents:
- S'ha descobert que amb un percentatge gran la velocitat d'aprenentatge també és més gran, però que sovint cau en l'oscil·lació al voltant del mínim error (s'ha posat en pràctica la teoria desenvolupada a <http://www.monografias.com/trabajos901/fundamentos-basicos-reconocimiento-voz/fundamentos-basicos-reconocimiento-voz.shtml>)



Funcionalitats: reconeixement

Xarxes neuronals

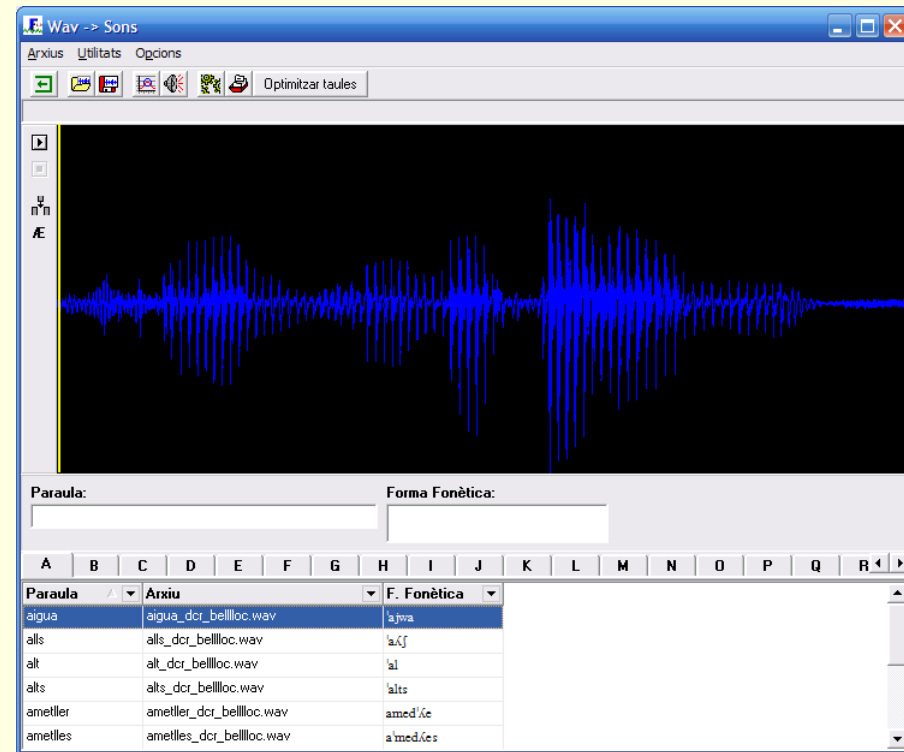
- A falta de proves empíriques evidents (continuació):
 - Es necessita un nombre molt gran d'iteracions per aconseguir un mínim d'error per a les xarxes de més de 2 capes, massa gran perquè sigui computacionalment acceptable.
 - Es treuen resultats prou bons en XN d'1 capa i 2 capes; a partir d'aquí, afegir més capes ocultes no ajuda a obtenir uns resultats més precisos

Historial

- Febrer 2006:

Es crea el programa WavFonemes, el qual permet:

- Emmagatzemar els arxius de so en format wav en una base de dades
- Reproduir els arxius wav i mostrar la representació de les seves ones

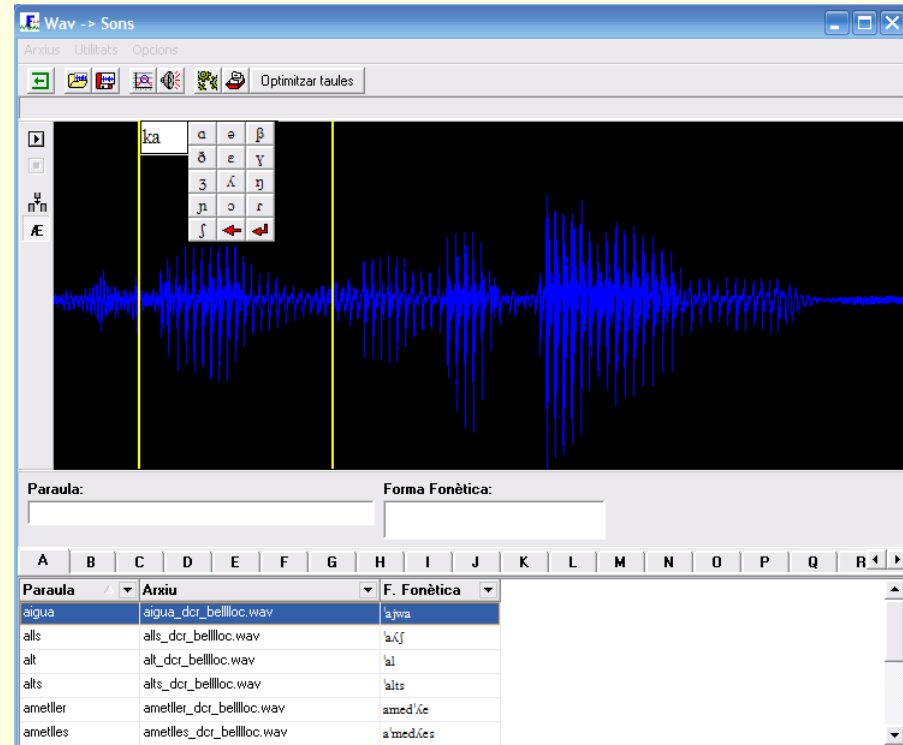


Historial

- Maig 2006:

