PROJECTE FINAL de CARRERA: AFINADOR de VEU a TEMPS REAL per MODIFICACIÓ de FREQÜÈNCIA

Enginyeria Tècnica de Telecomunicacions
especialitat en Imatge y Só

Autor – Carles X. Pons Prieto
Tutor - Javier Ruiz Hidalgo

Índex:

1.- Introducció:

1.1.- Motivació :

Aquest projecte pretén implementar un sistema que permeti afinar la veu cantada en els casos en els que aquesta veu presenti una desafinació respecte unes freqüències concretes, que anomenem 'notes musicals'.
Existeix una opinió errònia molt estesa al voltant de l'afinació vocal, o millor dit al voltant de la desafinació vocal, i és la creença de que només la gent que no sap cantar desafina. Es comú pensar que només les persones que tenen un gran entrenament vocal i musical, o que han tingut un aprenentatge en profunditat en cant són capaços d'afinar. Aquest projecte proposa la sentència que afirma que 'cap persona' sap afinar.
En aquest punt és necessari el plantejament i l'aclaració del significat del terme 'afinar', en quant a la veu es refereix. Segons el diccionari de l'acadèmia de la llengua, afinar s'entén com a '*Cantar o tocar entonant amb perfecció els sons' o 'Posar a to els instruments musicals'.* Així doncs, i en sentit estricte, afinar amb la veu significa cantar emetent unes freqüències exactes, que equivalen a les citades 'notes musicals'. Per tant ens trobem en una difícil situació en la que per a afinar ens hem d'ajustar a uns quants punts concrets en un entorn infinit que pertany al conjunt real positiu de freqüències. És en relació a aquesta dificultat que aquest projecte afirma que ningú no sap afinar.
Aleshores, com és que si ningú no afina coneixem veus que ens sonen o semblen tan afinades?...
Aquí s'ha de plantejar doncs de nou el terme 'afinar', però en sentit relatiu. El fet de que una veu no estigui estrictament afinada i ens soni com si ho estigués té el seu motiu en el funcionament fisiològic del sistema auditiu humà. I es que existeix un marge de freqüències al voltant d'una freqüència concreta 'freq. de pitch' dins el qual qualsevol freqüència que escoltem la nostra oïda la integra com si fos la citada freqüència central 'freq. de pitch'. Aquest marge de freqüències per suposat és variable en funció de la persona, y sobretot en funció de l'entrenament musical que tingui cada oïda. Així doncs, quan es diu d'algú que 'té oïda musical', se fa referència a que al tenir un entrenament musical o una capacitat natural superior a la mitjana, tenen un marge de freqüències indistingibles més petit del normal, i que en conseqüència sap discernir mes freqüències de les normals que diferenciaria un auditor mig.
Dins d'aquest marc d'afinació relativa és en el qual es defineix y pren importància aquest projecte, el qual intenta pal·liar els efectes d'aquestes diferencies en els conjunts de freqüències indistingibles en las persones, es a dir intenta pal·liar els efectes de les diferencies de 'oïda' de les persones, però no els efectes produïts per les diferencies en quant a les capacitats vocals en cant.

1.2.- Introducció al projecte:

Com s'ha comentat en el punt 1.1, aquest projecte pretén desenvolupar un sistema que afini la veu cantada. El funcionament general d'aquest sistema es basaria en diferents blocs que a continuació es presentaran. A grans pinzellades el nostre sistema es basarà en l'anàlisi d'un senyal entrant, típicament de veu encara que també podria ser de qualsevol altre instrument harmònic, i el vagi escalant en freqüència en funció d'un conjunt finit de notes (o el que és el mateix, de freqüències) vàlides, per tal de corregir qualsevol desviament freqüencial que pugui tenir aquesta entrada amb respecte les citades notes, i que a la sortida resulti un senyal 'afinat', com s'ha explicat anteriorment. El conjunt de notes vàlides es podrà seleccionar per escales musicals (escala do, re, …, cromàtica, …), o be nota per nota per tal de poder aconseguir sons que es moguin en només una, dos , o mes freqüències.
També es pretén afegir alguns paràmetres addicionals que controlin aquesta afinació com llindars d'afinació, temps d'atac i de release, detune, etc.
Finalment, aquest sistema s'implementarà en temps real. El sistema tindrà que efectuar també un overlap entre trames de factor 4 o superior per tal d'aconseguir que el senyal de sortida sigui sensiblement continu i sense artificis.
Així doncs mitjançant aquest sistema, qualsevol persona o instrumento musical harmònic podrà evitar els problemes sorgits degut a la desafinació, ja sigui constant o puntual en el temps, ajudant així a la perfecta unió dels diferents sons que composen una peça musical formant una suma de sons harmoniosa y agradable.
Cal dir que aquest sistema pretén aconseguir arribar a ser un afinador amb una exactitud suficientment gran com per a ser útil com a tal, ja que, com les desafinacions solen donar-se en marges freqüencials molt petits, serà necessari ajustar molt les mesures i les correccions per tal d'aconseguir ser eficient en el seu propòsit.
En els punts posteriors s'explicarà mes en detall el funcionament d'aquest sistema, i es discutirà l'estat de l'art per a cada part que l'integra, també es parlarà de l'estat de l'art del projecte conjunt, ja que aquest sistema es ja una realitat desenvolupada la qual ja es comercialitza per diferents empreses.

Enumeració dels diferents blocs:

Aquest sistema es composarà de diferents blocs que introduirem en aquest punt, i que seran discutits mes endavant.
En primer lloc l'usuari tindrà que configurar el sistema per tal que aquest es comporti de la manera esperada, oferint l'afinació adequada i actuant d'acord a la configuració dels diferents paràmetres. El senyal d'entrada entrarà en el sistema i passarà primer per un **convertidor A/D** el qual ens anirà donant el senyal per trames d'una duració coherent amb la freqüència de mostreig i la latència que s'imposi. Es recollirà aquest senyal trama per trama en un **buffer d'entrada,** es farà l'enfinestrament**, l'acondicionament**, i passarà després per un bloc **detector de pitch**, el qual ens permetrà conèixer la nota associada a cada trama.
En aquest punt tindrem una trama del senyal d'entrada sense modificar, la nota associada a la mateixa (pitch), i un conjunt de paràmetres de la interfície **de configuració**. Tota aquesta informació serà aleshores enviada cap a **l'escalador de freqüència**, que serà el bloc que modificarà el senyal, en cas de que aquesta no estigui ja afinada, es a dir, que la seva freqüència no coincideixi amb una de les introduïdes per l'usuari com a vàlides.
Així doncs en qualsevol cas en el punt de sortida d'aquest bloc s'obtindrà una trama ja afinada, i preparada per a ser extreta cap al **buffer de sortida,** després d'haver-li fet l'overlap adient entre trames, per a finalment ser convertit a senyal analògic de nou per el **convertidor D/A**.
A continuació es mostra un diagrama de blocs del sistema afinador:

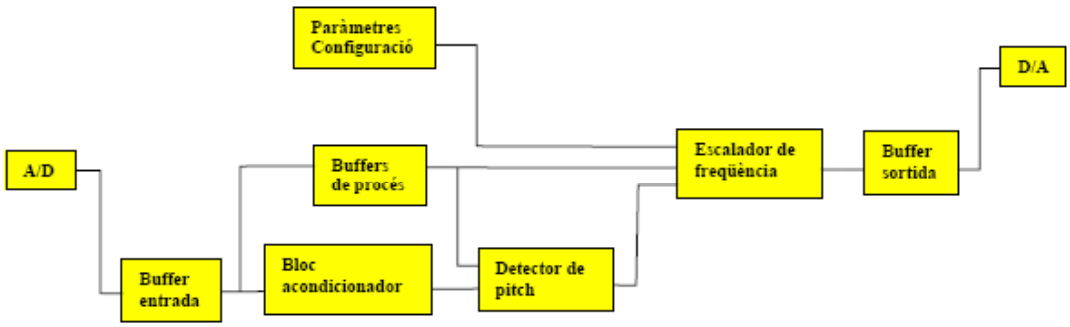


Diagrama de blocs del sistema afinador.

A continuació s'explicarà el funcionament de cada bloc per separat. S'obviarà el funcionament dels blocs convertidors i dels buffers en aquest punt, així com els blocs d'enfinestrament i condicionament degut al caràcter introductori d'aquest punt. Es pot obtenir informació amb mes detall en els següents punts d'aquesta memòria.

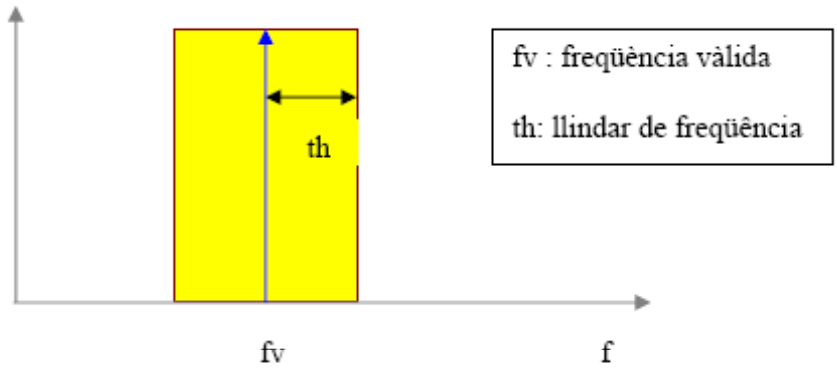
-Bloc Paràmetres de Configuració:

En aquest bloc es determinarà tota la part del funcionament del sistema susceptible de ser alterada per l'usuari:

-Selecció de freqüències vàlides: Com s'ha dit en la introducció, es podrà seleccionar des d'una sola nota vàlida (un to), fins els 12 semitons de qualsevol octava musical, passant per les set notes corresponents a una escala (p ex. Escala DO : do-re-mi-fa-sol-la-si, escala RE: re-mi-fa#-sol-la-si-do#, etc).

-Detune: Mitjançant aquest paràmetre es podrà variar la freqüència de referència per construir totes les notes. Per exemple a Espanya s'utilitza com a freq. De referència els 440Hz. associada a la nota LA, però al Japó s'utilitzen normalment escales de 445Hz (freq. de referència associada també a la nota LA). Així doncs, variant aquest paràmetre, que per defecte estarà centrat a l'escala de 440Hz., podrem variar aquesta freq. de referència. Una altre utilitat d'aquest paràmetre seria afinar relativament el sistema amb un o mes instruments que tinguin una afinació no exacta.

-Llindar de freqüència: Amb aquest paràmetre indicarem el marge de desviament admissible al senyal d'entrada respecte les freqs. vàlides, dins del qual el sistema no farà cap correcció. Aquest factor és necessari si volem assolir a la sortida un senyal que soni natural. El següent gràfic mostra l'efecte del llindar de freqüència :



-Temps d'atac: Defineix el temps entre el qual el sistema detecta un canvi de freqüència (suficient per ser estimat com a canvi de nota) fins que aquest actua o modifica la seva sortida. Mitjançant aquest paràmetre podem aconseguir un efecte tipus "cher" si seleccionem un temps d'atac molt petit, o un sistema que ens afini la veu d'una manera imperceptible o 'natural' mitjançant temps d'atac més llargs.

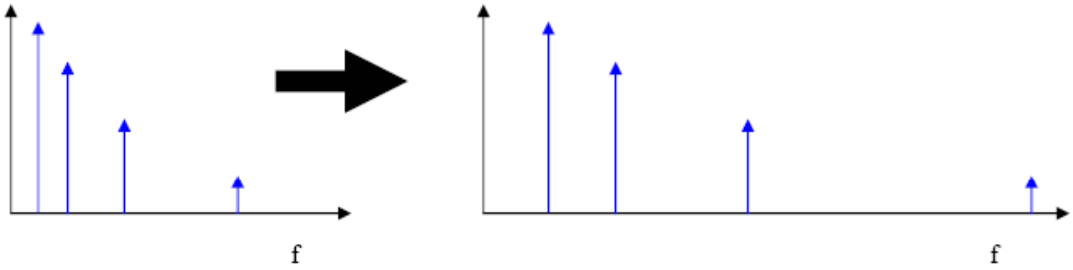
-LFO: Aquesta secció servirà per afegir una oscil·lació al senyal entrant que moduli l'amplitud de la mateixa. Aquest modul pretén simular l'efecte de vibrato de la veu dels cantants o dels instruments, i constarà de quatre paràmetres:
-tipus de senyal LFO: podrà ser sinusoïdal, quadrada o triangular
-freqüència LFO: determina la freq. del senyal modulador
-profunditat LFO: indicarà el nivell de variació d'amplitud resultant
-delay LFO: expressa el temps en ms. des que comença una nota fins que comença a actuar l'LFO.
En conclusió, amb tot aquest ventall de possibilitats de configuració podrem assolir efectes de tots tipus fent servir les múltiples combinacions que aquest sistema ens oferirà.

-Bloc detector de Pitch:

Aquest bloc es basa en l'anàlisi de l'fft de cada trama. El resultat d'aquest bloc és la freqüència fonamental de la nota que s'està introduint. El problema d'aquest bloc es troba en que aquesta freqüència no sempre es correspon amb la freq. del màxim de la fft, es a dir, de vegades els harmònics poden tenir més amplitud que la freq. fonamental. Aleshores cal fer un anàlisi de la relació entre harmònics per tal d'endevinar quina és la freqüència de la qual els altres màxims de la fft en son múltiples, i aquesta si que serà la freq. fonamental. La idea per aquest bloc es fer la seva implementació fent servir projectes d'altres estudiants, però de moment no hi ha hagut gaire èxit ja que si be he trobat moltes implementacions, cap d'elles resulta prou precisa com per fer aquest projecte. El problema principal que trobem en els codis existents és la manc de resolució freqüencial suficient per a aquesta aplicació. Aquest problema el trobem també en el bloc de l'escalat en freqüència.

-Bloc escalador de freqüència:

Finalment el senyal arriba al mòdul principal d'aquest disseny, en el qual serà transformat si és necessari per treure un flux de trames totes 'afinades' segons els criteris de configuració i el resultat del bloc detector de pitch.
Aquesta transformació es basa en una translació de l'espectre de les trames respecte la freqüència fonamental, fins a les freqüències vàlides més properes, tot conservant la relació òptima entre harmònics i sub-harmònics, es a dir, no es tracta d'un moviment simple de mostres de la fft sinó que per conservar aquesta relació òptima entre harmònics cal fer una redistribució de l'espectre per tal que aquest quedi eixamplat com mostra la següent figura:



El principal problema que es troba per la implementació d'aquest bloc consisteix en el que es pot anomenar 'resolució freqüencial', que consisteix en la qualitat en la aproximació que podem arribar a assolir en les mesures.

El següent capítol es centrarà en la discussió de l'estat de l'art dels blocs detallats en aquest punt, però es tractaran només els dos únics blocs que tenen una certa complexitat algorísmica, i que són la part més important d'aquest projecte. Aquests són els blocs de **detector de pitch** i **modificació de freqüència**.

1.2.- Objectius:

Com a objectius dins aquest projecte s'ha d'entendre el conjunt d'especificacions i funcionalitats que hi pugui tenir, incloent els formats de realització i tota la resta de característiques del mateix que a continuació s'especifiquen.

- Sistema a temps real en C++ / DirectX
- Latència inferior als 50ms
- Detector de pitch : tolerància d'error +/- 1Hz
- Simulació en matlab
- Possibilitat d'escollir les notes vàlides des de només una nota fns a tot el conjunt

2.- Estat de l'art:

2.1.- Estat de l'art dels blocs (pitch+shift freqüència):

Com s'ha dit al final del punt anterior, només es tractarà l'estat de l'art dels dos únics blocs més importants del sistema, el de detecció de pitch i el de modificació de freqüència. La resta de blocs no es tractarà en aquest bloc ja que la seva complexitat algorísmica és prou baixa com per no haver estat motiu d'anàlisi en quant a l'estat de l'art es refereix.
Aquests dos blocs que tractarem veurem que tots dos es poden implementar mitjançant tant un anàlisi del senyal en temps com mitjançant un anàlisi en freqüència, obtenint resultats diferents, però acomplint el mateix objectiu.

2.1.1.- *Detecció de pitch:*

La valoració de la freqüència fonamental (freq. de pitch), també designada com a detecció de pitch, ha estat un assumpte popular d'investigació per molts anys, i encara s'està investigant avui. En la conferència internacional 2002 de l'IEEE sobre el procés de l'acústica, del discurs i del senyal, havia una sessió completa sobre la valoració de la freq. de pitch. El problema bàsic és extreure la freqüència fonamental (freq. de pitch) d'un senyal de tons barrejats, la qual és generalment el component de la freqüència més baixa, que es relaciona bé amb la majoria dels altres harmònics. En una forma d'ona periòdica, la majoria dels harmònics es relacionen amb una relació de mulriplicitat, significant que la freqüència de la majoria dels harmònics és relacionada amb la freqüència de l'harmònic més baix per un quocient enter petit. La freqüència d'aquest harmònic més baix és la freq. de pitch de la forma d'ona.
La dificultat de trobar la freq. de pitch d'una forma d'ona depèn de la pròpia forma d'ona en sí mateix. Si la forma d'ona té pocs harmònics de més alta freqüència o l'energia dels harmònics més alts és petita, la freq. de pitch és més fàcil de detectar, com en els quadres 1 i 2. Si els harmònics tenen més energia que la freq. de pitch, llavors el període és més difícil de detectar, com en els quadres 3 i 4. El quadre 4 és un exemple del fenomen de la mancança d'freq. de pitch.

