

Fasea ahots-seinalearen eredu harmonikoetan: adierazpidea, prozesaketa eta aplikazioak

Ibon Saratxaga, Inma Hernaezek zuzenduta

Aholab Signal Processing Laboratory, Euskal Herriko Unibertsitatea UPV/EHU
{ibon,inma}@bips.bi.ehu.es

Abstract

In this paper we present the work and results of the PhD dissertation named “Phase in the harmonic models of the speech signal: strategies for representation, processing and applications”. In the thesis a novel phase representation for the speech signal is proposed, the Relative Phase Shift (RPS), which allows a deeper understanding of the phase structure of the signal. The properties of the RPS transformation are thoroughly explored and the representation is tested in several applications: ASR, speaker identification and impostor detection, where it improves the performance of the classical module based technologies. Finally the straightforward phase manipulation allowed by the RPS is used to test the hearing perceptual importance of the phase, in order to ponder the use of the phase in speech synthesis. These sections of the PhD dissertation are described in some detail in the paper, also presenting the most significant results of the research.

Laburpena

Artikulu honetan izenburu bereko doktorego-tesian landutako gaien eta emaitzen berri ematen da laburki. Tesian ahots-seinalearen fasea adierazteko errepresentazio berri bat proposatzen da, desfase erlatibo edo Relative Phase Shift (RPS) deritzona eta seinalearen fase egitura antzematea ahalbidetzen duena. Tesian RPS transformazioaren ezaugarriak sakonki ikertzen dira eta errepresentazioa hainbat aplikaziotan probatzen da: hizketa-ezagutze automatikoa (ASR), hiztunaren identifikazioan eta iruzurtien detekzioan aplikatzen da, guztietan emaitzak hobetzen dituela moduluetan oinarritutako teknologi klasikoekin alderatuz. Azkenik, RPSe fase maneiu zuzena errazten dutela baliatuz, faseek entzumenean duten garrantzi pertzeptuala ikertzen da, faseak ahots sintetikoa sortzeko erabili behar denetz aztertzeke. Tesiaren atal guzti hauen xehetasunak laburbiltzen dira artikuluan, ikerketaren emaitza esanguratsuenak aipatuz.

Keywords: Phase, harmonic models, polarity, ASR, speaker recognition, speech perception

Gako hitzak: Fasea, eredu harmonikoak, polaritatea, ASR, hiztun-ezagutze, ahots-pertzepzioa

1. Sarrera

Ahots teknologia askok ahots-seinalearen espektroaren fase informazioa ez ohi du aintzat hartzen, ordea, espektroaren modulutik eratorritako informazioa erabiltzen du bakarrik. Fase informazioak seinalearen prozesaketaren ikuspuntutik zailtasun batzuk dauzka bere ezaugarri bereziak direla eta, denborazko eta maiztasunezko dependentzia, adibidez. Gainera, fase-espektroak giza entzumenaren pertzepzio edo hautematean garrantzi txikia omen du, bere eragina bigarren mailakoa izanik seinalearen modulu-espektroarekin alderatuz.

Arrazoi hauek gutxi balira, modulu espektroa erraz ulertzen eta maneiatzen da eta harreman zuzena dauka hautematearekin, eta honek guztiak ahots teknologietan modulua ia bakarrik erabiltzera eramaten du, bai sintesia aplikazioetan (ahots sintesia, ahots bihurketa eta eraldaketa, e.a.), baita analisi aplikazioetan ere (hizketa eta hiztun ezagutza, besteak beste). Berrikus dezagun faseen papera bi eremu hauetan.

1.1. Faseak eta ahots sintesia

Ahots sintesiaren arloan faseak ahotsaren ezaugarri erabilgarri bat bezala baino, arazo bezala ikusten dira. Eredu sinusoidal [1][2] eta harmonikoen [3][4] kasuetan, esaterako, osagai sinusoidalen aldiuneko faseak kalkulatu dira. Aldiuneko fasea analisi unearen menpekota da, eta beraz datua erabilgarria izan dadin askotan prozesaketa gehiago beharrezko da bilbeen artean fase jauziak ager ez daitezzen, batez ere seinalea eraldatzen (pitch edo iraupen aldaketak, alegia) edo kateatzen denean. Arazo hauek konpontzeko, fase linealeko terminoa (analisiaren denboraren menpe dagoena) kendu behar zaio aldiuneko faseari. Teknika batzuk proposatu izan dira hau egiteko, hala nola analisisa era pitch sinkronoan egitea edota pitch periodoka puntu finko bat kalkulatzeko une horretan analisisa egiteko (eraso uneak [5], grabitate-zentroak [6] edo glotisaren ixte uneak).

