# Part III

# Tancament

# CAPÍTOL 6

# Conclusions i línies de futur

Aquesta tesi ha presentat un nou pas en l'ajust perceptiu de la funció de cost dels sistemes de conversió de text en parla basats en selecció d'unitats (CTP-SU) (Hunt i Black, 1996). El punt de partida d'aquesta tesi és la prova de viabilitat descrita a (Alías, 2006), que emprava computació evolutiva interactiva (IEC) basada en algorismes genètics interactius actius (aiGA) (Llorà *et al.*, 2005*b*) per ajustar de manera òptima la funció de cost d'un CTP-SU. Aquesta tesi ha presentat contribucions per tal de millorar la precisió, la robustesa, l'aplicabilitat i el consens dels aiGA en l'ajust perceptiu de sistemes CTP-SU. Addicionalment, s'ha analitzat la competitivitat de la proposta plantejada (basada en aiGA) respecte les altres tècniques representatives de l'estat de l'art (MLR/NNLS, GA, MOS-*Postmapping*) (Meron i Hirose, 1999; Chu i Peng, 2001; Toda *et al.*, 2006; Clark *et al.*, 2007) en un entorn real de selecció d'unitats. Aquest capítol realitza una síntesi del treball realitzat, els objectius assolits i les conclusions que se'n deriven.

## 6.1 Conclusions i compliment dels objectius

Com s'acaba de comentar en la introducció d'aquest apartat, el punt de partida d'aquesta tesi doctoral és la prova de viabilitat d'ajust perceptiu de la funció de cost d'un CTP-SU mitjançant aiGA realitzada per Alías (2006). El fet de passar d'una prova de viabilitat a la seva aplicabilitat en un entorn real de selecció d'unitats no és directe. Un cop superada la

fase de la prova de concepte, cal una revisió amb detall (teòrica i pràctica) de la metodologia d'ajust de pesos per tal de fer-la competitiva davant d'altres mètodes d'ajust en un entorn real de selecció d'unitats. Durant el transcurs de la tesi s'ha realitzat un treball de recerca considerant diferents aspectes de millora en aquest sentit. Tot seguit es detallen les conclusions obtingudes a partir del treball realitzat i agrupades segons els diferents àmbits de millora estudiats: precisió, robustesa, aplicabilitat i consens.

### 6.1.1 Precisió

La primera proposta d'ajust de pesos perceptiu mitjançant aiGA proposat per Alías *et al.* (2006*a*) només obté una ponderació (vector de pesos) per tot el corpus. No obstant, en síntesi és conegut que els subcostos prenen una importància diferent dins la funció de cost segons l'especificitat (fonètica i contextual) de la unitat a recuperar (Campillo *et al.*, 2005; Colotte i Beaufort, 2005). De fet, els mètodes automàtics d'ajust de pesos obtenen un vector de pesos en funció de la unitat (Hunt i Black, 1996; Meron i Hirose, 1999) o el seu context lingüístic-fonètic (Black i Taylor, 1997*a*).

En aquest sentit, aquesta tesi realitza una nova contribució (denominada *aiGAClustered*) que permet obtenir pesos a nivell perceptiu respectant l'especificitat dels diferents contextos fonètics i lingüístics en els que es troba la unitat en el corpus (nivell d'ajust definit per la subunitat contextualitzada). La proposta presentada supera la restricció que l'ajust perceptiu implica necessàriament ajustar els pesos a nivell global, cosa que implica obtenir un únic patró de pesos per a tot el corpus massa general. L'aproximació proposada per aquesta tesi es basa, en primer lloc, en realitzar un ajust automàtic (mitjançant MLR/NNLS i distàncies cepstrals) dels pesos de la funció de cost a nivell de subunitat contextualitzada. Posteriorment, en una segona etapa, s'identifiquen patrons de pesos i se'ls hi associa informació contextual (lingüística i fonètica). Un cop determinat un nombre assolible de clústers que s'han d'ajustar perceptivament (tenint en compte que un nombre alt de proves subjectives molt tediós en termes de fatiga i frustració) es realitza l'ajust perceptiu de cada clúster. Finalment, s'obté un vector de pesos per cada grup. Aquests vectors de pesos (un per cada grup) són els utilitzats pel sistema de síntesi. En certa manera aquesta aproximació, que combina l'especificitat contextual a nivell de subunitat amb l'ajust a nivell de grups, parteix del treball de Campillo *et al.* (2005) i Colotte i Beaufort (2005), pel fet de considerar que els subcostos adopten diferent importància en funció de l'especificitat concreta de la unitat a seleccionar. En una primera versió, la proposta considera que l'especificitat ve determinada per la informació fonètica de la unitat a recuperar (vocal-consonant, lloc d'articulació, mode d'articulació o sonoritat). Tanmateix, aquesta consideració resulta li-

mitada ja que no considera que els diferents contextos, lingüístics i també fonètics, de la unitat també poden determinar la importància relativa dels subcostos. Per aquest motiu, la proposta inicial descrita en el capítol 4 es millora en segon terme en el capítol 5 per tal d'obtenir pesos a nivell de subunitat contextualitzada. Aquesta aproximació permet obtenir pesos ajustats perceptivament en funció de l'especificitat de la unitat i els diferents contextos fonètics i lingüístics presents en el corpus.

Els experiments realitzats mostren una gran diferència entre els pesos obtinguts mitjançant ajustos automàtics (basats en distàncies cepstrals) i els pesos obtinguts mitjançant tècniques d'ajust perceptives. Aquest fet qüestiona la validesa dels clústers un cop s'obtenen els pesos perceptius. Tanmateix, els experiments perceptius demostren que no hi ha diferència, a nivell perceptiu, entre els clústers originals (segons els pesos ajustats automàticament) i els grups obtinguts de nou amb els pesos ajustats a perceptivament mitjançant aiGA, demostrant així la vigència dels grups malgrat obtenir pesos de diferent valor.

Per últim, s'ha observat que realitzar l'ajust de pesos incorporant un postprocessament del senyal (TD-PSOLA) després de realitzar la selecció, introdueix ambigüitat en l'ajust ja que emmascara les diferències que puguin percebre els usuaris.

### 6.1.2 Robustesa

#### Robustesa de les solucions de l'aiGA

La prova de viabilitat descrita a (Alías, 2006) fa una primera proposta per determinar la robustesa dels pesos obtinguts mitjançant aiGA. Concretament, proposa un indicador de consistència ($\kappa$) que determina el grau de consistència dels pesos ajustats per un usuari analitzant els cicles que aquest introdueix en el graf que modela les seves preferències. A través del treball desenvolupat en aquesta tesi s'ha detectat que aquesta mesura resultava insuficient per obtenir tota la informació necessària sobre la qualitat de les solucions.

Per exemple, un usuari pot ser consistent en no contradir-se en les avaluacions realitzades, però alhora, pot adoptar una actitud excessivament dubitativa o conservadora en l'evolució interactiva marcant un nombre elevat d'empats respecte la resta d'usuaris. Aquest no és l'únic motiu que pot inferir un alt nombre d'empats, sinó que aspectes com la convergència prematura o la manca de diferències reals entre els diferents fenotips també pot ocasionar que un usuari avaluï molts dels tornejos com a empats (no és capaç de discernir quina solució és millor perceptivament). En aquest sentit, l'indicador $\lambda$ proposat dóna un índex de la certesa (del que es pot deduir l'ambigüitat) que hi ha en el model de l'usuari.

Això permet aturar la prova de manera prematura si aquesta ja ha convergit, alhora que permet identificar avaluadors massa conservadors o determinar el grau de confusió que hi ha entre les diferents solucions presentades a l'usuari.

A nivell de genotip (els pesos) hi ha dos aspectes que també resulten essencials en l'anàlisi de la qualitat de les solucions. El primer aspecte és disposar d'informació de si una prova evoluciona cap a una solució única o bé existeixen diverses solucions que són igualment vàlides per l'usuari. Aquest fet es mesura amb l'índex de convergència intra-usuari $\rho$, que mesura la diversitat genètica (en termes de correlació lineal entre els pesos) que hi ha entre les millors solucions que ha escollit l'usuari. De manera similar, l'indicador $\tau$ analitza la similitud entre les millors solucions de tots els usuaris que han realitzat la mateixa prova. Aquesta mesura permet observar si hi ha disparitat de criteris entre els diferents usuaris.

En aquest sentit, s'han incorporat nous indicadors del procés evolutiu que analitzen altres aspectes més enllà de la consistència de les respostes de l'usuari. Concretament els nous aspectes que es monitoritzen en les proves evolutives són l'ambigüitat (indicador $\lambda$), la convergència intra-usuari (indicador $\rho$) i la correlació de les solucions obtingudes entre els diferents usuaris (indicador $\tau$). Aquests indicadors s'han presentat a Formiga *et al.* (2010)

En les proves de la metodologia d'ajust de pesos basada en l'aiGA en un entorn de selecció d'unitats real, aquests indicadors han demostrat que proporcionen una informació molt útil sobre la qualitat dels pesos obtinguts després de realitzar l'ajust. La primera conclusió obtinguda gràcies a aquests indicadors és que les frases amb poques unitats variables presenten una major dificultat a l'hora de ser ajustades de manera perceptiva. També és conclou que els clústers que presentaven més confusió entre ells en l'agrupament automàtic també presenten més d'una solució òptima segons els usuaris. A més, els models de consens han demostrat que resulta necessari realitzar fortes correccions en aquelles combinacions de pesos que obtenen mals indicadors evolutius.

#### Robustesa dels pesos automàtics

De tots els mètodes d'ajust de pesos automàtics que presenta la literatura, el més conegut en l'estat de l'art és l'ajust de pesos mitjançant una regressió lineal (MLR/NNLS) entre els diferents subcostos i les distàncies cepstrals (Meron i Hirose, 1999). D'altra banda, resulta imprescindible que els subcostos segueixin una funció de densitat propera a la distribució normal per obtenir una bona fiabilitat en els sistemes de regressió lineal (Chatterjee i Ha-

di, 2006). No obstant això, no s'ha trobat en la literatura quina és la millor normalització de dades a emprar per a calcular els diferents subcostos dins la funció de cost. En aquest sentit, aquesta tesi doctoral ha realitzat un estudi de diferents funcions per normalitzar els diferents subcostos prosòdics i espectrals. En l'estudi es conclou que la transformació d'arrel ($\sqrt{x}$ –Tukey (1957)) aporta millors índexs de normalitat que les transformacions logarítmiques, exponencials i sigmoidals dels subcostos en els corpus estudiats, i en conseqüència, millors models de regressió entre les distàncies i els subcostos.

Una altra conclusió a destacar, que també guarda relació amb la precisió, és que el fet de treballar a nivell de subunitat contextualitzada proporciona millors índex de fiabilitat ($R^2$, RMSE) que treballar a nivell d'unitat en l'ajust de pesos automàtic mitjançant regressió lineal. Tanmateix, en l'ajust automàtic mitjançant GA s'ha pogut observar que treballar a nivell de subunitat suposa pitjors índex de fiabilitat en els pesos obtinguts. Aquest fet no s'atribueix a l'aproximació evolutiva en sí mateixa sinó a la implementació del GA emprada per resoldre l'ajust de pesos (Alías i Llorà, 2003). En el disseny original d'aquell treball la funció de *fitness* simplement amitjana els costos obtinguts en ponderar mitjançant els pesos avaluats els subcostos per les 5 unitats cepstralment més properes. Aquesta avaluació del *fitness* dels pesos no considera la degradació real del *fitness* en funció de les distàncies cepstrals, fet que comporta una pèrdua d'informació útil pel procés evolutiu.

#### Robustesa dels patrons de pesos

L'aproximació proposada en aquesta tesi, l'ajust perceptiu de pesos mitjançant aiGA a nivell de clúster, necessita dotar als clústers de robustesa respecte a la distribució de les dades en les particions realitzades. Des del seu origen, l'agrupament de característiques més emprat en el camp dels CTP-SU ha estat l'arbre de decisió CART (Black i Taylor, 1997*a*). Tanmateix, realitzant una anàlisi amb més profunditat del problema, s'observa que el CART presenta dos problemes per dotar de robustesa als grups: el *clustering* predictiu i l'equilibrat dels clústers. L'aproximació final al problema proposada separa el problema de l'agrupament de clústers en una fase de *clustering* pròpiament dita i una fase d'assignació dels diferents contextos als diferents grups (Colotte i Beaufort, 2005). Després d'un estudi comparatiu de diferents metodologies de *clustering*, es conclou que l'algorisme d'esperança-maximització (EM) és el que realitza una millor partició dels pesos en grups segons les mètriques típiques d'anàlisi de la bondat dels grups. Convé destacar aquest fet ja que l'algorisme EM obté cada cop millors resultats en diferents camps relacionats amb el processament de la parla, tals com el reconeixement automàtic (Moon, 1996) o la síntesi mitjançant models ocults de Markov (HMM) (Tokuda *et al.*, 2003).

### 6.1.3 Aplicabilitat

A Alías (2006) es realitzava una prova de viabilitat de la metodologia aiGA per un corpus de mida reduïda (8 min.) i subcostos de tipologia acústica. En aquesta tesi s'ha estudiat la aplicabilitat de l'aiGA en l'ajust de pesos, juntament amb les millores proposades, en un context més real de sistema de síntesi basat en selecció d'unitats. A tal efecte s'ha treballat amb un corpus de mida mitjana (1.9h), ponderant els subcostos de diferent naturalesa (lingüística i acústica) i fent-lo competir amb altres tècniques d'ajust automàtic de la literatura (MLR/NNLS), incloent-hi tècniques d'ajust perceptives. Tanmateix, cal remarcar que el disseny perceptiu seguit en l'aiGA assumeix de manera heurística la independència entre els diferents pesos. En el transcurs de la tesi, es confirma aquest comportament quan s'ha analitzat el comportament (correlació) dels pesos entre sí quan l'optimització s'ha realitzat mitjançant tècniques de cerca evolutives (GA, iGA i aiGA).

L'ajust de pesos mitjançant MOS-*Postmapping* descrit per Chu i Peng (2001); Toda *et al.* (2006) es realitza a nivell global obtenint un vector de pesos únic per tot el corpus. Llavors, per poder comparar l'aiGA i el MOS-*Postmapping* en igualtat de condicions, s'ha introduït en la comparativa l'avaluació dels pesos obtinguts per l'aiGA a nivell global, sense considerar diferents patrons en funció del context o de la subunitat concreta. En altres paraules, es consideren les diferents proves aiGA com a proves d'ajust d'un únic clúster de pesos tal i com es va fer a Alías *et al.* (2006*a*). Aquesta configuració de pesos (*aiGAGlobal*), en termes de qualitat s'ubica entre l'*aiGAClustered* i el MOS-*Postmapping* permetent observar la diferència de qualitat en dos sentits: el primer, la millora de l'ajust a nivell de clúster respecte l'ajust global i segon la millora de la metodologia aiGA respecte la metodologia MOS-*Postmapping*.

Per tant, les proves realitzades permeten concloure la metodologia d'ajust de pesos mitjançant aiGA a nivell de subunitat (*aiGAClustered*) predomina com a tècnica preferida respecte les altres tècniques d'ajust (MLR, MOS-*Postmapping* i *aiGAGlobal*) en un corpus de mida mitjana i emprant subcostos acústics i lingüístics. Per tant, es confirma la importància de treballar amb diferents patrons de pesos en funció del context lingüístic i fonètic específic de la subunitat en la selecció.

A més, analitzant els patrons perceptius de pesos amb detall, s'observa que la diferència entre el valor dels pesos dels subcostos acústics i el valor dels pesos dels subcostos lingüístics no resulta tant accentuada com en la metodologia d'ajust de pesos automàtica (sobretot en la predominança del pes de durada). Per tant, es conclou que l'aiGA resulta idoni per aquest tipus d'ajust perquè permet que els pesos cooperin entre sí sense cap bi-

aix per culpa de la forta discretització dels seus valors en treballar amb notació simbòlica {0,0.5,1}. A més, es conclou que la tipologia de pesos no resulta incompatible (acústics vs. lingüístics) podent coexistir les dos tipologies per obtenir una selecció més acurada.

Per últim, és important destacar que els pesos globals obtinguts mitjançant MOS - *Postmapping* guarden certa similitud amb els pesos ajustats manualment per Clark *et al.* (2007). Aquest fet indica que no només l'expertesa i el coneixement de la naturalesa dels subcostos d'un CTP-SU resulten fonamentals per aconseguir obtenir una bona configuració global de pesos per CTP-SU.

### 6.1.4 Consens

La última línia d'investigació d'aquesta tesi doctoral s'ha centrat en l'estudi de la integració (consens) de les preferències dels diferents usuaris a l'hora d'ajustar els pesos. Degut a la manca d'una metodologia automàtica que obtingués aquest consens, a Alías *et al.* (2006*a*) resultava necessària una segona prova perceptiva per tal d'escollir els millors pesos d'entre les diferents solucions obtingudes per part dels usuaris que havien realitzat l'ajust. Aquesta tesi realitza una aportació en la integració de les preferències dels usuaris a nivell de genotip, és a dir, a nivell dels vectors de pesos. Malgrat que els creadors de l'aiGA (Llorà *et al.*, 2005*b*) proposen consensuar els models (grafs) a nivell de fenotip (Llorà *et al.*, 2008), no s'aborda el problema de trencar els cicles formats per diferents usuaris de manera no heurística (p.ex. ponderar cada fletxa del graf segons les vegades que l'han avaluat els usuaris). El consens a nivell de genotip permet realitzar una cerca de la millor solució considerant les relacions que prenen els diferents pesos entre sí en les avaluacions de tots els usuaris.

En aquest sentit, aquesta tesi proposa emprar models latents per agrupar les millors preferències dels diferents usuaris i extreure'n la millor configuració de pesos. La motivació d'elecció dels models latents ve motivada per la seva capacitat de trobar dependències més enllà de les lineals en les dades i la seva robustesa al soroll. A més, els mètodes tracten de manera implícita les contradiccions entre els diferents usuaris sense que s'hagin de resoldre de manera heurística. Després de realitzar les proves de preferència, es conclou que els mapes topogràfics generatius (GTM) aconsegueixen integrar de manera robusta (gràcies a l'adaptació EM) les preferències dels diferents usuaris. Paral·lelament, es confirma la predominança dels algorismes basats en esperança-maximització (EM) en les dues necessitats d'agrupament plantejades en aquesta tesi: *i*) la detecció de patrons de pesos i *ii*) el consens del criteri de diferents usuaris, ja que GTM és una versió més restringida d'EM.

## 6.2 Reflexions i línies de futur

En aquest apartat es realitza un repàs més general a les diverses conclusions que s'ha arribat durant el transcurs d'aquesta tesi per tal d'orientar el treball futur que s'en pugui derivar.

En el treball realitzat, l'aiGA ha aportat una nova perspectiva en l'ajust perceptiu dels pesos de la funció de cost d'un CTP-SU, alhora que l'ajust de pesos ha suposat un camp d'innovació pel propi aiGA. Concretament, l'adaptació de l'aiGA al problema de l'ajust perceptiu dels pesos ha revertit en dues contribucions cap al propi aiGA. La primera és la definició d'indicadors que determinin la qualitat dels resultats obtinguts tot monitoritzant el procés, i la segona és la introducció del GTM per consensuar i integrar de manera robusta els criteris perceptius de diferents usuaris recopilats per l'aiGA, a nivell de genotip (pesos).

La investigació realitzada en el marc d'aquesta tesi doctoral continua fent evident la importància d'una bona ponderació dels pesos de la funció de cost per tal d'obtenir veu sintètica d'alta qualitat. S'ha tornat a observar que les tècniques automàtiques d'ajust de pesos (MLR/NNLS i GA) obtenen veu sintètica de menor qualitat en comparació amb les tècniques d'ajust que compten amb intervenció humana. En un principi, s'atribuïa part del problema a la poca fiabilitat dels mètodes (Alías, 2006) però en aquesta tesi s'ha observat que encara que es millori la seva fiabilitat treballant a nivells d'ajust més precisos i assolint més normalitat en els subcostos, no aconsegueixen superar les tècniques d'ajust perceptiu.

De manera addicional, s'ha observat que considerar únicament una sola combinació de pesos per a tot el corpus resulta insuficient per obtenir una selecció d'unitats òptima. En aquesta tesi s'observa com els valors del vector de pesos no només depenen de la unitat que s'ha de recuperar sinó que també depenen del seu context fonètic i lingüístic. Aquesta conclusió resulta coherent amb els estudis realitzats prèviament en selecció d'unitats, on s'observa la importància del context de les unitats en la generació de veu sintètica. Aquests estudis inclouen la generació de prosòdia (Campbell i Black, 1997; Iriondo *et al.*, 2007), l'agrupament espectral de les unitats en funció del context (Black i Taylor, 1997*a*) o la generació de seqüències Mel-cepstrals en síntesi markoviana (Tokuda *et al.*, 2003).

Obtenir pesos de manera perceptiva per a cada context i unitat resulta pràcticament impossible per dos motius: l'elevat nombre de proves interactives que es poden donar i la dificultat de mantenir un criteri uniforme durant tot el procés d'ajust. En aquest sentit s'ha observat que obtenir patrons del comportament dels pesos, i tenir en compte aquests patrons en el disseny de l'ajust perceptiu, proporciona un nivell intermedi d'ajust que millora la qualitat sintètica respecte un ajust global. Els patrons s'obtenen a través dels

pesos calculats a partir de la minimització de la distàncies cepstral entre la unitat ideal i les unitat candidates, alineades mitjançant *Dynamic Time Warping* (DTW). No obstant això, aquestes distàncies resulten insuficients per determinar la qualitat de la veu sintètica (Campbell i Black, 1997), fet que deixa el camp obert en l'ajust automàtic de pesos amb la finalitat de cercar-hi patrons per l'ajust perceptiu. Com a línia futura, es proposa l'estudi de distàncies que mapin millor les percepcions humanes de la qualitat sintètica.

Tal com s'ha detallat en l'apartat 2.3.3 són diverses les parametritzacions que ofereixen informació sobre la diferència espectral entre les unitats. Partint del treball de (Toda *et al.*, 2006), es proposa com a línia de futur estudiar la distància òptima a través d'aiGA ponderant la informació d'una o diverses distàncies automàtiques: es podria definir una selecció d'unitats espectral que combinés diferents parametritzacions (MFCC, LSF, LPC, etc.) de manera ponderada, i així es determinés quina combinació d'elles s'adequa més a la percepció humana. Val a dir que la selecció d'unitats espectral esmentada ja s'empra avui en dia en sistemes híbrids HMM+CTP-SU (Lu *et al.*, 2009). La síntesi markoviana ofereix un entorn ideal per l'ajust de pesos mitjançant aiGA, en treballar directament en el terreny espectral.