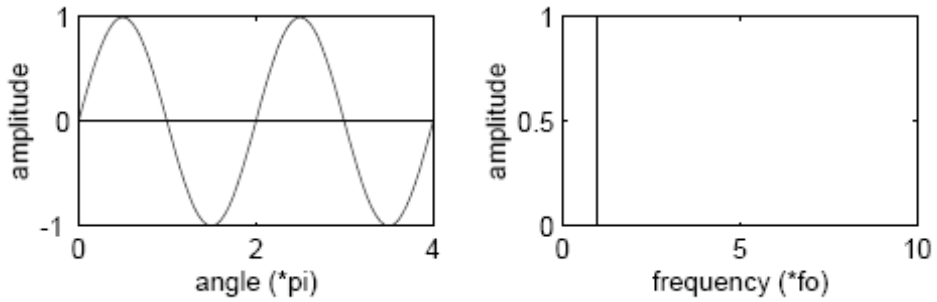


Figure 1: Waveform with no upper harmonics.

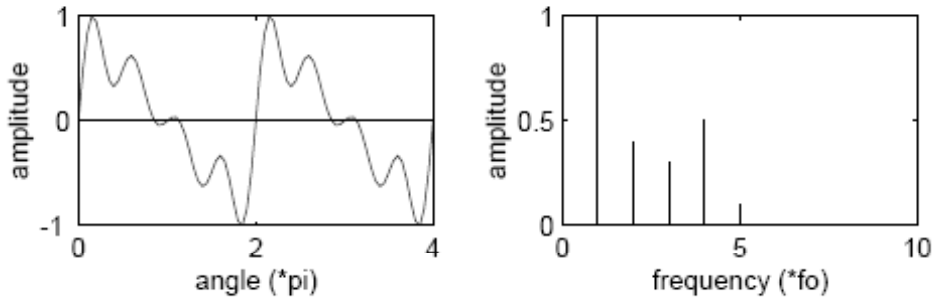


Figure 2: Waveform with lower power upper harmonics.

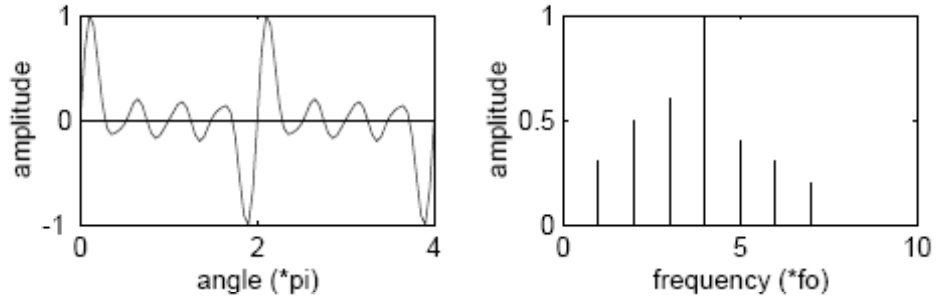


Figure 3: Waveform with higher power upper harmonics.

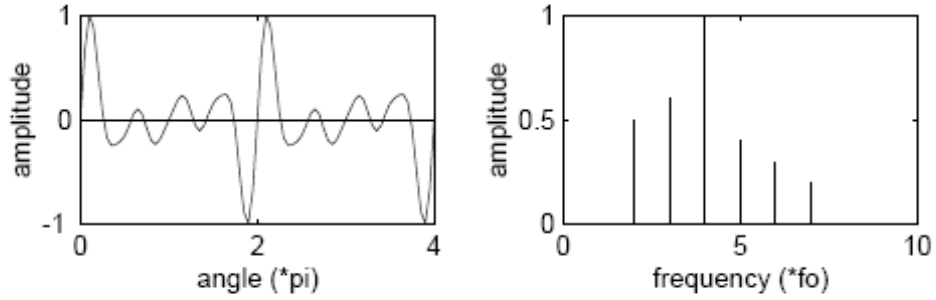


Figure 4: Waveform with high power upper harmonics and no fundamental.

2.1.1.a) Mètodes en el domini del Temps:

L'aproximació més bàsica per valorar l'freq. de pitch consisteix en mirar la forma d'ona, la qual representa el canvi en la pressió d'aire en un cert termini, i tractar de detectar la freq. de pitch en aquest tros de senyal.

1.-Detecció per repetició d'esdeveniments temporals:

Existeix una família de mètodes relacionats resultants de la valoració de l'freq. de pitch en el domini del temps que es basen en descobrir cada quant temps la forma d'ona es va repetint completament. La teoria d'aquests mètodes és que si una forma d'ona és periòdica, aquesta haurà d'incloure els esdeveniments temporals que la seva naturalesa implica, i el nombre d'aquests esdeveniments que succeeixin en un segon es relaciona inversament amb la freqüència. Cadascun d'aquests mètodes és útil per a classes particulars de formes d'ona. Si hi ha un esdeveniment específic que es coneix que nomes ha d'existir una vegada per període en la forma d'ona, aquest període pot ser doncs identificat i ser comtat.

Recompte de creuaments per zero (ZRC):

El ZCR és una mesura de com la forma d'ona passa per zero per unitat de temps. La idea és que el ZCR dóna la informació sobre el contingut espectral de la forma d'ona. El pensament és que el ZCR es pot relacionar directament amb el nombre de períodes que la forma d'ona fa per unitat de temps. Aviat es va veure clarament que hi ha problemes amb aquesta mesura de l'freq. de pitch. Si l'energia espectral de la forma d'ona es concentra al voltant de l'freq. de pitch, llavors creuarà la línia zero dues vegades per cicle, com en la figura 5a. No obstant això, si la forma d'ona conté components espectrals d'alta freqüència, com en la figura 5b, pot ser que la forma d'ona creuï la línia del zero més de dues vegades per cicle. Un detector d'freq. de pitch per ZCR es podria desenvolupar mitjançant un filtrat inicial per a eliminar els harmònics més alts que contaminen la mesura, però la freqüència de tall del filtre necessita ser triada curosament per a no eliminar l'freq. de pitch i eliminar tanta quantitat d'alta freqüència com sigui possible. Altra possibilitat del detector ZCR seria detectar patrons en les creuaments per zero, i aproximar un valor d'freq. de pitch basat en aquests patrons.

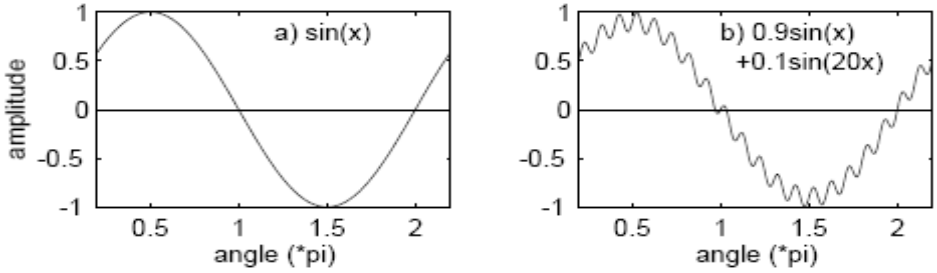


Figure 5: Influence of higher harmonics on zero crossing rate. (after [22])

Recompte de màxims:

Aquest mètode conta el nombre de pics positius per segon en la forma d'ona. En teoria, la forma d'ona tindrà un valor màxim i un valor mínim cada cicle, per tant només es necessita contar aquests valors màxims (o valors mínims) per a determinar la freqüència de la forma d'ona. En la pràctica, un detector de màxim local s'utilitza per a trobar on està localment la la forma d'ona més gran, i el nombre d'aquests màxims locals en un segon és la freqüència de la forma d'ona, a menys que cada període de la forma d'ona contingui més d'un màxim local. Alternatives similars estan disponibles per a aquest mètode, igual que estan disponibles per a la distància del detector de la tarifa de creuaments per zero.
La distància entre màxims locals dóna la longitud d'ona que és inversament proporcional a la freqüència.

Recompte d'events de l'envolupant:

Si una forma d'ona és periòdica, la costa de la forma d'ona també serà periòdica, i els becs o els zeros en la costa es poden extreure de la mateixa manera que el ZCR. En alguns casos, els zeros o els becs en la costa 5 van poder ser més informatius que zeros o becs en la forma d'ona original, o la detecció d'aquests esdeveniments va poder ser més robusta, depenent del domini del senyal.

Discussió:

La dificultat principal amb mètodes de detecció per taxes d'esdeveniments en el temps és que les formes d'ona espectralment complexes tenen rarament només un esdeveniment per cicle. Les formes d'ona amb espectres rics en harmònics poden creuar per zero moltes vegades o tenir molts pics en un cicle (quadre 5). Hi ha però alguns aspectes positius dels algorismes de detecció per recompte d'esdeveniments en el temps, ja que aquests són mètodes excessivament simples entendre i d'implementar, i consumeixen molt pocs recursos d'execució. Si es coneix la naturalesa del senyal, es pot incloure un mètode que s'adapti a la forma d'ona, reduint l'error. Els comptadors de màxims han estat la implementació escollida en la pràctica per als detectors de freqüència per molts anys, perquè el circuit és molt simple, i ajuntat amb un filtre passa baixes simple, proporcionen un mòdul bastant robust.

2.-Detecció per autocorrelació:

La correlació entre dues formes d'ona és una mesura de la seva semblança. Les formes d'ona es comparen en diversos intervals del temps, i el seu novell d'igualtat es calcula en cada interval. El resultat d'una correlació és una mesura de semblança en funció del retard de temps entre els principis de les dues formes d'ona. La funció de l'autocorrelació és la correlació d'una forma d'ona amb si mateixa. Així doncs podem assegurar una semblança màxima amb retard zero, i un progressiu descens en augmentar el retard.
L'autocorrelació de les formes d'ona periòdiques exhibeixen una característica interessant: la funció d'autocorrelació resultant és també periòdica. Mentre el retard de temps augmenta fins a la meitat del període de la forma d'ona, la correlació va disminuint fins arribar a un mínim.
Això és perquè la forma d'ona és fora de fase amb la seva còpia retardada en el temps. Quan el retard de temps supera el mig període, l'autocorrelació augmenta altra vegada de nou fins a un màxim, que es dona quan el retard de temps arriba a un període, perquè la forma d'ona i la seva còpia temps-retardada són aleshores en fase. El primer pic en l'autocorrelació indica el període de la forma d'ona. Els problemes amb aquest mètode es presenten amb l'autocorrelació de formes d'ona harmònicament complexes, o pseudo periòdiques.

2.1.1.b) Mètodes en el domini freqüencial:

Existeix molta informació en el domini freqüencial que es pot relacionar amb l'freq. de pitch del senyal. Els senyals de veu tendeixen a estar composats d'una sèrie de harmònics harmònicament relacionats, que es poden identificar i utilitzar per a extreure l'freq. de pitch. S'han fet moltes temptatives per extreure l'freq. de pitch d'un senyal d'aquesta manera.

Recompte dels components freqüencials:

El procediment original va començar amb una transformació espectral i una identificació dels harmònics en el senyal, fent us de la detecció de màxims. Per a cada parell d'aquests harmònics, l'algorisme troba el nombre d'harmònics més petits que correspondrien a una sèrie harmònica que inclourien aquests dos harmònics. Cadascun d'aquests parells harmònics del nombre llavors s'utilitza com hipòtesi per a la freqüència fonamental del senyal. Després que tots els parells de harmònics es considerin d'aquesta manera, la hipòtesi més fortament possible pels parells de harmònics es tria com la freqüència fonamental. Aquest mètode no requereix que la freqüència fonamental del senyal estigui present, i treballa bé amb harmònics inharmònics i amb la mancança de harmònics.

Mètodes basat en filtres de freqüència:

Un dels mètodes més utilitzats per a la valoració freq. de pitch és mitjançant diversos filtres amb diverses freqüències centrals, i comparant la seva sortida. Quan un pic espectral s'alinea amb la banda passant d'un filtre, el resultat és un valor més alt en la sortida del filtre que quan no s'alinea.

Filtre òptim en pinta

Es un algorisme robust però de còmput intensiu. Un filtre en pinta té moltes bandes de pas igualment espaiades. En el cas de l'algorisme òptim del filtre en pinta, la localització de les bandes de pas es basa en la localització del primer pas de banda. Per exemple, si la freqüència de centre del primer pas de banda és 10 Hz, llavors haurà bandes de pas estrets cada 10 Hz consecutivament, fins a la freqüència de shannon.
Si un sistema d'harmònics regularment espaiats està present en el senyal, després la sortida del filtre de la pinta serà la més gran quan les bandes de pas s'alineen amb els harmònics. Però si el senyal té solament un harmònic, el fonamental, el mètode fallarà perquè hi haurà moltes posicions que tindran la mateixa amplitud de sortida,

Filtres IIR:

Aquest mètode consisteix en un filtre passa banda comfigurable en selectivitat, que s'escombra a través de l'espectre de la freqüència. Quan el filtre està en línia amb una freqüència forta harmònic, una sortida màxima estarà present en la sortida del filtre, i el freq. de pitch es pot llavors llegir de la freqüència de centre del filtre. L'autor suggereix que un usuari experimentat d'aquest filtre pugui reconèixer la diferència entre un espectre uniformement espaiat, característic d'una sola nota ric harmònica, i un espectre que conté més que d'una nota distinta. Aquest mètode de la valoració de l'freq. de pitch està d'alguna manera relacionat a l'operació de l'estroboscopi, una eina usada pels sintonitzadors del piano.

L'anàlisi Cepstrum:

L'anàlisi Cepstrum és una forma d'anàlisi espectral on la sortida es la transformada de Fourier del logaritme de l'amplitud de l'espectre de la forma d'ona de l'entrada. Aquest procediment va ser desenvolupat en una temptativa de linealitzar un sistema que no era lineal. Els harmònics que apareixen en un espectre de freqüència són sovint lleument inharmònics, i les temptatives del cepstrum són corregir aquest efecte. El cepstrum conegut ve d'invertir les primeres quatre lletres en la paraula espectre, indicant un espectre modificat. La variable independent que es transforma amb el cepstrum s'ha cridat 'quefrency', i ja que aquesta variable es relaciona molt de prop amb el temps és acceptable referir-se a aquesta variable com temps. La teoria darrere d'aquest mètode confia en el fet que l'anàlisi de Fourier d'un senyal té generalment
un nombre de pics regularment espaiats, representant l'espectre harmònic del senyal. Quan es fa el logaritme de la magnitud de l'espectre, es redueixen aquests pics, i el resultat és una forma d'ona periòdica en el domini de la freqüència, el període del qual (la distància entre els pics) es relaciona amb la freqüència fonamental del senyal original.

El quadre 6 demostra el progrés de l'algorisme del cepstrum. La figura 6b demostra la representació espectral estàndard d'un senyal harmònic periòdica (amb la nota A4).
La figura 6c demostra el logaritme de la magnitud de l'espectre del mateix senyal. Observi la periodicitat d'ambdós espectres, i la naturalesa re-escalada de l'espectre de la magnitud del registre. El mètode del cepstrum assumeix que el senyal té els seus harmònics de freqüència regularment espaiat. Si aquest no és el cas, per exemple amb un espectre inharmònic d'una campana o l'espectre uni-harmònic d'una sinusoide, el mètode proveirà de resultats erronis. Com amb la majoria dels altres mètodes de la valoració freq. de pitch, aquest mètode s'adapta bé als tipus específics de senyals. Va ser desenvolupat originalment per a l'ús amb els senyals de veu parlada, que són espectralment rics i amb harmònics uniformement espaiats.

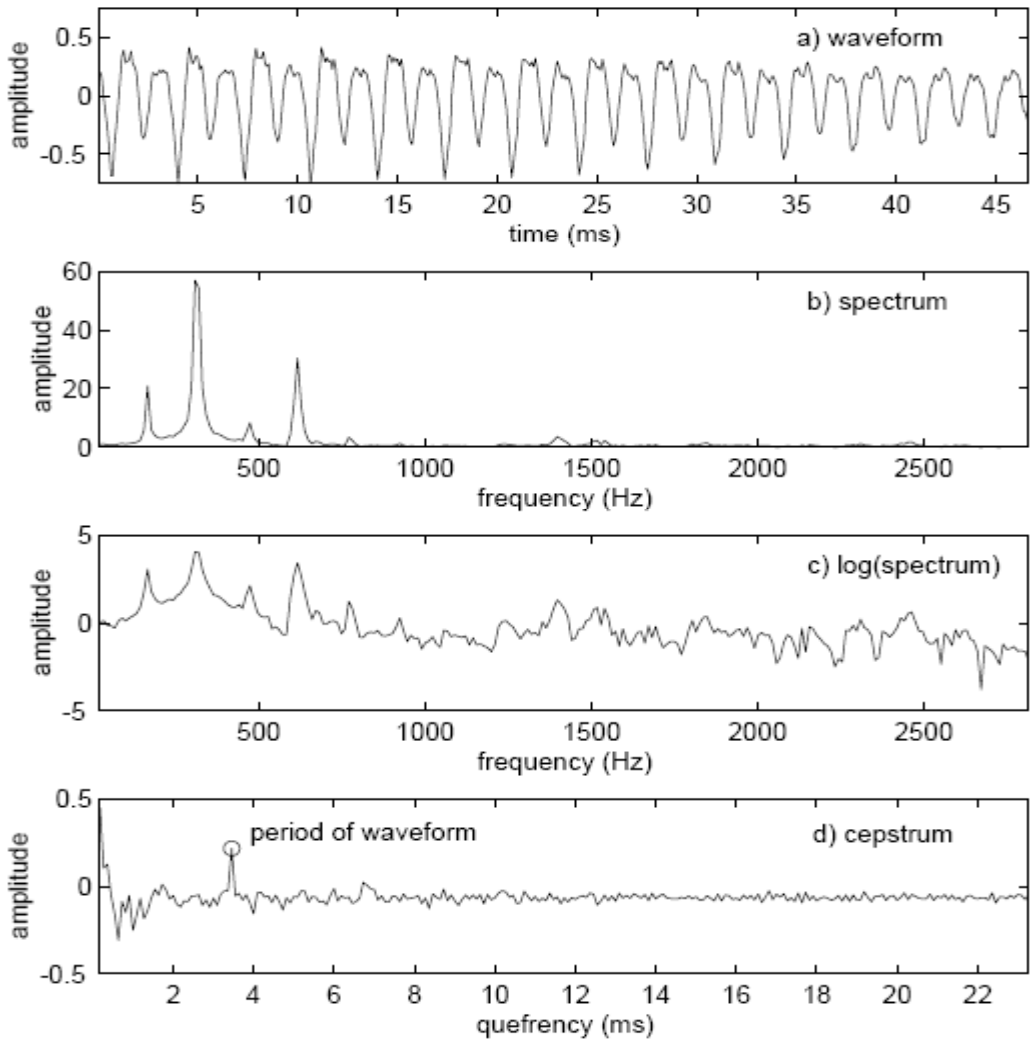


Figure 6: Stages in the cepstrum analysis algorithm.