Ahots sintesiko beste eredu batzuek, LPCtan oinarritutakoak kasu, zuzenean baztertzen dute fase informazioa eta fase artifiziala erabiltzen dute ahots sintetikoa sortzeko (normalean fase minimoa baina batzuetan ausazko edo zero fasea ere).

Bi aukera hauek eragin zuzena daukate seinale sintetikoaren uhin forman: jatorrizko faseak aldatzeak jatorrizko uhin-forma galtzera darama. Seinalearen forma aldatu gabe mantentzearen garrantzia ez dago argi. Autore batzuek [5], [7] uhin forma aldatzen ez duten metodoak proposatu dituzte, baina fasea aldatzearen eragina (eta ondorioz uhin formarena) ahotsaren kalitate pertzeptualean ez dago ondo ikertuta.

Faseen garrantzi pertzeptuala, aldiz, ondo ikertua izan da elektroakustika eta entzumenaren fisiologia bezalako arloetan, betiere soinu artifizialak edo musikalak erabiliz. Ahots teknologien arloan fasearen eragina ulergarritasunean ebaluatu izan dute zenbait autorek [8], [9]. Faseak vocoder-en ikuspuntutik analizatu dira, hain zuzen ere, zehazteko fase informazioa zein puntutaraino baztertu ahal den transkodedutako ahotsean eragin hautemangarriarik pairatu gabe [10], [11]. Baina oso ikerlari gutxik ebaluatu du faseen aldaketak dakarren efektua audio seinalearen kalitate orokorrean: egindako saiakerak normalean fase aldaketa handiekin izan dira (fase finkoa edo minimoa, alegia) eta seinale sintetikoak edota ez errealekin [12].

Laburbilduz, fasearen papera ahots sintesiaren arloan zalantzagarria da: faseak zailak dira maneiatzeko, askotan baztertzen dira eta ez dago argi ea nabari eragiten duten ahots sintetikoaren kalitatean. Tesi hau galdera hauek argitzen saiatzen da.

1.2. Faseak eta ahots analisisa

Epe laburreko espektr informazioak berebiziko garrantzia du ahots analisisan, hainbat ahots eredu teknikaren oinarrian baitago. Hala ere, hizketa edo hiztun ezagutze aplikazioetan normalean espektr magnitudea soilik erabiltzen da, fase informazioa alde batera uzten delarik. Faseak ez du ageriko egiturarik eta

honengatik ez da erabiltzen ereduak lortzeko.

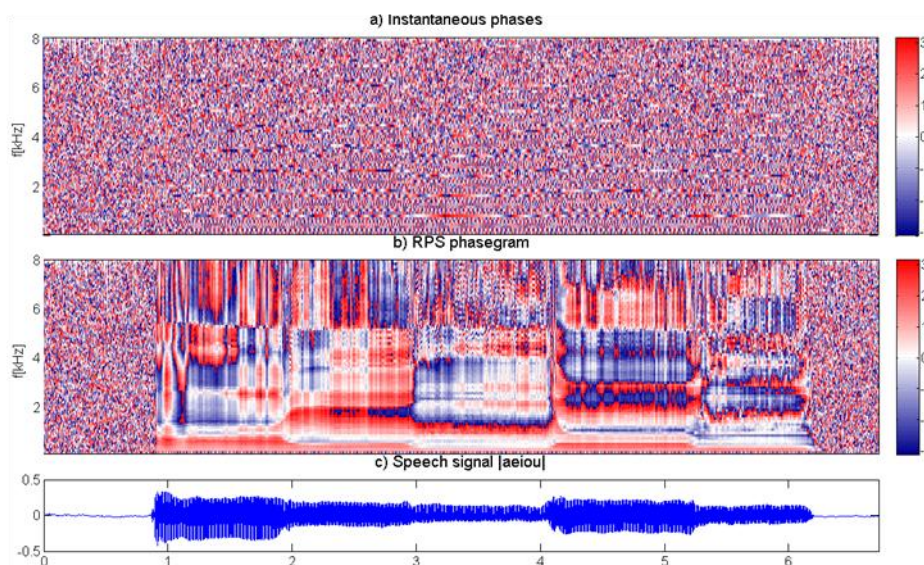
Dena den, egon dira saiakera batzuk fase informazioa ezagutza aplikazioetan erabiltzeko. Talde atzerapenetik eratorriko informazioa erabili izan da hizketa ezagutze automatiko (ASR) sistemetan [13], [14] eta hiztun ezagutze ereduetan [15], [16], [17]. Azken arlo honetan autore askok esaten dute emaitza onuragarriak lortzen direla fase informazioa MFCC magnitudean oinarritutako parametroekin konbinatzerakoan [18], [19], [20].

