Generació de claus criptogràfiques a partir de la veu.

Nom Estudiant: Cristian Requena Barreda.

Programa: Màster Universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC)

Nom Consultor: Antoni Martínez Ballesté.

Centre: Universitat Rovira i Virgili.

Data Lliurament: Juny de 2014.

FITXA DEL TREBALL FINAL

Títol del treball:	*Generació de claus criptogràfiques a partir de la veu.*
Nom de l'autor:	*Cristian Requena Barreda.*
Nom del consultor:	*Antoni Martínez Ballesté.*
Data de lliurament (mm/aaaa):	*Juny de 2014.*
Àrea del Treball Final:	*Seguretat en Sistemes Biomètrics.*
Titulació:	Màster Universitari en Seguretat de les Tecnologies de la Informació i de les Comunicacions (MISTIC)

Resum del Treball (màxim 250 paraules):

Aquest treball versa sobre la obtenció d'una eina que permeti la generació d'una clau criptogràfica derivada a partir de les característiques biomètriques de la veu.

Es tracten i executen aspectes com el control de la captura d'àudio, el tractament dels senyals d'entrada per obtenir un espectre se so que permeti relacionar les freqüències de so amb la seva amplitud relativa i l'aplicació d'un mètode per eliminar freqüències no relacionades amb la veu humana, de manera que només es considerin les bandes de freqüència relacionades amb la parla.

Abstract (in English, 250 words or less):

This dissertation explains the previous considerations and the creation of a computerized tool that generates a cryptographic key from the biometric features found in human voice.

Several points have been considered, like controlling the audio input from the microphone, handling the input data to obtain a sound sprectrum that relates the relative amplitude of each band of frequencies and the implementation of a method to remove frequencies not present in human voice to empower the related ones – which will be used to generate the key.

Paraules clau (entre 4 i 8):

Biometria, veu, generació, clau criptogràfica, espectre de so, MFCC, FFT.

Índex

1. Introducció

1.1 Context i justificació del Treball

Tot just durant els últims anys, l'autenticació d'usuaris mitjançant les seves característiques biomètriques ha començat a adinsar-se al *mainstream*, és a dir, ha guanyat popularitat, s'ha estès el seu ús i, en definitiva, ha acabat sent rellevant.

Molts dels seus usos comencen a ser quotidians, amb màquines de fitxatge de presència per empremta digital, desbloqueig d'accessos mitjançant l'escaneig de l'iris, i, des de fa uns mesos, fins i tot al palmell de la mà – els telèfons mòbils tot just comencen a incorporar autenticació biomètrica.

Seguint la tendència de l'estat de l'art d'aquesta àrea de coneixement de la seguretat de la informació, l'objectiu d'aquest treball final de màster se centra en l'anàlisi i desenvolupament d'una eina informàtica que permeti dur a terme l'autenticació d'usuaris mitjançant les característiques biomètriques de la veu.

1.2 Objectius del Treball

Creació d'una eina informàtica que generi claus criptogràfiques a partir d'una entrada de dades en forma de veu.

1.3 Enfocament i mètode seguit

Durant la primera fase d'aquest treball s'ha realitzat una cerca metòdica per estudiar la situació actual en l'ús de la veu com eina d'autenticació, i s'ha localitzat un ampli conjunt de papers descrivint diversos mètodes per generar claus criptogràfiques a partir de les característiques biomètriques d'un individu. D'aquests, s'han triat quatre de representatius i rellevants, que es relacionen a continuació.

Títol	Autor/s	Data i tipus de publicació
Cryptographic Key Generation from Voice	Fabian Monrose, Michael K. Reiter, Qi Li, Susanne Wetzel	2001, conferència IEEE.
Biometrics-based cryptographic key generation	Yao-Jen Chang, Wende Zhang, Tsuhan Chen	2004, conferència IEEE.
Multi-speaker voice cryptographic key generation	L. Paola Garcia-Perera, J. Carlos Mex-Perera, Juan A. Nolazco-Flores	2005, conferència IEEE.
You are the Key: Generating Cryptographic Keys from Voice Biometrics	Brent Carrara, Carlisle Adams	2010, conferència IEEE.

Taula 1: Papers de referència

D'altra banda, s'ha localitzat un producte comercial[1] que permet la realització de l'autenticació d'usuaris mitjançant la veu.

En una segona fase, es desenvoluparà l'eina descrita anteriorment. Com es tractarà d'un període de desenvolupament d'uns dos mesos, s'ha creat una planificació de treball per a definir el conjunt de fites que seran assolides, i que han estat entregades en forma i data.