Els mètodes de Multi-Resolució:

Una millora que es pot aplicar a qualsevol mètode espectral de valoració de l'freq. de pitch és utilitzar les resolucions múltiples. La idea és relativament simple: Si l'exactitud de cert algorisme en certa resolució és una mica sospitada, confirmi o negui qualsevol hipòtesi del perit freq. de pitch usant el mateix algorisme en una resolució més alta o més baixa. Així, utilitzi una finestra més gran o més petita del temps per a calcular l'espectre. Si un pic de freqüència es mostra en totes les o la majoria finestres, això es pot considerar una confirmació de la hipòtesi del perit freq. de pitch. No obstant això, cada nova resolució de l'anàlisi significa un cost més de càlcul.

Mètodes estadístics freqüencials:

El problema de la valoració automàtica de l'freq. de pitch es pot considerar, d'algunes maneres, estadístic. Cada marc d'entrada es classifica en només un d'un nombre de grups, representant l'estimació de l'freq. de pitch del senyal. Molts investigadors han pensat que els mètodes estadístics moderns es poden aplicar al problema de la valoració de l'freq. de pitch. Dos d'aquests mètodes es presenten a continuació.

Xarxes Neurals:

Els models de connectivitat, dels quals les xarxes neurals són un exemple, són patrons de semblança. Lògicament, consisteixen en una col·lecció de nodes, connectada per acoblaments amb els enllaços associats. En cada node, els senyals de tots els acoblaments entrants se sumen segons els pesos d'aquests acoblaments, i si la suma satisfà certa funció de transferència, un impuls s'envia a altres nodes amb acoblaments de la sortida. En l'etapa d'entrenament, l'entrada es presenta a la xarxa juntament amb una sortida suggerida, i els pesos dels acoblaments s'alteren per a produir la sortida desitjada. En l'etapa d'operació, la xarxa es presenta amb l'entrada i proporciona una sortida basada en els pesos de les connexions. Un model per al reconeixement del pitch pren com entrada un sistema de harmònics espectrals, o la forma d'ona en el domini del temps, o la representació espacial de la fase del senyal. A la sortida resultaria probablement una hipòtesi de la freqüència. Altra manera d'utilitzar els models de connectivitat per a la valoració de l'freq. de pitch és el modelar el sistema auditiu humà, on es presenta un model de la xarxa dels nervis basat en els mecanismes coclears de l'oïda humana.

Estimadors de màxima semblança:

L'intent d'aquesta tècnica és reconèixer i tractar la mancança lleu d'harmonia dels harmònics de freqüència que ocorren en un senyal. El model es basa en una observació dels harmònics en un anàlisi de Fourier a curt termini. Cada observació s'assumeix que ha estat produïda per un so amb una freqüència fonamental particular freq. de pitch, i que cada espectre conté tant la informació útil com harmònics inharmònics i no sinusoïdals (soroll). Per a un sistema de freqüències fonamentals, l'algorisme computa la probabilitat de cada freq. de pitch de totes les possibles, i troba el màxim.

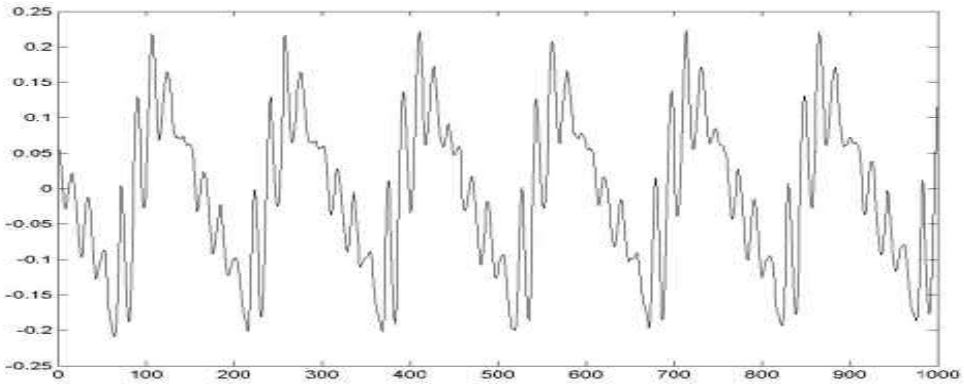
2.1.2.- Modificació del pitch:

Com s'ha comentat a l'inici del bloc de l'estat de l'art del detector de pitch, per a la modificació de freqüència es poden adoptar basicament dos tipus de estratègia: la modificació en el domini del temps, i la modificació en el domini freqüencial. Depenent del tipus de modificació requerida i del tipus de senyal a tractar resultarà més eficients un tipus d'estratègia concreta, com al final d'aquest bloc discutirem.

2.1.2.a) Modificació freqüencial en temps:

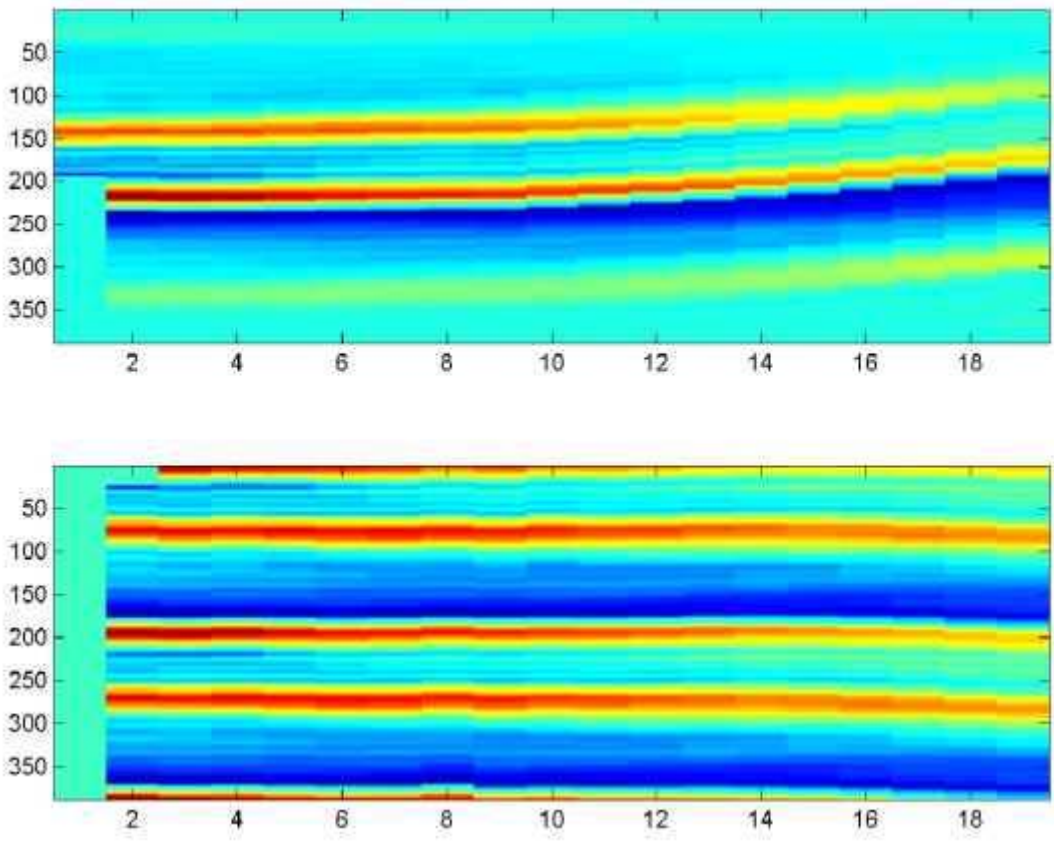
Els algorismes de modificació del pitch en el domini del temps tenen diversos avantatges sobre els del domini freqüencial. En primer lloc, els formants del senyal original es poden preservar, significant que el timbre del senyal d'entrada serà en gran part in-afectat, cosa que en la modificació freqüencial no succeeix. En segon lloc, la complexitat de càlcul és molt menor perquè no hi ha necessitat de transformar les dades. Dos algorismes diferents han estat creats i utilitzats amb una diferència principal que és la manera en la qual s'efectua l'overlap del senyal i la seva posterior suma. Amb aquests sistemes el senyal conserva molt millor la seva forma original. Per ambdós algorismes el senyal original és analitzat en finestres i amb un overlap d'una grandària especificada. Llavors, cada finestra es computa i s'utilitza el període detectat. Després de la construcció de cada finestra, que es descriu més lluny sota dos acostaments, les finestres es superposen i es sumen per a crear el nou pitch en el senyal de sortida.

Quan l'algorisme de la detecció decideix que una trama es correspon a un senyal sord (és a dir sense cap freqüència fonamental), còpia la finestra tal com està sense modificar. S'utilitza una finestra de Hanning per a filtrar les discontinuïtats.



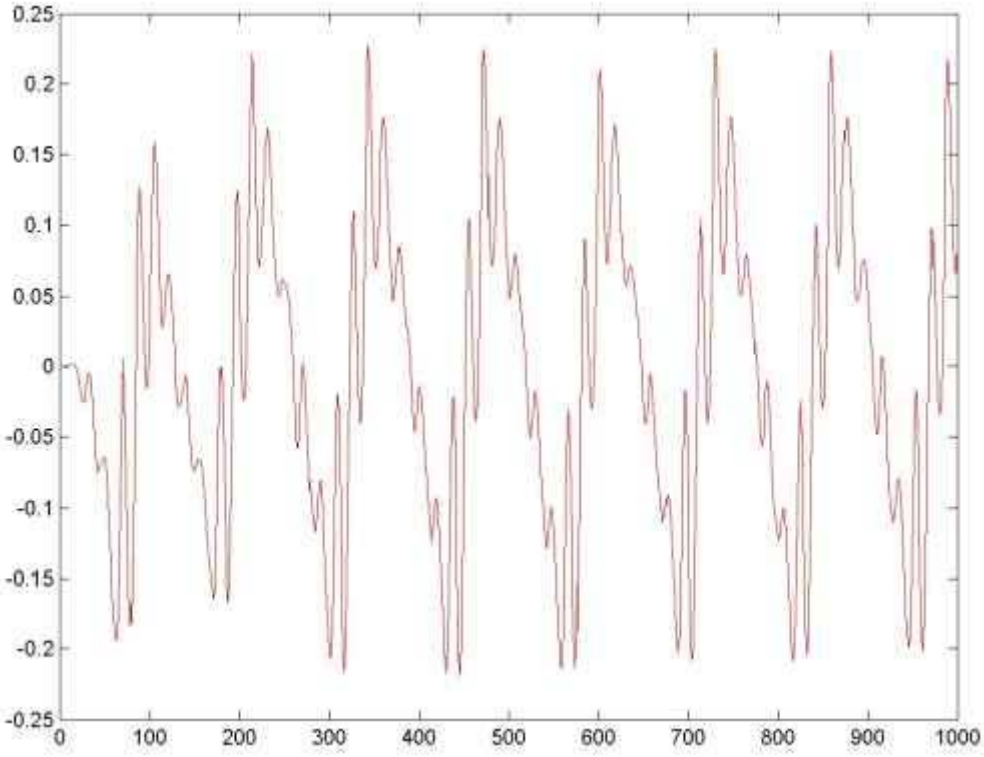
PSOLA : Pitch-Synchronous Overlap-Add

La clau del sistema PSOLA és la determinació i la utilització de marcadors de pitch en els senyals originals. La idea és que aquests marcadors estiguin igualment espaiats a través del senyal (en els intervals iguals al període fonamental detectat), però també deuen ser col·locats en una localització per a la qual el senyal tingui un valor màxim (un pic). Aquests dos requeriments sovint presenten conflictes, especialment degut a que la nostra assumpció de que el període fonamental és constant per a la finestra sencera no és del tot cert. Conseqüentment, si fem un seguiment dels pics màxims del senyal per a cada període haurem de relaxar el requisit de que els marcadors estiguin igualment espaiats. Per altra banda, si seguim solament els pics màxim sense tenir en compte el període fonamental, els nostres marcadors no tindran relació amb el pitch i no ens seran útils. Per a avaluar aquest compromís, creem una matriu a la qual cada columna conté dos períodes del senyal, la fila del centre comença en 0, i els increments són d'un període per cada columna. Llavors utilitzem un algorisme dinàmic (creat per Vladimir Goncharoff i Patrick Gries de la universitat de Xicago en Illinois) per a trobar una trajectòria que vagi passant a través dels pics màxims tant com sigui possible, però que no va excedir d'uns paràmetres donats mentre que passa a través de la matriu.



Marcadors de pitch per finestres

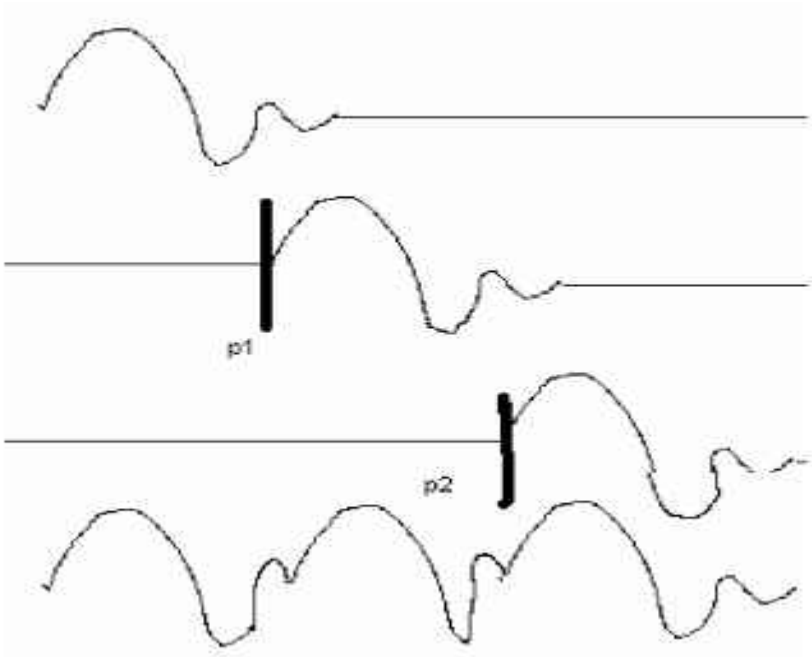
La matriu descrita es representa gràficament en el gràfic de la figura de sobre (on el cian és zero, blau marí és negatiu, i el vermell és positiu). Com es pot veure, els pics semblen moure's a través de la matriu en una línia recta, significant que quan superposem i sumem aquests segments, els pics són agregats un damunt de l'altre. Això redueix problemes de la fase amb interferència constructiva i destructiva entre els pics. Marcant els límits de les regions a extreure del senyal original, les seves noves localitzacions necessiten ser definides. Es crea aleshores un vector dels marcadors nous del pitch, que comença amb el primer marcador antic de pitch, que és la compensació de la fase, i després s'espaien igualment en els intervals iguals al període fonamental desitjat. Per a cada marcador nou, el marcador més proper del senyal original es troba i els dos períodes centrats al voltant d'aquest marcador són enfinestrats i copiats al senyal de sortida, centrats sobre el marcador nou. El resultat de tot això és un senyal que la seva forma d'ona conserva la forma de l'original, però té un període més curt o més llarg (depenent de la quantitat de canvi i en quina adreça). Per tant, el pitch es canvia de lloc sense alterar les qualitats de la veu que va produir el so.