Ahots analisisaren eremuan gai irekia den atal bat iruzurti sintetikoaren detekzioa da, alegia, hiztun-egiaztatze sistema batean baimenduta dagoen hiztun baten ahotsa antzeratzen duen ahots sintetikoa. Egungo hiztun egiaztatze sistemek arazo larriak dituzte kalitate handiko ahots sintetikoa antzematen eta, eskatzaile faltsua izaki, atzera botatzen [21], [22]. Fase informazioa lagungarria izan daiteke horrelako iruzurti sintetikoak bereizten, batez ere ahotsaren jatorrizko faseak baztertuak edo oso aldatuta daudenean.

2. Motibazioa eta helburuak

Ahotsaren eredu harmonikoen testuinguruan fase informazioarako adierazpide berri bat proposatu dugu, faseen prozesamendua errazten duten ezaugarri anitz dituen. Desfase erlatibo edo Relative Phase Shift (RPS) transformazio honek, espektraren aldiuneko fasetik abiatuz osagai harmoniko bakoitzaren eta oinarritzko maiztasunaren arteko desfase erlatiboa lortzen du.

Transformazio honek fasearen menpekotasuna maiztasun eta analisi unearekiko ezabatzen du. RPSak hasierako fasearekiko menpekoak dira soilik eta beraz, bere interpretazioa eta manipulazioa aldiuneko fasearena baino askoz zuzenagoa da. Adierazpide hau egokia da fase aldaketak egiteko eredu harmonikoetan,



1. Irudia: Aldiuneko fase (a) eta RPS (b) fasegramak /aeiou/ seinalearentzako.

baina, garrantzitsuagoa dena, seinalearen fase egitura ezkutua agerrarazten du, 1. irudian ikus daitekeenez.

1. irudian /aeiou/ bokalak dituen seinale baten bi fasegrama erakusten dira (espektrogramaren analogoa, baina moduluaren ordezkari fasea irudikatuz). (a) irudian ikusten den aldiuneko faseen egiturarik eza eta (b) irudian RPSek azaltzen dituzten faseen patroia nabariaren arteko ezberdintasuna argia da.

Tesi honen motibazioa faseen adierazpide berri honen propietateak eta balizko potentzialak ikertzean datza. Horregatik tesiaren helburuak horrela laburbildu daitezke:

2.1. Helburuak

Tesi honen helburu nagusia fase informazioaren potentziala analizatu eta ulertzea izan da, RPSen bidez adierazita, ahots seinalearen prozesamenduaren hainbat eremutan. Helburu orokor hau puntu ezberdinetan bana daiteke:

- Faseak ahots seinalearen karakterizazioan: RPSen interpretazioa eta ahotsaren ezaugarriekin duten erlazioa: formanteak, polaritatea, iturri eta ahots traktuaren arteko banaketa.
- Faseen eredutzea: Fasearen parametrizazio bat lortzea eredu estatistikoak eraikitzeke egokia dena.
- Faseak ahots seinalearen analisisan: Fonema eta hiztunen ezaugarrien eragina fasean ikertzea, eta honen aplikazioa ahots seinalearen analisisiko ataza batzuetan: hizketaren eta hiztunen ezagutzea eta iruzurti sintetikoen detekzioa.