Treballar a nivell de subunitat contextualitzada ha comportat un empitjorament de la fiabilitat de l'ajust de pesos mitjançant GA. Llavors, en un futur resultaria necessari investigar com dissenyar una nova funció de *fitness* dels pesos, capaç de discriminar la millor unitat més enllà de l'amitjanat de les $N$ unitats cepstralment més properes. Es proposa, doncs, emprar una funció de *fitness* inspirada en la classificació per *ranking* que tingui la diferència de posicions entre l'ordenació produïda per les distàncies cepstrals i l'ordenació produïda per la ponderació dels subcostos. S'hauria d'estudiar, en un futur, si obtenint uns pesos fiables ajustats mitjançant GA a nivell de subunitat s'aconseguiria minimitzar la gran diferència que hi ha entre els pesos automàtics i els pesos perceptius o són millors respecte l'ajust MLR/NNLS per detectar-hi patrons de pesos.

Malgrat que la metodologia d'agrupament de patrons s'ha proposat per l'ajust perceptiu mitjançant aiGA, aquesta es pot aplicar també a altres metodologies d'ajust perceptiu (p.ex. MOS-*Postmapping*) degut a que les premisses per ajustar els pesos a nivell de clúster són: ajustar els pesos automàticament, agrupar-los en funció de la seva especificitat contextual (fonètica i lingüística) i realitzar l'ajust de les unitats a ajustar mitjançant frases portadores. En un futur es podria realitzar la comparativa aiGA i MOS-*Postmapping* a nivell de clúster.

L'aiGA ha mostrat la seva aplicabilitat per un entorn real de selecció d'unitats. No obstant això, actualment existeixen a l'estat de l'art sistemes CTP-SU amb corpus de més

de cinc hores d'extensió que treballen amb més de 30 subcostos diferents (Kaszczuk i Osowski, 2009) els quals comportarien ajustar 30 pesos diferents. Per tant, s'hauria d'estudiar la idoneïtat de l'aiGA per aquests nous escenaris. A més, cal remarcar, com ja s'ha mencionat en l'apartat d'aplicabilitat, que el disseny proposat en l'aiGA assumeix heurísticament la independència entre els diferents pesos en no observar indicis de dependència quan s'han emprat tècniques evolutives (GA, iGA i aiGA). No obstant això, cal prendre consciència dels següents aspectes en un futur: *i* Incrementar substancialment el nombre de pesos implicaria augmentar el nombre d'avaluacions per part de l'usuari. *ii*) Un alt nombre d'avaluacions per part de l'usuari augmentaria el risc de fatiga i frustració en l'optimització interactiva provocant una pèrdua d'eficiència de l'aiGA. Per tant, *iii*) resultaria necessari un nou disseny de la metodologia aiGA proposada per tal de detectar dependències en els pesos prèvies a l'evolució interactiva i així replantejar l'algorisme genètic emprat i el model de *fitness* induït.

Com alternativa o complementació de l'aiGA, la literatura presenta els models gràfics probabilístics (*probabilistic graphical models*)(Jordan, 2004) proporcionen metodologies de cerca eficients, que es troben a cavall entre la teoria de la probabilitat i la teoria de grafs. Aquests models han demostrat ser mètodes robustos per la cerca (optimització) evolutiva (Larrañaga i Lozano, 2002). En aquest sentit, es podrien emprar els models gràfics per obtenir ordenacions més acurades del *ranking* de les solucions presentat a l'usuari emprant tècniques com la moralització o l'absorció (Barber, 2003). A més, els esmentats models poden proporcionar mètriques de consistència i ambigüitat més robustes per solventar el problema de l'ajust de pesos per sistemes de CTP-SU. No obstant això, cal considerar que els indicadors definits per l'aiGA busquen simplificar la comprensió del procés evolutiu de manera ad hoc per millorar la qualitat de les solucions (pesos) obtingudes, mentre que les mesures de consistència dels models gràfics proporcionen informació sobre la consistència de l'espai de cerca i el seu aprenentatge. És a dir, mentre que les mètriques proposades en aquesta tesi persegueixen obtenir bones solucions finals independentment de la qualitat del model de *fitness* sintètic construït, les mètriques de consistència que proposen els models gràfics proporcionen indicadors sobre la qualitat del model de *fitness* sintètic pròpiament dit, qüestió interessant per l'objectiu que es persegueix. Per tant, resultaria idoni en un futur estudiar com integrar les mètriques de consistència dels models gràfics en el procés de l'aiGA.

En termes de consens, el GTM ha mostrat ser un model vàlid per integrar les percepcions de l'usuari a nivell de genotip. No obstant això, queda pendent un estudi comparatiu més detallat sobre la idoneïtat de consensuar les preferències dels usuaris a nivell de ge-

notip, com proposa aquesta tesi, o bé realitzar un consens a nivell de fenotip (com proposa (Llorà *et al.*, 2008)) unint les relacions dels diferents grafs de cada usuari en un sol graf general de consens. Tanmateix, existeixen altres tècniques basades en grafs que permeten obtenir un graf de consens (p.ex *Median-Graph* (Ferrer *et al.*, 2005)), que també seria interessant estudiar en un futur.

Quedaria pendent demostrar l'aplicabilitat de l'aiGA en un entorn real de síntesi expressiva (Schröder, 2004), on la selecció d'unitats esdevé més complexa (Steiner *et al.*, 2010). Tanmateix, l'aiGA ha mostrat la seva viabilitat, en l'entorn de la síntesi de la parla, en generar prosòdia expressiva (Alm i Llora, 2006).

Per últim, cal notar que en aquesta tesi s'ha treballat en tot moment amb les variants lineals de la funció de cost (segons la distància de Hsamming o la distància euclídea). Tanmateix, en un futur, es podria estudiar la programació genètica interactiva (Johanson i Poli, 1998), que oferiria noves possibilitats per evolucionar interactivament una nova funció de cost de manera perceptiva, deixant oberta la possibilitat de combinar les aportacions de l'aiGA en la programació genètica interactiva.

# Part IV

# Annexos i Bibliografia

# APÈNDIX A

## Proves de normalitat de dades

## A.1 El test de Kolmogorov-Smirnov

El test de Kolmogorov-Smirnov (Beightler *et al.*, 1979) compara els valors d'un vector de dades $X$ amb els valors d'una distribució normal tipificada, amb mitjana zero i variància 1. La hipòtesi nul·la pel test de Kolmogorov-Smirnov és que aquest conjunt de valors $X$ provenen d'una distribució normal tipificada. La hipòtesi alternativa és que els valors de $X$ no provenen de la distribució normal.

Per a cada valor potencial $x \in X$, el test de Kolmogorov-Smirnov compara la proporció de valors menors a $x$ amb el nombre esperat predit per una distribució normal tipificada. Si es vol comprovar la normalitat de les dades i no es disposa dels paràmetres previs necessaris (mitjana i variància de la distribució) per dur a terme la prova, cal emprar el test de Lilliefors que es descriu en l'apartat A.2.

## A.2 El test de Lilliefors

El test de Lilliefors (Lilliefors, 1967) prova el grau d'adaptació d'un conjunt de dades $X$ a una distribució normal. El test de Lilliefors avalua la hipòtesi que $X$, com a conjunt de dades segueixi una distribució normal amb una mitjana i una variància no especificades,

davant de la hipòtesi de que no provingui de cap distribució normal de dades. Aquest test compara la distribució donada $X$ amb una distribució normal, que tingui la mateixa mitjana i variància que $X$. Es tracta d'una adaptació del test de Kolmogorov-Smirnov.

El procediment és el següent, extret de Lilliefors (1967):

Donada una mostra de $N$ observacions, es determina $D = max_X |F^*(X) - S_N(X)|$, on $S_N(X)$ és la funció de distribució acumulada (*Cumulative Distribution Function* – cdf) de la mostra i $F^*(x)$ és la funció de distribució acumulada normal amb $\mu = \bar{X}$, la mitjana de la mostra, i $\sigma^2 = s^2$, la variància de la mostra, que es defineix amb un denominador $N$-1, on $N$ és la mida de la mostra. Si el valor de $D$ excedeix el valor crític en la taula, es rebutja la hipòtesi que la distribució observada prové d'una població normal.

Els valors en la taula es varen obtenir mitjançant simulacions de Monte Carlo per diferents mides de distribucions. Per cada valor de $N$, es varen generar 1000 o més distribucions aleatòries de mida $N$, estimant el seu índex $D$.

Les simulacions van mostrar que la relació entre el llindar i la mida de la mostra $N$ és constant per distribucions de mida gran $N > 30$. La taula determina que per a distribucions grans, el valor $D$ ha de ser inferior a $\frac{1.031}{\sqrt{N}}$ amb un nivell de significança de ($p = 0.01$). En aquest cas, cal remarcar que augmentar el nivell de significança (augmentar $p$) significa un criteri més estricte, i per tant més susceptible a l'error, en rebutjar la hipòtesi nul·la. Llavors, si les dades són normals amb una significança de $p = 0.20$, també ho seran amb una $p = 0.01$, però no necessàriament viceversa.

# APÈNDIX B

## Proves de significança

## B.1 Prova de hipòtesi estadística

La prova d'hipòtesi estadística permet contrastar si una mostra de valors segueix una funció de densitat (pdf) determinada. La hipòtesi estadística és simple si els paràmetres de la pdf original són coneguts ($\mu$ i $\sigma$). En canvi si es vol contrastar la hipòtesi sobre una fdp de $\mu$ i $\sigma$ desconeguts llavors es diu que la hipòtesi es composta.

El funcionament és el següent: es formula una hipòtesi nul·la ($H_0$), que la seva refutació permet acceptar la hipòtesi original ($H_1$), anomenada hipòtesi alternativa. Concretament, la hipòtesi nul·la és que les dades observades no segueixen una funció de densitat determinada, i la hipòtesi alternativa és que les dades sí segueixen una funció de densitat concreta. Llavors, s'accepta o es refuta la hipòtesi nul·la en funció d'un interval d'acceptació o de confiança. Si la probabilitat $p$ obtinguda sobre la veracitat de la hipòtesi nul·la és inferior a un determinat valor $\alpha$ (típicament $\alpha = \{0.05, 0.01, 0.001\}$)), llavors es rebutja la hipòtesi nul·la prenent per bona la hipòtesi alternativa ( $p < \alpha$) que les dades sí segueixen una pdf determinada). En canvi, si la probabilitat obtinguda és superior ($p > \alpha$) no es pot refutar la hipòtesi nul·la i per tant no es pot prendre per bona la hipòtesi alternativa (en aquest cas, no es pot dir que les dades segueixin un pdf determinada).

A l'interval $[0 - \alpha]$ s'anomena interval d'acceptació de la hipòtesi nul·la, per tant si $p$ és superior a l'interval d'acceptació es refuta la hipòtesi nul·la dient que les dades seguei-

xen una pdf determinada de manera significativa. No obstant, cal observar que es poden refutar hipòtesis nul·les verdaderes en un $100 \cdot \alpha\%$ de casos. Al valor $\alpha$ s'anomena nivell de significança o coeficient de risc.

## B.2 Anàlisi de la variància (ANOVA)

Si es disposa de dos mostres $x$ i $y$ de mida $n$ que segueixen dos variables normals $x \in N_1(\mu, \sigma)$ i $y \in N_2(\mu, \sigma)$, es demostra que la diferència de les mitjanes $(\bar{x} - \bar{y})$ és:

$$N\left(0, \epsilon = \frac{\sigma\sqrt{2}}{\sqrt{n}}\right) \tag{B.1}$$

En el cas que les mostres tinguin diferent mida $n_1$ i $n_2$ la diferència és:

$$N\left(0, \epsilon = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) \tag{B.2}$$

A més, si les mostres tenen diferent desviació típica $\sigma_1$ i $\sigma_2$, llavors la diferència és:

$$N\left(0, \epsilon = \sqrt{\frac{\sigma_2^2}{n_1} + \frac{\sigma_1^2}{n_2}}\right) \tag{B.3}$$

El fet de calcular la diferència de dues mostres permet resoldre el problema de contrastar les hipòtesis que dues mostres provinguin de variables aleatòries normals amb la mateixa mitjana. Considerant que es disposa de dues mostres simples obtingudes aleatòriament de mida $n_1$ i $n_2$ i que es vol endevinar si aquestes mostres provenen de la mateixa distribució normal amb desviació típica $\sigma$ coneguda, es pot dir que si la diferència de mitjanes $\bar{x} - \bar{y}$ el valor és superior a $3\epsilon$, la diferència és molt significativa, si és superior a $2\epsilon$ és significativa i si es inferior a $2\epsilon$ la diferència no és significativa. En cas de no conèixer a priori $\sigma_1$ i $\sigma_2$, es poden prendre com a valors aproximats $\sigma_1 = s_x$ i $\sigma_2 = s_y$, essent $s$ la desviació típica de la mostra (aproximada) (Ríos, 2000) sempre i quan les mostres siguin suficientment grans ($n > 100$). Aquest procés es coneix com a anàlisi de la variància o ANOVA.

## B.3 Prova de diferències $t$-Student

En canvi, si les mostres són petites ($n < 100$) la diferència real entre $\sigma$ i $s$ pot esdevenir molt gran i no es pot aplicar la metodologia ANOVA explicada. Llavors, en aquest cas s'aplica la prova $t$ de Student que mesura si un estadístic $U$ obtingut a partir de les dues mostres segueix una distribució $t$ de Student (Leon-Garcia, 1994).

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{n_1 s_1^2 + n_2 s_2^2}} \sqrt{\frac{n_1 + n_2 - 2}{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{B.4}$$

Llavors, l'acceptació o no de la hipòtesi nul·la ve donada pels valors llindar $t$ de la taula $t$ de Student determinats per $m$ graus de llibertat $m = n_1 + n_2 - 2$ i el nivell de significança que es vulgui obtenir ($\alpha = \{0.05, 0.01, 0.001\}$). Llavors, si $u < t$ s'accepta la hipòtesi nul·la i en canvi, si $u > t$ es pot refutar la hipòtesi nul·la, donant per bona la hipòtesi alternativa.

La prova de diferències $t$-Student també es pot realitzar per comparar més de dues mostres ($k = \{3, 4, 5, \ldots\}$). En aquest cas s'aplica la correcció de Bonferroni (Hochberg, 1988) per corregir el nivell de significança. En aquest sentit, es divideix el nivell de significança $\alpha$ per el nombre de comparacions que es realitzen:

$$\alpha_{BONF} = \frac{2\alpha}{k \cdot (k-1)} \tag{B.5}$$

on $k$ és el nombre de mostres que es comparen. Alternativament, també s'expressa com a correcció en la significança $p$ obtinguda:

$$p_{BONF} = \frac{p \cdot k \cdot (k-1)}{2} \tag{B.6}$$

## B.4 Prova de signe Wilcoxon

No obstant, els contrastos d'hipòtesi explicats fins ara assumeixen que les dades segueixen una distribució normal. Llavors, si no es pot garantir normalitat en les dades es realitza la prova de signe o altrament dita de Wilcoxon o Mann-Whitney (Mann i Whitney, 1947).

La prova de signa considera si dos mostres de dades diferents segueixen la mateixa distribució. Si les dues mostres són d'igual mida es realitza la prova de Wilcoxon (Hollander i Wolfe, 1973), i si són de mida diferent es realitza la prova de Mann-Whitney (Mann i Whitney, 1947).

La idea del test de Wilcoxon és la següent: si les dues distribucions són iguals, la diferència de valors obtinguts per parelles de cada distribució s'ha de distribuir de manera igual en ambdós costats de 0 i a més, a mesura que s'augmenti la distància de les dades a un costat de zero (positiu o negatiu) també ha d'augmentar la distància en l'altre costat.

Si s'obtenen dues mostres $x$ i $y$ mitjançant parelles ($x_1$,$y_1$), ($x_2$,$y_2$), ..., ($x_n$,$y_n$) de dues poblacions de densitat $f_1(x)$ i $f_2(x)$. Llavors es procedeix de la manera següent:

1. Sigui $R_i$ la posició dins l'ordre (*Rank*) de la diferència absoluta ($|D_i| = |x_i - y_i|$) ordenades de menor a major, exceptuant les iguals. Per exemple, si $x_4 - y_4$ presenta la menor diferència de totes les comparacions $|x_i - y_i|$ llavors $R_4 = 1$

2. Llavors s'assigna a cada ordre $R_i$ el signe de la diferència $D_i = x_i - y_i$. En l'exemple anterior, si $x_4 > y_4$ llavors $R_4 = 1$, en canvi si $x_4 < y_4$ llavors $R_4 = -1$.

3. Es calculen ambdós valors:

$$W_+ = \sum R_i | R_i > 0 \tag{B.7}$$
$$W_- = \sum R_i | R_i < 0 \tag{B.8}$$

4. Si $\min(W_+, W_-)$ és inferior al valor crític de la taula de Wilcoxon segons el nombre de mostres i un nivell de significança $\alpha$, llavors es pot refutar la hipòtesi nul·la $H_0$ que les dues distribucions són iguals. En canvi, si el valor és superior al valor crític s'accepta la hipòtesi nul·la ($H_0$), que les dues mostres són iguals.

## B.5 Prova U de Mann-Whitney

Si la mida de les mostres és diferent o les distribucions no es poden obtenir de manera aparellada, llavors es realitza la prova de Mann-Whitney (Mann i Whitney, 1947) que segueix una idea semblant:

1. Generar la distribució $xy$ unint les mostres $x$ i $y$:

$$xy = \{xy_1, xy_2, \ldots, xy_n\} = \{x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n\} \tag{B.9}$$

2. Sigui $R_i$ la posició dins l'ordre (*Rank*) de $xy_i$ dins $xy$ ordenades de menor a major.

3. Llavors es calcula:

$$W_1 = \sum R_i : \text{ Considerant els } R_i \text{ dels valors de la mostra 1} \quad \text{(B.10)}$$

$$W_2 = \sum R_i : \text{ Considerant els } R_i \text{ dels valors de la mostra 2} \quad \text{(B.11)}$$

4. S'ajusten els valors $W_1$ i $W_2$ en funció de la mida de la seva mostra:

$$U_1 = W_1 - \frac{n_1(n_1+1)}{2} \quad \text{(B.12)}$$

$$U_2 = W_2 - \frac{n_2(n_2+1)}{2} \quad \text{(B.13)}$$

5. Si $\min(U_1, U_2)$ és inferior al valor crític de la taula de Mann-Whitney segons el nombre de mostres i un nivell de significança $\alpha$, llavors es pot refutar la hipòtesi nul·la $H_0$ que les dues distribucions són iguals. En canvi, si el valor és superior al valor crític s'accepta la hipòtesi nul·la ($H_0$) que les dues mostres són iguals.

# APÈNDIX C

## Descripció del corpus *url_fer_ct*

En aquest annex es proporciona informació addicional sobre el corpus de veu *url_fer_ct* d'un locutor masculí (Ferran). Convé recordar que el corpus emprat es va emprar en els treballs anteriors de Guaus i Iriondo (2000*a*,*b*), referenciat com a *bdp2* . Aquest corpus es va dissenyar per formar part d'un sintetitzador per difonemes (de segona generació segons (Taylor, 2009)) en català. El corpus està composat per 1207 unitats, de les quals 895 són difonemes i 312 són trifonemes. Aquest conjunt inicial es va estendre amb 313 frases (de 8 minuts de durada) balancejades fonèticament per a obtenir més variabilitat d'unitats, i així disposar d'una primera aproximació de corpus on realitzar una selecció d'unitats (difonemes i trifonemes).

A la taula C.1 es mostra la tipificació fonètica (tipus, lloc d'articulació, mode d'articulació i sonoritat) de cada al·lòfon del corpus. A la taula C.2 es mostra la presència de cada al·lòfon dins del corpus. A la taula C.3 es mostra la presència de les unitats en el corpus.

A la figura C.1 es mostren els histogrames de la prosòdia i la seva derivada, a la figura C.2 es mostren les comparatives quartil-quartil (*qqplot*) de les distribucions de prosòdia i la seva derivada respecte la distribució normal. A les figures C.3 i C.4 es mostren els *boxplots*, histogrames i *qqplots* dels subcostos de *target* i concatenació analitzats per la unitat /@l/ del corpus. Finalment, es tornen a mostrar els *boxplots*, histogrames i *qqplots* dels subcostos transformats segons la transformació sigmoide clàssica (figures C.5 i C.6) i sigmoide linear (figures C.7 i C.8).

Taula C.1: Descripció dels diferents al·lòfons en notació SAMPA (Wells *et al.,* 1992) que composen el corpus *url_fer_ct*.

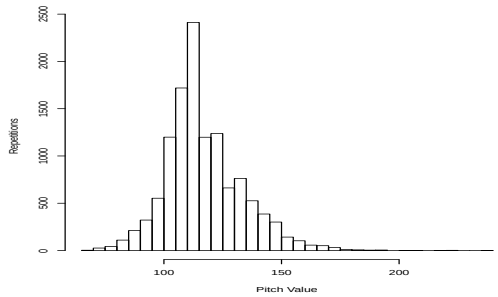
| Al·lòfon | Tipologia | Lloc d'articulació | Mode d'articulació | Sonoritat |
|---|---|---|---|---|
| /@/ | Vocal | Central | Semi-oberta | Sonora |
| /B/ | Consonant | Bilabial | Fricativa | Sonora |
| /D/ | Consonant | Interdental | Fricativa | Sonora |
| /E/ | Vocal | Anterior | Semi-oberta | Sonora |
| /G/ | Consonant | Velar | Fricativa | Sonora |
| /J/ | Consonant | Palatal | Nasal | Sonora |
| /L/ | Consonant | Palatal | Líquida-Lateral | Sonora |
| /M/ | Consonant | Labiodental | Nasal | Sonora |
| /N/ | Consonant | Velar | Nasal | Sonora |
| /O/ | Vocal | Posterior | Semi-oberta | Sonora |
| /R/ | Consonant | Alveolar | Líquida-Vibrant | Sonora |
| /S/ | Consonant | Prepalatal | Fricativa | Sorda |
| /T/ | Consonant | Interdental | Fricativa | Sorda |
| /Z/ | Consonant | Prepalatal | Fricativa | Sonora |
| /_/ | Silenci | Silenci | Silenci | Sorda |
| /a/ | Vocal | Anterior | Oberta | Sonora |
| /b/ | Consonant | Bilabial | Oclusiva | Sonora |
| /d/ | Consonant | Dental | Oclusiva | Sonora |
| /e/ | Vocal | Anterior | Semi-tancada | Sonora |
| /f/ | Consonant | Labiodental | Fricativa | Sorda |
| /g/ | Consonant | Velar | Oclusiva | Sonora |
| /i/ | Vocal | Anterior | Tancada | Sonora |
| /j/ | Semivocal | Anterior | Tancada | Sonora |
| /k/ | Consonant | Velar | Oclusiva | Sorda |
| /l/ | Consonant | Alveolar | Líquida-Lateral | Sonora |
| /m/ | Consonant | Bilabial | Nasal | Sonora |
| /n/ | Consonant | Alveolar | Nasal | Sonora |
| /o/ | Vocal | Posterior | Semi-tancada | Sonora |
| /p/ | Consonant | Bilabial | Oclusiva | Sorda |
| /r/ | Consonant | Alveolar | Líquida-Vibrant | Sonora |
| /s/ | Consonant | Alveolar | Fricativa | Sorda |
| /t/ | Consonant | Dental | Oclusiva | Sorda |
| /u/ | Vocal | Posterior | Tancada | Sonora |
| /v/ | Consonant | Labiodental | Fricativa | Sonora |
| /w/ | Semivocal | Posterior | Tancada | Sonora |
| /x/ | Consonant | Velar | Fricativa | Sorda |
| /z/ | Consonant | Alveolar | Fricativa | Sonora |

Taula C.2: Distribució dels al·lòfons en notació SAMPA (Wells *et al.*, 1992) en català a través del corpus *url_fer_ct*.