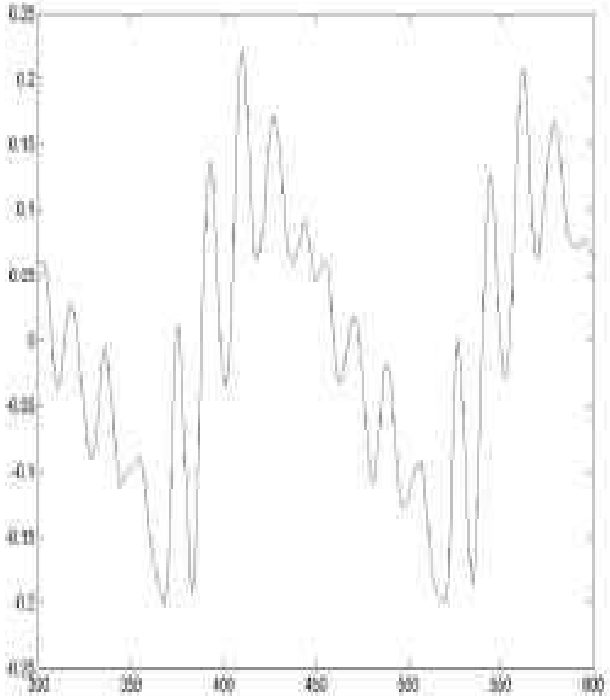
Trama mostrada al començament del bloc modificada amb PSOLA

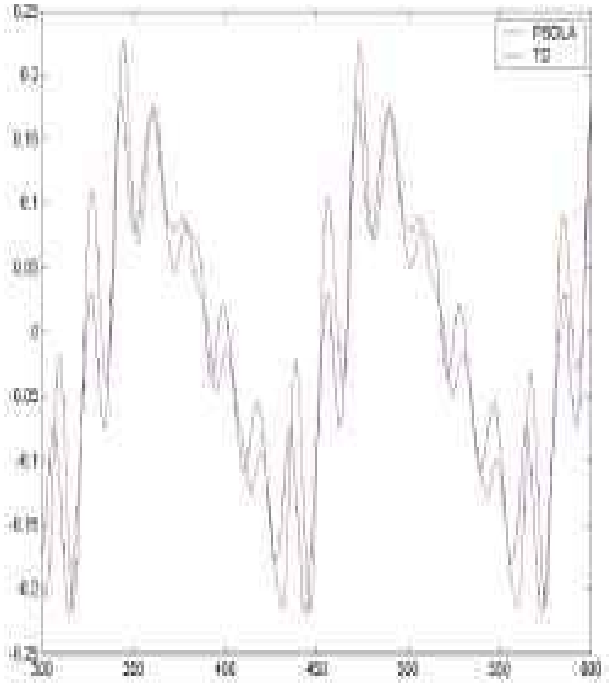
Time Shifting:

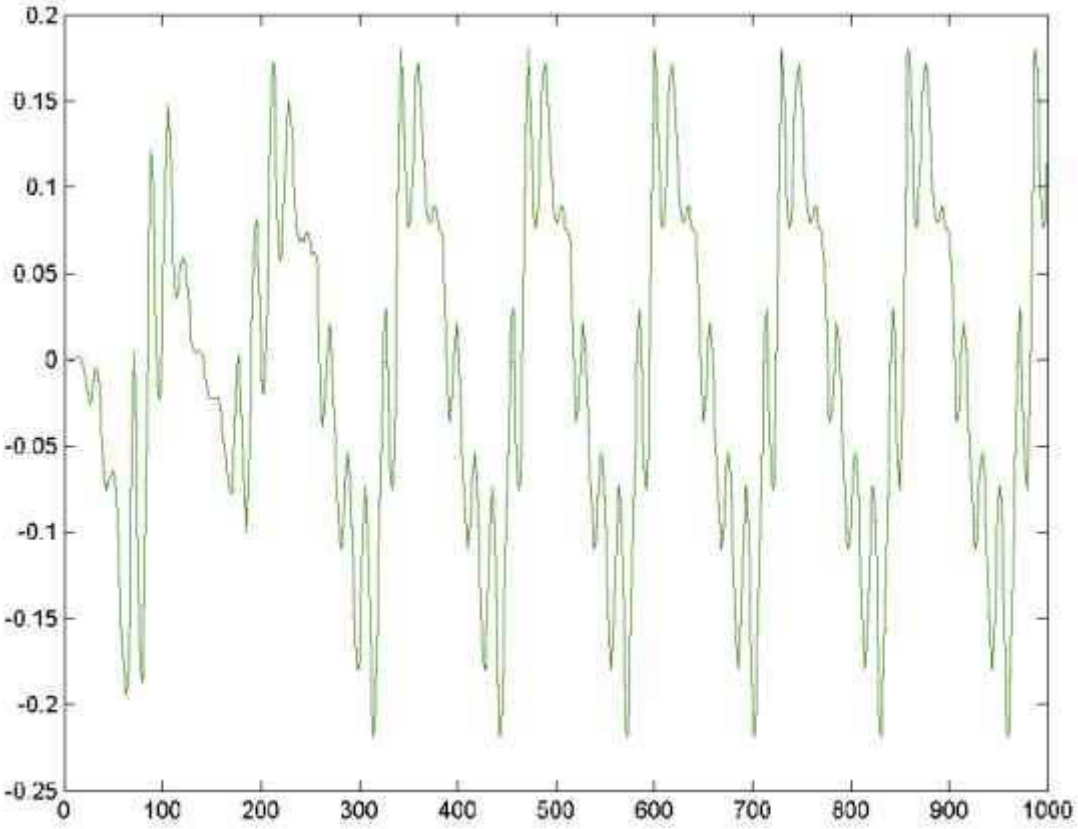
Per començar, es localitzen els dos primers períodes del senyal original (amb el nostre coneixement de la freqüència detectada per a la trama). Després apliquem una finestra de Hanning a aquests dos períodes i els copiem en els intervals de la nova freqüència desitjada. Això és molt similar a PSOLA, excepte que no es col·loquen marcadors del pitch a través del senyal original i no es localitzen els més propers a la sortida. En lloc, s'utilitzen els primers dos períodes de la trama i es van copiant, sota la suposició que cada trama serà molt semblant en tota la seva durada, ja que cobreix solament alguns mil·lisegons. Una vegada més el resultat és una forma d'ona de bon tros semblant a l'original (almenys en general), però amb la freqüència fonamental modificada.



La figura mostrada a sota ofereix una comparació visual d'aquests dos algorismes. El gràfic a l'esquerra és prop de dos períodes del senyal original, mentre que el gràfic a la dreta demostra el senyal de sortida durant el mateix interval del temps per a ambdós sistemes, el PSOLA (vermell) i l'algorisme de time-shift (blau). Mentre que ambdós algorismes produeixen sortides similars, l'algorisme de PSOLA s'assembla bastant més a la forma del senyal original. Una prova d'escolta subjectiva informal confirma que l'algorisme de PSOLA sona millor.







Trama de començament de bloc modificada mitjançant time_shifting

2.1.2.b) Modificació en freqüència:

I.- Introducció al 'Phase Vocoder:

El 'phase vocoder' és una eina establerta per a l'escalament del temps i el canvi de pitch dels senyals àudio via la modificació de la seva transformada de Fourier a curt termini (STFT). En contrast amb l'escalament temporal, el 'phase vocoder' es considera generalment que ofereix resultats d'alta qualitat, especialment per a factors grans de modificació i/o els senyals polifònics. No obstant això, el 'phase vocoder' també es caracteritza per introduir un artefacte perceptiu característic, descrit sovint com a "soroll de fase", "reverberació", o "pèrdua de presència". Aquí s'examinarà en profunditat el problema del "soroll de fase". S'introdueixen dues extensions a l'algorisme estàndard del 'phase vocoder', i la qualitat de so resulta millorada perceptiblement. Per altra banda, el 'phase vocoder' disminueix el cost de càlcul envers les tècniques de modificació en temps per un factor 2.

La majoria de les implementacions comercials d'escalament freqüencial en el domini del temps prenen parts del senyal i el van solapant cada cert període per aconseguir la freqüència desitjada. Aquests mètodes són atractius per al seu cost de còmput relativament baix i perquè donen bons resultats en alguns casos especials. No obstant això, aquestes tècniques tendeixen per a realitzar-se malament quan estan aplicades en senyals complexes o polifòniques, o quan es necessiten factors grans de modificació (factors majors del 20% o 30%). En aquests casos, apareixen uns sorolls típics tipus 'rumble' a la freqüència de destí. En contrast, les tècniques basades en el domini freqüencial tals com el 'phase vocoder' no es limiten als factors de modificació, i estan lliures de molts dels sorolls típics de les tècniques en domini temporal. No obstant això, el cost de càlcul del 'phase vocoder' és molt més alt que el de les tècniques en domini temporal i introdueix els seus el propis sorolls distintius. Son aquests artefactes que plantegen la barrera principal a un ús més extens del 'phase vocoder. El soroll, descrit anteriorment s'escolta com a una coloració característica del senyal que sona sovint com si l'altaveu estigués molt més lluny del micròfon que en l'enregistrament original. Aquest projecte proposa una explicació per a la presència del 'phasiness'.

II.- L'algorisme bàsic del 'Phase Vocoder:

L'essència d'aquesta tècnica es basa en tres fases, anàlisi del senyal, modificació, i síntesi del senyal, sempre treballant en domini freqüencial.

a) Fases d'anàlisi i síntesi:

Durant la fase d'anàlisi, el senyal es separat per trames de llargària constant cada t_a^u segons amb un overlap típicament del 50% o del 75%. Aleshores per a cada trama es calcula la seva transformada de Fourier, i resulta la representació següent del senyal:

$$X\left(t_a^u, \Omega_k\right) = \sum_{n=-\infty}^{\infty} h(n)x\left(t_a^u + n\right)e^{-j\Omega_k n} \quad (1)$$

A on x és el senyal original, h(n) és la finestra d'anàlisi, $\Omega_k = (\frac{2\Pi k}{N})$ és la freqüència central de la mostra k (canal k) i N és el número de mostres de la fft. A la pràctica h(n) també té un número limitat de mostres (típicament N mostres) i el sumatori (1) es converteix en una suma finita de termes. $X(t_a^u, \Omega_k)$ aleshores és tant una funció depenent del temps (variable u) com de la freqüència (variable Ω_k).

La fase de síntesi per la seva part a cada instant t_s^u (normalment uniforme i igual a t_a^u) recupera un senyal de curta durada i de nou en domini temporal, cadascuna de les quals serà aleshores multiplicada per una nova finestra opcional de síntesi w(n) , i que totes sumades resultaran el nou senyal modificat y(n).

$$y(n) = \sum_{u=-\infty}^{\infty} w\left(n - t_s^u\right) y_u\left(n - t_s^u\right) \quad (2)$$

A on :

$$y_u(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y\left(t_s^u, \Omega_k\right) e^{j\Omega_k n}.$$

b) Fase de modificació del senyal:

Quan tenim les trames superposades i transformades en freqüència, la seva modificació s'efectuarà canviant les fases de les mostres en funció d'una fórmula donada a continuació. Aquestes modificacions estan basades en un model de comportament sinusoïdal, encara que no es fa cap estimació sinusoïdal implícitament.
D'acord al model en el que es basa, el senyal d'entrada és una suma de un nombre I(t) de sinusoides amb amplituds variants amb el temps i freqüències instantànies $\omega_i(t)$,

$$x(t) = \sum_{i=1}^{I(t)} A_i(t) e^{j\phi_i(t)}$$

$$\phi_i(t) = \phi_i(0) + \int_0^t \omega_i(\tau)\, d\tau \quad (3)$$

En la qual $\phi_i(t)$ i $\omega_i(t)$ s'anomenen fase instantània i freqüència de la sinusoide iéssima.

Basant-nos en (3), per a una modificació constant de factor α tal que t_s^u = α·t_a^u, la fase $\phi_s(t_s^u)$ de la sinusoide modificada i seria:

$$\begin{aligned}\phi_s(t_s^u) &= \phi_s(0) + \int_0^{t_s^u} \omega_i(\tau/\alpha)\,d\tau \\ &= \phi_s(0) + \alpha\int_0^{t_a^u} \omega_i(\tau)\,d\tau \\ &= \phi_s(0) + \alpha\left[\phi_i(t_a^u) - \phi_i(0)\right]\end{aligned} \quad (4)$$

$\phi_s(0)$ = fase arbitraria inicial

El 'Phase vocoder' modifica la FFT per tal d'aconseguir les sinusoides a dalt expressades. L'evolució temporal de les amplituds de les sinusoides és modificada simplement aplicant la igualtat $\left|Y(t_s^u,\Omega_k)\right| = \left|X(t_a^u,\Omega_k)\right|$. La modificació de les fases resultarà més complicada.

Per calcular la fase de $Y(t_s^u,\Omega_k)$, la tècnica standard del 'phase vocoder' necessita fer l'anomenat 'phase unwrapping', un procés pel qual s'utilitzen trames consecutives per a calcular la freqüència instantània de cada sinusoide en cada canal. Aquesta freq. Instantània $\varpi_k(t_a^u)$ s'estima calculant primer l'increment de fase heterodí:

$$\Delta\Phi_k^u = \angle X\left(t_a^u,\Omega_k\right) - \angle X\left(t_a^{u-1},\Omega_k\right) - R_a\Omega_k$$

Agafant després el seu equivalent dins el marge [-π,+π] anomenat $\Delta_p\phi_k^u$ i derivant la freq. Instantània $\varpi_k(t_a^u)$ de la sinusoide més propera fent servir la següent formula:

$$\hat{\omega}_k\left(t_a^u\right) = \Omega_k + \frac{1}{R_a}\Delta_P\Phi_k^u. \quad (5)$$

Aquest procés, com s'ha comentat abans, s'anomena 'phase unwrapping'. L'increment de fase heterodí $\Delta\phi_k^u$ és doncs simplement el petit canvi de fase resultant de $\varpi_k(t_a^u)$, esdevenint proper però no necessàriament igual a Ω_k.

Un cop s'ha calculat la freq. Instantània en l'instant t_a^u, la fase de la FFT modificada en l'instant t_s^u es comportarà en relació a la següent fórmula de propagació de fase:

$$\angle Y\left(t_s^u, \Omega_k\right) = \angle Y\left(t_s^{u-1}, \Omega_k\right) + R_s \hat{\omega}_k\left(t_a^u\right). \quad (6)$$

L'equació (6) garanteix el que s'anomena 'coherència horitzontal de fase', es a dir, per a sinusoides de freq. constant hi haurà un encavalcament perfecte. Una altre manera de dir això es que es tindrà coherència dintre de cada canal en el transcurs del temps.
Per a sinusoides de freq. constant, la tècnica de 'phase unwrapping'(5) dona una bona estimació de la freq. instantània si el canal k està només influenciat per una sola sinusoide. A la pràctica però, per a finestres standard d'anàlisi (tipus hanning o haming), es necessari un overlap del 50% o del 75%.
Un punt important és que l'equació de propagació de fase (6) no indica com les fases de la FFT han de ser inicialitzades a l'instant t_a^0. Com es mostrarà més tard, l'elecció de la fase inicial pot influir significativament en el resultat final del senyal de sortida.
La tècnica explicada fins aquí conforma la més bàsica dins el mètode de 'phase vocoder'.Es proposen variants d'aquest mètode.

c) Problemes de fase amb la tècnica de modificació per freq. 'phase vocoder':

1) Coherència de fase:

Com els errors de propagació conformen el principal problema d'aquesta tècnica, és important entendre com les fases són modificades pel 'phase vocoder'.
Aquest algorisme assegura la consistència de fase dins de cada canal al llarg del temps, com abans s'ha esmentat, però també és important tenir consistència a través dels canals dintre de cada trama donada. Aquest requeriment s'anomena 'coherència vertical de fase' .
Ambdues coherències (horitzontal i vertical) cal que siguin preservades en el procés de síntesi per a que la fase esdevingui vàlida.
Si la coherència de fase no és preservada, l'equació de síntesi (2) formarà un senyal que no és proper a $Y(t_s^u, \Omega_k)$. Aquest nou senyal mostrarà aleshores cops deguts a les discontinuïtats de fase, i es produirà l'anomenat efecte de reverberació.
Com es pot reconèixer la coherència vertical de fase en una FFT?
Doncs, si tenim una sinusoide de freq. i amplitud constants i la finestra d'anàlisi és simètrica al voltant del cero, existeix una relació molt simple entre els canals adjacents de la transformada. Si una sinusoide f constant ω_i cau dins el canal k, i la finestra d'anàlisi h(n) és simètrica, és pot demostrar que els canals adjacents a aquest que es veuen influenciats per aquesta sinusoide tenen una fase d'anàlisi igual a la del canal k.

A la pràctica, la finestra d'anàlisi h(n) és normalment no nul-la per 0<n<L i és simètrica al voltant del seu punt mig, però aquests canvis canvia les coses només una mica: Si el tamany de la transformada de fourier és igual al tamany de la finestra h(n), aleshores canals adjacents mostraran diferències de fase de $\pm\pi$.
Per a senyals més complicats, desafortunadament, no existeix cap relació de comparació senzilla. Per a una sinusoide de freq. lentament variant les fases del canals adjacents es mantenen gairebé iguals, però resulta complicat desenvolupar una fórmula analítica, conseqüentment no es té una manera de calcular la coherència vertical.

2) Fase de sortida envers fase d'entrada:

En aquesta secció es tractarà de relacionar la fase de la FFT modificada en el canal k amb la fase de la FFT original en el mateix canal. Si s'assumeix una modificació constant de factor α, i donada una fase inicial de transformació $\alpha_s(o,k)$, es pot utilitzar l'equació de propagació de fase (6) per expressar la fase de la FFT de sortida en qualsevol instant donat y_s^u
Iterant amb (6) per a successius valor de u (començant a u=0) s'obté:

$$\begin{aligned}\angle Y(t_s^u,\Omega_k) &= \angle Y(t_s^0,\Omega_k) + \sum_{i=1}^{u} R_s\hat{\omega}_k(t_a^i) \\ &= \phi_s(0,k) + \sum_{i=1}^{u} R_s\hat{\omega}_k(t_a^i)\end{aligned}$$

A on $\varpi_k(t_a^i)$ es la freqüència instantània estimada en l'instant t_a^i dins el canal k . Ara fent us de (5) s'obté:

$$\angle Y(t_s^u,\Omega_k) = \phi_s(0,k) + \sum_{i=1}^{u}\left[R_s\Omega_k + \frac{R_s}{R_a}\Delta_p\Phi_k^i\right]$$

I usant la definició de $\Delta_p\phi_k^u$ s'obté:

$$\angle Y(t_s^u,\Omega_k) = \phi_s(0,k) + \alpha\sum_{i=1}^{u}\left[\angle X(t_a^i,\Omega_k) - \angle X(t_a^{i-1},\Omega_k) + 2m_k^i\pi\right]$$

A on m_k^i és el factor de 'unwrapping' en l'instant d'anàlisi $t_a^i : 2m_k^i\pi = \Delta_p\phi_k^i - \Delta\phi_k^i$. Això implica:

$$\angle Y\left(t_s^u, \Omega_k\right) = \phi_s(0,k) + \alpha\left[\angle X\left(t_a^u, \Omega_k\right) - \angle X(0, \Omega_k)\right] + \alpha \sum_{i=1}^{u} 2m_k^i \pi. \quad (8)$$

L'equació implica dona l'expressió de la fase de síntesi de la FFT a l'instant t_s^u, la fase inicial, el factor de modificació α, i les series de nombres de 'phase unwrapping' m_k^i.