Faseak eta hautematea: Fase aldaketen eragina hautematean ebaluatzea du xedetzat, ahots sintesian fasea sartzea merezi duen edo ez zehazteko.

3. Tesiaren ekarpen nagusiak

3.1. Relative Phase Shift adierazpidea

Tesiaren lehenengo ekarpena Relative Phase Shift adierazpide bera da. Fourierren transformatuaren bidez kalkulaturako aldiuneko fasea eta pitch-a erabiliz, proposaturako RPS transformazioak RPS balioak lortzea ahalbidetzen du era pitch asinkronoan.

Adierazpide berri hau erabilgarria da ahots seinale baten iraupena eta pitch-a aldatzeko eredu harmonikoak erabiliz. RPSak aldatutako seinaleen faseak era zuzenean kalkulatzeko ahalbidetzen du, anplitudeekin egiten den antzeko eran. Horrela ez da beharrezkoa doiketa berezirik egitea faseetan edota analisisa pitch periodoaren puntu espezifiko batzuetara mugatzea, ohikoak diren metodoekin bezala.

RPS adierazpidea faseen aldaketa errazteko teknika baino gehiago da, hala ere. Propietate komenigarri antzak ditu, bereziki seinalearen uhin formarekin duen erlazio zuzena, eta ahotsaren fase egitura agerian uzten du.

RPSen harremana ahots formanteekin eta iturri eta ahots traktuaren banaketarekin ere aztertu eta deskribatu da tesian.

3.2. DCT-mel-RPS parametrizazioa

RPS fasegrametan agertzen diren faseen egiturak erabilgarriak izan daitezke gizakiontzat, baina ez dira oso egokiak aplikazio automatikoentzako. Horregatik RPSen parametrizazio bat garatu dugu, eredutze estatistikorako aproposagoa. Parametrizazio honek RPSen zailtasun batzuk ekiditen ditu, hala nola, fase balioen periodikotasuna (wrapping delakoa), balio kopuru aldakorra eta handiegia maiztasun ezberdinetan bilbe bakoitzean, e.a. DCT-mel-RPS parametrizazioak mel iragazketa eta kosinu diskretuko transformazioa (DCT) aplikatzen du, jatorrizko RPS balioen dimentsionalitatea murrizteko informazio galera txikienarekin. Parametrizazio hau HMMetan oinarritutako ereduarekin probatu da ASR sistema batean eta baita GMMetan oinarritutako ereduarekin hiztun-identifikazio sistema batean. Bi esperimentuaren parametrizazioak eraginkortasuna frogatu du eredu estatistiko hauetan.

Era berean esperimenduek frogatu dute fase informazioa erabilgarria dela ezagutze sistemetan, bai bakarka (hiztun identifikazioaren kasuan) baita espektro-moduluarekin, MFCC forman adierazita, (ASRren kasuan). Kasu bietan fase informazioaren sarrerak modulutan oinarritutako ohiko sistemen emaitzak hobetzen du.

3.3. Polaritatearen detekzioa

RPSen ezaugarri aipagarri bat seinalearen polaritatea bereizteko gaitasuna da. Polaritatearen inbertsioak harmoniko bikoitien RPS baliotan π radianeko aldaketa eragiten du. Hau erabiliz polaritatea bereizteko algoritmo bat proposatu dugu, bilbe bakoitzaren RPSen maiztasun-inguratzailaren kizkurdura. Algoritmo hau ebaluatu izan da beste batzuekin konparatuz eta oso emaitza onak lortu ditu baita seinale laburrekin ere.

3.4. Iruzurti sintetikoen detekzioa

Fase informazioa erabili izan dugu iruzurti sintetikoak detektatzeko hitzuna egiaztatze sistema batean emaitza bikainekin. Tesi honetan fasea proposatu izan da lehendabiziko aldiz sintetizagailu batekin sortutako ahots iruzurtiak bereizteko, eta honek etorkizun oparoko ikerketa lerro bat zabaltzen du, beste metodoek zailtasun antzak zeuzkaten arlo batean.