1 http://www.nuance.com/landing-pages/products/voicebiometrics/

1.4 Planificació del Treball

El període complet per dur a terme les tasques de desenvolupament de l'eina és de dos mesos, que han estat dividits en un conjunt de fites a assolir:

- Fita 1: Configuració de l'entorn i proves.
  Accés a l'entrada d'àudio del sistema.
  És convenient fer proves i detectar els possibles problemes o incompatibilitats, amb l'objectiu de, si és necessari, poder reaccionar i substituir la biblioteca.

- Fita 2: Mètode d'entrada bàsic.
  Sota petició de l'usuari, realitzar la gravació d'àudio durant un lapse de temps fix, i realitzar operacions bàsiques sobre el mateix; per exemple, la detecció del soroll de fons.

- Fita 3: Mètode d'entrada intel·ligent.
  Millorar el codi desenvolupat per la fita 2 i, en comptes d'establir un límit temporal, detectar quan l'usuari comença i acaba de parlar, per tal d'enregistrar l'àudio durant el temps requerit.

- Fita 4: Realització d'operacions bàsiques.
  Sobre l'àudio obtingut amb el mètode d'entrada desenvolupat per la fita 3, executar operacions bàsiques per començar a tractar les característiques de la veu. Per exemple, obtenir la freqüència màxima i mínima de la veu, i, si escau, dades estadístiques simples.

- Fita 5: Selecció de *centroids*.
  Seguint l'algorisme descrit per la referència bibliogràfica més referenciada [1], realitzar la tria dels punts centrals de l'espectre de so enregistrat a la fita 3.

- Fita 6: Autenticació a partir dels descriptors
  Implementació de l'algorisme per autenticar els usuaris.

Es planifica l'entrega de les diverses fites a partir del següent cronograma:

	Setmana 1	**Setmana 2**	**Setmana 3**	**Setmana 4**
Abril	Fita 1	Fita 2		Fita 3
Maig	Fita 4		Fita 5	
Juny	Fita 6			Memòria

Taula 2: Planificació de fites.

És a dir:

- Fita 1: Diumenge, 6 d'abril de 2014.
- Fita 2: Diumenge, 13 d'abril de 2014.
- Fita 3: Diumenge, 27 d'abril de 2014.
- Fita 4: Diumenge, 4 de maig de 2014.
- Fita 5: Diumenge, 18 de maig de 2014.
- Fita 6: Diumenge, 8 de juny de 2014.
- Memòria: Diumenge, 22 de juny de 2014.

1.5 Sumari de productes obtinguts

S'ha creat un programa informàtic que obté el so capturat per un micròfon d'un computador, el tracta i genera una clau criptogràfica que varia en funció de les característiques de la veu.

1.6 Descripció dels altres capítols de la memòria

La resta de capítols de la memòria expliquen els diferents aspectes treballats, així com els problemes localitzats i les solucions adoptades.

2. Resta de capítols

2.1 Mètode de captura de so.

A l'anàlisi previ del treball es va considerar d'emprar una biblioteca anomenada OpenAL, que permet l'ús dels dispositius d'àudio presents a un computador. Els aspectes que es van considerar rellevants van ser que estava portada a múltiples plataformes i es podia emprar des de força llenguatges de programació.

Una vegada escollida la plataforma sobre la que es realitzaria el desenvolupament (Java), es va fer una recerca per veure si existia cap mètode òptim per evitar l'ús d'altres biblioteques. Com es pot veure a l'eina desenvolupada, Java proveeix de biblioteques natives que permeten la captura i el tractament del so capturat per les interfícies corresponents, pel que es va descartar l'ús d'OpenAL.

El funcionament de la captura de so a un computador es realitza mesurant la pressió sonora captada pel micròfon al llarg del temps. Aquesta pressió sonora es comporta com una ona sinusoïdal al llarg del temps, pel que es fa necessari capturar un conjunt de mostres en petits intervals de temps que permetin disposar del detall suficient sense perdre informació rellevant. En aquest sentit s'han seleccionat dues constants estàndard per a capturar el so: es prenen 44100 mesures de pressió per segon (44100Hz.) amb una resolució de 16 bits. Tot i això, per facilitat les proves inicials, durant les primeres fases es va treballar amb resolucions de 8 bits, però posteriorment es va fer l'ampliació a 16.

2.2 Soroll de fons.

Durant el desenvolupament del treball, i sobretot durant les primeres proves funcionals, es va comprovar que tot trobar-se en silenci, el micròfon del computador sempre capturava un petit soroll de fons.

Per evitar que aquest soroll afectés, d'una banda, al so capturat i, de l'altra, a la detecció d'inici i finalització de la gravació, es va decidir d'anul·lar-lo, permetent establir el llindar mínim que havia de disposar el so per poder ser capturat.

2.3 Detecció d'inici i finalització de l'entrada de so.

A partir del considerat al punt anterior, es va realitzar un càlcul estadístic per cada conjunt de mostre de so, de manera que és possible d'avaluar com varia la pressió sonora en un determinat lapse de temps.