Diverses conclusions es poden extreure d'aquesta equació:

- Indica que la fase de síntesi depèn de la fase d'anàlisi només en l'instant de temps d'anàlisi i a l'origen de temps. Això significa que si una fase d' anàlisi s'estima incorrectament en qualsevol instant, l'error no provocarà errors de fase en trames consecutives donat que el factor d'unwrapping roman correcte.
- D'altre banda, les series de factors de 'phase unwrapping' m_k^i compleixen un efecte acumulatiu: si un factor de unwrapping és calculat errròniament totes les trames següents mostraran un error acumulat.
- Els errors potencials de 'phase unwrapping' en múltiples de 2απ que són sumats a la fase de síntesi. Si α és un nombre enter aleshores els errors de 'phase unwrapping' seran transparents degut a que seran múltiples de 2π. Com a resultat, operacions de transformació amb factors enters es poden efectuar sense 'phase unwrapping' només utilitzant (8), a on el factor $\sum_{i=1}^{u} 2m_k^i \pi$ es desprecia, reduint així significativament el cost computacional de les operacions.

L'equació (8) també proporciona un fundament sòlid essencial per a veure la mancança de coherència vertical de fase en les implementacions standard dels 'phase vocoder'.

3) Pèrdua de coherència vertical de fase:

D'acord amb (8), la coherència de fase vertical depèn de dos factors: el valor de fase inicial i els errors d'acumulació de 'phase engualdrapin.
Per veure això es suposa per començar un factor de modificació α enter, amb el qual els errors de 'phase unwrapping' no influiran en el senyal modificat. Ara es considera una sinusoide que la seva freq. instantània varia amb el temps de tal manera que pot passar d'un canal a un altre. Re-agrupant els termes en (8), la fase de síntesi del canal amb el pic de la FFT es pot expressar com:

$$Y\left(t_s^u, \Omega_k\right) = \alpha \angle X\left(t_a^u, \Omega_k\right) + \theta \qquad \text{h}$$
$$\theta_k = \phi_s(0,k) - \alpha \angle X(0, \Omega_k) \qquad (9)$$

a on la suma dels factors d'unwrapping ha estat despreciada, ja que és múltiple de 2π. Aquesta expressió difereix de la fase ideal de la equació (4) en que ϕ_k no és necessàriament una constant, i que varia en funció del canal k. El fet de que ϕ_k pugui no ser una constant comporta dos conseqüències adverses:

1) Les fases de síntesi en canals adjacents poden ser molt diferents, si els valors de ϕ_k varien considerablement d'un canal a l'altre.
2) Quan la freq. instantània de la sinusoide migra del canal k_0 en l'instant t_a^u al canal $k_0 + 1$ en l'instant de temps t_a^{u+1}, la fase de síntesi experimenta un salt de igual a $\phi_{k_0+1} - \phi_{k_0}$, que resulta sorollós.

Com s'ha mostrat en (9), els valors de ϕ_k per a canals successius només depenen de les fases d'anàlisi i síntesi en l'instant inicial. Si per exemple una àrea de l'espectre està dominada per un soroll en un moment donat, els valors de $LX(0,\Omega_k)$ i conseqüentment per a ϕ_k seran aleatoris per als corresponents canals, y probablement mostraran grans variacions de canal a canal, excepte si s'estableix la fase inicial $\phi_s(0,k)$ tal que:

$$\theta_k = \phi_s(0,k) - \alpha \angle X(0, \Omega_k) = C \qquad (10)$$

a on C és la constant depenent del canal. Si aquesta equació és satisfà aleshores cap del dos problemes mencionats abans ocurriran, i la fase de síntesi esdevindrà idèntica a la fase de síntesi ideal en (4).

Ara es considera el cas més general en el qual el factor de modificació no sigui enter, i de nou es suposa que hi ha una component sinusoïdal en la proximitat del canal k_0.

L'equació (8) indica que, sempre que (10) es compleixi, la coherència de fase està garantitzada només si les sumes dels factor d'unwrapping $\sum_{i=1}^{u} 2m_k^i \pi$ són iguals (mòdul 2 π) en els canals propers. No hi ha perill d'errors de 'phase unwrapping' en els canals propers al de la freq. instantània de la sinusoide degut a que l'overlap entre mostres es prou gran. El problema més gran és que segurament cap component sinusoïdal d'un so romandrà sense talls durant tota la seva durada. En conseqüència inevitablement hi haurà moment en els que un canal afectarà els seus canals veïns. És durant aquest instants que les diferències de 'phase unwrapping' s'acumularan necessàriament, fent que els termes $\sum_{i=1}^{u} 2m_k^i \pi$ siguin diferents per canals diferents; com a resultat, la coherència vertical es perdrà definitivament i ja no es tornarà a recuperar.
Considerant aquest punts anteriors, sembla increïble que el 'phase vocoder' pugui funcionar…
Per a factors de modificació enters, es pot gairebé assegurar que la coherència vertical es perdrà excepte si cada component sinusoïdal roman en el mateix canal tot el temps. Aquesta premissa és violada clarament per la majoria de senyals d'interès, incloent veu i música.

III.- Antigues i noves estratègies per a corregir el soroll de fase:

A) Reconstrucció d'amplitud:

El 'phase vocoder' no és l'única tècnica en el domini freqüencial que pot ser aplicat per al problema del canvi de pitch. El 'phase unwrapping' pot ser obviat completament mitjançant dos criteris: reconstruir el senyal mitjançant les amplituds de la FFT o suposant la FFT com a una explícita suma de sinusoides.
Els algorismes per reconstrucció d'amplitud de la FFT requereixen un gran número d'iteracions del cicle d'anàlisi i síntesi per aconseguir una FFT de sortida consistent. Això fa que aquesta tècnica esdevingui molt cara en termes de càlcul. A més, encara que la convergència d'aquests mètodes ha estat provada, no es pot assegurar l'assoliment d'un mínim global. Més recentment s'ha descobert que mitjançant un càlcul acurat de les fases inicials es pot arribar a reduir considerablement el nombre de iteracions. Tot i això encara continua essent una tècnica molt més costosa que el 'phase vocoder'.
L'altra alternativa al 'phase vocoder' en domini freqüencial és l'anomenat el model sinusoïdal. Aquesta aproximació és també més costosa que el 'phase vocoder' i a més introdueix els seus propis artefactes.

B) Loose phase locking:

La limitació fonamental (i el principal atractiu) de l'esquema d'aquesta tècnica és que evita qualsevol determinació explícita de l'estructura del senyal: es fa el mateix procés per a cada canal independentment del seu contingut. El resultat és que les fases de síntesi en els canals al voltant d'una freq. donada només aproximen gradualment la coherència vertical de fase. Si realment es vol reconstruir la coherència vertical s'haurà de prendre una mesura propera a un model de suma de sinusoides. Es presenten dues versions d'una nova tècnica de 'phase locking', inspirada en la tècnica explicada de 'Loose phase locking' però basada en la identificació explícita dels pics (i per tant presumiblement sinusoides) de l'espectre. Aquesta nova tècnica comença per trobar els màxims locals entre canals. En la implementació més simple un canal que la seva amplitud sigui superior a la dels seus quatre veïns més propers es considera un pic.; aquest criteri és simple i costosament efectiu, encara que una mica primitiu. Aleshores les sèries de pics divideixen l'espectre en regions d'influència localitzades al voltant de cada pic. La idea basicament és actualitzar les fases només pels canals dels pics d'acord a la teoria standard de propagació de fase (6). Les fases de la resta de canals seran aleshores lligades en certa manera a la fase del pic de la seva regió d'influència. Una manera de escollir els límits de les regions d'influència és escollir la freq. intermitja entre dos pics. Així doncs, el límit superior de la regió d'influència del pic Ω_{k1} serà $\frac{\Omega_{k1}+\Omega_{k2}}{2}$.
Un altre criteri raonable seria escollir com a límit el canal amb amplitud mínima entre dos pics.

Mètode 1 - Identity Phase Locking:

Millor que imposar una restricció d'igualtat de fase (basant-se en la hipòtesi que cada pic representa una sinusoide és d'amplitud i freqüència constant), es poden restringir les fases de síntesi de cada regió de influència en funció de la del seu pic corresponent: les diferències de fase entre canals successius d'una mateixa àrea d'influència són idèntiques en la FFT de sortida a les de la FFT d'anàlisi.

Si Ω_{kl} és la freq. central del pic dominant aleshores tenim (12) per a tots els canals k de la regió d'influència :

$$\angle Y\left(t_s^u, \Omega_k\right) = \angle Y\left(t_s^u, \Omega_{k_l}\right) + \angle X\left(t_a^u, \Omega_k\right) - \angle X\left(t_a^u, \Omega_{k_l}\right) \quad (12)$$

Aquest mecanisme millora significativament la consistència de les sèries resultants de la FFT i redueix importantment el soroll de fase del senyal modificat. Aquesta tècnica també te dos avantatges computacionals importants.

Primer, degut a que el 'phase unwrapping' només s'aplica en els canals de pic, la freq. instantània de la sinusoide associada serà segurament la freq. central del canal en qüestió. Això significa que la restricció del 'phase unwrapping' $R_a \omega_h < \pi$ pot no ser tant estricta, i es poden utilitzar valors més petits d'overlap. A la pràctica un overlap d'entrada del 50% és possible sense generar errors de fase, amb el qual es pot reduir el cost de càlcul a la meitat respecte el 'phase vocoder', el qual necessita un 75% d'overlap.

Segon, aquesta nova tècnica requereix càlculs trigonomètrics només per els canals de pic: una vegada s'ha determinat la fase de síntesi del canal de pic es pot calcular l'angle θ
Necessari per rotar $X\left(t_a^u, \Omega_{kt}\right)$ en $Y\left(t_s^u, \Omega_{kt}\right)$ de la manera següent:

$$\theta = \angle Y\left(t_s^u, \Omega_{k_t}\right) - \angle X\left(t_a^u, \Omega_{k_t}\right) \qquad (13)$$

I calculant aleshores el fasor $Z = e^{j\vartheta}$ i obtenir els canals veïns utilitzant la següent simple regla d'àlgebra:

$$Y\left(t_s^u, \Omega_k\right) = ZX\left(t_a^u, \Omega_k\right) \qquad (14)$$

Amb la qual es pot fàcilment satisfer l'equació (12) amb només una multiplicació complexa.
A continuació es presenta l'esquema de l'Identity Phase Locking per punts:

1. per cada trama FFT, localitzar els pics
2. per cada pic, calcular la freq. instantània fent us de la fórmula de coherència horitzontal de fase, i calcular la fase actualitzada de síntesi
3. Calcular la fase de rotació θ amb l'equació (13) i calcular el fasor $Z = e^{j\vartheta}$
4. aplicar la rotació resultant a tots els canals de l'àrea d'influència de cada pic segons (14)
5. repetir els punts anteriors per el següent pic, fins que tots els pics hagin estat processats
6. procedir amb la següent trama

Mètode 2 – Scaled phase locking:

Una millora de la tècnica precedent esdevé del reconeixement de que si un pic canvia del canal k_0 en la trama u -1 al canal k_1 en la trama u l'equació d'unwrapping (5) estaria basada en $LX\left(t_a^u,\Omega_{k1}\right)-LX\left(t_a^{u-1},\Omega_{k0}\right)$ enlloc de $LX\left(t_a^u,\Omega_{k1}\right)-LX\left(t_a^{u-1},\Omega_{k1}\right)$. Aleshores l'equació de propagació de fase (6) seria:

$$\angle Y\left(t_s^u,\Omega_{k_1}\right)=\angle Y\left(t_s^{u-1},\Omega_{k_0}\right)+R_s\hat{\omega}_{k_1}\left(t_a^u\right) \qquad (15)$$

A on l'increment de fase $R_s\varpi_{k1}\left(t_a^u\right)$ s'acumula en $LY\left(t_s^{u-1},\Omega_{k0}\right)$ en lloc de en $LY\left(t_s^{u-1},\Omega_{k1}\right)$. Es fàcil demostrar que en aquest cas la fase de síntesi en el canal de pic corresponent a la sinusoide i en l'instant t_a^u es $C+\alpha\phi_i\left(t_a^u\right)$, el qual no és necessàriament el cas en la tècnica precedent.

El problema aleshores és determinar quin pic a la trama u-1 es correspon al pic Ω_{k1} en la trama u. Una manera senzilla de fer això es agafar el pic de la regió a la qual el canal Ω_{k1} pertany en la trama u-1. En conseqüència, per calcular la fase de síntesi del canal k_1 en la trama u, simplement cal buscar el pic dominant en la regió del canal k_1 en la trama u-1, i utilitzar les seves fases d'anàlisi i síntesi al aplicar (5) i (6).
Fent això la resta de canals de la zona d'influència es poden sincronitzar amb la de pic i l'equació de Identity Phase Locking pot ser generalitzada com a:

$$\angle Y\left(t_s^u,\Omega_k\right)=\angle Y\left(t_s^u,\Omega_{k_i}\right)+\beta\left[\angle X\left(t_a^u,\Omega_k\right)-\angle X\left(t_a^u,\Omega_{k_i}\right)\right] \qquad (16)$$

A on β es el factor d'escalat de fase. La Identity Phase Locking es dóna quan β =1.
La Identity Phase Locking pot ser altament millorada establint β amb un valor entre 1 i α. Quan s'utilitzen factors de modificació enters en la implementació standard del phase vocoder, i si es fa ús de la inicialització de (11), és fàcil verificar que les diferències de fase estan també escalades per $\beta=\alpha$. A mes, proves d'escolta han determinat que establint $\beta=2/3+\alpha/3$ es pot reduir considerablement el soroll de fase. Cal remarcar que les fases $LH\left(t_a^u,\Omega_k\right)$ han de ser desfetes (unwrap) a través dels canals k al voltant del canal de pic abans d'aplicar (16), per tal d'evitar salts de canal $2\beta\pi$ en les fases de síntesi.

En contrast a la tècnica anterior, aquesta no permet calcular els valors del canals veïns de la FFT mitjançant una multiplicació complexa, en conseqüència aquesta implementació esdevé amb un major cost computacional, però d'altre banda, la qualitat resultant amb aquesta tècnica és molt millor que amb l'anterior.

A continuació es presenta l'esquema del scaled Phase Locking per punts:

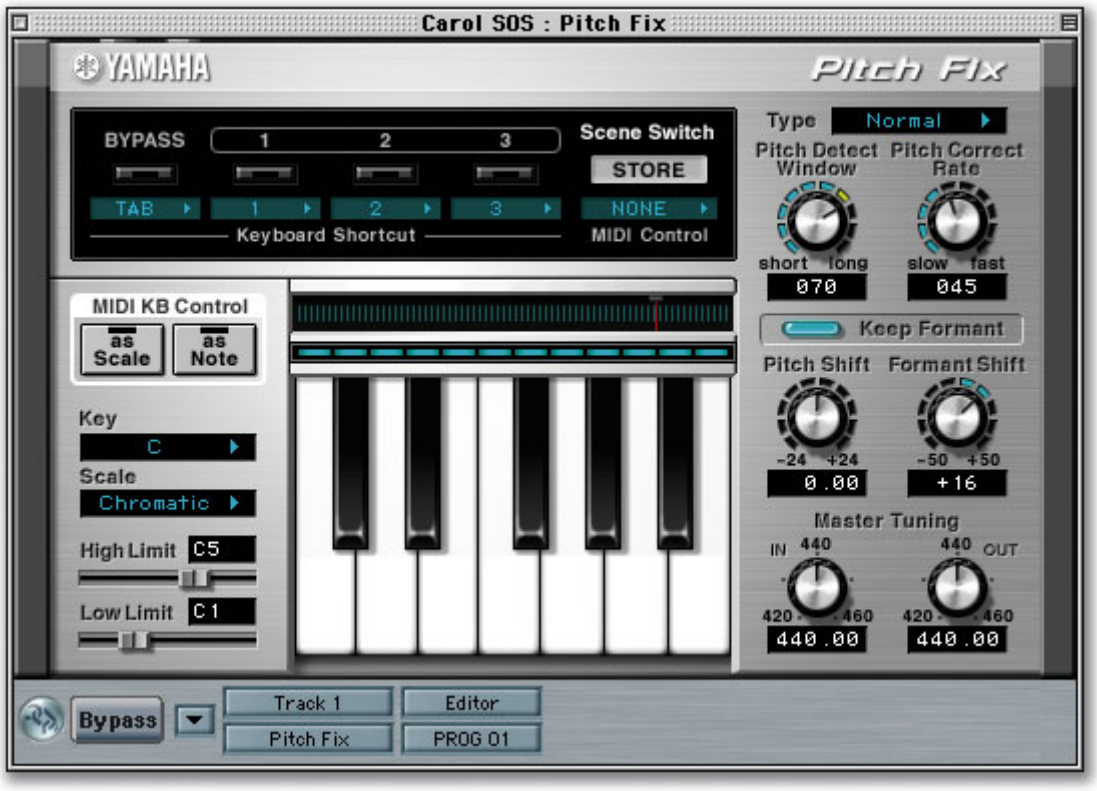
1. per cada trama FFT, localitzar els pics
2. per cada pic, calcular la freq. instantània fent us de la fórmula de coherència horitzontal de fase, i calcular la fase actualitzada de síntesi segons (15)
3. desfer les fases de tots els canals en la regió d'influència.
4. per a cada canal al voltant del canal de pic, calcular la diferència de fase d'anàlisi entre el pic i el canal en procés, i calcular la fase de síntesi utilitzant (16)
5. repetir els punts anteriors per el següent pic, fins que tots els pics hagin estat processats
6. procedir amb la següent trama

2.2.- Estat de l'art del projecte:

En els punts anteriors, ja hem parlat de l'estat de l'art dels mòduls que composen aquest projecte per separat, ara tractarem l'estat de l'art de tot el projecte en conjunt.
Hi ha en l'actualitat diverses empreses que es dediquen a comercialitzar aquest sistema de correcció/afinació de veu sota diferents noms i en formats diferents, totes però incorporen prestacions semblants. Tots presenten un selector de notes musicals o d'escales musicals, uns selectors d'atac o velocitat d'afinació, i de release o velocitat de finalització d'actuació, un regulador de detune o afinació general, i un selector per indicar quin tipus de senyal s'introdueix. Aquest últim punt és recomanable en la mesura que ajuda al programa a fer la recerca freqüencial, així si li especifiquem que el senyal d'entrada es correspon a un senyal de home greu, l'anàlisi es centrarà en una banda que serà diferent de si l'indiquem que el senyal serà d'una dona o d'un nen.
Cal comentar que aquest afinadors també serveixen per afinar qualsevol tipus d'instrument harmònic tipus trompeta, violí etc…
En quant als formats en que es presenta aquest sistema podem destacar els dos més importants i gairebé únics en el mercat, que son en format plug-in., ja sigui mitjançant tecnologia VST, directX, RTA (pro tools)…, com en format rack.