Garatu dugun iruzurti detekzio sistemak, egiaztatze sistema baten azkeneko etapa bezala lan egiten du, lehenak errorez onartu dituen seinale iruzurtiak arbuizatu saiatuz. Ahots-sintetiko detektagailuak (SSD, synthetic speech detector) hiztuneko bina eredu sortzen ditu, bat ahots naturalerako eta bestea ahots sintetikorako, horretarako DCT-mel-RPS parametroak

erabiliz, eta ahots seinalea bi ereduak konparatuz sailkatzen du.

3.5. Fase hautematearen ebaluazioa

Tesiaren hasierako ideietako bat izan zen fasearen tratamendua hobetzea sintesi sistemetan, non ohikoa baita faseak arbuia edo asko aldatzea. Alabaina, lehenik, jakin behar genuen ea sintesian faseez arduratzeak merezi zuen, hau da, ea fasearen manipulazioak seinalearen andeatze gisa hautematen ziren. Galdera honi erantzuteko asmoz, fase eraldaketak jasatako ahots seinaleak ebaluatu ditugu, jatorrizko seinale naturalekin alderatuz. Esperimentu hauek egin izan dira beste audio seinale mota batzuekin (musika, tonu sintetikoak, e.a.) baina, guk dakigula, fasearen eragina ahots seinalearen kalitate orokorrean preseski analizatzen duen ikerketarik ez dago.

Emaitzek erakusten dute giza entzumena ahots seinalearen faseekiko sentikorra dela, eta fase aldaketa handiak edo naturaltasun gutxieneakoak andeatzea eragiten dutela ahotsaren hautematean. Dena den, sintesi sistemei jatorrizko faseen informazioa gehitzeak merezi duen galderari dagokionez, emaitzak ez dira kategorikoak. Zenbait fase aldaketak sortutako degradazioa ukalezina bada ere ebaluaketaren emaitzen arabera, azpimarratu behar da ebaluazio baldintzak kontu handiz diseinatu zirela andeatze hauen hautematea indartzeko, beraz sistema errealean bere efektuak ez lirateke ebaluazioaren emaitzek aditzera ematen duten bezain nabariak izango.

Zentzu honetan, aipatu behar dira DCT-mel-RPS parametrizazioak lortu dituen emaitza onak, parametro hauekin sortutako ahots seinaleek andeatze ia hautemanezina baitute.

4. Ondorioak eta irekitako ikerketa-lerroak

Tesi honen ondorioz ekarpenak egin dira ahots seinaleen fase adierazpidean eta berauen aplikazioan ahots prozesamenduaren hainbat arlotan, aurreko atalean azaldu denez. Laburbilduz, tesi hau adierazpide honen erabilgarritasuna ahots prozesamenduaren arlo ezberdinetan azpimarratzen saiatzen da. RPS adierazpideak fase informazioaren denborarekiko menpekotasuna ezabatzen du, uhin formarekin duen lotura argituz. Ezaugarri hauek oraino espektro modulua bakarrik erabiltzen zen aplikazioetan erabiltzeko egoki bihurtzen dituzte RPSak. Guk uste dugu tesi honetan azaltzen diren esperimenduek, aurretiazkoak badira ere, etorkizunean RPSak aplikatzeko ikerketa sakonagoa garatzeko atea zabaltzen dutela.

Horrela, tesiaren emaitzetariko bat ikerketa lerro berrien irekitzea da, hiru ataletan laburbildu daitezkeenak:

4.1. RPS analisi hobetua

RPSen kalkulua hobetu daiteke sendoagoa izan dadin seinale zaratsuen aurrean, pitch errorean aurrean edota seinalearen aurre-prozesaketaren aurrean. Hobekuntza honek berebiziko garrantzia du RPSak seinale telefoniko eta kodetuen analisisan erabiltzeko.