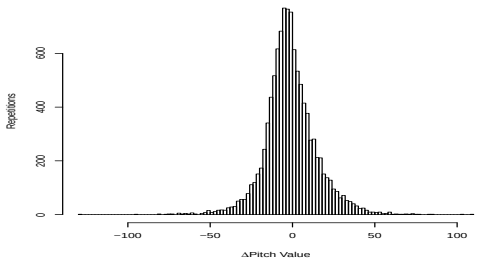
| Fonemes | Repeticions | Percentatge |
|---|---|---|
| /@/ | 1859 | 15.38% |
| /i/ | 870 | 7.20% |
| /s/ | 721 | 5.97% |
| /l/ | 694 | 5.74% |
| /n/ | 644 | 5.33% |
| /u/ | 566 | 4.68% |
| /t/ | 561 | 4.64% |
| /a/ | 519 | 4.29% |
| /k/ | 472 | 3.91% |
| /r/ | 407 | 3.37% |
| /R/ | 404 | 3.34% |
| /m/ | 388 | 3.21% |
| /z/ | 350 | 2.90% |
| /e/ | 332 | 2.75% |
| /o/ | 319 | 2.64% |
| /p/ | 316 | 2.61% |
| /d/ | 310 | 2.57% |
| /E/ | 252 | 2.09% |
| /D/ | 220 | 1.82% |
| /B/ | 204 | 1.69% |
| /O/ | 188 | 1.56% |
| /b/ | 175 | 1.45% |
| /w/ | 170 | 1.41% |
| /f/ | 150 | 1.24% |
| /Z/ | 110 | 0.91% |
| /g/ | 107 | 0.89% |
| /G/ | 102 | 0.84% |
| /L/ | 97 | 0.80% |
| /N/ | 96 | 0.79% |
| /J/ | 92 | 0.76% |
| /j/ | 82 | 0.68% |
| /S/ | 77 | 0.64% |
| /T/ | 63 | 0.52% |
| /x/ | 63 | 0.52% |
| /_/ | 57 | 0.47% |
| /v/ | 29 | 0.24% |
| /M/ | 19 | 0.16% |

Taula C.3: Distribució de les diferents unitats en català (en notació SAMPA (Wells *et al.*, 1992)) a través del corpus *url_fer_ct*, les repeticions (Rep.) s'ordenen de major a menor aparició.

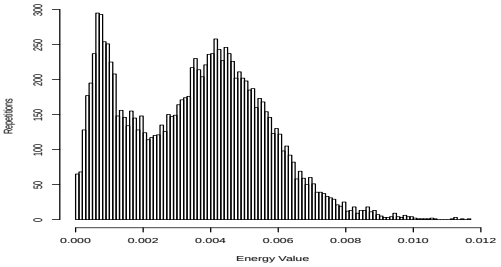
| **Unitat** | **Rep.** | **‰** | **Unitat** | **Rep.** | **‰** | **Unitat** | **Rep.** | **‰** |
|---|---|---|---|---|---|---|---|---|
| /@l/ | 282 | 20.57‰ | /ik/ | 62 | 4.52‰ | /mp/ | 35 | 2.55‰ |
| /@s/ | 220 | 16.05‰ | /en/ | 61 | 4.45‰ | /no/ | 35 | 2.55‰ |
| /si/ | 171 | 12.47‰ | /me/ | 57 | 4.16‰ | /il/ | 35 | 2.55‰ |
| /l@/ | 170 | 12.40‰ | /nd/ | 56 | 4.09‰ | /@i/ | 35 | 2.55‰ |
| /@n/ | 155 | 11.31‰ | /ta/ | 55 | 4.01‰ | /Nk/ | 34 | 2.48‰ |
| /t@/ | 153 | 11.16‰ | /zd/ | 54 | 3.94‰ | /lz/ | 34 | 2.48‰ |
| /k@/ | 125 | 9.12‰ | /ni/ | 52 | 3.79‰ | /@G/ | 33 | 2.41‰ |
| /D@/ | 122 | 8.90‰ | /@r/ | 49 | 3.57‰ | /lk/ | 32 | 2.33‰ |
| /n@/ | 121 | 8.83‰ | /pu/ | 47 | 3.43‰ | /ez/ | 32 | 2.33‰ |
| /ns/ | 111 | 8.10‰ | /ka/ | 47 | 3.43‰ | /@R/ | 32 | 2.33‰ |
| /@k/ | 108 | 7.88‰ | /li/ | 47 | 3.43‰ | /nu/ | 31 | 2.26‰ |
| /d@/ | 107 | 7.81‰ | /mb/ | 46 | 3.36‰ | /zi/ | 31 | 2.26‰ |
| /i@/ | 106 | 7.73‰ | /sp/ | 46 | 3.36‰ | /@f/ | 31 | 2.26‰ |
| /@m/ | 103 | 7.51‰ | /ri/ | 46 | 3.36‰ | /ts/ | 29 | 2.12‰ |
| /@z/ | 100 | 7.30‰ | /r@/ | 46 | 3.36‰ | /lt/ | 29 | 2.12‰ |
| /s@/ | 97 | 7.08‰ | /it/ | 46 | 3.36‰ | /iB/ | 29 | 2.12‰ |
| /@D/ | 91 | 6.64‰ | /um/ | 46 | 3.36‰ | /aD/ | 29 | 2.12‰ |
| /un/ | 91 | 6.64‰ | /sk/ | 45 | 3.28‰ | /tu/ | 28 | 2.04‰ |
| /st/ | 84 | 6.13‰ | /@p/ | 44 | 3.21‰ | /dz/ | 28 | 2.04‰ |
| /z@/ | 84 | 6.13‰ | /at/ | 44 | 3.21‰ | /zm/ | 28 | 2.04‰ |
| /p@/ | 83 | 6.05‰ | /ul/ | 42 | 3.06‰ | /zu/ | 28 | 2.04‰ |
| /ti/ | 83 | 6.05‰ | /tr@/ | 40 | 2.92‰ | /ra/ | 28 | 2.04‰ |
| /ku/ | 80 | 5.84‰ | /Bi/ | 40 | 2.92‰ | /En/ | 28 | 2.04‰ |
| /@B/ | 79 | 5.76‰ | /ls/ | 40 | 2.92‰ | /di/ | 27 | 1.97‰ |
| /nt/ | 76 | 5.54‰ | /B@/ | 39 | 2.85‰ | /kt/ | 27 | 1.97‰ |
| /in/ | 76 | 5.54‰ | /R@/ | 39 | 2.85‰ | /ia/ | 27 | 1.97‰ |
| /@@/ | 75 | 5.47‰ | /im/ | 39 | 2.85‰ | /uB/ | 27 | 1.97‰ |
| /m@/ | 74 | 5.40‰ | /mi/ | 38 | 2.77‰ | /la/ | 26 | 1.90‰ |
| /an/ | 73 | 5.33‰ | /iD/ | 38 | 2.77‰ | /b@/ | 25 | 1.82‰ |
| /@t/ | 72 | 5.25‰ | /Di/ | 37 | 2.70‰ | /se/ | 25 | 1.82‰ |
| /al/ | 70 | 5.11‰ | /on/ | 37 | 2.70‰ | /es/ | 25 | 1.82‰ |
| /is/ | 69 | 5.03‰ | /nz/ | 36 | 2.63‰ | /ur/ | 24 | 1.75‰ |
| /io/ | 63 | 4.60‰ | /us/ | 36 | 2.63‰ | Altres | 3845 | 28.05% |



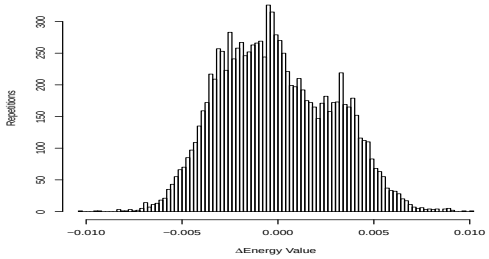

(a) Histograma de la distribució de la $F_0$ a través de tot el corpus.



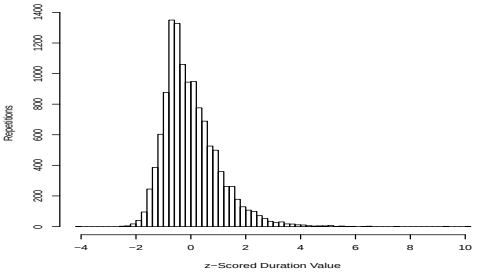

(b) Histograma de la distribució de la $\Delta F_0$ a través de tot el corpus.



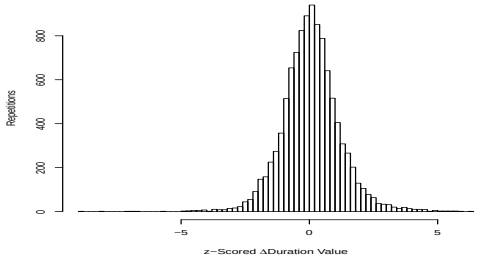

(c) Histograma de la distribució de l'energia a través de tot el corpus.




(d) Histograma de la distribució de la $\Delta$ d'energia a través de tot el corpus.

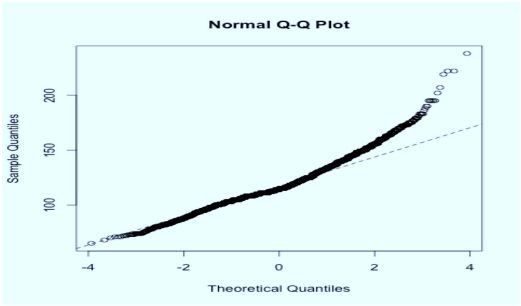



(e) Histograma de la distribució de la durada (*z-score*) a través de tot el corpus.

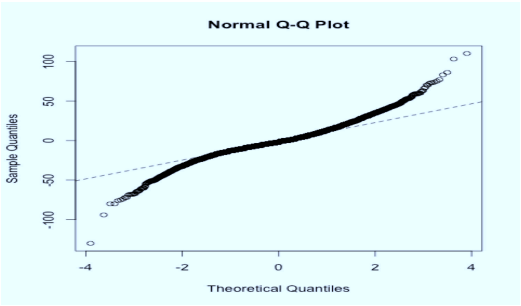



(f) Histograma de la distribució de la $\Delta$ de durada (*z-score*) a través de tot el corpus.

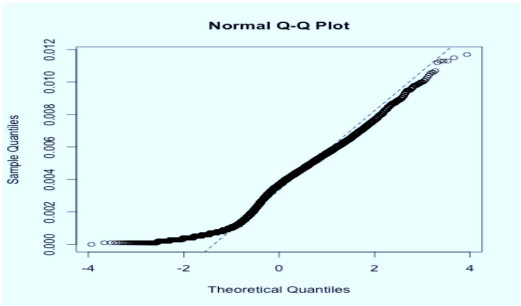
Figura C.1: Histogrames de la prosòdia i la seva derivada per tot el corpus *url_fer_ct*.



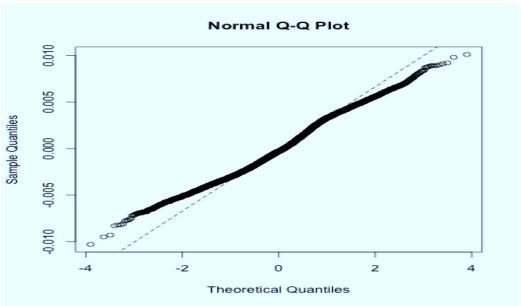

(a) *Qqplot* de la distribució de la $F_0$ a través de tot el corpus.



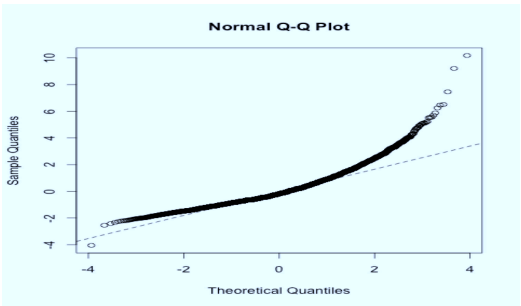

(b) *Qqplot* de la distribució de la $\Delta F_0$ a través de tot el corpus.



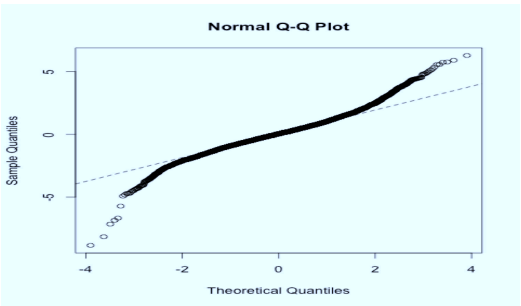

(c) *Qqplot* de la distribució de l'energia a través de tot el corpus.




(d) *Qqplot* de la distribució de l'$\Delta$ d'energia a través de tot el corpus.

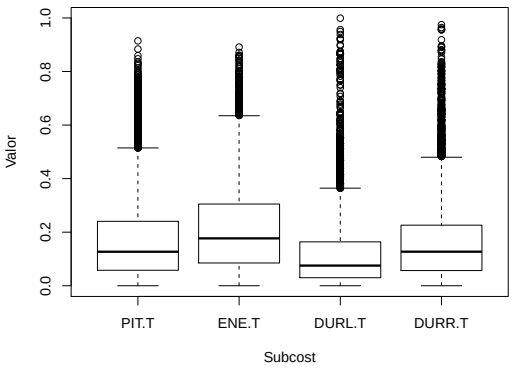



(e) *Qqplot* de la distribució de la durada (*z-score*) a través de tot el corpus.

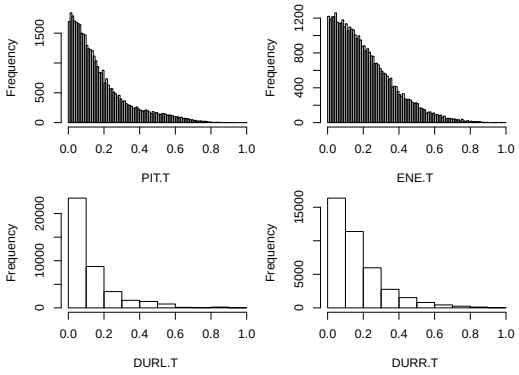



(f) *Qqplot* de la distribució de l'$\Delta$ de durada (*z-score*) a través de tot el corpus.

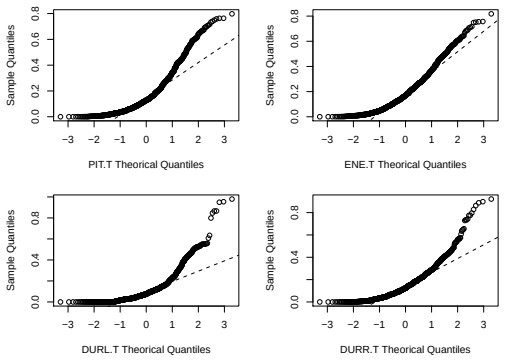
Figura C.2: Comparació quartil-quartil (*qqplot*) de la prosòdia i la seva derivada per tot el corpus *url_fer_ct*.




(a) *Boxplot* dels diferents subcostos de *target* considerats.

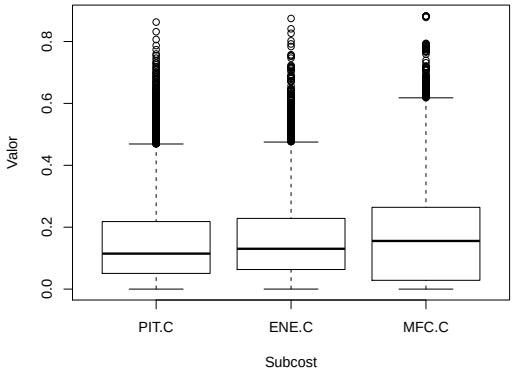



(b) Histogrames dels diferents subcostos de *target* considerats.

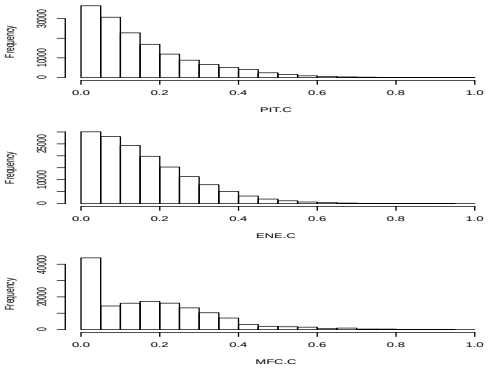



(c) *Qqplot* dels diferents subcostos de *target* considerats.

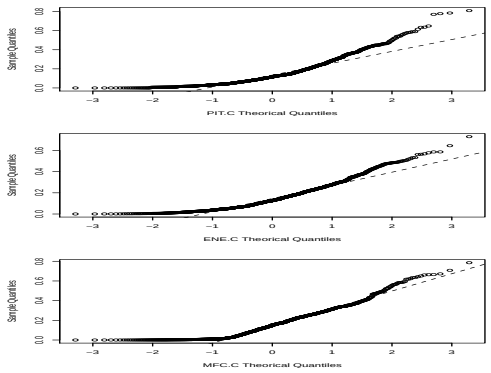
Figura C.3: *Boxplot*, histograma i *qqplot* dels subcostos de *target*, normalitzats segons la funció *max-min*, analitzats per la unitat /@l/ del corpus *url_fer_ct*.




(a) *Boxplot* dels diferents subcostos de concatenació considerats.

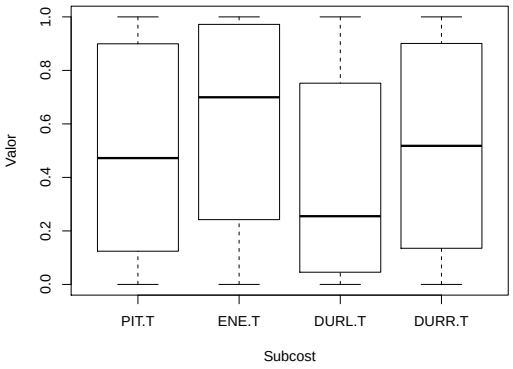



(b) Histogrames dels diferents subcostos de concatenació considerats.

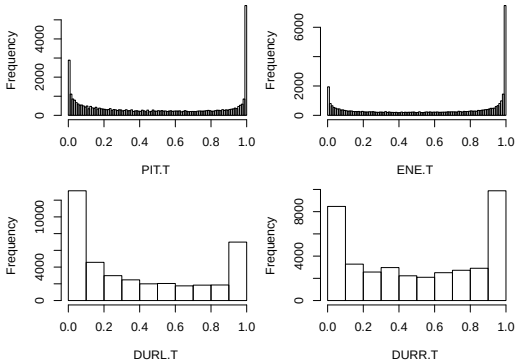



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

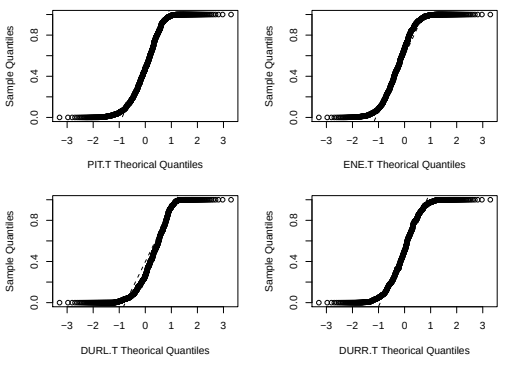
Figura C.4: *Boxplot*, histograma i *qqplot* dels subcostos de concatenació, normalitzats segons la funció *max-min*, analitzats per la unitat /@l/ del corpus *url_fer_ct*.



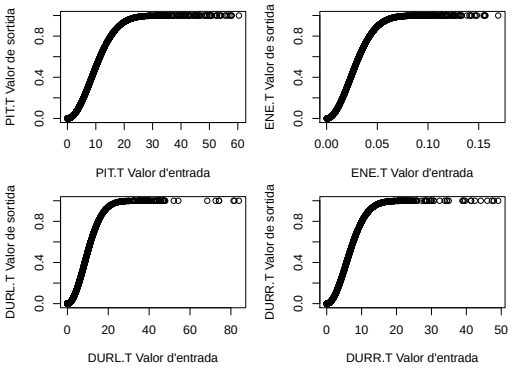

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

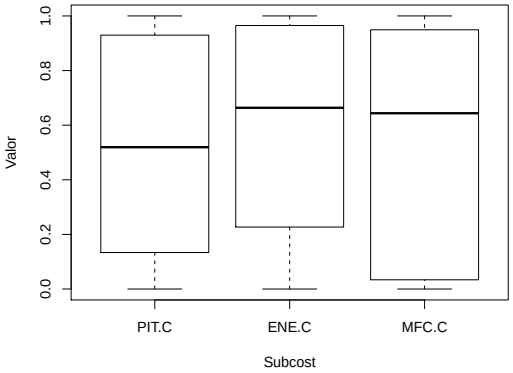



(c) *Qqplot* dels diferents subcostos de *target* considerats.