Tots dos formats solen implementar el sistema a temps real.
A continuació es presenten unes imatges dels diferents productes més coneguts del mercat en ambdós formats.


antares auto-tune 4
01
Auto
Graphical
CORRECTION MODE
SELECT PITCH REF.
OPTIONS
ANTARES
Input Type
192 kHz
Soprano Voice
PLAYBACK
Track Pitch
Correct Pitch
SCALE VIBRATO
AMOUNT
Make Vibrato Curve
RETUNE
Slow
Fast
TRACKING
Relaxed
Choosy
Undo
Redo
Cut
Copy
Paste
Select All
Make Curve
Make Auto
Cursor Pitch
C#3 +22
All
Tie


Carol SOS : Pitch Fix
YAMAHA
Pitch Fix
BYPASS
1
2
3
Scene Switch
STORE
TAB
NONE
Keyboard Shortcut
MIDI Control
Type
Normal
Pitch Detect Window
Pitch Correct Rate
short
long
slow
fast
070
045
MIDI KB Control
as Scale
as Note
Keep Formant
Key
C
Scale
Chromatic
Pitch Shift
Formant Shift
-24
+24
-50
+50
0.00
+16
High Limit
C5
Low Limit
C1
Master Tuning
IN
OUT
440
420
460
440.00
Bypass
Track 1
Pitch Fix
Editor
PROG 01


ANTARES


POWER
VOICEONE
VOICE PITCH AND MODELING TOOL
TC HELICON
LEVELS
INPUT
OUTPUT
MULTI SPECTRAL LCD DISPLAY
23 MEGA Voice
PITCH
CORRECT
SHIFT
VOICE MODELING
INFLECTION
SPECTRAL
GROWL
VIBRATO
BREATH
RESONANCE
FUNCTIONS
I/O
RECALL
MIDI/UTIL
STORE
CONTROL
ENTER
BYPASS

3.- Descripció tècnica del projecte:

3.1.- Tècniques escollides i paràmetres:

Per a la realització d'aquest projecte s'han tingut en compte totes les tècniques descrites en el capítol anterior de l'estat de l'art, per tal d'implementar la que millors resultats ofereixi. Així doncs s'ha efectuat un procés de comparació subjectiva dels diferents algorismes fins arribar a una decisió final, que no sempre ha esta una tècnica concreta sinó també una barreja de diferents tècniques, com més endavant es detallarà.

3.1.1.- Detecció de pitch

Per a implementar el detector de pitch es van provar diferents tècniques en funció del seu cost computacional i del seu resultat, i es van avaluar subjectivament mitjançant una aplicació que, introduint-li el senyal a analitzar, executava la implementació desitjada de detector de pitch i els resultats de freq. de pitch de cada trama els convertia en un to de la durada de la trama, i finalment enfinestrava totes les trames i les superposava per generar una melodia tonal que, si el detector de pitch funcionava correctament, sonava igual que el senyal d'àudio introduït però amb tons. D'aquesta manera es podia avaluar d'una bona manera una magnitud perceptual com és el pitch. També s'utilitzava la tècnica de test consistent en introduir tons en el detector de freq. conegudes per comprovar la seva fiabilitat d'una manera més rigorosa. Es va doncs començar implementant els sistemes més senzills, que són els temporals. Primer es va implementar un detector de pitch per recompte de màxims i mínims, però aquest resultava molt poc fiable amb senyals rics en harmònics. Més tard es va implementar un altre detector de pitch per recompte de creuaments per zero, i encara que va millorar una mica el seu comportament, amb senyals complexes es perdia constantment. Aleshores, abans de passar a les tècniques en freqüència, es va implementar una tècnica personal basada en la de recerca de creuaments per zero, però que utilitzava la transformada FFT. Es va anomenar 'tècnica espai-freqüencial'.

Tècnica espai-freqüencial:

Aquesta nova tècnica personal bàsicament consisteix en buscar un màxim en l'espectre de la transformada que contingui la freq. de pitch del senyal mitjançant una de les possibles tècniques existent per acotar la decisió de la freq. de pitch en un marge freqüencial petit (10~100Hz), corresponent a l'increment de Hz per mostra de la FFT. La manera més senzilla seria cercar la mostra màxima del mòdul de la FFT o també es podria agafar el primer pic, però són estratègies poc robustes que en pocs casos garanteixen una bona elecció de mostra o canal freqüencial. Altres mètodes seran discutits més endavant. Una vegada es té una decisió del canal que conté la freq. de pitch i es té aquesta freq. de pitch acotada entre dos valors relativament propers es pot també acotar el període mínim i màxim de la trama en número de mostres.
Aleshores es passa de nou al domini temporal i es busca el primer pas per zero. Quan es localitza la mostra amb el primer pas per zero, enlloc de buscar el següent a partir de la següent mostra, es fa un salt en el marge de búsqueda de tamany igual al número de mostres corresponent al període mínim de la hipòtesi feta en el domini freqüencial., i es busca el proper pas per zero a partir d'aquella mostra. Quan es troba aquest següent creuament per zero es calcula l'increment entre aquesta mostra i la corresponent a la de l'anterior creuament per zero i s'emmagatzema. Aquest procés es repeteix fins el final de la trama obtenint com a resultat un conjunt de valors corresponents a possibles períodes tots dins el marge d'acotació establert. Aleshores s'analitzen aquests valors per veure si hi ha cap valor que difereix molt de la resta, i si és el cas s'elimina. Amb la resta de valors es fa una mitja aritmètica de tots i s'obté la freq. de pitch desitjada.
Cal comentar que aquesta tècnica es pot implementar també amb cerca de màxims i mínims, o amb una barreja de màxims, mínims i creuaments per zero, obtenint així més valors per decidir i fer la mitja de la que resultarà la freq. de pitch.
El procés d'aquesta tècnica es detalla per punts a continuació.

1. Buscar el canal de l'espectre que conté la freq. de pitch (procés no determinat)
2. Calcular el període mínim i màxim de pitch, en mostres, en funció del marge freqüencial que englobi el canal freq. de pitch trobat en (1).
3. Buscar el primer event (creuament per zero, màxim o mínim, en funció de la decisió d'anàlisi) en la trama temporal.
4. Fer un salt en la cerca igual al període mínim trobat en (2) en mostres des de la mostra a on hem trobat el primer event destacable en (3).
5. Tornar a buscar el mateix event que s'estigui analitzant (ZRC, màxim…) a partir de la mostra resultant de (4).
6. Calcular la diferència entre la mostra d'event trobada en (5) i la trobada en (3) i emmagatzemar-la.

7. Repetir els punts 4,5 i 6 fins exhaurir les mostres de la trama temporal, però en el punt 5 fent referència al event immediatament anterior i no al primer.
8. Analitzar tots els valors resultants, i despreciar aquells que difereixin en més d'un marge de tolerància α de la mitjana aritmètica de tots ells.
9. Amb els valors resultants, tornar a fer la mitja aritmètica.
10. En cas d'analitzar més d'un tipus d'event, repetir els punts 1 a 9 per a cada event i finalment fer la mitja de tots els valors resultants.
11. Repetir els punts 1 a 10 per a les següents trames fins a exhaurir totes les trames del senyal.

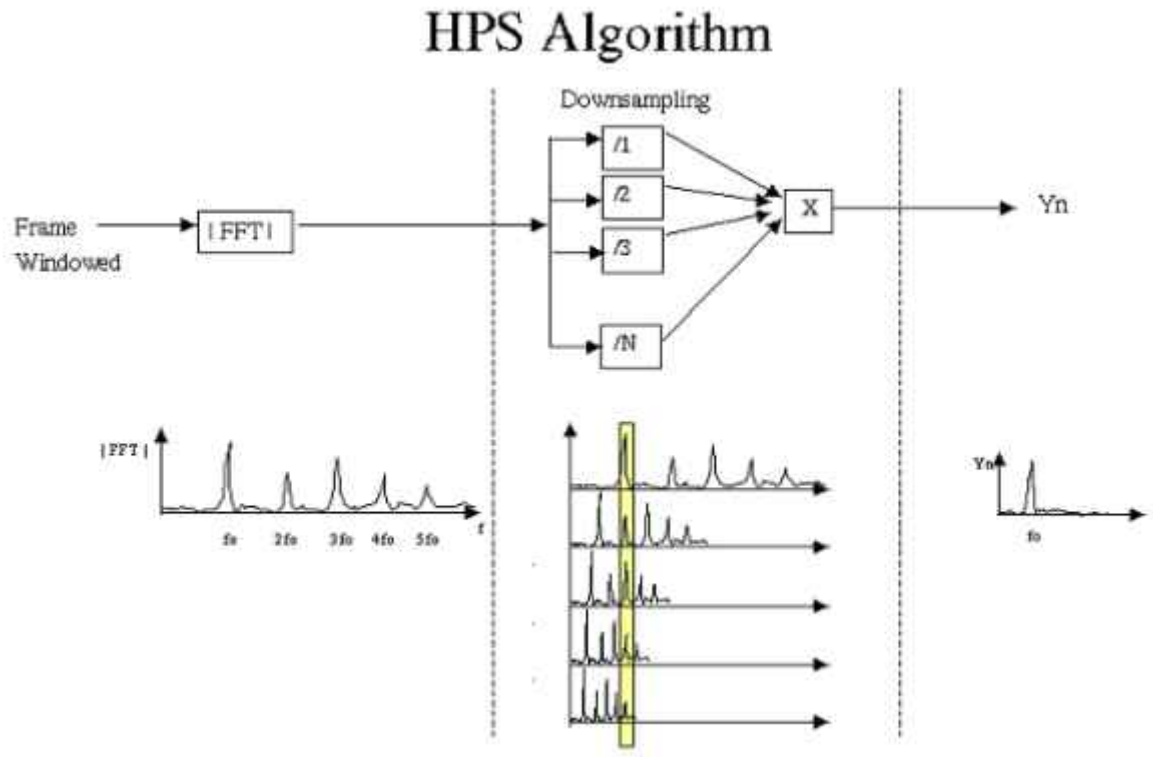
Aquesta tècnica resulta bastant fiable inclòs en senyals sorolloses, ja que es va aproximant progressivament i el marge d'error es fa molt petit. Cal remarcar que la fiabilitat d'aquest mètode va en funció de la decisió de canal de pitch que es faci en la transformada, decisió la qual aquest mètode no estableix, deixant-la a elecció personal del lector segons el tipus de senyal que es vulgui analitzar.
Aquesta tècnica finalment no ha esta escollida per a desenvolupar aquest projecte ja que no es va trobar una tècnica prou fiable de cerca del canal de pitch en la transformada, i queda pendent d'estudi amb caràcter de superació personal de l'autor d'aquesta tècnica i d'aquest projecte.

- Finalment es va continuar implementant les tècniques d'anàlisi del pitch en l'entorn freqüencial que ha esdevingut el tipus de tècnica utilitzat per aquest bloc en aquest projecte.

Per a poder calcular el pitch d'un senyal en el domini freqüencial es necessari trobar amb exactitud quin és el canal de la FFT (un canal s'entén com a una mostra de la FFT) que transporta la freq. d'aquest pitch, per tant en aquest punt es va començar a buscar el mètode més òptim per a desenvolupar aquesta tasca.
Es va començar per la tècnica del recompte de màxims freqüencials explicat en l'apartat de l'estat de l'art, però va resultar poc fiable ja que en molts casos la decisió del canal de freq. de pitch era errònia ja que en molts senyals complicats el mòdul del canal que transporta la freq. de pitch és d'amplitud molt petit, en contra del que es pugui suposar, donant així un valor resultant enganyós.
Finalment es va implementar un sistema de cerca de canal de pitch que s'anomena 'producte dels harmònics de l'espectre' o HPS.

Tècnica HPS:

Com a l'entrada tenim un senyal harmònic, el mòdul del seu espectre consisteix en un conjunt de pics espaiats corresponents a la freq. de pitch i als seus harmònics i resta de soroll existent en el senyal. Si es comprimeix l'espectre un nombre enter de vegades i es compara amb l'espectre original, aleshores es podrà veure quins d'aquests pics corresponen a harmònics de la freq. de pitch i quin és podrà analitzar amb més criteri quin és el canal que transporta la freq. de pitch.
Si aquest procés es repeteix per diferents factors de compressió es podrà determinar quin és aquest canal amb una bona fiabilitat com es pot observar en la següent figura.



Quan un pic en l'espectre original coincideix amb un segon de l'espectre comprimit en un factor 2, si a més coincideix amb un tercer pic de l'espectre comprimit per un factor 3, i així consecutivament, es pot veure clarament, que aquest pic de l'espectre original és el que transporta la freq. de pitch, ja que la resta de pics són múltiples d'aquest per definició del mètode.
El mètode simplement consisteix en multiplicar els espectres comprimits pels diferents factors amb l'espectre original consecutivament i mostra a mostra com es veu en la figura, i de l'espectre resultant buscar el seu màxim, la posició del qual correspondrà al canal que transporta la freq. de pitch.

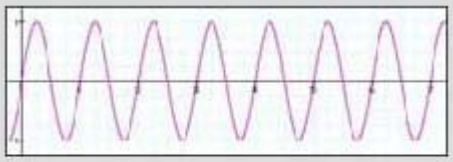
El procés es mostra detallat per punts a continuació:

1. Fer la FFT de la trama i calculem el seu mòdul
2. El mòdul de la FFT resultant en (1) es sotmet a un re mostreig en el qual es desprecia una de cada dos mostres (FACTOR 2), s'agrupa tot el resultat consecutiu al començament d'un vector de llargària igual a la de l'espectre en mostres, i aquest vector es farceix de zeros des del final dels resultats fins al final (segona meitat).
3. Multiplicar el resultat de (2) amb l'espectre original i s'emmagatzema el resultat.
4. Tornar a repetir el punt (2) però amb un factor superior de compressió i de manera anàloga.
5. Multiplicar el resultat de (3) amb el resultat anterior i emmagatzemar.
6. Repetir els passos 4 i 5 fins a completar les operacions amb tots els factor desitjats.
7. Buscar la mostra màxima del resultat obtingut en (6), que serà la mostra que transporta la freq. de pitch en l'espectre original.

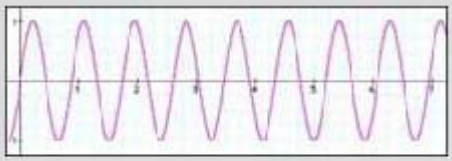
Aquest mètode és molt robust enfront el soroll i els sons polifònics, i es pot adaptar al tipus de so que es vulgui analitzar mitjançant una bona elecció en quant a quins factors de compressió s'utilitzaran i quins no, i és el mètode de cerca de canal de pitch que s'ha implementat per a la realització d'aquest projecte.

- Un cop s'ha pres la decisió del canal de pitch expressada anteriorment, s'ha d aplicar un nou algorisme per tal de calcular quina és la freq. instantània exacta del canal de pitch de la trama en aquell instant. Per a complir amb aquesta tasca es podria haver aplicat perfectament la 'tècnica espaio-freqüencial' expressada anteriorment i de desenvolupament personal, ja que en aquest punt ja es disposa d'una bona decisió del canal de pitch de la transformada. Finalment però es va provar una altre tècnica basada en el 'phase vocoder', però aplicada només a un canal (el de pitch).
Aquesta tècnica doncs aplicant el 'phase vocoder' al canal de pitch tracta de trobar la freq. instantània del canal (pitch) mitjançant l'increment entre la fase del canal actual i la del mateix canal en la trama anterior. Aquesta tècnica suposa que el senyal està format per un model de sumes sinusoïdals de tants termes com punts tingui la transformada. Aleshores dins cada canal k viatja una sinusoide que pot tenir qualsevol freq. dins el marge $\left[(k-1)\cdot fm/nfft , k\cdot fm/nfft \right]$, a on fm indica la freq. de mostreig del sistema, i nfft el nombre de punts de la transformada.

Com es pot veure en els següents gràfics, es presenten dos senyals que no tenen exactament la mateixa freqüència, el senyal del gràfic 3.2 té una freq. major que el de la gràfica 3.1

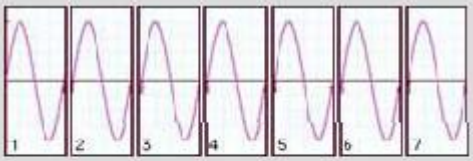


Graph 3.1: Let this be the waveform of a measured signal with a frequency that is exactly that of a bin (click to enlarge)

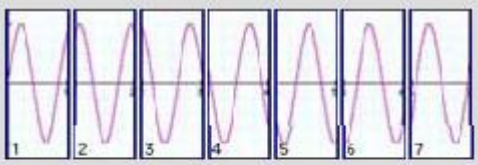


Graph 3.2: Let this be the waveform of a measured signal with a frequency that is not on a bin frequency (click to enlarge)

Si prenem instants curts per a analitzar el senyal:



Graph 3.3: Our signal of 3.1 now split into 7 frames. Each frame will be passed to our transform and be analyzed (click to enlarge)



Graph 3.4: Our signal of 3.2 now split into 7 frames. Each frame will be passed to our transform and be analyzed. We see that while the signal in 3.3 nicely fits into the frames, the second signal appears to have a *phase shift* in each window (click to enlarge).