4.2. RPSen emaitzak ezagutze sistema aurreratueta

Tesian egin diren esperimenduak RPS informazioa ASR eta hiztun ezagutze sistemetan erabiltzeko, fase informazioa ataza hauetan esanguratsua ote zen baieztatzeko pentsatu ziren, eta beraz, ezagutze sistema klasikoak erabili ziren. Proba hauen emaitzak positiboak izan direnez, ikerketa lan gehiago egin beharko lirateke ikusteko zenbateraino RPSen erabilerak hobetu ditzakeen gaur egungo teknika aurreratuak erabiltzen dituzten ezagutze sistemen asmatze tasak. Era berean ebaluaziorako erabiltzen diren datu-base estandarrekin ere probatu beharko lirateke.

4.3. RPSen aplikazio berriak

Badaude ahots-prozesamenduaren eremuan arlo gehiago non fase informazioa aplikatu ahal den. Hautemate testen emaitzek iradokitzen dutenez, DCT-mel-RPS ereduak kalitate handiko sintesi estatistikoan probatu beharko lirateke. Bestalde, faseen eragin pertzeptuala txikia denez, erabil litezke informazioa gehitzeko ahots seinalean esteganografia edo datu-ezkatuketa egiteko adibidez.

5. Argitalpenak

Tesiaren emaitza nagusiak argitaratu dira aldizkari eta biltzarretan. RPS adierazpidea aldizkari artikuluan batean deskribatu da [25], eta baita beste artikuluan batean iruzurti sintetikoaren detekzioa [26]. Beste ekarpen batzuk nazioarteko biltzarretan argitaratu dira: besteak beste, polaritatea detektatzeko algoritmoa [27], RPSen aplikazioa ASR sistemetan [28], hiztun ezagutza [29], iruzurti sintetikoaren inguruko beste esperimenduak [30] eta faseen hautematearen ebaluaketak [31].

Beste ikerketa lanak tesiaren inguruan, pitch detekzio algoritmoarenak (CDP) bereziki, [23] eta [32]-n argitaratu ziren. Azkenengo artikulua honek ikasle baten artikulua onenatariko saria jaso zuen 5. JTH biltzarrean.

6. Aipamenak

1. McAulay, R.J., Quatieri, T.F.: Speech analysis/synthesis based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech and Signal Processing. 34, 744–754 (1986).
2. George, E.B., Smith, M.J.T.: Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model. IEEE Transactions on Speech and Audio Processing. 5, 389–406 (1997).