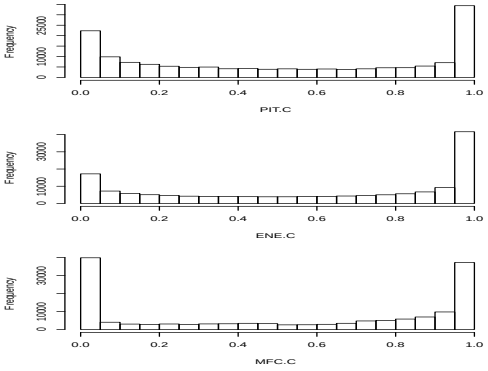



(d) Funció de transferència dels diferents subcostos de *target* considerats.

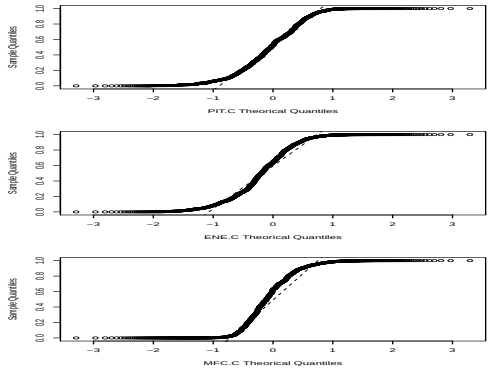
Figura C.5: *Boxplot*, histograma, *qqplot* i funció de transferència dels subcostos de *target* analitzats per la unitat /@l/ del corpus *url_fer_ct* un cop aplicada la normalització sigmoide clàssica.



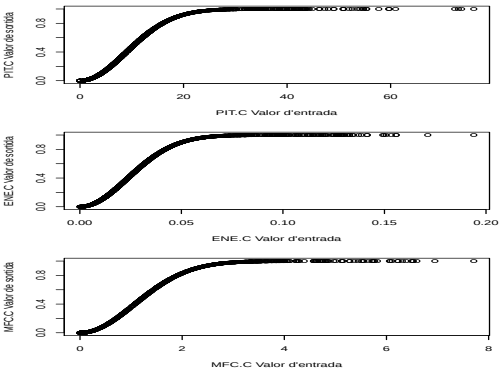

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.

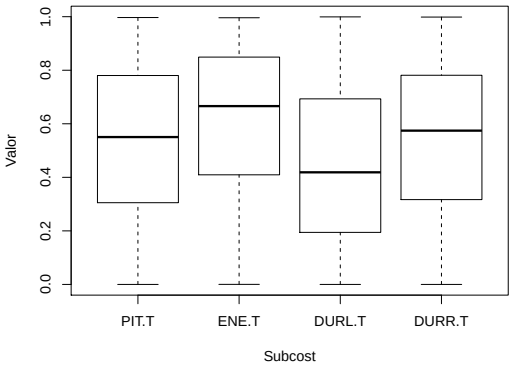



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

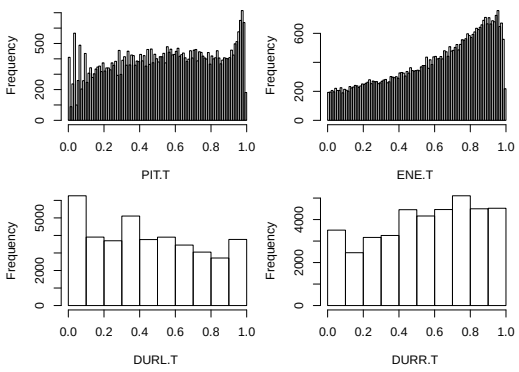



(d) Funció de transferència dels diferents subcostos de concatenació considerats.

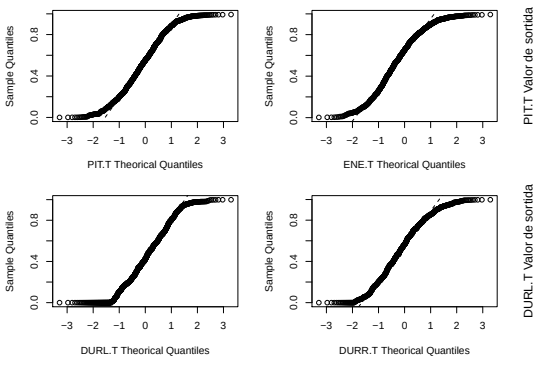
Figura C.6: *Boxplot*, histograma, *qqplot* i funció de transferència dels subcostos de concatenació analitzats per la unitat /@l/ del corpus *url_fer_ct* un cop aplicada la normalització sigmoide clàssica.



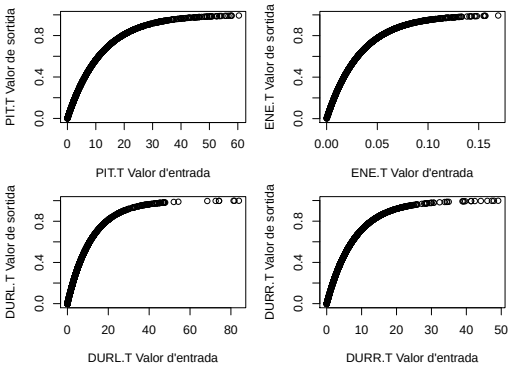

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

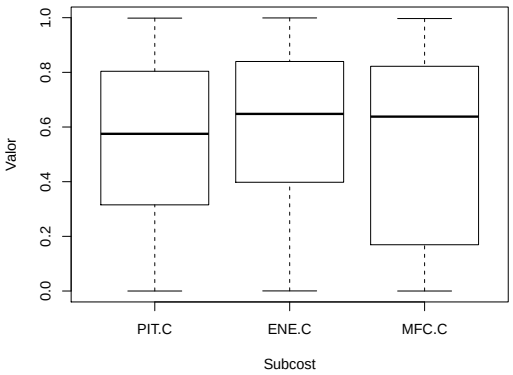



(c) *Qqplot* dels diferents subcostos de *target* considerats.

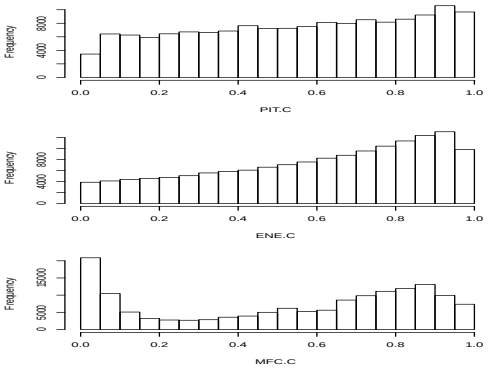



(d) Funció de transferència dels diferents subcostos de *target* considerats.

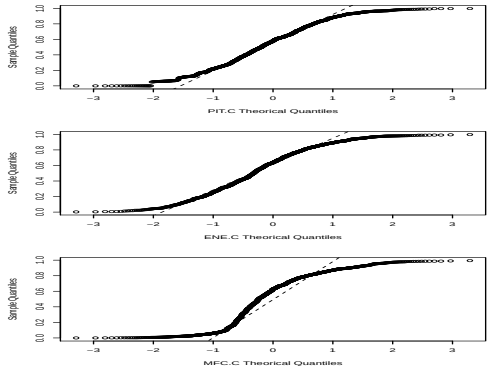
Figura C.7: *Boxplot*, histograma, *qqplot* i funció de transferència dels subcostos de *target* analitzats per la unitat /@l/ del corpus *url_fer_ct* un cop aplicada la normalització sigmoide lineal.



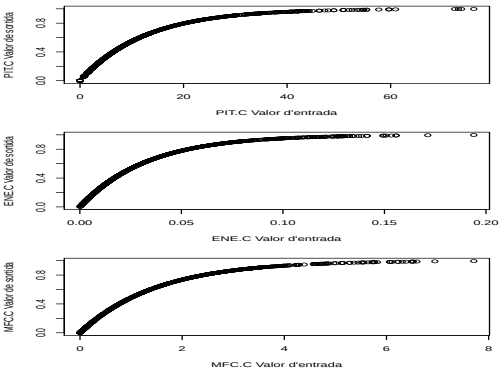

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.




(c) *Qqplot* dels diferents subcostos de concatenació considerats.




(d) Funció de transferència dels diferents subcostos de concatenació considerats.

Figura C.8: *Boxplot*, histograma, *qqplot* i funció de transferència dels subcostos de concatenació analitzats per la unitat /@l/ del corpus *url_fer_ct* un cop aplicada la normalització sigmoide lineal.

# APÈNDIX D

## Descripció del corpus *uvig_dav_es*

En aquest annex es proporciona informació addicional sobre el corpus *uvig_dav_es* (Méndez Pazó *et al.*, 2010).

El corpus està dissenyat i enregistrat per l'Universitat de Vigo i té una durada d'aproximadament dues hores (1.9h) d'un locutor masculí (David). Val a dir que el corpus s'ha dissenyat expressament per un CTP basat en selecció d'unitats (CTP-SU). El corpus es composa de 1217 locucions formades per 17797 paraules, les quals proporcionen una cobertura per a un vocabulari de 5465 paraules diferents.

A la taula D.1 es mostra la tipificació fonètica (tipus, lloc d'articulació, mode d'articulació i sonoritat) de cada al·lòfon del corpus. A la taula D.2 es mostra la presència de cada al·lòfon dins del corpus. A la taula D.3 es mostra la presència de les unitats en el corpus.

A la figura D.1 es mostren els histogrames de la prosòdia i la seva derivada, a la figura D.2 es mostren les comparatives quartil-quartil (*qqplot*) de les distribucions de prosòdia i la seva derivada respecte la distribució normal. A les figures D.3 i D.4 es mostren els *boxplots*, histogrames i *qqplots* dels subcostos de *target* i concatenació analitzats per la unitat /D-e/ del corpus. Finalment, es tornen a mostrar els *boxplots*, histogrames i *qqplots* dels subcostos transformats segons la transformació sigmoide clàssica (figures D.5 i D.6), sigmoide linear (figures D.7 i D.8), logarítmica (figures D.9 i D.10) i d'arrel (figures D.11 i D.12).

Taula D.1: Descripció dels diferents al·lòfons en notació SAMPA (Wells *et al.,* 1992) que composen el corpus *uvig_dav_es*.

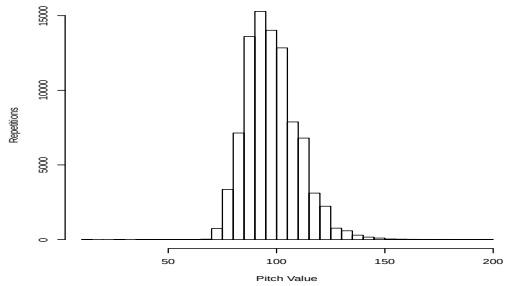
| Al·lòfon | Tipologia | Lloc d'articulació | Mode d'articulació | Sonoritat |
|---|---|---|---|---|
| /B/ | Consonant | Bilabial | Fricativa | Sonora |
| /C/ | Consonant | Alveolar | Africada | Sorda |
| /D/ | Consonant | Dental | Fricativa | Sonora |
| /G/ | Consonant | Velar | Fricativa | Sonora |
| /J/ | Consonant | Palatal | Nasal | Sonora |
| /L/ | Consonant | Palatal | Líquida-Lateral | Sonora |
| /M/ | Consonant | Labiodental | Nasal | Sonora |
| /N/ | Consonant | Velar | Nasal | Sonora |
| /R/ | Consonant | Alveolar | Líquida-Vibrant | Sorda |
| /SIL/ | Silenci | Silenci | Silenci | Sorda |
| /T/ | Consonant | Interdental | Fricativa | Sorda |
| /a/ | Vocal | Anterior | Oberta | Sonora |
| /b/ | Consonant | Bilabial | Oclusiva | Sonora |
| /d/ | Consonant | Alveolar | Oclusiva | Sonora |
| /e/ | Vocal | Anterior | Semi-tancada | Sonora |
| /f/ | Consonant | Labiodental | Fricativa | Sorda |
| /g/ | Consonant | Velar | Oclusiva | Sonora |
| /i/ | Vocal | Anterior | Tancada | Sonora |
| /j/ | Semivocal | Palatal | Aproximant | Sonora |
| /k/ | Consonant | Velar | Oclusiva | Sorda |
| /l/ | Consonant | Alveolar | Líquida-Lateral | Sonora |
| /m/ | Consonant | Bilabial | Nasal | Sonora |
| /n/ | Consonant | Alveolar | Nasal | Sonora |
| /o/ | Vocal | Posterior | Semi-tancada | Sonora |
| /p/ | Consonant | Bilabial | Oclusiva | Sorda |
| /r/ | Consonant | Alveolar | Líquida-Vibrant | Sorda |
| /s/ | Consonant | Alveolar | Fricativa | Sorda |
| /t/ | Consonant | Alveolar | Oclusiva | Sorda |
| /u/ | Vocal | Posterior | Tancada | Sonora |
| /w/ | Semivocal | Velar | Aproximant | Sonora |
| /x/ | Consonant | Velar | Fricativa | Sorda |

Taula D.2: Distribució dels al·lòfons en notació SAMPA (Wells *et al.,* 1992) en castellà a través del corpus *uvig_dav_es*.

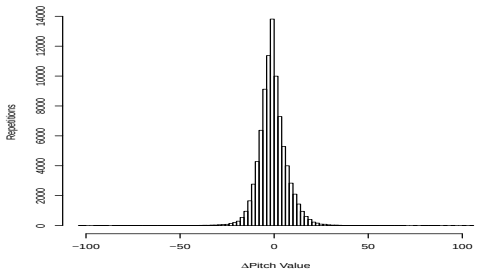
| Fonemes | Repeticions | Percentatge |
|---|---|---|
| /e/ | 11330 | 12.62% |
| /a/ | 11115 | 12.38% |
| /o/ | 8233 | 9.17% |
| /s/ | 6778 | 7.55% |
| /n/ | 5324 | 5.93% |
| /r/ | 4986 | 5.55% |
| /l/ | 4787 | 5.33% |
| /SIL/ | 4290 | 4.78% |
| /i/ | 4247 | 4.73% |
| /t/ | 4215 | 4.69% |
| /D/ | 3960 | 4.41% |
| /k/ | 3442 | 3.83% |
| /m/ | 2505 | 2.79% |
| /j/ | 2295 | 2.56% |
| /p/ | 2266 | 2.52% |
| /B/ | 1837 | 2.05% |
| /u/ | 1737 | 1.93% |
| /T/ | 1707 | 1.9% |
| /G/ | 806 | 0.9% |
| /w/ | 764 | 0.85% |
| /f/ | 723 | 0.81% |
| /R/ | 653 | 0.73% |
| /x/ | 529 | 0.59% |
| /N/ | 426 | 0.47% |
| /L/ | 247 | 0.28% |
| /C/ | 168 | 0.19% |
| /J/ | 144 | 0.16% |
| /d/ | 124 | 0.14% |
| /M/ | 111 | 0.12% |
| /b/ | 26 | 0.03% |
| /g/ | 13 | 0.01% |

Taula D.3: Distribució de les diferents unitats en castellà (en notació SAMPA (Wells *et al.*, 1992)) a través del corpus *uvig_dav_es*, ordenades de major a menor aparició.

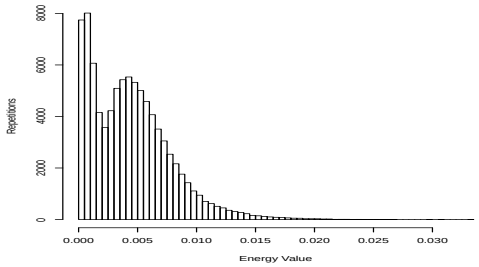
| **Unitats** | **Rep.** | **‰** | **Unitats** | **Rep.** | **‰** | **Unitats** | **Rep.** | **‰** |
|---|---|---|---|---|---|---|---|---|
| /D-e/ | 1734 | 19.58‰ | /n-e/ | 541 | 6.11‰ | /i-n/ | 352 | 3.97‰ |
| /o-s/ | 1642 | 18.54‰ | /n-a/ | 539 | 6.09‰ | /D-a/ | 350 | 3.95‰ |
| /e-l/ | 1518 | 17.14‰ | /'e-r/ | 533 | 6.02‰ | /j-o/ | 350 | 3.95‰ |
| /e-s/ | 1450 | 16.37‰ | /s-SIL2/ | 532 | 6.01‰ | /a-p/ | 345 | 3.9‰ |
| /l-a/ | 1332 | 15.04‰ | /n-D/ | 529 | 5.97‰ | /e-D/ | 341 | 3.85‰ |
| /a-s/ | 1187 | 13.4‰ | /e-r/ | 517 | 5.84‰ | /a-B/ | 339 | 3.83‰ |
| /n-t/ | 1174 | 13.25‰ | /k-a/ | 517 | 5.84‰ | /m-a/ | 333 | 3.76‰ |
| /e-n/ | 1000 | 11.29‰ | /e-k/ | 511 | 5.77‰ | /a-SIL2/ | 328 | 3.7‰ |
| /s-t/ | 966 | 10.91‰ | /p-r/ | 503 | 5.68‰ | /N-k/ | 327 | 3.69‰ |
| /k-o/ | 890 | 10.05‰ | /a-n/ | 498 | 5.62‰ | /e-m/ | 324 | 3.66‰ |
| /s-e/ | 858 | 9.69‰ | /k-e/ | 489 | 5.52‰ | /R-e/ | 315 | 3.56‰ |
| /t-e/ | 813 | 9.18‰ | /s-a/ | 482 | 5.44‰ | /m-o/ | 315 | 3.56‰ |
| /T-j/ | 783 | 8.84‰ | /t-o/ | 479 | 5.41‰ | /j-a/ | 314 | 3.55‰ |
| /l-o/ | 756 | 8.54‰ | /a-k/ | 474 | 5.35‰ | /w-'e/ | 306 | 3.45‰ |
| /r-a/ | 756 | 8.54‰ | /a-r/ | 468 | 5.28‰ | /SIL4-l/ | 305 | 3.44‰ |
| /t-a/ | 740 | 8.35‰ | /n-o/ | 468 | 5.28‰ | /'o-r/ | 304 | 3.43‰ |
| /a-l/ | 724 | 8.17‰ | /o-r/ | 468 | 5.28‰ | /o-m/ | 302 | 3.41‰ |
| /'a-r/ | 701 | 7.91‰ | /t-'a/ | 461 | 5.2‰ | /a-t/ | 297 | 3.35‰ |
| /'e-n/ | 678 | 7.65‰ | /s-p/ | 439 | 4.96‰ | /o-k/ | 295 | 3.33‰ |
| /r-o/ | 670 | 7.56‰ | /p-o/ | 437 | 4.93‰ | /s-o/ | 283 | 3.2‰ |
| /o-n/ | 665 | 7.51‰ | /l-e/ | 419 | 4.73‰ | /r-t/ | 280 | 3.16‰ |
| /'o-n/ | 614 | 6.93‰ | /i-k/ | 415 | 4.69‰ | /t-i/ | 276 | 3.12‰ |
| /a-D/ | 600 | 6.77‰ | /s-k/ | 409 | 4.62‰ | /o-l/ | 271 | 3.06‰ |
| /j-'o/ | 588 | 6.64‰ | /s-D/ | 395 | 4.46‰ | /e-e/ | 270 | 3.05‰ |
| /r-e/ | 581 | 6.56‰ | /s-i/ | 393 | 4.44‰ | /n-T/ | 269 | 3.04‰ |
| /'a-n/ | 574 | 6.48‰ | /a-m/ | 385 | 4.35‰ | /'a-s/ | 267 | 3.01‰ |
| /r-'a/ | 567 | 6.4‰ | /o-D/ | 380 | 4.29‰ | /e-p/ | 267 | 3.01‰ |
| /D-o/ | 559 | 6.31‰ | /a-T/ | 372 | 4.2‰ | /k-'a/ | 266 | 3‰ |
| /t-r/ | 559 | 6.31‰ | /'a-l/ | 370 | 4.18‰ | /n-l/ | 263 | 2.97‰ |
| /'e-s/ | 553 | 6.24‰ | /m-p/ | 363 | 4.1‰ | /n-s/ | 263 | 2.97‰ |
| /'a-D/ | 544 | 6.14‰ | /o-SIL2/ | 359 | 4.05‰ | /i-m/ | 260 | 2.94‰ |
| /j-'e/ | 542 | 6.12‰ | /s-SIL4/ | 358 | 4.04‰ | Altres | 37373 | 42.20% |



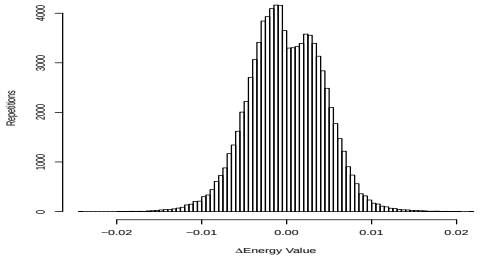

(a) Histograma de la distribució de la $F_0$ a través de tot el corpus.



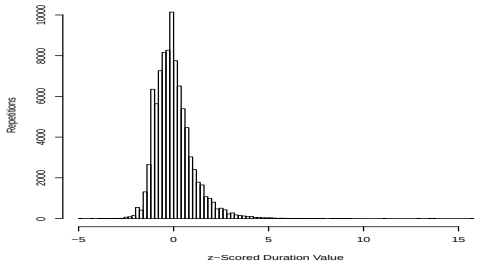

(b) Histograma de la distribució de la $\Delta F_0$ a través de tot el corpus.



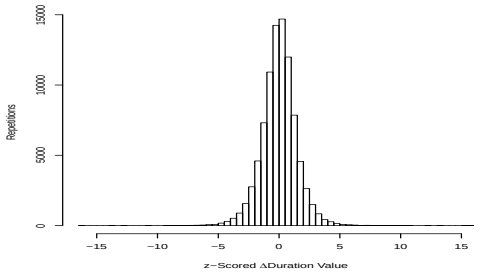

(c) Histograma de la distribució de la energia a través de tot el corpus.




(d) Histograma de la distribució de la $\Delta$ d'energia a través de tot el corpus.

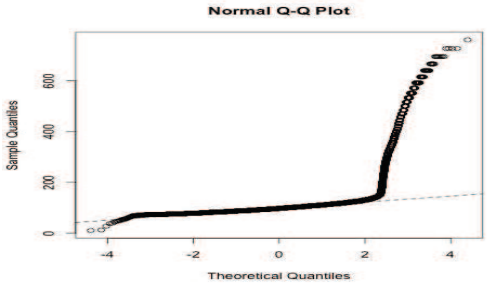



(e) Histograma de la distribució de la durada (*z-score*) a través de tot el corpus.