Com es pot observar, al gràfic 3.4 la sinusoide té una fase diferent en cada trama mentre que en el gràfic 3.3 sempre comença amb la mateixa fase. L'increment de fase entre dos trames ens dona la desviació en freq respecte a la freq. més propera de la FFT, aleshores mitjançant aquest canvi de fase podem calcular la freq. real de la sinusoide. Matemàticament, la derivada diu quant ha canviat l'ultima mesura, aleshores existeix una relació de proporcionalitat entre la derivada de la fase de la mostra de la FFT y la separació de la freq. real respecte la freq. central de la mostra de la FFT.
Per a poder discriminar be la derivada de la fase sense ambigüitat és necessari fer solapament entre trames. El factor típic de solapament és al menys de factor 4, es a dir, el solapament entre dos trames adjacents ha de ser de com a mínim un 75%. Si escollim un factor més gran el marge en el que pot variar la freq. verdadera es més gran.

Tenint coneixement previ del canal que transporta el pitch de la trama transformada, el procés s'esquematitza per punts a continuació:

1. Extreure la fase de la mostra de la FFT que correspon al canal de pitch en la trama actual
2. Extreure la fase de la mostra de la FFT que correspon al canal de pitch en la trama anterior
3. Calcular l'increment de fase entre (2) i (1).
4. Multiplicar (3) pel factor d'overlap per a conèixer quin seria l'increment si no es tingués encavalcament de trames (si l'encavalcament és del 75% el factor d'overlap serà doncs 4 i anàlogament)
5. Dividir el resultat de (4) per 2π, per a obtenir el nombre de voltes complertes (conversió d'increment de fase a increment de rotacions complertes)
6. Extreure de (5) la part entera, es a dir el nombre enter de voltes completades de fase i que resti només un increment decimal comprés entre 0 i 1 corresponent a una fase entre 0 i 2 π.
7. Multiplicar el resultat de (6) per l'increment corresponent per mostra de la FFT descrit per $fm/nfft$. Aquest resultat correspondrà a l'increment de freq. respecte la freq. del canal de pitch.
8. Prendre la freq. corresponent al canal de pitch k donada per $(k-1)\cdot fm/nfft$ i: sumar-li el resultat de (7) si aquest és menor de 0.5, o restar-li si és menor. Aquest criteri es pren degut a que la freq. que es descriu com la de canal es correspon amb la freq. central del mateix i per tant, en funció de si és superior o inferior de 0.5 s'haurà de sumar o restar l'increment de freq. respecte aquesta freq. central. El resultat d'aquest punt conforma la freq. de pitch de la trama actual.
9. Repetir els punts 1 a 8 per a totes les trames.

Cal remarcar que aquest sistema té el desavantatge de que si la freq. de la sinusoide d'un canal no és correspon exactament amb la freq. central o freq. de canal, aquesta s'expandeix per tot l'espectre, influint en la mesura de la resta. Si això es generalitza a que gairebé tots els canal estan fora de la seva freq. central de canal, la situació serà que cada es veurà influenciat per la resta de canals. De totes maneres les proves subjectives d'oïda donen molt bons resultats amb aquesta tècnica.

3.1.2.- Modificació del pitch:

Per a desenvolupar aquest bloc, des d'un bon començament es va tenir el convenciment de voler utilitzar les tècniques en domini freqüencial, ja que les tècniques temporals són molt poc indicades per canvis grans de freqüència i per sons polifònics i les proves realitzades ens han confirmat un canvi molt notable en el timbre de la veu en fer el canvi de pitch, que és el que comunment es coneix com a 'veu de barrufet'.
Així doncs es va començar implementant la tècnica bàsica del 'phase vocoder', en la qual es buscava la freq. instantània de cada canal freqüencial en funció de l'increment de fase de cada un d'ells. El resultat no va ser gaire òptim, degut a que el senyal transformat tenia massa soroll de fase (referir-se l'apartat 'phase vocoder' de l'estat de l'art), tot i que s'apreciava clarament el canvi de pitch.
Aleshores es va decidir implementar el sistema de 'Scaled phase locking', explicada anteriorment en el capítol de l'estat de l'art. Aquesta és una tècnica avançada de transformació freqüencial, i també molt costosa a nivell computacional, però que ofereix uns millors resultats.
La implementació d'aquesta tècnica es detalla per punts a continuació:

1. Calcular per la trama actual el factor de modificació de freq. desitjat mitjançant $\alpha = {fd}/{f0}$, a on α és el factor de modificació, fd és la freq. de destinament desitjada i freq. de pitch és la freq. de pitch.
2. Trobar els pics del mòdul de la transformada de la trama actual. Un pic s'ha definit com aquella mostra que sigui superior a les dues mostres adjacents superiors i a les dues inferiors.
3. Dividir l'espectre en zones d'influència al voltant dels pics resultants en (2). Els límits de les zones d'influència s'han establit mitjançant el criteri de la mostra mitja entre dos pics consecutius.
4. Calcula la freq. dels canals de pic creant una paràbola amb la mostra de pic i les seves dues mostres veïnes i calculant el seu vèrtex.
5. Generar un nou vector FFT inicialitzat a cero.
6. Generar un nou vector amb longitud igual a les mostres de la FFT inicialitzat cada posició amb el nombre de mostra que li correspon, es a dir començant des de 1 a la primera posició, fins al valor nfft a la última mostra. Aquest vector servirà per indicar com es fan els canvis de les mostres de la FFT original
7. Per a cada pic
    - 7.1. Busquem el pic amb mateixa numeració en la trama anterior(pic primer, segon…)
    - 7.2. Acumular la quantitat de fase associada al canvi de freq. desitjat, a la fase del pic trobat en (7.1)
    - 7.3. Restar del vector FFT generat a (5) els valors de la FFT original de la trama actual corresponents a la zona d'influència del pic actual.
    - 7.4. Extreure la part sencera del factor de modificació α i la part decimal.

7.5. Sumar la part sencera del factor de pitch trobat en (7.4) a les posicions corresponents a la zona d'influència actual en el vector generat en (6).
7.6. Sumar els valors de les posicions de la zona d'influència actual de la FFT actual al vector FFT generat a (5) en les noves posicions trobades a (7.5).

8. Sumar la FFT original de la trama actual al vector resultant generat en (5) i modificat posteriorment, i multiplicar-la després per l'increment de fase calculat en 7.2.
9. Transformar inversament la nova trama FFT resultant en (8).
10. Repetir els punts 1 a 8 per a totes les trames.
11. Reconstruir el senyal modificat en temps.

Amb aquesta tècnica aconseguim un senyal modificat en freqüència mitjançant uns factors de variació variables amb uns resultats acceptables comparant amb els resultats observats amb la tècnica anterior. A més la tècnica anterior funcionava relativament be quan el canvi de freq. era constant per a tot el senyal, però disminuïa molt la qualitat de transformació quan el factor canviava d'una trama a una altre, en canvi amb aquesta tècnica aconseguim uns resultats relativament bons independentment de si el factor és constant o no.
Cal remarcar que, encara que aquest sistema millora bastant la qualitat del senyal, encara afegeix artefactes indesitjats de fase proporcionalment perceptibles amb l'augment de factor de canvi de freqüència.

3.1.3.- Selecció de notes vàlides:

Per a saber quin és el canvi de freqüència que s'ha d'aplicar a cada trama primer s'ha de calcular el pitch de cada una, i després s'ha de calcular el citat canvi en funció de un conjunt de notes vàlides com s'ha explicat en la introducció dels blocs d'aquesta memòria de projecte. Per a poder escollir be el conjunt de notes que es desitja utilitzar, s'ha implementat una aplicació la qual inclou tot el conjunt de les notes musicals dins el registre freqüencial de la veu humana, i a la qual per seleccionar un conjunt concret, se li ha de passar com a paràmetre un vector de dotze posicions, cada una de les quals representarà la selecció d'un semitò, dins l'escala musical, en totes les escales del registre freqüencial. Cada posició serà doncs inicialitzada amb ceros i uns, de manera que si la posició corresponent a un semitò concret està inicialitzada a 1 s'estarà seleccionant aquest semitò al conjunt de freqüències vàlides per a la correcció, i si en canvi aquesta posició està inicialitzada a cero significarà que aquesta posició serà despreciada a l'hora de crear el conjunt de freqs. vàlides.

3.1.3.- Paràmetres escollits:

Per a la realització d'aquest projecte s'ha hagut de prendre decisions en quant a diferents paràmetres que fixen el funcionament del sistema global. Aquests han estat dissenyats pensant en el funcionament final d'aquest projecte en temps real i han resultat els següents.

1. **Freqüència de mostreig: 44,1 KHz**
   Aquest valor ha estat escollit per ser una de les freqs. de mostreig més típiques dels sistemes electrònics musicals.
2. **Mostres per trama: 1024**
   Amb aquest valor i la freq. de mostreig escollida en (1) s'aconsegueix una latència de $1024/44100 = 23.22ms$, que és perfectament acceptable per a una sistema de veu a temps real, i en canvi és un nombre suficientment gran com per a poder fer els càlculs de pitch amb èxit.
3. **Marges de freqüència de pitch: 80 – 1500 Hz**
   Aquests dos valors s'adapten als marges del registre vocal humà.
4. **Nombre de punts de les transformades: 4192**
   Aquest valor s'ha escollit per a tenir una millor resolució freqüencial o, el que és el mateix, per tenir un increment més petit de freq. entre mostra i mostra. Aquesta major resolució freqüencial és necessària per al correcte funcionament de l'algorisme de detecció de pitch, ja que per a la implementació del bloc de desplaçament freqüencial s'ha escollit un nombre de punts igual al de les mostres per trama, es a dir 1024, que es pot aconseguir tornant a fer una nova FFT per trama o fent una delmació de la FFT de 4096 mostres en un factor 4.
5. **Factor d'overlap: 4**
   Aquest valor equival a un 75% d'encavalcament entre trames, que és el valor recomanat per a la implementació de totes les tècniques freqüencials, ja siguin per la detecció de pitch, com per el seu desplaçament.

3.2.- Simulació en Matlab:

Per investigar i provar els algorismes dels blocs que composen aquest sistema afinador, s'ha fet us del programa Matlab degut a la seva versatilitat i a la seva facilitat de programació. El fet de que Matlab sigui un llenguatge de tant alt nivell permet d'una manera molt ràpida la depuració i perfeccionament dels diferents processos.
Així doncs es presenten sis arxius de Matlab (extensió m), que simulen els diferents blocs d'aquest projecte:

1. **tono.m**: funció que genera un to, en un vector de sortida, d'una freqüència i una quantitat de temps o de mostres especificades per paràmetre. També es pot especificar una fase inicial i escollir si es reprodueix el to generat o si no.
2. **correctfreq.m**: funció que donat un vector amb les freqs. de pitch per cada trama d'un arxiu d'àudio i un vector amb les notes vàlides seleccionades, genera un vector amb la quantitat de canvi necessària per trama i un altre amb les freqs. finals de cada trama.
3. **pitch2.m**: funció que donat un arxiu d'àudio i especificant una freq. de mostreig, un número de punts FFT, un factor d'overlap, i un nombre de mostres per trama, separa l'arxiu en trames i per cada trama calcula el seu pitch. Finalment genera u vector de sortida freq. de pitch de llargària igual al nombre de trames amb totes les freqs. de pitch de cada trama.
4. **tone_melody.m**: aquesta funció amb un arxiu d'àudio, un vector amb les freqs. de pitch per trama, el nombre de mostres per trama, el factor d'overlap i el nombre de punts de la FFT, genera un arxiu de so que simula el so original però amb tons, i genera un altre arxiu de so amb la suma de l'arxiu original i l'arxiu de tons generat per comparar si les freqs. de pitch trobades són correctes.
5. **shift_fft_2**: aquesta funció rep un arxiu de veu, un vector amb les freqs. de pitch per trama, un altre vector amb les freqs. de destinament, una freq. de mostreig, un nombre de mostres per trama i un factor d'overlap, i genera un altre arxiu de so amb el pitch modificat en funció dels dos vector de freqs. rebuts, el de pitch i el de modificació.
6. **correct_tone.m**: aquest és l'arxiu principal que fa les crides a tots els demés per a que implementin en conjunt un afinador de veu, objectiu d'aquest projecte. Com a paràmetres d'entrada té un arxiu amb el so a transformar, un vector amb el conjunt de notes vàlides escollides, una freq. de mostreig, un nombre de mostres per trama, un nombre de punts FFT i un factor d'overlap, i genera l'arxiu de sortida afinat a les notes seleccionades.

Càlculs de temps:

Els càlculs de temps en Matlab no són massa útils, ja que al ser un llenguatge de tant alt nivell mai optimitza la magnitud temporal de procés, però si que pot donar una idea aproximada sobre si el temps de procés pot ser o no tolerable per al tipus d'aplicació que volem implementar. Per a això, caldrà tenir en compte que la implementació en C++ s'executarà segurament en un temps molt inferior al temps que trigui Matlab, amb el qual si una rutina en Matlab s'executa en el mateix temps que dura l'arxiu, o inclòs en el doble o el triple, es podrà gairebé assegurar que una vegada implementada en C++, aquesta rutina es podrà adaptar en temps real, i si no es compleix aquesta condició temporal en Matlab, no es podrà assegurar res respecte el seu comportament futur en C++.
Els càlculs de temps de les funcions secundàries son tan petits comparats amb la durada dels arxius originals que s'obviaran en aquest estudi, el qual es centrarà en les dues funcion més important, la de detecció de pitch i la de desplaçament.

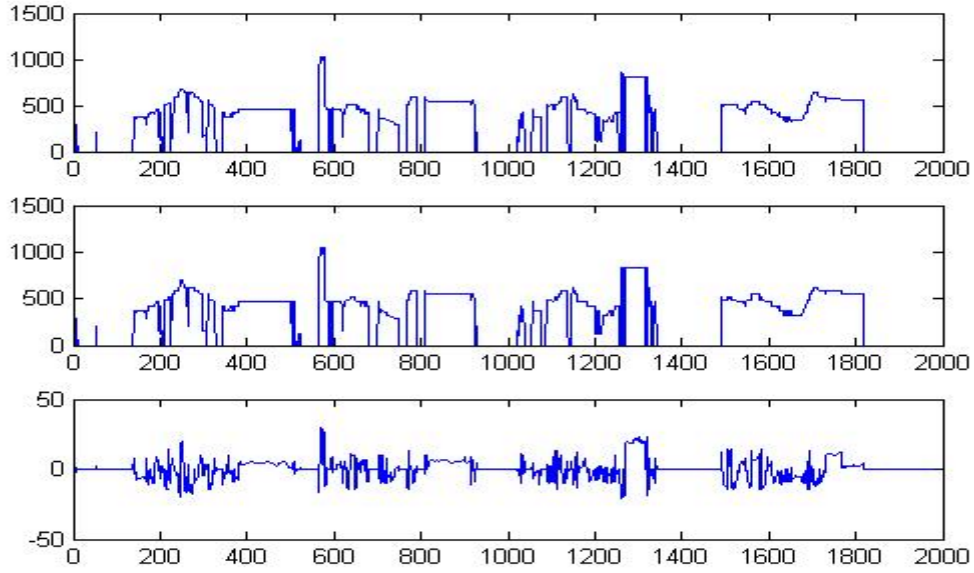
Detecció de pitch:

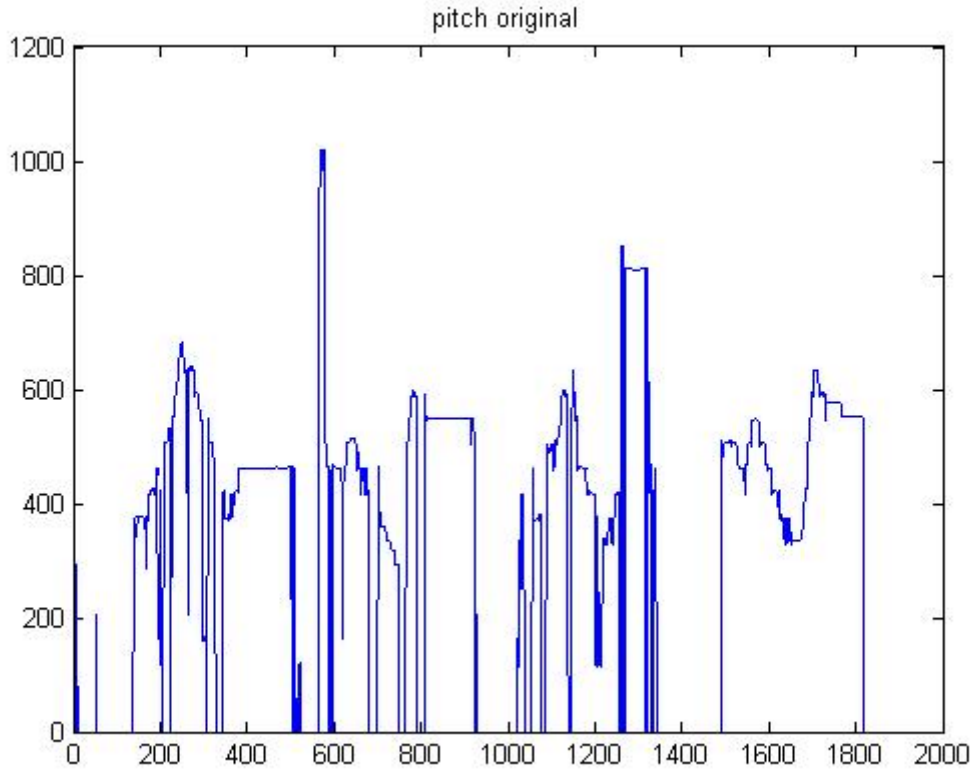
Les dades temporals d'aquesta funció resulten positius, ja que per a un arxiu de poc més de 8 segons, la rutina s'executa en poc més de 5 segons, amb el qual la futura implementació en C++ serà molt més ràpida. La conclusió és que l'algorisme de detecció de pitch és apte per a una implementació a temps real.