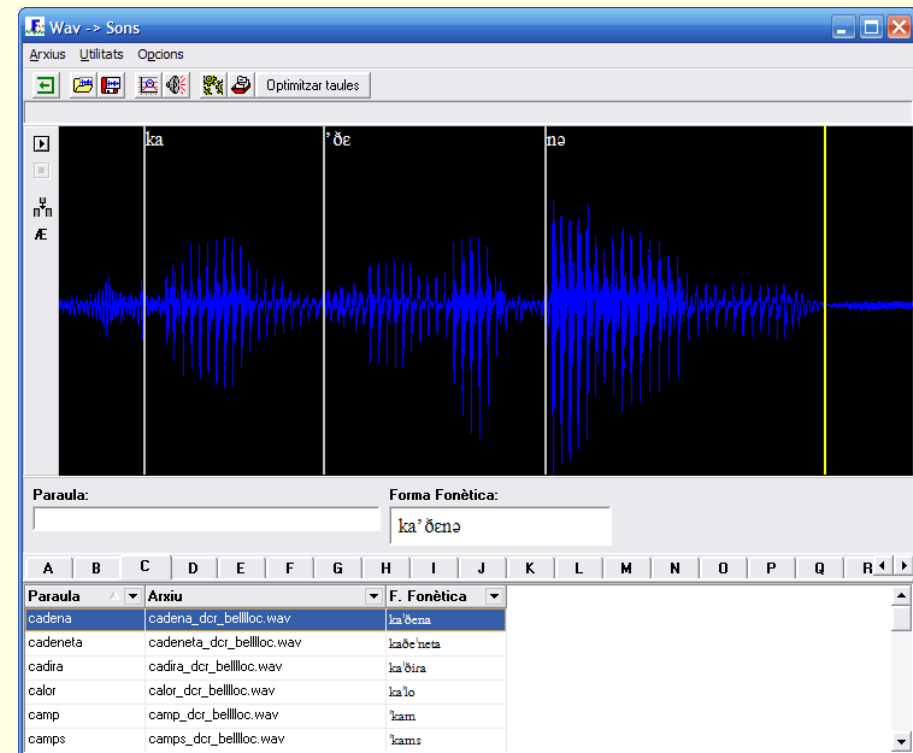
S'afegeixen noves funcionalitats:

- La possibilitat de poder retallar parts d'un arxiu wav a partir de la seva ona.
- Assignar a cada part la sí·l·laba fonètica corresponent.



Historial

- Juny 2006:
 - Aplicació de la fórmula de Fourier (FFT) per detectar el pes de cada freqüència en el so representat



$$\text{FFT: } f_j = \sum_{k=0}^{n-1} x_k e^{-\frac{2\pi i}{n} jk} \quad j = 0, \dots, n-1.$$

Historial

■ Juny 2006:

Donada la gran variabilitat dels sons dels locutors (estats d'ànim, gèneres, edats, etc) cal unificar els fonemes i dissenyar un fonema representatiu (patró), dividint la tasca en 2 etapes:

- **Normalització (etapa d'entrenament)**
- **Reconeixement (etapa de comparació)**

Opcions de Normalització i Reconeixement del fonema

NORMALITZACIÓ DELS FONEMES

Freqüències

Mínima	20	Hz	Màxima	10000	Hz
--------	----	----	--------	-------	----

Valors (pesos de la freqüència)

Recollida dels valors

80	%
----	---

RECONeixEMENT DELS FONEMES

Interval respecte al fonema normalitzat

Marqe d'error per sota	2	%	Marqe d'error per sobre	2	%
------------------------	---	---	-------------------------	---	---

Reconeix els fonemes al moment d'obrir l'arxiu de so

Acceptar
Cancel·lar

Es prenen com a representatives una freqüència mínima i màxima (normalment s'accepta l'interval de freqüències que estigui dins de la capacitat de l'oïda humana, que va dels 20 a 10000 Hz).

Historial

- Abril - Juny 2008:

S'elabora el WavFonemes v2, el qual inclou, a part de les funcions de la primera versió, les característiques següents :

- Incorporació de mètodes estadístics amb l'objectiu de tenir en compte el comportament de les dades recollides amb les diferents característiques fonètiques
- Estudi empíric dels resultats dels diferents mètodes, mitjançant gràfiques de fiabilitat

Historial

- Setembre 2009 fins ara:
 - Es treballen (i actualment en fase desenvolupament) amb les xarxes neuronals (algoritme de Back Propagation) per la seva gran capacitat d'aprenentatge, i per tant, per suplir els problemes que hi pot haver entre fonemes iguals però produïts per diferents locutors (variabilitat a causa a les diferències de longitud del tractament de les vocals i consonants, accents depenent del dialecte, el to de veu, etc.).

- Expectatives:

- La part de les xarxes neuronals queda pendent d'incorporació al programa WavFonemes v2, ja que s'ha desenvolupat en una aplicació nova.
- A partir d'ara s'hauria d'incorporar la informació obtinguda en la fase d'investigació als resultats de les xarxes neuronals, amb l'objectiu de matisar i completar aquests resultats.