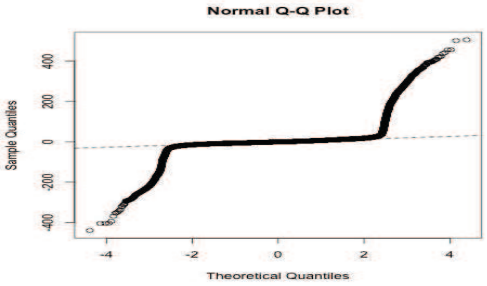



(f) Histograma de la distribució de la $\Delta$ de durada (*z-score*) a través de tot el corpus.

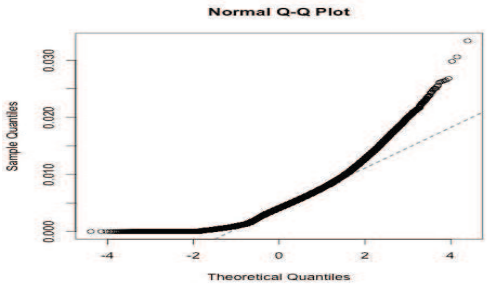
Figura D.1: Histogrames de la prosòdia i la seva derivada per tot el corpus *uvig_dav_es*.



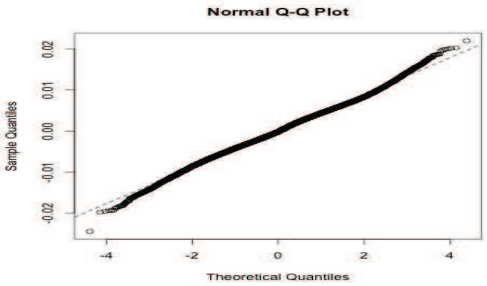

(a) *Qqplot* de la distribució de la $F_0$ a través de tot el corpus.



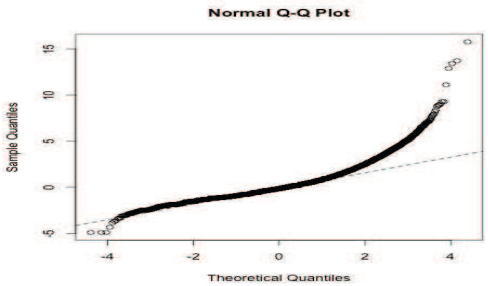

(b) *Qqplot* de la distribució de la $\Delta F_0$ a través de tot el corpus.



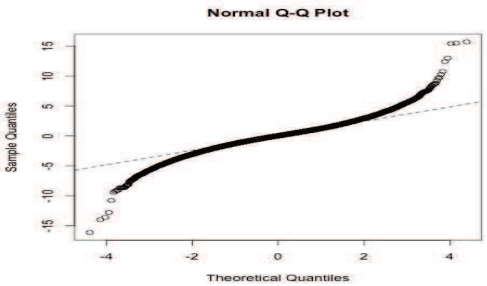

(c) *Qqplot* de la distribució de la energia a través de tot el corpus.




(d) *Qqplot* de la distribució de la $\Delta$ d'energia a través de tot el corpus.

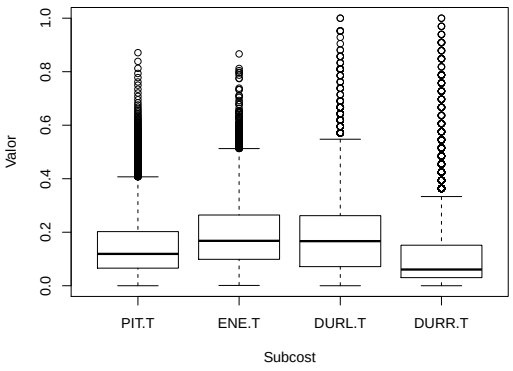



(e) *Qqplot* de la distribució de la durada (*z-score*) a través de tot el corpus.

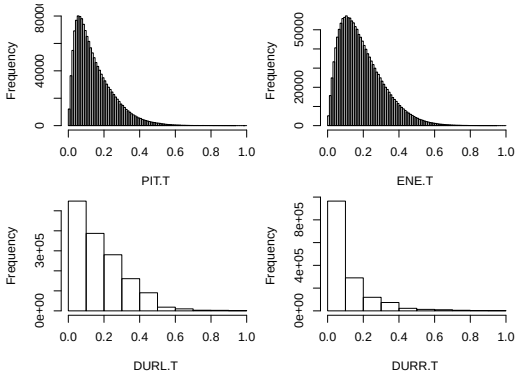



(f) *Qqplot* de la distribució de la $\Delta$ de durada (*z-score*) a través de tot el corpus.

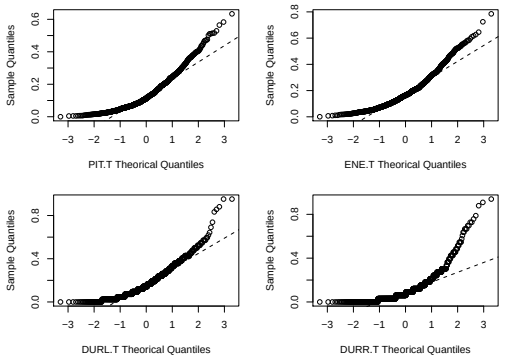
Figura D.2: Comparació quartil-quartil (*qqplot*) de la prosòdia i la seva derivada per tot el corpus *uvig_dav_es*.




(a) *Boxplot* dels diferents subcostos de *target* considerats.

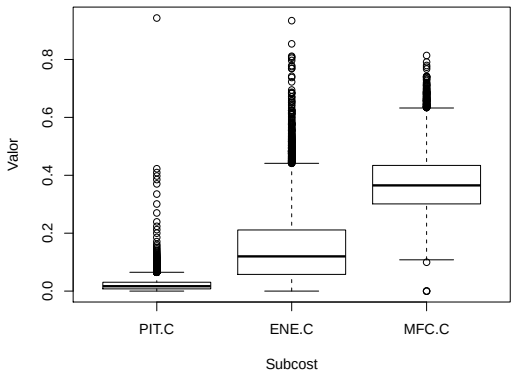



(b) Histogrames dels diferents subcostos de *target* considerats.

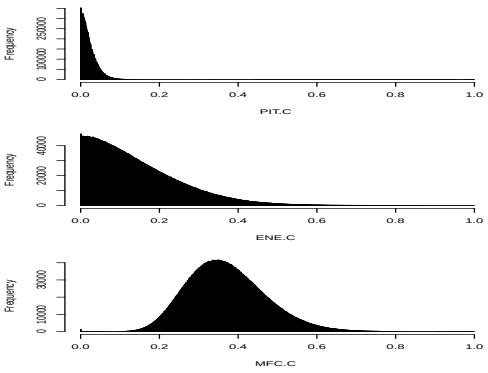



(c) *Qqplot* dels diferents subcostos de *target* considerats.

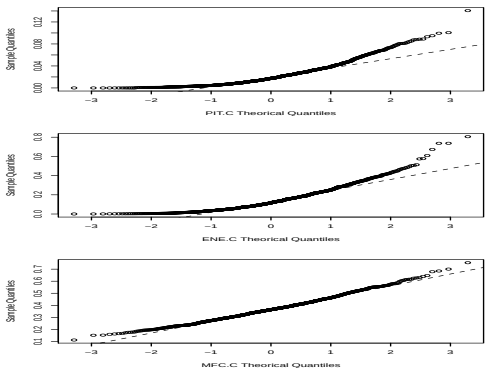
Figura D.3: *Boxplot*, histograma i *qqplot*, normalitzats segons la funció *max-min*, dels subcostos de *target* analitzats per la unitat /D-e/ del corpus *uvig_dav_es*.




(a) *Boxplot* dels diferents subcostos de concatenació considerats.

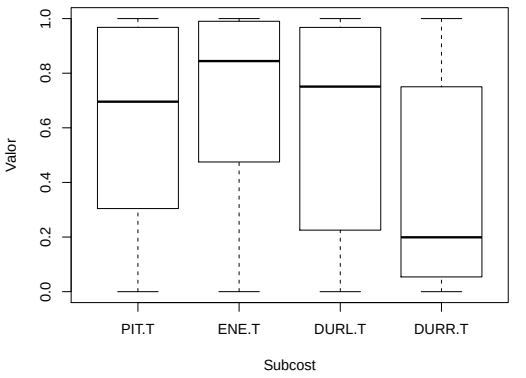



(b) Histogrames dels diferents subcostos de concatenació considerats.

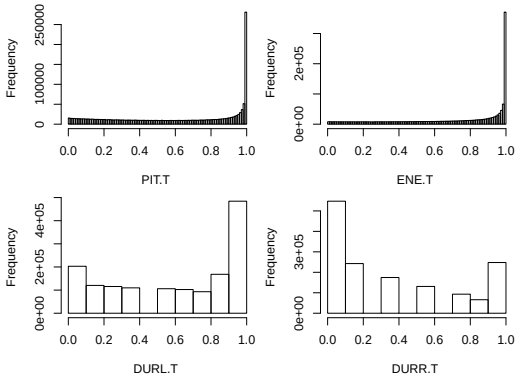



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

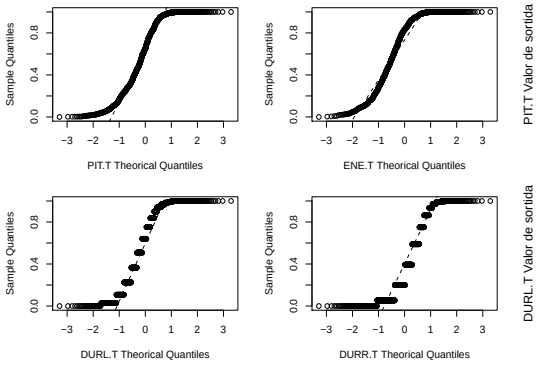
Figura D.4: *Boxplot*, histograma i*qqplot*,normalitzats segons la funció *max-min*, dels subcostos de concatenació analitzats per la unitat /D-e/ del corpus *uvig_dav_es*.



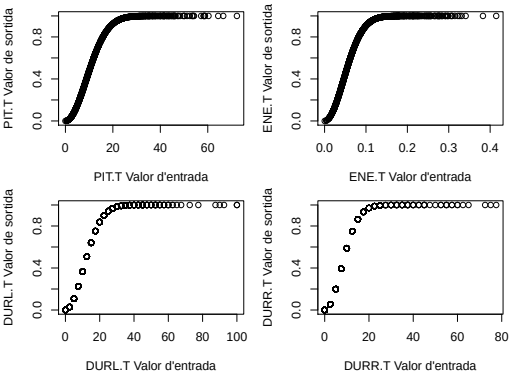

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

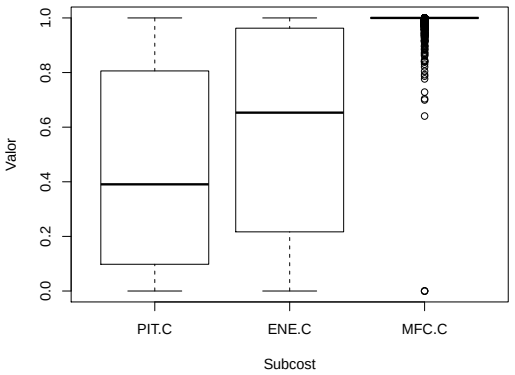



(c) *Qqplot* dels diferents subcostos de *target* considerats.

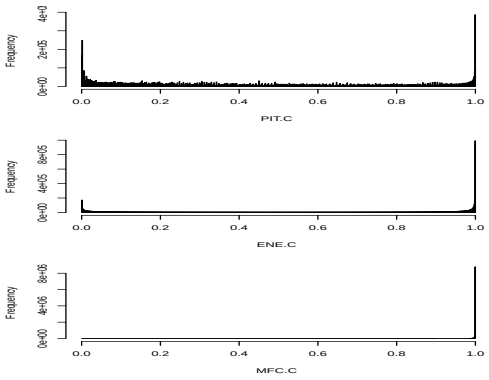



(d) Funció de transferència dels diferents subcostos de *target* considerats.

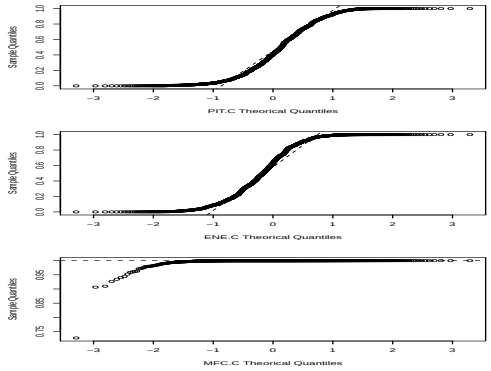
Figura D.5: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de *target* analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització sigmoide clàssica.



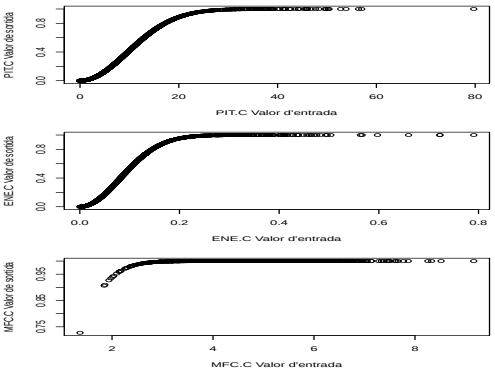

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.

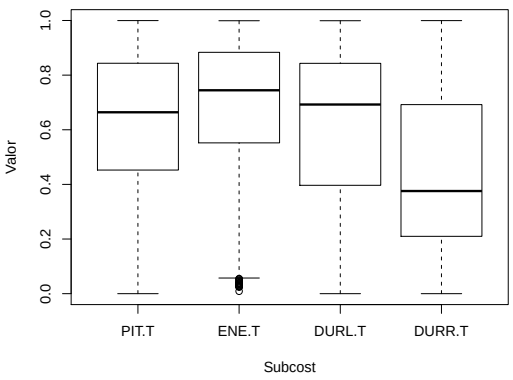



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

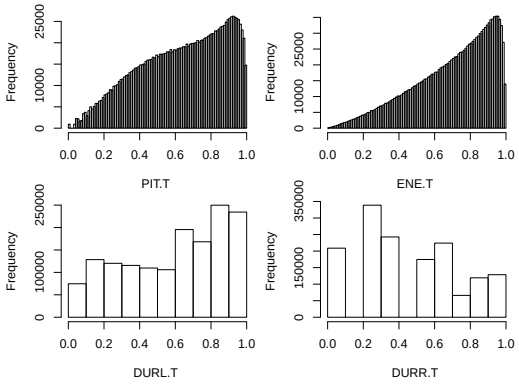



(d) Funció de transferència dels diferents subcostos de concatenació considerats.

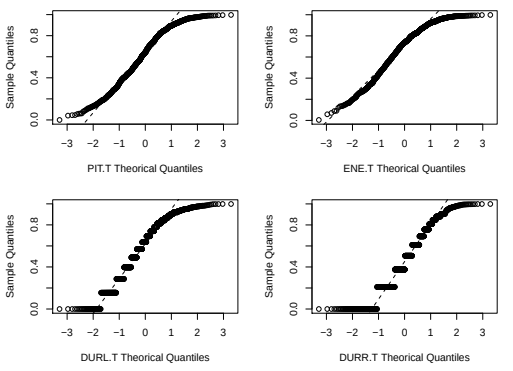
Figura D.6: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de concatenació analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització sigmoide clàssica.



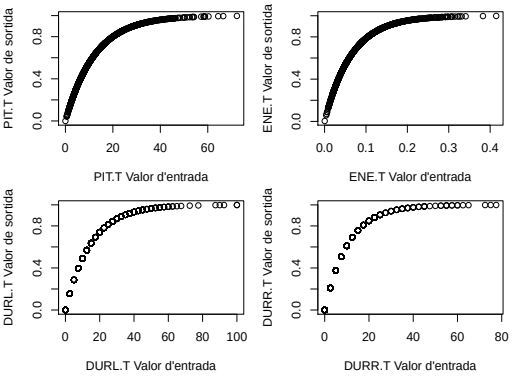

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

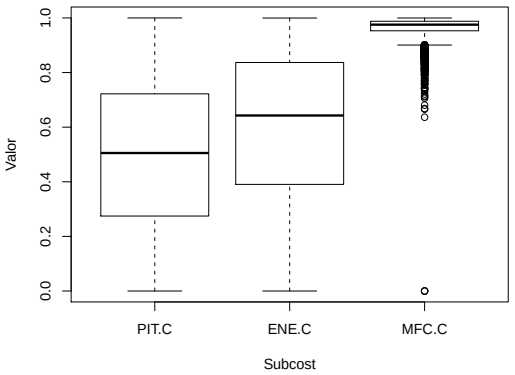



(c) *Qqplot* dels diferents subcostos de *target* considerats.

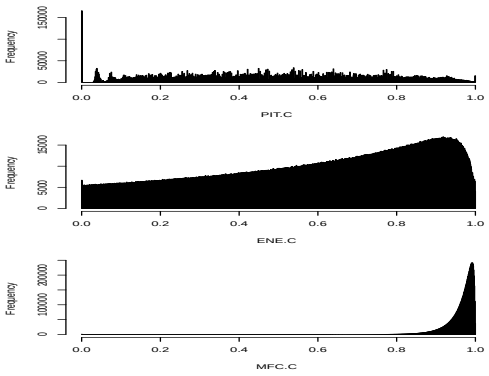



(d) Funció de transferència dels diferents subcostos de *target* considerats.

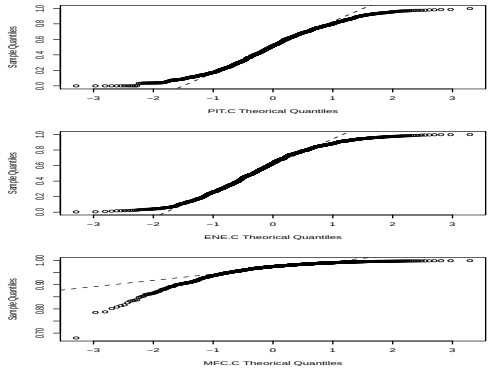
Figura D.7: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de *target* analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització sigmoide linear.



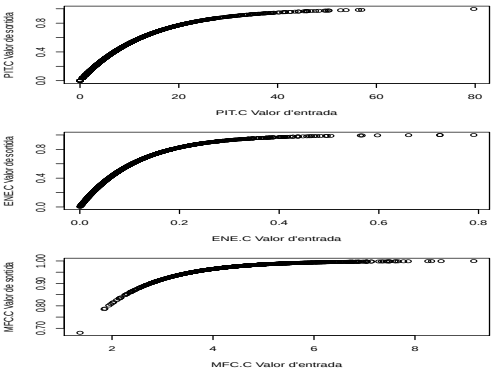

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.

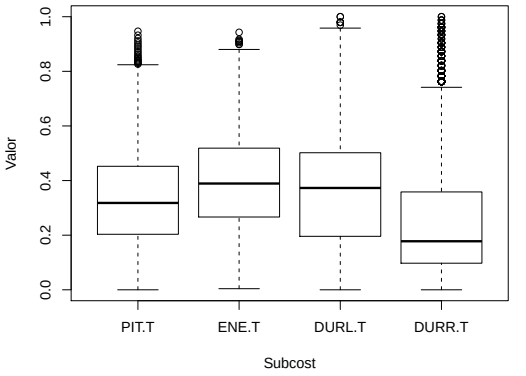



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

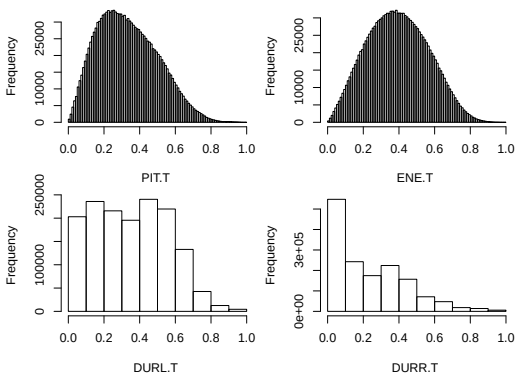



(d) Funció de transferència dels diferents subcostos de concatenació considerats.

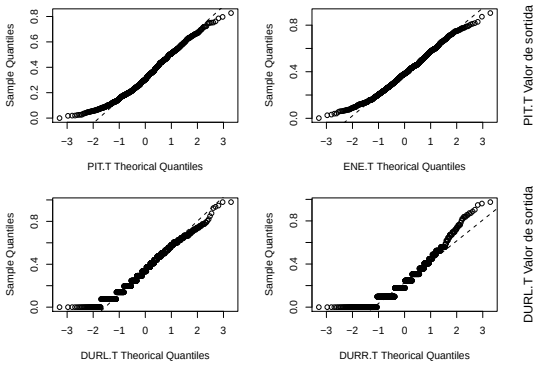
Figura D.8: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de concatenació analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització sigmoide lineal.



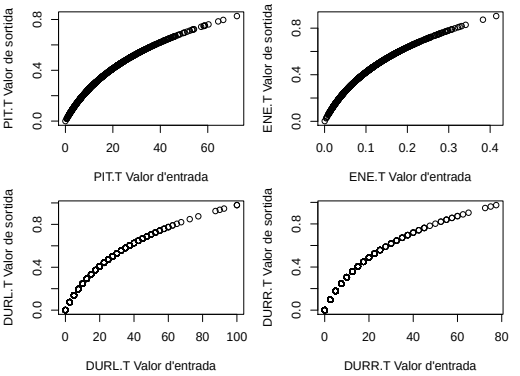

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

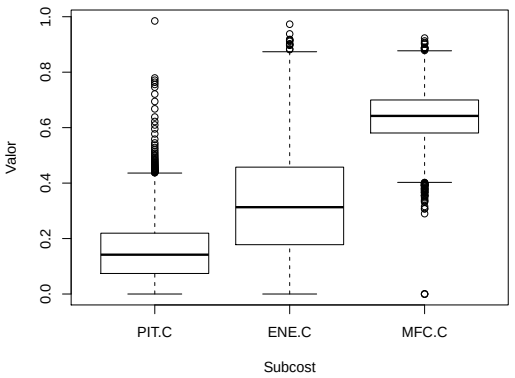



(c) *Qqplot* dels diferents subcostos de *target* considerats.