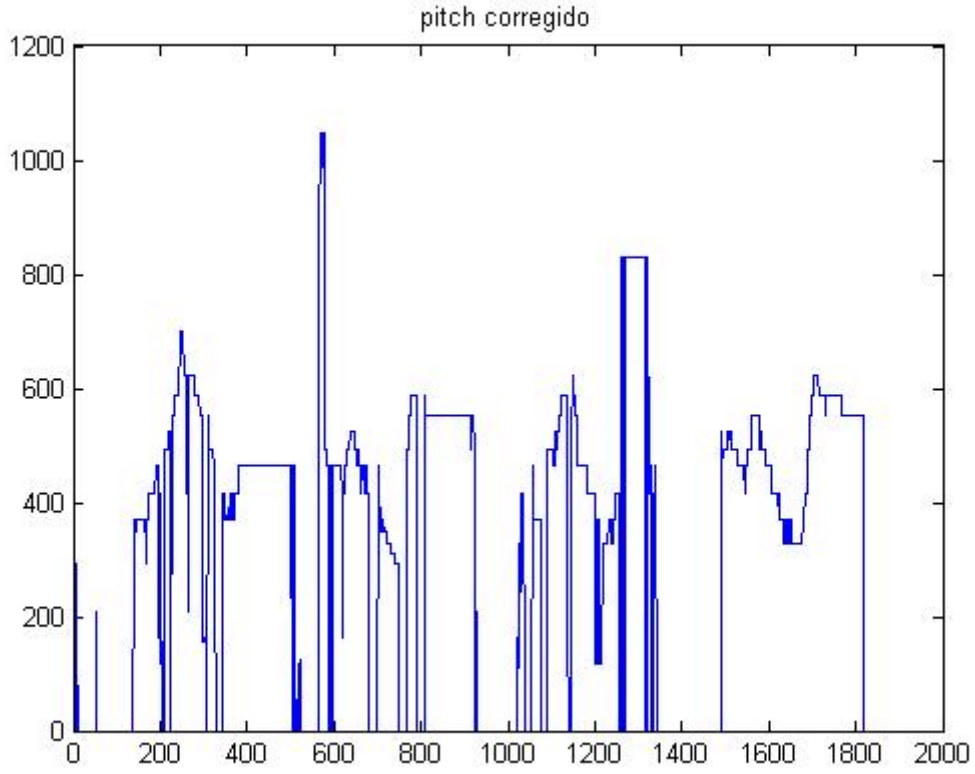
Desplaçament de pitch:

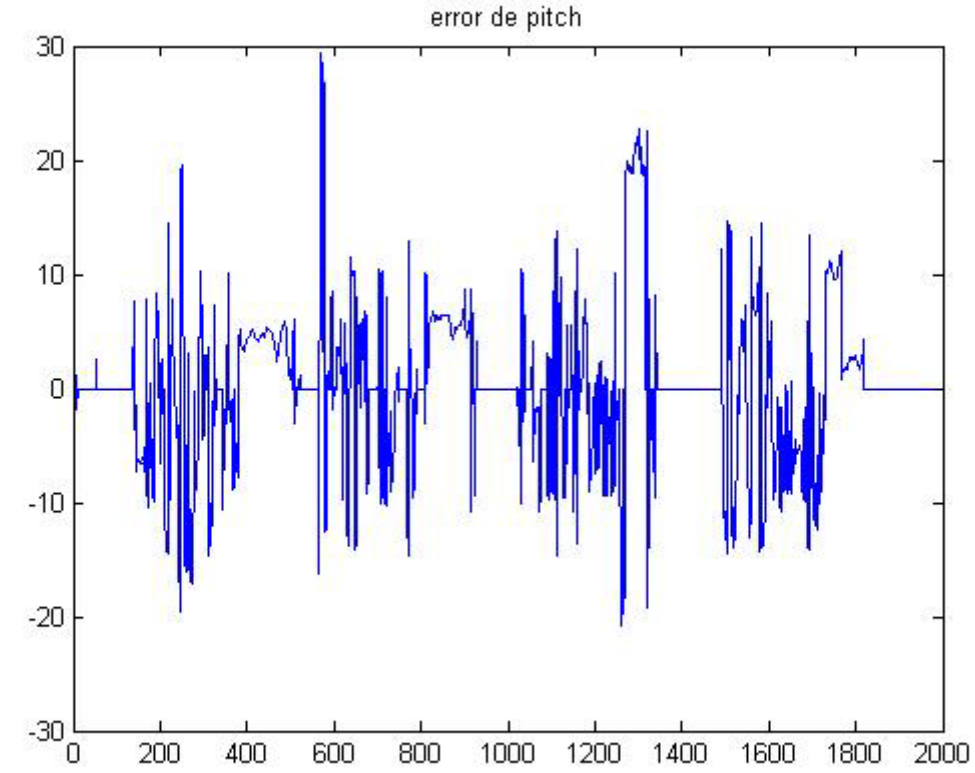
Les probes temporals amb aquesta tècnica no han resultat tan bones com en el cas del detector de pitch. Era de preveure que aquesta tècnica seria bastant més exigent en termes de càlcul que la primera, però no es pensava a priori que pogués trigar tant de temps, ja que per aquest mateix arxiu de so de poc més de 8 segons, l'algorisme ha trigat poc més de 150 segons, que és un temps que supera en més d'un ordre de magnitud el temps de duració de l'arxiu, amb el qual no es pot assegurar res del seu funcionament en temps real un cop implementat en C++.

A continuació es mostres uns gràfics representatius del funcionament del sistema. Es poden veure tres gràfics diferents, un primer que representa les freqüències de pitch detectades pel bloc de pitch referents a un arxiu de so abans de ser modificades, un segon gràfic que representa les mateixes freqüències però després de ser modificades en freqüència, i un tercer gràfic que indica l'error de pitch en cada trama o la desafinació produïda o, el que és el mateix, la correcció que s'ha fet de cada trama.






pitch corregido
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
1200
1400
1600
1800
2000


error de pitch
30
20
10
0
-10
-20
-30
0
200
400
600
800
1000
1200
1400
1600
1800
2000

3.2.- Implementació en C++ / Direct X:

La implementació del sistema en temps real es va decidir fer-la mitjançant el llenguatge C++, i fent us de les llibreries de Microsoft de DirectX 9.0. Mitjançant aquestes dues eines es poden controlar tots els perifèrics d'un ordinador personal amb sistema operatiu Windows així com prendre els control de les targes de vídeo, de so, de xarxa, etc… L'objectiu doncs d'aquest projecte és prendre el control de la tarja de so del PC i adquirir trames de so d'un micròfon, processar-les per a que esdevinguin mostres afinades i, en temps real, reproduir-les. Aquesta tasca finalment no s'ha pogut completar degut a que la investigació dels algorismes i la seva simulació en Matlab ha portat molt del temps disponible per a la execució d'aquest projecte, i quan s'ha començat a implementar en C++ en temps real han falta coneixements com per fer-ho d'una manera prou ràpida com per acabar el projecte en el temps establert, ja que s'havien d'aprendre nous coneixements quan quedava poc temps.
Així doncs s'ha pogut implementar l'algorisme de pitch amb èxit ja que no requeria una entrada i sortida de dades en mode full-dúplex, però no s'ha pogut completar l'algorisme de desplaçament freqüencial, el qual es deixa gairebé implementat, i també es presenta.
Així doncs es presenten els següents projectes en visual C 6.0:

1. **pitch.dsw**: projecte que donat un arxiu d'àudio, genera un arxiu de text amb totes les freq. de pitch de cada trama.
2. **shift.dsw**: projecte inacabat que forma la base de l'algorisme de desplaçament freqüencial, pren com a entrades un arxiu de so, i dos arxius de text, un amb les freqs. de pitch per cada trama generat per (1), i un altre amb les freqs. de destinament. Aquest segon arxiu de text es pot substituir per un factor de modificació constant per ta de provar millor l'algorisme mentre s'edita.

Càlculs de temps:

Com només s'ha implementat l'algorisme de detecció de pitch, només es farà referència en aquest a l'hora de calcular els temps, i el resultat ha estat molt bo, ja que per al mateix arxiu avaluat en l'apartat de Matlab de duració poc superior al 8 segons, el programa s'ha executat amb èxit en poc menys de 2 segons, resultat que deixa un gran marge de temps restant per al procés de les altres tasques, especialment la de transformació freqüencial.

Condicions de mesura:

Aquestes probes temporals han estat fetes amb el següent equip:
Pentium IV 1.7GHz , 512Mb de memòria RIMM
I per tant aquest valors mesurats queden sotmesos a la potència de procés amb la que han estat fetes, i són susceptibles per tant de canviar si es fan amb altres equips diferents

4.- Conclusions:

4.1.- Limitacions :

Si be aquest projecte ha aconseguit el seu objectiu d'implementar una afinador de veu, aquest presenta certes limitacions que podrien amb temps ser evitades o minvades notablement. El problema principal amb el que s'ha treballat és indubtablement el soroll de fase en el bloc de desplaçament freqüencial, però també s'han presentat altres limitacions que a continuació es detallen:

1. **Soroll de fase**: Com s'ha comentat, aquest suposa el principal inconvenient que aquest projecte ha presentat, i s'ha treballat per tal de minimitzar aquest efecte fins el límit de generar un algorisme que difícilment es pot encabir en un sistema en temps real, però no s'ha aconseguit evitar. Es podria haver evitat mitjançant una tècnica de desplaçament en el domini temporal, que són molt més senzilles d'implementar, amb menys cost computacional, i que no presenten aquest efecte, però aleshores la limitació hauria estat la quantitat de desplaçament possible, i aquest és un punt al que s'ha donat més importància o que s'ha ponderat més a l'hora d'establir les directrius del projecte.
2. **Latència mínima:** El fet d'utilitzar una freq. de mostreig de 44.1KHz i 1024 mostres per trama limita la latència del sistema en la seva implementació en temps real a poc més de 23ms. Encara que aquest temps es podria disminuir augmentant la freq. de mostreig o disminuint les mostre per trama, segueix existint la limitació de que per a calcular el pitch amb èxit es necessita un temps semblant al de la latència d'aquest sistema per a poder detectar be les freqs. greus de períodes grans.
3. **Baixa freqüència**: La tècnica utilitzada és més eficient amb les altes freqs que a baixes freqs. pel motiu expressat en (2), i es que per detectar trames amb períodes grans de temps es necessiten molts mil-lisegons de senyal, mes temps de senyal quant més baixa sigui la freqüència, ja que per a un mateix temps de trama, quant més gran és el període, menys cicles sencers tenim per trama, situació que complica la determinació del pitch de la mateixa.

4.2.- Resultats :

Quan es va proposar aquest projecte, es van establir uns certs criteris d'anàlisi i uns certs paràmetres de funcionament que es proposaven com a premisses de disseny del mateix. Aquests criteris de disseny han estat els valors de màxima ponderació a l'hora de crear els diferents blocs i algorismes, ja que determinen el correcte funcionament del sistema, però alguns d'aquests criteris han hagut de ser modificats en el transcurs del desenvolupament del projecte, o hi ha hagut d'altres que no han pogut ser assolits. En aquest capítol es recull tot el conjunt de criteris i paràmetres de disseny i tots els resultats del projecte.

Com a criteris de disseny es van plantejar els següents:

1. ***Capacitat per a la detecció i modificació del pitch amb tolerància petita d'error:***
   Degut a que es pretén fer un sistema afinador, es de preveure que els desviaments freqüencials puguin ser molt petits si el senyal d'entrada està gairebé afinat. Degut a això, aquest sistema ha de ser capaç de detectar aquests mínims desviaments i de corretgir-los, per tant un dels principals criteris de disseny ha estat la resolució freqüencial del sistema tant en detecció com en correcció.
   Els resultats en aquest punt han esdevingut força satisfactoris ja que la resolució per ambdós algorismes ha resultat amb un a tolerància d'error de menys de 1Hz, per a les trames detectades correctament que són aproximadament un 95% del total de trames analitzades en un arxiu. Aquests marges d'error tan petits són deguts al sistema de càlcul per fase escollit, que tot i presentar el problema de la interferència entre canals, ofereix una mesura quantificada amb 32 bits, ja que la FFT es calcula amb aquesta quantificació, i l'increment de fase i el posterior càlcul de freq. instantània igual, tos són valors 'double'.

2. ***Possibilitat de seleccionar el conjunt de notes vàlides per a la correcció:***
   Per a poder actuar el sistema com a afinador, es va posar com a criteri de disseny que es pugui interactuar amb l'afinador per tal que aquest s'adapti a l'escala en la que s'està cantant, així es poden eliminar unes quantes freqüències de tot el conjunt, facilitant la tasca de l'afinador ja que no ha de discernir entre tantes freqs. i fent que l'error de decisió de quina és la freqüència de destinament sigui més fiable, ja que al seleccionar les notes vàlides s'imposen salts en freq. reduint les opcions de decisió.

3. ***Sistema a temps real:***

   Aquest sistema es va pensar per a ser utilitzar en temps real mitjançant un micròfon, i obtenint una correcció automàtica, escoltant al moment la veu del cantant afinada, però com s'ha comentat en el capítol de la implementació en c++, aquest projecte no s'ha pogut finalment implementar en temps real ja que no ha donat temps per a l'assimilació de tots els nous conceptes de C++ i DirectX necessaris per a implementar aquesta aplicació. També s'ha comentat que la part de simulació en Matlab va ocupar més temps del previst i quan es va voler passar a temps real quedava ja poc temps i mancaven coneixements. Però cal remarcar que l'algorisme de pitch si que ha estat implementat i preparat per al sistema real com a bloc independent.

   Pel que fa als algorismes també s'ha tingut l'inconvenient anteriorment comentat, i es que el bloc de desplaçament freqüencial triga per sobre de deu vegades més en executar-se que el temps de duració dels arxius que analitza. Aquesta però és una mesura en Matlab, i quedaria pendent de conèixer si es pot arribar a encabir a temps real amb les optimitzacions que ofereix C++.

   Així doncs aquesta part no ha pogut ser completada, i queda pendent de ser finalitzada a títol personal.

5.- Bibliografia i referències:

REFERÈNCIES:

[1] M. Dolson, “The phase vocoder: A tutorial,” *Comput. Music J.*, vol. 10, pp. 14–27, 1986.
[2] H. J. S. Ferreira, “An odd-DFT based approach to time-scale expansion of audio signals,” to be published.
[3] E. B. George and M. J. T. Smith, “Analysis-by-synthesis/Overlap-add sinusoidal modeling applied to the analysis and synthesis of musical tones,” *J. Audio Eng. Soc.*, vol. 40, pp. 497–516, 1992.
[4] D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time fourier transform,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 236–243, Apr. 1984.
[5] J. Laroche, “Time and pitch scale modification of audio signals,” in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Boston, MA: Kluwer, 1998.
[6] R. J. McAulay and T. F. Quatieri, “Speech analysis/synthesis based on a sinusoidal representation,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 744–754, Aug. 1986.
[7] E. Moulines and J. Laroche, “Non parametric techniques for pitch-scale and time-scale modification of speech,” *Speech Commun.*, vol. 16, pp.175–205, Feb. 1995.
[8] S. H. Nawab, T. Quatieri, and J. S. Lim, “Signal reconstruction from short-time fourier transform magnitude,” *IEEE Trans. Acoust., Speech,Signal Processing*, vol. ASSP-31, pp. 986–998, Aug. 1983.
[9] R. Portnoff, “Time-scale modifications of speech based on short-time Fourier analysis,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol.29, pp. 374–390, 1981.
[10] M. S. Puckette, “Phase-locked vocoder,” *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1995.
[11] T. F. Quatieri, R. B. Dunn, and T. E. Hanna, “A subband approach to time-scale expansion of complex acoustic signals,” *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 515–519, Nov. 1995.
[12] T. F. Quatieri and J. McAulay, “Shape invariant time-scale and pitch modification of speech,” *IEEE Trans. Signal Processing.*, vol. 40, pp. 497–510, Mar. 1992.
[13] S. Roucos and A. M. Wilgus, “High quality time-scale modification of speech,” in *Proc. IEEE ICASSP’85*, Tampa, FL, pp. 493–496.
[14] B. Sylvestre and P. Kabal, “Time-scale modification of speech using an incremental time-frequency approach with waveform structure compensation,” in *Proc. IEEE ICASSP’92*, pp. 81–84.

[15] Alain de Cheveign´e and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4), 2002.
[16] Phillipe Depalle, Guillermo Garc´ıa, and Xavier Rodet. Tracking of partials for additive sound synthesis using Hidden Markov Models. In International Conference on Acoustics, Speech and Signal Processing, volume I, pages 225–228. IEEE, 1993.
[17] Erkan Dorken and S. Hamid Nawab. Improved musical pitch tracking using principal decomposition analysis. In International Conference on Acoustics, Speech and Signal Processing, volume II, pages 217–220. IEEE, 1994.
[18] Boris Doval and Xavier Rodet. Estimation of fundamental frequency of musical sound signals. In International Conference on Acoustics, Speech and Signal Processing, pages 3657–3660. IEEE, 1991.
[19] Boris Doval and Xavier Rodet. Fundamental frequency estimation and tracking using maximum likelihood harmonic matching and HMMs. In International Conference on Acoustics, Speech and Signal Processing, volume I, pages 221–224. IEEE, 1993.
[20] John M. Eargle. Music, Sound and Technology. Van Nostrand Reinhold, Toronto, 1995.
[21] James L. Flanagan. Speech Analysis, Synthesis and Perception. Springer-Verlag, New York, 1965.

LINKS:

http://cnx.rice.edu/content/m11714/latest/
http://cnx.rice.edu/content/m12539/latest/
http://iie.fing.edu.uy/ense/asign/sisdsp/
http://www.uspto.gov/patft/index.html