3. Stylianou, Y.: Harmonic plus noise models for speech, combined with statistical methods, for speech and speaker modification, PhD Thesis, (1996).
4. Erro, D.: Intra-lingual and cross-lingual voice conversion using harmonic plus stochastic models, PhD Thesis, (2008).
5. Quatieri, T.F., McAulay, R.J.: Shape invariant time-scale and pitch modification of speech. *IEEE Transactions on Signal Processing*, 40, 497–510 (1992).
6. Stylianou, Y.: Removing linear phase mismatches in concatenative speech synthesis. *IEEE Transactions on Speech and Audio Processing*, 9, 232–239 (2001).
7. O'Brien, D., Monaghan, A.: Shape Invariant Pitch and Time-Scale Modification of Speech Based on a Harmonic Model. *Improvements in Speech Synthesis*. pp. 64–75. Wiley Online Library (2002).
8. Liu, L., He, J., Palm, G.: Effects of phase on the perception of intervocalic stop consonants. *Speech Communication*, 22, 403–417 (1997).
9. Alsteris, L.D., Paliwal, K.K.: Evaluation of the modified group delay feature for isolated word recognition. *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, 2005. pp. 715–718. IEEE (2005).
10. Pobloth, H., Kleijn, W.B.: On phase perception in speech. 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. *Proceedings*. pp. 11–14 (1999).
11. Kim, D.: On the perceptually irrelevant phase information in sinusoidal representation of speech. *IEEE Trans. Speech and Audio Processing*, 9, 900–905 (2001).
12. Banno, H., Lu, J., Nakamura, S., Shikano, K., Kawahara, H.: Efficient representation of short-time phase based on group delay. *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98*. pp. 861–864. IEEE (1998).
13. Alsteris, L.D., Paliwal, K.K.: Short-time phase spectrum in speech processing: A review and some experimental results. *Digital Signal Processing*, 17, 578–616 (2007).
14. Murthy, H.A., Gadde, V.R.R.: The modified group delay function and its application to phoneme recognition. 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, *Proceedings (ICASSP'03)*. pp. 2–5 (2003).
15. Padmanabhan, R., Parthasarathi, S., Murthy, H.A.: Robustness of phase based features for speaker recognition. *Proc. Interspeech 2009*. pp. 3–6 (2009).
16. Kua, J.M.K., Epps, J., Ambikairajah, E., Choi, E.: LS regularization of group delay features for speaker recognition. *Proc. Interspeech 2009*. pp. 2887–2890 (2009).
17. Hegde, R.M., Murthy, H.A., Gadde, V.R.R.: Application of the modified group delay function to speaker identification and discrimination. 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, p. I–517–20. IEEE (2004).
18. Murty, K., Yegnanarayana, B.: Combining evidence from residual phase and MFCC features for speaker recognition. *IEEE Signal Proc. Letters*, 13, 52–55 (2005).
19. Zheng, N., Lee, T., Ching, P.C.: Integration of Complementary Acoustic Features for Speaker Recognition. *IEEE Signal Proc. L*, 14, 181–184 (2007).
20. Wang, L., Minami, K., Yamamoto, K., Nakagawa, S.: Speaker identification by combining MFCC and phase information in noisy environments. *Acoustics Speech and Signal Processing (ICASSP)*, 2010 IEEE International Conference on. pp. 4502–4505. IEEE (2010).
21. Masuko, T., Hitotsumatsu, T., Tokuda, K., Kobayashi, T.: On the security of HMM-based speaker verification systems against imposture using synthetic speech. *Proceedings of the European Conference on Speech Communication and Technology*. pp. 1223–1226. Citeseer (1999).
22. De Leon, P.L., Apsingekar, V.R., Pucher, M., Yamagishi, J.: Revisiting the security of speaker verification systems against imposture using synthetic speech. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 1798–1801. IEEE (2010).
23. Luengo, I., Saratxaga, I., Navas, E., Hernández, I., Sanchez, J., Sainz, I.: Evaluation of pitch detection algorithms under real conditions. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*. pp. 1057–1060 (2007).
24. ITU-R: ITU-R BS 1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, (1994).
25. Saratxaga, I., Hernández, I., Erro, D., Navas, E., Sanchez, J.: Simple representation of signal phase for harmonic speech models. *Electronics Letters*, 45, 381–383 (2009).
26. De Leon, P.L., Pucher, M., Yamagishi, J., Hernaez, I., Saratxaga, I.: Evaluation of Speaker Verification Security and Detection of HMM-Based Synthetic Speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 20, 2280–2290 (2012).
27. Saratxaga, I., Erro, D., Hernández, I., Sainz, I., Navas, E.: Use of Harmonic Phase Information for Polarity Detection in Speech Signals. *Proc. Interspeech 2009*. 1075–1078 (2009).
28. Saratxaga, I., Hernández, I., Odriozola, I., Navas, E., Luengo, I., Erro, D.: Using Harmonic Phase Information to Improve ASR Rate. *Proc. Interspeech 2010*. pp. 1185–1188 (2010).
29. Hernández, I., Saratxaga, I., Sanchez, J., Navas, E., Luengo, I.: Use of The Harmonic Phase in Speaker Recognition. *Proc. Interspeech 2011*. pp. 2757–2760 (2011).
30. De Leon, P.L., Hernández, I., Saratxaga, I., Pucher, M., Yamagishi, J.: Detection of synthetic speech for the problem of imposture. *Acoustics, Speech and Signal Processing (ICASSP)*, 2011 IEEE International Conference on. pp. 4844–4847 (2011).
31. Saratxaga, I., Hernaez, I., Pucher, M., Navas, E., Sainz, I.: Perceptual Importance of the Phase Related Information in Speech. *Proc. Interspeech 2012. ISCA, Portland, OR* (2012).
32. Saratxaga, I., Luengo, I., Navas, E., Hernández, I., Sánchez, J., Sainz, I.: Detección de pitch en condiciones adversas. *Proc. V Jornadas en Tecnología del Habla*. pp. 13–18 (2006).