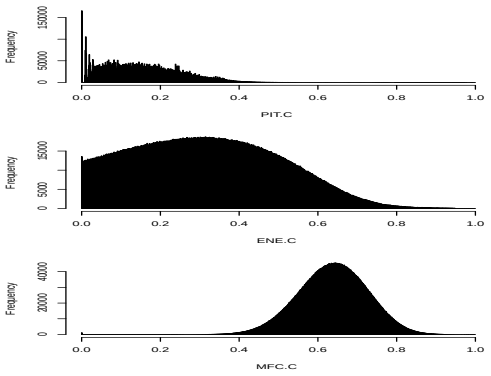



(d) Funció de transferència dels diferents subcostos de *target* considerats.

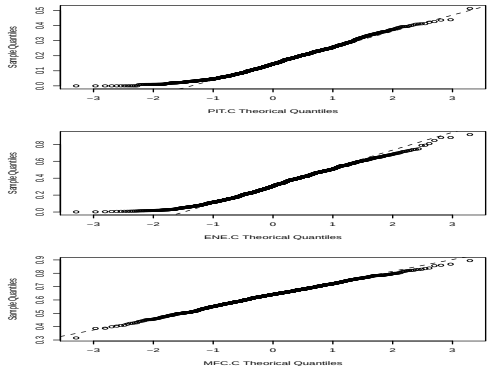
Figura D.9: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de *target* analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització logarítmica.



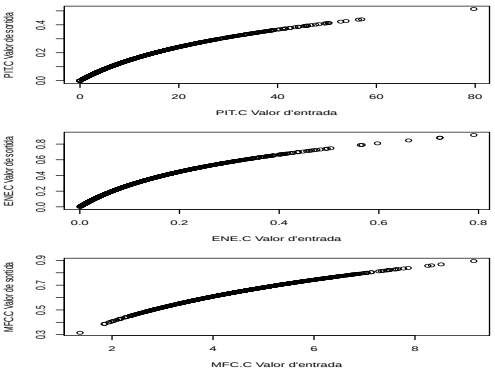

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.

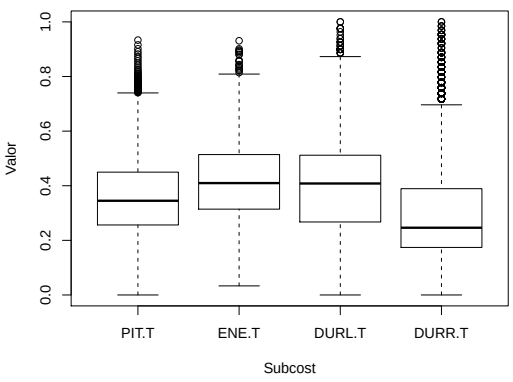



(c) *Qqplot* dels diferents subcostos de concatenació considerats.

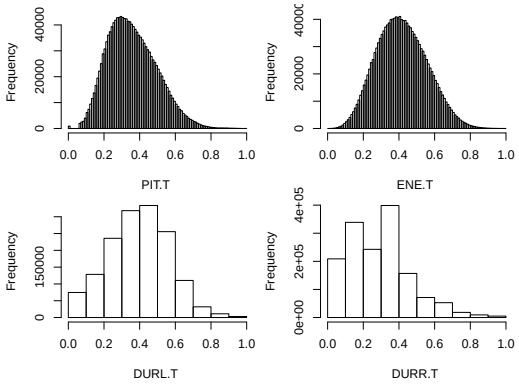



(d) Funció de transferència dels diferents subcostos de concatenació considerats.

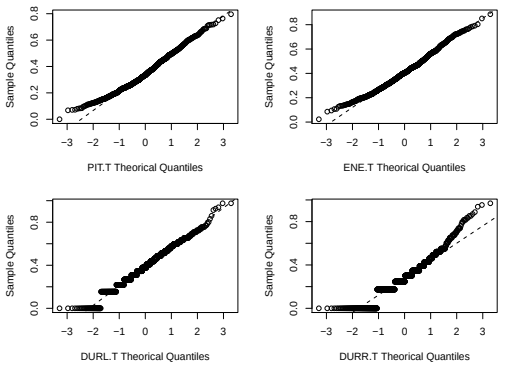
Figura D.10: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de concatenació analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització logarítmica.



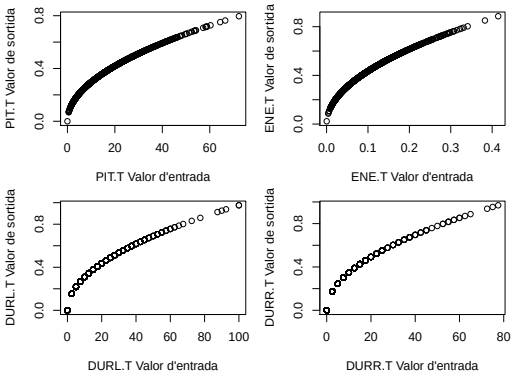

(a) *Boxplot* dels diferents subcostos de *target* considerats.




(b) Histogrames dels diferents subcostos de *target* considerats.

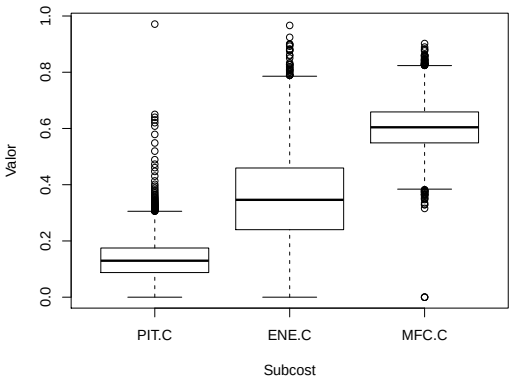



(c) *Qqplot* dels diferents subcostos de *target* considerats.

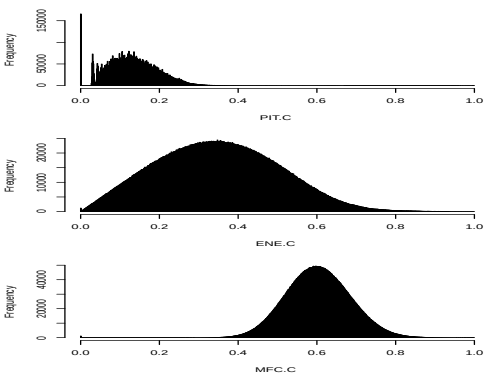



(d) Funció de transferència dels diferents subcostos de *target* considerats.

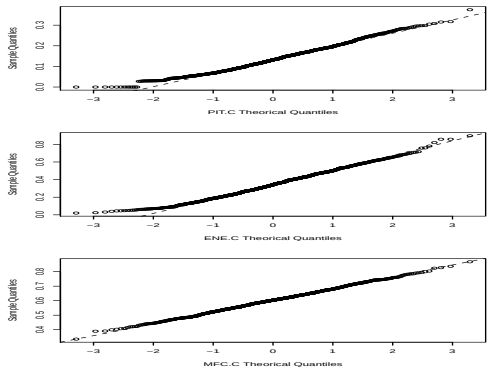
Figura D.11: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de *target* analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització *SQRT*.



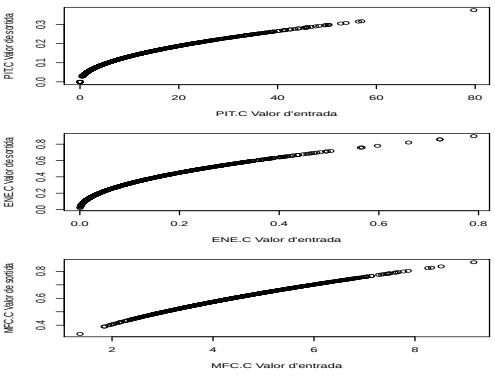

(a) *Boxplot* dels diferents subcostos de concatenació considerats.




(b) Histogrames dels diferents subcostos de concatenació considerats.




(c) *Qqplot* dels diferents subcostos de concatenació considerats.




(d) Funció de transferència dels diferents subcostos de concatenació considerats.

Figura D.12: *Boxplot*, histograma, *qqplot* i funció de transformació dels subcostos de concatenació analitzats per la unitat /D-e/ del corpus *uvig_dav_es* un cop aplicada la normalització *SQRT*.

# Bibliografia


Adell, J.; Bonafonte, A. i Escudero, D. (2006), "Disfluent Speech Analysis and Synthesis: a Preliminary Approach", a *Proc. of 3rd International Conference on Speech Prosody*, ISCA, Dresden (Alemanya), paper 152.

Alías, F. (2006), *Conversión de texto en habla multidominio basada en selección de unidades con ajuste subjetivo de pesos y marcado robusto de pitch*, Ph.D. thesis, La Salle - Universitat Ramon Llull, Barcelona (Espanya).

Alías, F.; Formiga, L. i Llorá, X. (2011), "Efficient and Reliable Perceptual Weight Tuning for Unit-Selection Text-to-Speech Synthesis based on active interactive Genetic Algorithms: A Proof-of-Concept". *Speech Communication*, vol. to appear.
URL `doi:10.1016/j.specom.2011.01.004`

Alías, F. i Iriondo, I. (2002), "La evolución de la Síntesis del Habla en Ingeniería La Salle", a *II Jornadas en Tecnología del Habla*, (RTTH), Red Temática en Tecnologías del Habla, Granada (Espanya).

Alías, F.; Iriondo, I.; Formiga, L.; Gonzalvo, X.; Monzo, C. i Sevillano, X. (2005), "High Quality Spanish restricted-Domain TTS Oriented to a Weather Forecast Application", a *Proc. of the 9th International Conference on Speech Communication and Technology (InterSpeech)*, ISCA, Lisboa (Portugal), (pp. 2573–2576).

Alías, F. i Llorà, X. (2003), "Evolutionary Weight Tuning Based on Diphone Pairs for Unit Selection Speech Synthesis", a *Proc of the 8th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Ginebra (Suïssa), (pp. 1333–1336).

Alías, F.; Llorà, X.; Formiga, L.; Sastry, K. i Goldberg, D. E. (2006*a*), "Efficient Interactive Weight Tuning for TTS Synthesis: Reducing User Fatigue by Improving User Consistency", a *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. I, IEEE, Toulouse (França), (pp. 865–868).

Alías, F.; Llorà, X.; Iriondo, I. i Formiga, L. (2003), "Ajuste subjetivo de pesos para selección de unidades a través de algoritmos genéticos interactivos". *Procesamiento del Lenguaje Natural*, vol. 31:pp. 75–82.

Alías, F.; Llorà, X.; Iriondo, I.; Sevillano, X.; Formiga, L. i Socoró, J.C. (2004), "Perception-Guided and Phonetic Clustering Weight Tuning Based on Diphone Pairs for Unit Selection TTS", a *Proc. of the 8th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Jeju Island (Corea del Sud), (pp. 1221–1224).

Alías, F.; Monzo, C. i Socoró, J. C. (2006*b*), "A Pitch Marks Filtering Algorithm based on Restricted Dynamic Programming", a *Proc. of InterSpeech - International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA), paper 1625.

Alm, C.O. i Llora, X. (2006), "Evolving Emotional Prosody", a *Proc. of 9th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh, PA (EUA), (pp. 1826–1829).

Alvarez, Y.V. i Huckvale, M. (2002), "The Reliability of the ITU-T P. 85 Standard for the Evaluation of Text-to-Speech Systems", a *Proc. of the 7th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Denver, Colorado (EUA), (pp. 329–332).

Angluin, D. (1988), "Queries and Concept Learning". *Machine Learning*, vol. 2(4):pp. 319–342.

Bahl, L.R.; Cocke, J.; Jelinek, F. i Raviv, J. (1974), "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate". *IEEE Transactions on Information Theory*, vol. 20(2):pp. 284–287.

Bailly, G.; Benoit, C. i Sawallis, T.R. (1992), "Tree-Based Modelling of Segmental Durations". *Talking Machines: Theories, Models, and Designs*, vol. 1:pp. 265–273.

Baker, James E. (1985), "Adaptive Selection Methods for Genetic Algorithms", a *Proc. of the 1st International Conference on Genetic Algorithms*, L. Erlbaum Associates Inc., Hillsdale, NJ (EUA), (pp. 101–111).

Balestri, M.; Paechiotti, A.; Quazza, S.; Salza, P. L. i Sandri, S. (1999), "Choose the Best to Modify the Least: a new Generation Concatenative Synthesis System", a *Proc. of the 6th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 5, Budapest (Hongria), (pp. 2291–2294).

Baluja, S. i Caruana, R. (1995), "Removing the Genetics from the Standard Genetic Algorithm", a A Prieditis i S. Russel (editors), *Proc. of the International Conference on Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA (EUA), (pp. 38–46).

Barber, D. (2003), "Probabilistic Modelling and Reasoning The Junction Tree Algorithm", a *Course Note for Probabilistic Modelling and Reasoning, University of Edinburgh*.

Baum, EB (1991), "Neural Network Algorithms that Learn in Polynomial Time from Examples and Queries". *IEEE Transactions on Neural Networks*, vol. 2(1):pp. 5–19.

Beckman, M.E. i Hirschberg, J. (1994), "The ToBI Annotation Conventions", Linguistics Dept - Ohio State University (EUA).

Beightler, C.S.; Phillips, D.T. i Wilde, D.J. (1979), *Foundations of Optimization*, Prentice-Hall, Englewood Cliffs, NJ (EUA).

Bellegarda, J. (2009), "A Novel Approach to Cost Weighting in Unit Selection TTS", a *Proc. of InterSpeech - International Conference on Spoken Language Processing (ICSLP)*, Brighton (Regne Unit), (pp. 744–747).

Bellman, R. (1954), "Some Problems in the Theory of Dynamic Programming". *Econometrica: Journal of the Econometric Society*, vol. 22(1):pp. 37–48.

Beutnagel, M.; Conkie, A. i Syrdal, A. (1998), "Diphone Synthesis using Unit Selection", a *Proc. of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Jenolan Caves (Austràlia), (pp. 185–190).

Bishop, C. M.; Svensen, M. i Williams, C. K. I. (1998), "GTM: The Generative Topographic Mapping". *Neural Computation*, vol. 10(1):pp. 215–234.

Black, A. W. (2002), "Perfect Synthesis for All of the People All of the Time", a *Proc. of the IEEE Workshop on Speech Synthesis*, Santa Monica (EUA), (pp. 167–170).

Black, A. W. i Campbell, N. (1995), "Optimising Selection of Units from Speech Databases for Concatenative Synthesis", a *Proc. of the 4th European Conference on Speech Communication and Technology (Eurospeech)*, vol. 1, Madrid (Espanya), (pp. 581–584).

Black, A. W. i Taylor, P. (1997*a*), "Automatically Clustering Similar Units for Unit Selection in Speech Synthesis", a *Proc. of the 5th European Conference on Speech Communication and Technology (EuroSpeech)*, Rodas (Grècia), (pp. 601–604).

Black, A. W. i Taylor, P. (1997*b*), *The Festival Speech Synthesis System: System documentation*, Human Communciation Research Centre - University of Edinburgh, (United Kingdom).

Black, A. W. i Tokuda, K. (2005), "Blizzard Challenge - 2005: Evaluating Corpus-Based Speech Synthesis on Common Datasets", a *Proc. of the Blizzard Challenge 2005 Workshop*, Lisboa (Portugal), (pp. 77–80).

Boersma, P. i Weenink, D. (2010), "Praat: Doing Phonetics by Computer (version 5.1.37)", http://www.fon.hum.uva.nl/praat/, consultat l'1 de juliol de 2010.

Bollen, K.A. (2002), "Latent Variables in Psychology and the Social Sciences". *Annual Review of Psychology*, vol. 1:pp. 605–635.

Bolshakova, N. i Azuaje, F. (2006), "Estimating the Number of Clusters in DNA Microarray Data". *Methods of Information in Medicine*, vol. 45(2):pp. 153–157.

Bonafonte, A.; Höge, H.; Kiss, I.; Moreno, A.; Ziegenhain, U.; van den Heuvel, H.; Hain, H.U.; Wang, X.S. i Garcia, M.N. (2006), "TC-STAR: Specifications of Language Resources and Evaluation for Speech Synthesis", a *Proc. of Language Resources and Evaluation Conference*, Gènova (Itàlia), (pp. 311–314).

Bonafonte, A.; Moreno, A.; Adell, J.; Agüero, P.D.; Banos, E.; Erro, D.; Esquerra, I.; Pérez, J. i Polyakova, T. (2008), "The UPC TTS System Description for the 2008 Blizzard Challenge", a *Proc. of the Blizzard Challenge 2008 Workshop*, Brisbane (Austràlia), paper 017.

Box, G.E.P. i Draper, N.R. (1987), *Empirical Model-Building and Response Surfaces*, Wiley New York, (EUA).

Breen, A. i Jackson, P. (1998), "Non-Uniform Unit Selection and the Similarity Metric within BT's LAUREATE TTS System", a *Proc of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Jenolan Caves (Austràlia), (pp. 201–206).

Breiman, L.; Friedman, J. H.; Olshen, R. A. i J., Stone C. (1984), *Classification and Regression Trees*, The Wadsworth & Brooks/Cole Advanced & Books Software, (EUA).

Breuer, R. i Abresch, J. (2004), "Phoxsy: Multi-Phone Segments for Unit Selection Speech Synthesis", a *Proc. of the 8th International Conference on Spoken Language Processing (ICSLP)*, Jeju Island (Corea del Sud), (pp. 1217–1220).

Brindle, A. (1981), *Genetic Algorithms for Function Optimization*, Ph.D. thesis, Dept. of Computing Science - University of Alberta, Edmonton (Canadà).

Bulyko, I. (2002), *Flexible Speech Synthesis Using Weighted Finite State Transducers*, Ph.D. thesis, Signal, Speech and Language Interpretation Lab - University of Washington, (EUA).

Cabral, J.P. i Oliveira, L.C. (2006), "EmoVoice: A System to Generate Emotions in Speech", a *Proc. of the 9th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA).

Campbell, N. i Black, A. (1997), *Progress in Speech Synthesis*, cap. Prosody and the Selection of Source Units for Concatenative Synthesis, Springer-Verlag, Berlín (Alemanya), (pp. 279–292).

Campbell, W.N. (1990), "Evidence for a Syllable-Based Model of Speech Timing", a *Proc. of 1st International Conference on Spoken Language Processing (ICSLP)*, ISCA, Kobe (Japó), (pp. 9–12).

Campillo, F.; Alba, J. L. i Rodríguez-Banga, E. (2005), "A Neural Network Approach for the Design of the Target Cost Function in Unit-Selection Speech Synthesis", a *Proc. of the 9th International Conference on Speech Communication and Technology (InterSpeech)*, ISCA, Lisboa (Portugal), (pp. 2533–2536).

Campillo, F. i Rodríguez-Banga, E. (2003), "On the Design of Cost Functions for Unit-Selection Speech Synthesis", a *Proc. of the 8th European Conference on Speech Communication and Technology (EuroSpeech)*, Ginebra (Suïssa), (pp. 289–292).

Campillo, F. i Rodríguez-Banga, E. (2006), "A Method for Combining Intonation Modelling and Speech Unit Selection in Corpus-Based Speech Synthesis Systems". *Speech Communication*, vol. 48(8):pp. 941–956.

Campillo, F. L. (2005), *Síntesis de voz basada en selección de unidades acústicas y prosódicas*, Ph.D. thesis, Universidade de Vigo, Vigo (Espanya).

Ceccaroni, L.; Martínez, P.; Hernández, J. Z. i Verdaguer, X. (2005), "IntegraTV-4all: an Interactive Television for All", a Thomson (editor), *Proc. of the 1st International Symposium on Ubiquitous Computing and Ambient Intelligence (UCAmI)*, (Espanya), (pp. 345–352).

Chang, C. i Lin, Ch. (2001), *LIBSVM: a library for support vector machines*, software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Chatterjee, S. i Hadi, A.S. (2006), *Regression Analysis by Example*, Wiley Series in Probability and Mathematical Statistics, Wiley-Interscience, New Jersey (EUA).

Chazan, D.; Hoory, R.; Cohen, G. i Zibulski, M. (2000), "Speech Reconstruction from Mel Frequency Cepstral Coefficients and Pitch Frequency", a *Proc of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3, IEEE, Istambul (Turquía), (pp. 1299–1302).

Chen, S.F. (2003), "Conditional and Joint Models for Grapheme-to-Phoneme Conversion", a *Proc. of the 8th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Ginebra (Suïssa), (pp. 2033–2036).

Chu, M. i Peng, H. (2001), "An Objective Measure for Estimating MOS of Synthesized Speech", a *Proc. of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, ISCA, Aalborg (Dinamarca), (pp. 2087–2090).

Clark, R. A. J.; Richmond, K. i King, S. (2004), "Festival 2 - Build Your Own General Purpose Unit Selection Speech Synthesiser", a *Proc. of the 5th ISCA Speech Synthesis Workshop*, Pittsburgh (EUA), (pp. 173–178).

Clark, R. A. J.; Richmond, K. i King, S. (2005), "Multisyn Voices from ARCTIC Data for the Blizzard Challenge", a *Proc. of the Blizzard Challenge 2009 Workshop*, ISCA, Lisboa (Portugal), (pp. 101–104).

Clark, R.A.J. i King, S. (2006), "Joint Prosodic and Segmental Unit Selection Speech Synthesis", a *Proc. del 9th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA), paper 1262.

Clark, R.A.J.; Richmond, K. i King, S. (2007), "Multisyn: Open-Domain Unit Selection for the Festival Speech Synthesis System". *Speech Communication*, vol. 49(4):pp. 317–330.

Coello-Coello, Carlos A. (December, 1998), "An Updated Survey of GA-Based Multiobjective Optimization Techniques", Tech. Rep. Lania-RD-09-08, Laboratorio Nacional de Informática Avanzada (LANIA), Xalapa, Veracruz (Mèxic).

Cohn, D.; Atlas, L. i Ladner, R. (1994), "Improving Generalization with Active Learning". *Machine Learning*, vol. 15(2):pp. 201–221.

Cohn, D.A.; Ghahramani, Z. i Jordan, M.I. (1996), "Active Learning with Statistical Models". *Journal of Artificial Intelligence*, vol. 4:pp. 129–145.

Colotte, V. i Beaufort, R. (2005), "Linguistic Features Weighting for a Text-to-Speech System Without Prosody Model", a *Proc. of the 9th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Lisboa (Portugal), (pp. 2549–2552.).

Conkie, A. i Isard, S. (1996), *Progress in Speech Synthesis*, cap. Optimal coupling of diphones, Springer-Verlag, Berlín (Alemanya).

Coorman, G.; Fackrell, J.; Rutten, P. i Coile, B. Van (2000), "Segment Selection in the L&H RealSpeak Laboratory TTS System", a *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP)*, vol. 2, Pequín (Xina), (pp. 395–398).

Cox, R.V.; Kamm, C.A.; Rabiner, L.R.; Schroeter, J. i Wilpon, J.G. (2002), "Speech and Language Processing for Next-Millennium Communications Services". *Proceedings of the IEEE*, vol. 88(8):pp. 1314–1337.

Cristianini, N. i Shawe-Taylor, J. (2000), *An Introduction to Support Vector Machines*, Cambridge Press, (Regne Unit).

Cui, D.; Huang, D.; Dong, Y.; Cai, L. i Wang, H. (2007), "Script Design Based on Decision Tree with Context Vector and Acoustic Distance for Mandarin TTS", a *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, IEEE, Istambul (Turquía), (pp. 713–716).

Davis, L. (1991), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, Nova York (EUA).

Deb, K.; Agrawal, S.; Pratab, A. i Meyarivan, T. (2000), "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II", a Springer-Verlag (editor), *Proc. of the 6th International Conference on Parallel Problem Solving from Nature (PPSN)*, vol. 1, Londres (Regne Unit), (pp. 849–858), Indian Institute of Technology. KanGAL report 200001.

Dempster, A.P.; Laird, N.M.; Rubin, D.B. *et al.* (1977), "Maximum Likelihood from Incomplete Data Via the EM Algorithm". *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39(1):pp. 1–38.

Ding, W. i Campbell, N. (1997), "Optimising Unit Selection with Voice Source and Formants in the CHATR Speech Synthesis System", a *Proc. of the 5th European Conference on Speech Communication and Technology (EuroSpeech)*, Rodas (Grècia), (pp. 537–540).

Donovan, R. (2000), "Segment Preselection in Decision-Tree Based Speech Synthesis Systems", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, IEEE, Istambul (Turquía), (pp. 937–940).

Donovan, R. E. (2001), "A New Distance Measure for Costing Spectral Discontinuities in Concatenative Speech Synthesizers", a *Proc of the 4th ISCA Workshop on Speech Synthesis*, Perthshire (Escòcia), paper 123.

Donovan, R. E. i Eide, E. M. (1998), "The IBM Trainable Speech Synthesis System", a *Proc. of International Conference on Spoken Language Processing (ICSLP)*, vol. 5, ISCA, Sydney (Austràlia), (pp. 1703–1706).

Donovan, R. E.; Franz, M.; Sorensen, J. S. i Roukos, S. (1999), "Phrase Splicing and Variable Substitution using the IBM Trainable Speech Synthesis System", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, IEEE, Phoenix (EUA), (pp. 373–376).

Donovan, R.E. i Woodland, PC (1999), "A Hidden Markov-Model-Based Trainable Speech Synthesizer". *Computer Speech and Language*, vol. 13(3):pp. 223–242.

Dudoit, S. i Fridlyand, J. (2002), "A Prediction-Based Resampling Method for Estimating the Number of Clusters in a Dataset". *Genome Biology*, vol. 3(7):pp. 1–21.

Durant, E.; Wakefield, G.; Van Tasell, D. i Rickert, M. (2004), "Efficient Perceptual Tuning of Hearing Aids with Genetic Algorithms". *Transactions of IEEE on Speech and Audio Processing*, vol. 12(2):pp. 144–155.

Dutoit, T. (1997), *An Introduction to Text-to-Speech Synthesis*, Kluwer Academic Publishers, Dordrecht (Països Baixos).

Escudero, D. i Cardeñoso, V. (2003), "Modelado estadístico de entonación con funciones de Bézier: Aplicaciones a la conversión texto-voz en Español". *Procesamiento del Lenguaje Natural*, (30):pp. 125–126.

Estruch, M.; Garrido, J.M.; Llisterri, J. i Riera, M. (1996), "Una aproximación fonética al estudio de la entonación". *Philologia Hispalensis*, vol. 1:pp. 281–293.

Febrer, A. (2001), *Síntesi de la parla per concatenació basada en la selecció*, Ph.D. thesis, Dept. de Teoria del Senyal i Comunicacions - Universitat Politècnica de Catalunya, Barcelona (Espanya).

Feldman, R. i Sanger, J. (2008), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, vol. 34, Cambridge University Press, (Regne Unit).

Ferrer, M.; Serratosa, F. i Sanfeliu, A. (2005), "Synthesis of Median Spectral Graph". *Lecture Notes in Computer Science - Pattern Recognition and Image Analysis*, vol. 3523:pp. 139–146.

Fessant, F. i Midenet, S. (2002), "Self-organising Map for Data Imputation and Correction in Surveys". *Neural Computing and Applications*, vol. 10(4):pp. 300–310.

Fogel, L.J.; Owens, A.J. i Walsh, M.J. (1966), *Artificial Intelligence through Simulated Evolution*, John Wiley & Sons Inc, Nova York (EUA).

Formiga, L. (2003), *Ajust de pesos de selecció per a síntesi de la parla a través d'algorismes genètics interactius*, Treball Final de Carrera. Enginyeria Tècnica en Informàtica de Sistemes. La Salle - Universitat Ramon Llull, Barcelona (Espanya).

Formiga, L. (2005), *Reducció de la fatiga i la ambigüitat en l'ajust subjectiu de pesos per a síntesi de la parla*, Projecte Final de Carrera. Enginyeria Superior en Informàtica, La Salle - Universitat Ramon Llull, Barcelona (Espanya).

Formiga, L. i Alías, F. (2007), "Extracting User Preferences by GTM for aiGA Weight Tuning in Unit Selection Text-to-Speech Synthesis". *Lecture Notes in Computer Science - Computational and Ambiental Intelligence*, vol. 4507:pp. 654–661, proc. of the 9th International Work-Conference on Artificial Neural Networks (IWANN).

Formiga, L.; Alías, F. i Llorà, X. (2010), "Evolutionary Process Indicators for active iGAs Applied to Weight Tuning in Unit Selection TTS Synthesis", a *Proc. of the IEEE Conference on Evolutionary Computation (CEC)*, IEEE, Barcelona (Espanya), (pp. 2322–2329).

Fornells, A. (2007), *Marc integrador de les capacitats de Soft-Computing i de Knowledge Discovery dels Mapes Autoorganitzatius en el Raonament Basat en Casos*, Ph.D. thesis, La Salle - Universitat Ramon Llull, Barcelona (Espanya).

Fraley, C. i Raftery, A.E. (1998), "How Many Clusters? Which Clustering Method? Answers Via Model-based Cluster Analysis". *The Computer Journal*, vol. 41(8):p. 578.

Fujisaki, H. (1992), "The Role of Quantitative Modeling in the Study of Intonation", a *Symposium on Research on Japanese and its Pedagogical Applications*, Nava (Japó).

Garrido, J.M. (2001), "La estructura de las curvas melódicas del español: propuesta de modelización". *Lingüística Española Actual*, vol. 23(2):pp. 173–209.

Gerhard, D. (2003), "Pitch Extraction and Fundamental Frequency: History and Current Techniques", Tech. Rep. TR-CS 2003-06, University of Regina, Regina, Saskatchewan (Canadà).

Gersho, A. i Gray, R.M. (1992), *Vector Quantization and Signal Compression*, Springer-Verlag, (Països Baixos).

Gibson, W. (1959), "Three Multivariate Models: Factor Analysis, Latent Structure Analysis, and Latent Profile Analysis". *Psychometrika*, vol. 24:pp. 229–252.

Goldberg, D. E. (1989), *Genetic Algorithms in Search Optimization and Machine Learning*, Addison-Wesley Longman Publishing, Boston (EUA).

Goldberg, D. E. (2002), *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, Boston (EUA).

Goldberg, D. E.; Korb, B. i Deb, K. (1989), "Messy Genetic Algorithms: Motivation, Analysis and First Results". *Complex Systems*, vol. 3(5):pp. 493–530.

Goldberg, D.E. i Deb, K. (1991), "A Comparative Analysis of Selection Schemes Used in Genetic Algorithms". *Foundations of Genetic Algorithms*, vol. 1:pp. 69–93.

Goldberg, D.E. i Voessner, S. (1999), "Optimizing Global-Local Search Hybrids", a *Genetic and Evolutionary Computation Conference (GECCO)*, vol. 1, Orlando (EUA), (pp. 220–228).

Goubanova, O. i Taylor, P. (2000), "Using Bayesian Belief Networks for Model Duration in Text-to-Speech Systems", a *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pequín (Xina), CD-ROM.

Grefenstette, J. J. i Baker, J. E. (1989), "How Genetic Algorithms Work: a Critical Look at Implicit Parallelism", a *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers Inc., San Francisco, CA (EUA), (pp. 20–27).

GTM (2008*a*), "Cuenta Cuentos 2.0", http://www.salle.url.edu/portal/departaments/home-depts-DTM-projectes-info?id_projecte=51.

GTM (2008*b*), "Reune-T", http://www.salle.url.edu/portal/departaments/ home-depts-DTM-projectes-info?id_projecte=50.

Guaus, R. i Iriondo, I. (2000*a*), "Diphone-Based Unit Selection for Catalan TTS Synthesis". *Lecture Notes in Computer Science*, vol. 1902:pp. 43–60, Proceedings of the International Conference on Text, Speech and Dialogue (TSD).

Guaus, R. i Iriondo, I. (2000*b*), "Unit Selection based on Diphones for Catalan Text-to-Speech Conversion", a *Workshop on Developing Language Resources for Minority Languages*, Atenes (Grècia).

Günter, S. i Bunke, H. (2003), "Validation Indices for Graph Clustering". *Pattern Recognition Letters*, vol. 24(8):pp. 1107–1113.

Haas, W. i Thallinger, G. (2005), "SALERO: Semantic Audiovisual Entertainment Reusable Objects", a *Proc. of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT)*, Londres (Regne Unit), (pp. 383–384).

Harik, G. R.; Lobo, F. G. i Goldberg, D. E. (1999), "The Compact Genetic Algorithm". *IEEE Transactions on Evolutionary Computation*, vol. 3(4):pp. 287–297.

Harris, F.J. (1978), "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform". *Proceedings of the IEEE*, vol. 66(1):pp. 51–83.

Hartigan, J.A. i Wong, M.A. (1979), "A *k-means* Clustering Algorithm". *Journal of the Royal Statistical Society, Serie C*, vol. 28:pp. 100–108.

Hirai, T. i Tenpaku, S. (2004), "Using 5 ms Segments in Concatenative Speech Synthesis", a *Proc. of the 5th ISCA Workshop on Speech Synthesis*, ISCA, Pittsburgh (EUA).

Hirschberg, J. i Nakatani, C.H. (1998), "Acoustic Indicators of Topic Segmentation", a *Proc. of the 5th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Sydney (Austràlia).

Hochberg, Y. (1988), "A Sharper Bonferroni Procedure for Multiple Tests of Significance". *Biometrika*, vol. 75(4):pp. 800–802.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*, Univesity of Michigan Press, (EUA).

Hollander, M. i Wolfe, D.A. (1973), *Nonparametric Statistical Inference*, John Wiley & Sons, Inc, Nova York (EUA).

Holmes, J.N. i Holmes, W.J. (2001), *Speech Synthesis and Recognition*, Taylor and Francis, Inc., Bristol, PA (EUA).

Horn, J.; Nafpliotis, N. i Goldberg, D. E. (1994), "A Niched Pareto Genetic Algorithm for Multiobjective Optimization", a *Proc. of the 1st IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, vol. 1, Piscataway, New Jersey (EUA), (pp. 82–87).

Hripcsak, G. i Rothschild, A.S. (2005), "Agreement, the f-Measure, and Reliability in Information Retrieval". *Journal of the American Medical Informatics Association*, vol. 12(3):pp. 296–298.

Hunt, A. i Black, A. W. (1996), "Unit Selection in a Concatenative Speech Synthesis System Using a Large Speech Database", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Atlanta (EUA), (pp. 373–376).

Hunt, M.J.; Zwierzyriski, D.A. i Can, R.C. (1989), "Issues in High Quality LPC Analysis and Synthesis", a *Proc. of the 1st European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Paris (França), (pp. 2348–2351).

Iriondo, I. (2008), *Producción de un corpus oral y modelado prosódico para la síntesis del habla expresiva*, Ph.D. thesis, La Salle - Universitat Ramon Llull, Barcelona (Espanya).

Iriondo, I.; Planet, S.; Socoró, J.C.; Martínez, E.; Alías, F. i Monzo, C. (2008), "Automatic Refinement of an Expressive Speech Corpus Assembling Subjective Perception and Automatic Classification". *Speech Communication*, vol. 51(9):pp. 744–758, Special Issue on Non-Linear and Conventional Speech Processing.

Iriondo, I.; Socoró, J. C. i Alías, F. (2007), "Prosody Modelling of Spanish for Expressive Speech Synthesis", a *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 4, Honolulu (EUA), (pp. 821–824).

Itakura, F. (1975), "Line Spectrum Representation of Linear Predictor Coefficients of Speech Signals". *The Journal of the Acoustical Society of America*, vol. 57(S1):p. S35.

ITU-T (1996), "Methods for Subjective Determination of Transmission Quality", Recommendation ITU-T P.800.

Iwahashi, N.; Kaiki, N. i Sagisaka, Y. (1993), "Speech Segment Selection for Concatenative Synthesis Based on Spectral Distorsion Discrimination". *IEICE Transactions Fundamentals*, vol. E76-A(11):pp. 1942–1948.

Jain, A.; Nandakumar, K. i Ross, A. (2005), "Score Normalization in Multimodal Biometric Systems". *Pattern Recognition*, vol. 38(12):pp. 2270–2285.

Jain, A.K.; Murty, M.N. i Flynn, P.J. (1999), "Data Clustering: a Review". *ACM Computing Surveys (CSUR)*, vol. 31(3):pp. 264–323.

Jiménez, F. (2008), *Inteligencia Artificial*, cap. Computación Evolutiva, McGraw-Hill, (Espanya), (pp. 433–469).

Johanson, B. i Poli, R. (1998), "GP-Music: An Interactive Genetic Programming System for Music Generation with Automated Fitness Raters", Tech. Rep. CSRP-98-13, School of Computer Science - The University of Birmingham, (Regne Unit).

Johnson, S.C. (1967), "Hierarchical Clustering Schemes". *Psychometrika*, vol. 32(3):pp. 241–254.

Jordan, M. I. (2004), "Graphical models". *Statistical Science*, vol. 19(1):pp. 140–155.

Jurafsky, D.; Martin, J.H.; Kehler, A.; Vander Linden, K. i Ward, N. (2000), *Speech and Language Processing*, vol. 934, Prentice Hall, Nova York (EUA).

Kang, G. i Fransen, L. (1987), "Experimentation with Synthesized Speech Generated from Line-Spectrum Pairs". *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 35(4):pp. 568–571.

Kaszczuk, M. i Osowski, L. (2009), "The IVO Software Blizzard Challenge 2009 Entry: Improving IVONA Text-To-Speech", a *Proc. of the Blizzard Challenge 2009 Workshop*, Edinburgh (Regne Unit), paper 010.

King, S.; Black, AW; Taylor, P.; Caley, R. i Clark, R. (2002), "Edinburgh Speech Tools Library".
URL `www.cstr.ed.ac.uk/projects/speech_tools/manual-1.2.0/`

Klabbers, E. i Veldhuis, R. (1998), "On the Reduction of Concatenation Artefacts in Diphone Synthesis", a *Proc. of the International Conference on Speech Language Processing*, Sydney (Austràlia), (pp. 1983—1986).

Klatt, D.H. (1979), "Synthesis by Rule of Segmental Durations in English Sentences". *Frontiers of Speech Communication Research*, vol. 1:pp. 287–300.

Kohonen, T. (1982), "Analysis of a Simple Self-Organizing Process". *Biological Cybernetics*, vol. 44(2):pp. 135–140.

Kohonen, T. (1990), "The Self-Organizing Map". *Proceedings of the IEEE*, vol. 78(9):pp. 1464–1480.

Kohonen, T. (1995), *Self-Organizing Maps*, vol. 30 de *Information Series*, Springer-Verlag, (Alemanya).

Koishida, K.; Tokuda, K.; Kobayashi, T. i Imai, S. (1995), "CELP Coding Based on Mel-Cepstral Analysis", a *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, Detroit (EUA).

Kominek, J. i Black, A. W. (2004), "The CMU ARCTIC Speech Databases", a *Proc. of the 5th ISCA Speech Synthesis Workshop*, Pittsburgh (EUA), (pp. 223–224).

Kominek, J. i Black, A.W. (2005), "Measuring Unsupervised Acoustic Clustering through Phoneme Pair Merge-and-Split Tests", a *Proc. of the 9th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Lisboa (Portugal), (pp. 689–692).

Kosorukoff, A. i Goldberg, DE (2002), "Genetic Algorithm as a Form of Organization", a *Proc. of the Genetic and Evolutionary Computation Conference (GECCO)*, Nova York (EUA), (pp. 965–972).

Koza, J. i Poli, R. (1992), *Search Methodologies*, cap. Genetic Programming, Springer, (Alemanya), (pp. 127–164).

Larrañaga, P. i Lozano, J. A. (2002), *Estimation of Distribution Algorithms. A New Tool for Evolutionary Computation*, Kluwer Academic Publishers.

Law, K.M. i Lee, T. (2000), "Using Cross-Syllable Units for Cantonese Speech Synthesis", a *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP)*, vol. 2, ISCA, Lisboa (Portugal), (pp. 407–410).

Lawson, C.L. i Hanson, R.J. (1995), *Solving Least Squares Problems*, Classics In Applied Mathematics, Society for Industrial Mathematics, Englewood Cliffs, NJ (EUA).

Lee, M.; Lopresti, D.P. i Olive, J.P. (2003), "A Text-to-Speech Platform for Variable Length Optimal Unit Searching using Perception Based Cost Functions". *International Journal of Speech Technology*, vol. 6(4):pp. 347–356.

Lenzo, K. i Black, A. W. (2000), "Diphone Collection and Synthesis", a *Proc. of the 6th International Conference on Spoken Language Processing (ICSLP)*, vol. 3, Pequín (Xina), (pp. 306–309).

Leon-Garcia, A. (1994), *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, (EUA).

Lilliefors, H. (1967), "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown", a *Journal of the American Statistical Association*, vol. 62, (pp. 399–402).

Lin, S. i Kernighan, BW (1973), "An Effective Heuristic Algorithm for the Traveling-Salesman Problem". *Operations Research*, vol. 21(2):pp. 498–516.

Linden, A. i Weber, F. (1993), "Implementing Inner Drive through Competence Reflection", a *From animals to animats 2: Proc. of the 2nd International Conference on Simulation of Adaptive Behavior*, The MIT Press, Honolulu (EUA), (pp. 321–326).

Llorà, X.; Alías, F.; Formiga, L.; Sastry, K. i Goldberg, D.E. (2005*a*), "Evaluation Consistency in iGAs: User Contradictions as Cycles in Partial-Ordering Graphs", Tech. Rep. 2005022, Illinois Genetic Algorithms Laboratory - University of Illinois at Urbana-Champaign, (EUA).

Llorà, X.; Sastry, K.; Goldberg, D. E.; Gupta, A. i Lakshmi, L. (2005*b*), "Combating User Fatigue in iGAs: Partial Ordering, Support Vector Machines, and Synthetic Fitness". *Proc. of the Genetic and Evolutionary Computation Conference (GECCO)*:pp. 1363–1371, també a IlliGAL Report No. 2005009.

Llorà, X.; Yasui, N. I. i Goldberg, D.E. (2008), "Graph-theoretic Measure for active iGAs: Interaction Sizing and Parallel Evaluation Ensemble", a *Proc. of the 10th Conference on Genetic and Evolutionary Computation (GECCO)*, ACM, Nova York (EUA), (pp. 985–992).

Lu, H.; Ling, Z.H.; Lei, M.; Wang, C.C.; Zhao, H.H.; Chen, L.H.; Hu, Y.; Dai, L.R. i Wang, R.H. (2009), "The USTC System for Blizzard Challenge 2009", a *Proc. of the Blizzard Challenge 2009 Workshop*, SynSIG - Speech Synthesis Special Interest Group, Edinburgh (Regne Unit), paper 017.

Macon, M. W.; Cronk, A. E. i Wouters, J. (1998), "Generalization and Discrimination in Tree-Structured Unit Selection", a *Proc. of the 3rd ESCA/COCOSDA Workshop on Speech Synthesis*, Jenolan Caves (Austràlia), (pp. 195–200).

Macon, M.W. (1996), *Speech Synthesis Based on Sinusoidal Modeling*, Ph.D. thesis, School of Electrical and Computer Engineering - Georgia Institute of Technology, Atlanta (EUA).

Mann, H.B. i Whitney, D.R. (1947), "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other". *The Annals of Mathematical Statistics*, vol. 18(1):pp. 50–60.

Matousek, J.; Hanzlícek, Z. i Tihelka, D. (2005), "Hybrid Syllable/Triphone Speech Synthesis", a *Proc. of the 9th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Lisboa (Portugal), (pp. 2529–2532).

McLachlan, G.J. i Krishnan, T. (1997), *The EM Algorithm and Extensions*, Wiley-Interscience New York, Nova York (EUA).

Melssen, W.; Wehrens, R. i Buydens, L. (2006), "Supervised Kohonen Networks for Classification Problems". *Chemometrics and Intelligent Laboratory Systems*, vol. 83(2):pp. 99–113.

Méndez Pazó, F.; Docío-Fernández, L.; Arza Rodríguez, M. i Campillo Díaz, F. (2010), "The Albayzín 2010 Text-to Speech Evaluation", a *Proc. of Fala 2010, VI Jornadas en Tecnología del Habla*, ISCA, Vigo (Espanya), (pp. 317–321).

Meron, Y. i Hirose, K. (1999), "Efficient Weight Training for Selection based Synthesis", a *Proc. of the 6th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 5, Budapest (Hongria), (pp. 2319–2322).

Miao, Qi; Kain, Alexander i van Santen, Jan P. H. (2009), "Perceptual Cost Function for Cross-fading Based Concatenation", a *Proc. of InterSpeech - International Conference on Spoken Language Processing (ICSLP)*, Brighton (Regne Unit), (pp. 732–735).

Michaelis, D.; Gramss, T. i Strube, HW (1997), "Glottal-to-Noise Excitation Ratio - a New Measure for Describing Pathological Voices". *Acustica*, vol. 83(4):pp. 700–706.

Michalewicz, Z. (1992), *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, (Alemanya).

Miller, B.L. i Goldberg, D.E. (1995), "Genetic Algorithms, Tournament Selection, and the Effects of Noise". *Complex Systems*, vol. 9(3):pp. 193–212, també a IlliGAL Report No. 95006.

Mobius, B. i Von-Santen, J. (1996), "Modeling Segmental Duration in German Text-to-Speech Synthesis", a *Proc. of the 4th International Conference on Spoken Language (ICSLP)*, vol. 4, ISCA, Philadelphia (EUA), (pp. 2395–2398).

Möhler, G. i Conkie, A. (1998), "Parametric Modeling of Intonation using Vector Quantization", a *Proc. of the 3rd ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*, ISCA, Jenolan Caves (Austràlia), (pp. 311–316).

Montero, J. M.; Gutiérrez-Arriola, J.; Colás, J.; Enríquez, E. i Pardo, J. M. (1999), "Analysis and Modelling of Emotional Speech in Spanish", a *Proc. of the 14th International Congress of Phonetic Sciences (ICPhS)*, vol. 2, San Francisco (EUA), (pp. 957–960).

Monzo, C.; Alías, F.; Iriondo, I.; Gonzalvo, X. i Planet, S. (2007), "Discriminating Expressive Speech Styles by Voice Quality Parameterization", a *Proc. of the 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrüken (Alemanya), (pp. 2081–2084).

Monzo, C.; Alías, F.; Morán, J.A. i Gonzalvo, X. (2006), "Transcripción fonética de acrónimos en castellano utilizando el algoritmo C4.5". *Procesamiento del Lenguaje Natural*, vol. 37:pp. 275–282.

Moon, T.K. (1996), "The Expectation-Maximization Algorithm". *Signal Proccessing Magazine, IEEE*, vol. 13(6):pp. 47–60.

Moulines, E. i Charpentier, F. (1990), "Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis using Diphones". *Speech Communication*, (9):pp. 453–467.

Moulines, E. i Verhelst, W. (1995), *Speech Coding and Synthesis*, cap. Time-Domain and Frequency-Domain Techniques for Prosodic Modification of Speech, Elsevier Science Inc., Nova York (EUA), (pp. 519–555).

Mühlenbein, H. (1989), "Parallel Genetic Algorithms, Population Genetics, and Combinatorial Optimization", a *Proc. of the Workshop on Evolutionary Models and Strategies, Workshop on Parallel Processing: Logic, Organization, and Technology: Parallelism, Learning, Evolution*, Springer-Verlag, Londres (Regne Unit), (pp. 398–406).

Nakajima, S. i Hamada, H. (1988), "Automatic Generation of Synthesis Units Based On Context Oriented Clustering", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Nova York (EUA), (pp. 659–662).

Navas, E.; Hernáez, I. i Ezeiza, N. (2002*a*), "Assigning Phrase Breaks using CARTs for Basque TTS", a *Proc. of the International Conference Speech Prosody*, Aix-en-Provence (França), (pp. 527–530).

Navas, E.; Hernáez, I. i Sánchez, J.M. (2002*b*), "Modelo de duración para conversión de texto a voz en euskera". *Procesamiento del Lenguaje Natural*, vol. 29:pp. 147–152.

Nelder, JA i Mead, R. (1965), "A Simplex Method for Function Minimization". *The Computer Journal*, vol. 7(4):pp. 308–313.

Netter, J.; Wasserman, W. i Kutner, M.H. (1990), *Applied Linear Statistics Models*, Richard D. Irving, Chicago (EUA).

Ney, H. (1982), "A Time Warping Approach to Fundamental Period Estimation". *IEEE Transactions on Systems, Man and Cybernetics*, vol. 12(3):pp. 383—388.

Nuttall, A.H. (1981), "Some Windows with Very Good Sidelobe Behavior". *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 29(1):pp. 84–91.

Ohsaki, M. i Takagi, H. (1998), "Improvement of Presenting Interface by Predicting the Evaluation Order to Reduce the Burden of Human Interactive EC Operators", a *Proc.*

*of the IEEE International Conference on Systems, Man, and Cybernetics*, vol. 2, San Diego (EUA), (pp. 1284–1289).

Oja, M.; Kaski, S. i Kohonen, T. (2003), "Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum". *Neural Computing Surveys*, vol. 3:pp. 1–156.

Ostendorf, M. i Bulyko, I. (2002), "The Impact of Speech Recognition on Speech Synthesis", a *Proc. of the IEEE Workshop on Speech Synthesis*, Santa Monica (EUA).

Oura, K.; Wu, Y.J. i Tokuda, K. (2009), "Overview of NIT HMM-Based Speech Synthesis System for Blizzard Challenge 2009", a *Proc. of the Blizzard Challenge 2009 Workshop*, Syn-SIG - Speech Synthesis Special Interest Group, Edinburgh (Regne Unit), paper 013.

Pareto, V. (1896), *Cours d'Economie Politique, volume I and II*, F. Rouge, Lausanne (Suïssa).

Park, S. S.; Kim, C. K. i Kim, N. S. (2003), "Discriminative Weight Training for Unit-Selection Based Speech Synthesis", a *Proc. of the 8th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 1, Ginebra (Suïssa), (pp. 281–284).

Pearson, S.; Kibre, N. i Niedzielski, N. (1998), "A Synthesis Method Based on Concatenation of Demisyllables and a Residual Excited Vocal Tract Model", a *Proc. of the 5th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Sydney (Austràlia), paper 0648.

Peng, H.; Zhao, Y. i Chu, M. (2002), "Perpetually Optimizing the Cost Function for Unit Selection in a TTS System with One Single Run of MOS Evaluation", a *Proc of International Conference on Speech Language Processing*, Denver (EUA), (pp. 1341–1344).

Planet, S.; Iriondo, I.; Martínez, E. i Montero, J.A. (2008), "TRUE: an Online Testing Platform for Multimedia Evaluation", a *Proc. of the 2nd International Workshop on EMOTION: Corpora for Research on Emotion and Affect at the 6th Conference on Language Resources & Evaluation (LREC)*, Marrakech (Marroc), (pp. 61–65).

Portele, T.; Stöber, K.H.; Meyer, H. i Hess, W. (1996), "Generation of Multiple Synthesis Inventories by a Bootstrapping Procedure", a *Proc. of the 4th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA), (pp. 2391–2394).

Prieto, P. (2004), *Fonètica i fonologia*, UOC, Barcelona (Espanya).

Qian, G.; Sural, S.; Gu, Y. i Pramanik, S. (2004), "Similarity between Euclidean and Cosine Angle Distance for Nearest Neighbor Queries", a *Proc. of the ACM Symposium on Applied Computing*, ACM, Nicosia (Xipre), (pp. 1232–1237).

Rabiner, L. i Juang, B. (1993), *Fundamentals of Speech Recognition*, Prentice Hall, (EUA).

Rabiner, L. R. i Schafer (1978), *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ (EUA).

Rafel, J. (1980), "Dades sobre la freqüència de les unitats fonològiques en català". *Estudis Universitaris Catalans*, vol. 24:pp. 473–496.

Rechenberg, I. (1973), *Evolutionsstrategie: Optimierung technischer Systeme nack Prinzipien der biologischen Evolution*, Frommann-Holzboog.

Rennison, E. (1994), "Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes", a *ACM Symposium on User Interface Software and Technology*, Marina del Rey (EUA), (pp. 3–12).

Ríos, S. (2000), *Iniciación Estadística*, Editorial Paraninfo (ITP - An International Thomson Publishing Company), Madrid (Espanya).

Sagisaka, Y. (1988), "Speech Synthesis by Rule Using an Optimal Selection of Non-Uniform Synthesis Units", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Nova York (EUA), (pp. 679–682).

Sagisaka, Y.; Kaiki, N.; Iwahashi, N. i Mimura, K. (1992), "ATR - $\nu$-TALK Speech Synthesis System", a *Proc. of the International Conference on Speech Language Processing*, vol. 1, Banff (Canadà), (pp. 483–486).

Saito, T.; Hashimoto, Y. i Sakamoto, M. (1996), "High-Quality Speech Synthesis Using Context-Dependent Syllabic Units", a *Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 1, IEEE, Atlanta (EUA), (pp. 381–384).

van Santen, J. P. H. i Buchsbaum, A. L. (1997), "Methods for Optimal Text Selection", a *Proc. of the 5th European Conference on Speech Communication and Technology (EuroSpeech)*, Rodas (Grècia), (pp. 553–556).

Saon, G.; Dharanipragada, S. i Povey, D. (2004), "Feature Space Gaussianization", a *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, IEEE, Montreal (Canadà), (pp. I–329–332).

Sastry, K. i Goldberg, D. E. (2002), "Genetic Algorithms, Efficiency Enhancement, and Deciding Well with Fitness Functions with Differing Bias Values", a *Proc. of the Genetic and Evolutionary Computation Conference (GECCO)*, Nova York (EUA), (pp. 536–543), també a IlliGAL Report No. 2002003.

Sastry, K.; Pelikan, M. i Goldberg, D. E. (2004), "Efficiency Enhancement of Genetic Algorithms Via Building-Block-Wise Fitness Estimation", a *Proc. of the IEEE International Congress on Evolutionary Computation (CEC)*, vol. 1, IEEE, Portland (EUA), (pp. 720–727), illiGAL Report No. 20040010.

Sato, Y. (1997), "Voice Conversation Using Evolutionary Computation of Prosodic Control", a *Proc. of the 2nd International Conference on Intelligent Processing of Manufacturing of Materials*, Honolulu (EUA), (pp. 342–348).

Schmidhuber, J. i Storck, J. (1993), "Reinforcement Driven Information Acquisition in Non-deterministic Environments", a *Proc. of the International Conference on Neural Networks (ICANN)*, Amsterdam (Holanda), (pp. 159–164).

Schröder, M. (2004), "Dimensional Emotion Representation as a Basis for Speech Synthesis with Non-Extreme Emotions". *Lecture Notes in Artificial Intelligence*, vol. 3068:pp. 209–220, proc. of the Tutorial and Research Workshop on Affective Dialog Systems.

Schröder, M.; Pammi, S. i Türk, O. (2009), "Multilingual MARY TTS Participation in the Blizzard Challenge 2009", a *Proc. of the Blizzard Challenge 2009 Workshop*, Edinburgh (Regne Unit), paper 007.

Schweitzer, A. i Möbius, B. (2004), "Exemplar-Based Production of Prosody: Evidence from Segment and Syllable Durations", a *Proc onf the International Conference Speech Prosody*, ISCA, Nara (Japó), (pp. 459–462).

Sebag, M. i Ducoulombier, Q. (1998), "Extending Population-Based Incremental Learning to Continuous Search Spaces". *Lecture Notes in Computer Science*, vol. 1498:pp. 418–427.

Sharan, R.; Maron-Katz, A. i Shamir, R. (2003), "CLICK and EXPANDER: a System for Clustering and Visualizing Gene Expression Data". *Bioinformatics*, vol. 19(14):pp. 1787–1799.

Shawe-Taylor, J. i Cristianini, N. (2004), *Kernel Methods for Pattern Analysis*, Cambridge University Press, Regne Unit.

Sityaev, D.; Knill, K. i Burrows, T. (2006), "Comparison of the ITU-T P. 85 Standard to Other Methods for the Evaluation of Text-to-Speech Systems", a *Proc of the 9th International Conference on Spoken Language Processing (ICSLP)*, Pittsburgh (EUA), paper 1233.

Soong, F. i Juang, B. (1984), "Line Spectrum Pair (LSP) and Speech Data Compression", a *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 9, San Diego (EUA), (pp. 37–40).

Srinivas, N. i Deb, K. (1994), "Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms". *Journal on Evolutionary Computation*, vol. 2(3):pp. 221–248.

Steiner, I.; Schroeder, M.; Charfuelan, M. i Klepp, A. (2010), "Symbolic vs. Acoustics-Based Style Control for Expressive Unit Selection", a *Proc. of the 7th ISCA Speech Synthesis Workshop*, vol. L 4.1, Kyoto (Japó), (pp. 114–119).

Stöber, K.; Portele, T.; Wagner, P. i Hess, W. (1999), "Synthesis by Word Concatenation", a *Proc. of the 6th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 2, Budapest (Hongria), (pp. 619–622).

Strehl, A. (2002), *Relationship-Based Clustering and Cluster Ensembles for High-Dimensional Data Mining*, Ph.D. thesis, Faculty of the Graduate School - University of Texas at Austin, (EUA).

Strom, V.; Clark, R.A.J. i King, S. (2006), "Expressive Prosody for Unit-Selection Speech Synthesis", a *Proc of the 9th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA), paper 1522.

Strom, V. i King, S. (2008), "Investigating Festival's Target Cost Function Using Perceptual Experiments", a *Proc. of the 9th International Conference of the International Speech Communication Association (Interspeech)*, (pp. 1873–1876).

Stylianou, Y. (2001), "Applying the Harmonic Plus Noise Model in Concatenative Speech Synthesis". *IEEE Transactions on Speech and Audio Processing*, vol. 9(1):pp. 21–29.

Stylianou, Y. i Syrdal, A. K. (2001), "Perceptual and Objective Detection of Discontinuities in Concatenative Speech Synthesis", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, Salt Lake City (EUA), (pp. 987–990).

Suh, J.Y. i Van Gucht, D. (1987), "Incorporating Heuristic Information into Genetic Search", a *Proc. of the 2nd International Conference on Genetic Algorithms and their Application*, L. Erlbaum Associates Inc., Hillsdale (EUA), (pp. 100–107).

Sywerda, G. (1989), "Uniform Crossover in Genetic Algorithms", a *Proc. of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufmann Publishers Inc., Virginia (EUA), (pp. 2–9).

Takagi, H. (2001), "Interactive Evolutionary Computation: Fusion of the Capabilities of the EC Optimization and Human Evaluation". *Proceedings of the IEEE*, vol. 89(9):pp. 1275–1296.

Taylor, D.W. i Faust, W.L. (1952), "Twenty Questions: Efficiency in Problem Solving as a Function of Size of Group". *Journal of Experimental Psychology*, vol. 44(5):pp. 360–368.

Taylor, P. (2006), "The Target Cost Formulation in Unit Selection Speech Synthesis", a *Proc. of the 9th International Conference on Spoken Language Processing (ICSLP)*, ISCA, Pittsburgh (EUA), paper 1455.

Taylor, P. (2009), *Text-to-Speech Synthesis*, Cambridge University Press, (Regne Unit).

Taylor, P. i Black, A. W. (1999), "Speech Synthesis by Phonological Structure Matching", a *Proc of the 6th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 4, Budapest (Hongria), (pp. 1531–1534).

Taylor, P. i Black, A.W. (1998), "Assigning Phrase Breaks from Part-of-Speech Sequences". *Computer Speech and Language*, vol. 12:pp. 99–117.

Taylor, P.A.; Nairn, I.A.; Sutherland, A.M. i Jack, M.A. (1991), "A Realtime Speech Synthesis System", a *Proc. of the 2nd European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Gènova (Itàlia), (pp. 341–344).

Technosite (2009), "Interfaces de Relación entre el Entorno y las personas con Discapacidad", http://www.inredis.es/.

Thrun, S.B. i Möller, K. (1993), "Active Exploration in Dynamic Environments". *Advances in Neural Information Processing Systems*, vol. 4:pp. 531–531.

Tihelka, D. (2005), "Symbolic Prosody Driven Unit Selection for Highly Natural Synthetic Speech", a *Proc. of 9th European Conference on Speech Communication and Technology (EuroSpeech)*, Lisboa (Portugal), (pp. 2525–2528).

Tihelka, Daniel i Romportl, Jan (2009), "Exploring Automatic Similarity Measures for Unit Selection Tuning", a *Proc. of InterSpeech - International Conference on Spoken Language Processing (ICSLP)*, Brighton (Regne Unit), (pp. 736–739).

Toda, T. (2003), *High-Quality and Flexible Speech Synthesis with Segment Selection and Voice Conversion*, Ph.D. thesis, Department of Information Processing - Nara Institute of Science and Technology, (Japó).

Toda, T.; Kawai, H. i Tsuzaki, M. (2003), "Optimizing Integrated Cost Function for Segment Selection in Concatenative Speech Synthesis Based on Perceptual Evaluations", a *Proc. of the 8th European Conference on Speech Communication and Technology (EuroSpeech)*, vol. 1, Ginebra (Suïssa), (pp. 297–300).

Toda, T.; Kawai, H. i Tsuzaki, M. (2004), "Optimizing Sub-Cost Functions for Segment Selection Based on Perceptual Evaluations in Concatenative Speech Synthesis", a *Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Montreal (Canadà), (pp. 657–660).

Toda, T.; Kawai, H.; Tsuzaki, M. i Shikano, K. (2002), "Perceptual Evaluation of Cost for Segment Selection in Concatenative Speech Synthesis", a *Proc. of the IEEE Workshop on Speech Synthesis*, Santa Monica (EUA).

Toda, T.; Kawai, H.; Tsuzaki, M. i Shikano, K. (2006), "An Evaluation of Cost Functions Sensitively Capturing Local Degradation of Naturalness for Segment Selection in Concatenative Speech Synthesis". *Speech Communication, Elsevier*, vol. 48(247):pp. 45–56.

Todoroki, Y. i Takagi, H. (2000), "User Interface of an Interactive Evolutionary Computation for Speech Processing", a *Proc. of the 6th International Conference on Soft Computing (IIZUKA)*, (pp. 112–118).

Tokuda, K.; Zen, H. i Black, A.W. (2003), "An HMM-Based Speech Synthesis System Applied to English", a *Proc. of 2002 IEEE Workshop on Speech Synthesis*, IEEE, (pp. 227–230).

Trask, R.L. (1996), *A Dictionary of Phonetics and Phonology*, Burns & Oates, Londres (Regne Unit).

Tsuzaki, M. (2001), "Feature Extraction by Auditory Modeling for Unit Selection in Concatenative Speech Synthesis", a *Proc. of the 7th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Aalborg (Dinamarca).

Tsuzaki, M. i Hisashi, H. (2002), "Feature Extraction for Unit Selection in Concatenative Speech Synthesis: Comparision between AIM, LPC and MFCC", a *Proc. of International Conference on Speech Language Processing*, vol. 1, Denver (EUA), (pp. 137–140).

Tukey, J.W. (1957), "On the Comparative Anatomy of Transformations". *The Annals of Mathematical Statistics*, vol. 28(3):pp. 602–632.

Vapnik, V. N. (1999), *The Nature of Statistical Learning Theory*, Springer-Verlag, Nova York (EUA).

Veldhuis, R. i Klabbers, E. (2003), "On the Computation of the Kullback-Leibler Measure for Spectral Distances". *IEEE Transactions on Speech and Audio Processing*, vol. 11(1):pp. 100–103.

Vellido, A. (2006), "Missing Data Imputation through GTM as a Mixture of t-distributions". *Neural Networks*, vol. 19(10):pp. 1624–1635.

Vepa, J.; King, S. i Taylor, P. (2002), "Objective Distance Measures for Spectral Discontinuities in Concatenative Speech Synthesis", a *Proc. of International Conference on Speech Language Processing*, vol. 4, Denver (EUA), (pp. 2604–2608).

Vincent, D.; Rosec, O. i Chonavel, T. (2005), "Estimation of LF Glottal Source Parameters based on an ARX Model", a *Proc of the 9th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Lisboa (Portugal).

Viterbi, A. (1967), "Error Bounds for Convolutional Codes and an Asymtotically Optimum Decoding Algorithm". *IEEE Transactions on Information Theory*, vol. 13:pp. 260–267.

Vosnidis, C. i Digalakis, V. (2001), "Use of Clustering Information for Coarticulation Compensation in Speech Synthesis by Word Concatenation", a *Proc. of the 7th European Conference on Speech Communication and Technology (EuroSpeech)*, ISCA, Aalborg (Dinamarca).

Wang, K.; Wang, B. i Peng, L. (2009), "CVAP: Validation for Cluster Analyses". *Data Science Journal*, vol. 8:pp. 88–93.

Watanabe, T. i Takagi, H. (1995), "Recovering Sytem of the Distorted Speech Using Interactive Genetic Algorithms", a *Proc. of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, vol. 1, Vancouver (Canadà), (pp. 684–689).

Wells, J.; Barry, W.; Grice, M.; Fourcin, A. i Gibbon, D. (1992), "Standard Computer-Compatible Transcription", Tech. Rep. Sen.3 SAM UCL-037, Phonetics and Linguistics Department - University College London, (Regne Unit), final Report. ESPRIT PROJECT 2589 (SAM) Multilingual Speech Input/Output Assessment, Methodology and Standardisation.

Whitehead, S.D. i Ballard, D.H. (1991), "A Study of Cooperative Mechanisms for Faster Reinforcement Learning", Tech. Rep. 365, Computer Science Department - University of Rochester, Rochester (EUA).

Wouters, J. i Macon, M.W. (1998), "A Perceptual Evaluation of Distance Measures for Concatenative Speech Synthesis", a *Proc. of International Conference on Speech Language Processing*, ISCA, Sydney (Austràlia).

Wu, C.H. i Chen, J.H. (2001), "Automatic Generation of Synthesis Units and Prosodic Information for Chinese Concatenative Synthesis". *Speech Communication*, vol. 35:pp. 219–237.

Yamagishi, J.; Tachibana, M.; Masuko, T. i Kobayashi, T. (2004), "Speaking Style Adaptation Using Context Clustering Decision Tree for HMM-based Speech Synthesis", a *Proc. of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 1, Montreal (Canadà), (pp. 5–8).

Yi, J. R.-W. (2003), *Corpus-Based Unit Selection for Natural-Sounding Speech Synthesis*, Ph.D. thesis, Department of Electrical Engineering and Computer Science - Massachusetts Institute of Technology, (EUA).

Yu, A-T i Wang, H-C (2004), "New Harmonicity Measures for Pitch Estimation and Voice Activity Detection", a *Proc. of the 8th International Conference on Spoken Language Processing (ICSLP)*, Jeju Island (Corea del Sud), (pp. 2429–2243).

Universitat Ramon Llull

Aquesta Tesi Doctoral ha estat defensada el dia ____ d _____ de ____

al Centre _____

de la Universitat Ramon Llull

davant el Tribunal format pels Doctors sotasignants, havent obtingut la qualificació:

President/a

_____

Vocal

_____

Vocal

_____

Vocal

_____

Secretari/ària

_____

Doctorand/a



C. Claravall, 1-3
08022 Barcelona
Tel. 936 022 200
Fax 936 022 249
E-mail: urlsc@sec.url.es
www.url.es