Concretament, per permetre la detecció de l'inici i final de la parla, es comprova l'evolució de la desviació estàndard del conjunt de mostres: si aquest valor és 0 (cal recordar que s'està obviant el soroll de fons, que sempre és present), significa que no s'està detectant cap variació de pressió, és a dir, que no s'està parlant. En aquest cas, s'atura la captura de so en curs.

Per un altre costat, si es parteix de la situació d'aturada anterior, quan es detecta una variació de la desviació estàndard, s'inicia la captura i avaluació del so.

2.4 Obtenció de l'espectre sonor de les mostres.

L'espectre sonor és una representació de la pressió sonora efectuada durant un temps determinat en diferents bandes, cadascuna de les quals cobreix un rang de freqüències diferents.

En línia amb el paper de referència de Fabian Monrose et al. [1], s'ha implementat un algorisme de càlcul de la transformada discreta de Fourier, concretament l'anomenat ràpid (FFT, de *Fast-Fourier Transform*). Aquest algorisme permet obtenir l'espectre sonor d'una manera computacionalment òptima.

El resultat d'aplicar aquest algorisme és, com s'ha dit, l'espectre sonor. Amb aquest producte s'efectuen dues tasques: d'un costat, es dibuixa un diagrama on es relaciona un rang de freqüències amb la seva pressió relativa; de l'altre, s'empra per realitzar la resta de tasques rellevants del treball.

2.5 Obtenció de l'espectre invers (*cepstrum*) i canvi d'implementació.

Seguint el paper de referència de Fabian Monrose et al. [1], després de l'obtenció de l'espectre de so es genera un espectre invers d'aquest espectre. Concretament, es tracta d'aplicar l'algorisme de càlcul de la derivada discreta de Fourier, però amb exponents inversos, de manera que el resultat és una relació de la variabilitat de cada franja de freqüències.

Després d'haver implementat i comprovat aquest punt, em vaig adonar que hi havia una gran variabilitat de les dades, pel que era necessari elaborar un conjunt gran de captures de so amb una mostra gran de població per detectar quines d'aquestes franges de freqüències eren rellevants i aplicar-les-hi una quantificació vectorial en funció d'aquests resultats.

Com aquesta tasca excedia les pretensions del treball, es van considerar altres implementacions, com la de L. Paola Garcia-Perera et al. [2], que en comptres de treballar amb aquest espectre invers, realitza un altre processament anomenat MFCC (Mel Frequency Cepstral Coefficients) que, aplicat a les mostres de so, obté, per totes elles, un vector de 13 dimensions, que caracteritza les bandes d'una manera més adient a la veu humana. Les freqüències sonores de la veu que distingeixen el timbre de la mateixa són tractades amb molt de pes, mentre que les freqüències a partir de 3500Hz (on la parla humana no arriba) no són rellevants.

Així doncs, s'ha implementat un algorisme per obtenir el MFCC a partir de l'espectre de so. El vector resultant d'aquest algorisme serà el que s'emprarà per obtenir les dades de les quals es derivarà la clau criptogràfica, ja que són aquestes les que varien d'una persona a una altra.

En efecte, la veu humana no es composa d'una freqüència de so pura, sinó que genera un conjunt d'armònics que varien d'un subjecte a un altre. Aquests armònics són els que maximitza l'MFCC, obviant les freqüències no rellevants de la veu, de manera que un mateix fonema disposarà d'un vector diferent en funció del subjecte.

3. Conclusions

He assolit un nivell de comprensió del funcionament del so digitalitzat que no tenia anteriorment. Realment desconeixia, per exemple, perquè el so de qualitat de CD es digitalitza amb 44100 mostres per segon – ara he comprès que al aplicar una transformada de Fourier cal com a mínim el doble de mostres per obtenir el rang de freqüències audibles complet de 20000Hz.

D'altra banda, en el terreny de les conclusions pròpies del projecte, he vist que la veu és, segurament, una de les característiques biomètriques més complexa de tractar. Per exemple, una ditada és més o menys constant: en pot variar l'orientació o pot ser parcial, però no canvia d'un moment a l'altre. Una mateixa cara sempre té uns elements que, tot i variar-ne l'orientació, sempre són més o menys reconeixibles. La veu, al seu torn, és molt complexa de treballar. En efecte, el nostre aparell fonador té moltíssimes formes de pronunciar un mateix fonema (de fet, hi ha llengües, com el xinès, en les quals la forma de pronunciar-lo en canvia el significat), pel que és molt complex de reconèixer unes característiques concretes que identifiquin un individu.

Vaig dedicar un esforç temporal elevat a seguir la implementació del paper de referència de Fabian Monrose et al. [1], però donada la necessitat de canviar aquesta implementació, he comprés que pot ser necessari deixar un camí si aquest és bloquejant i continuar per altres camins.

Crec que la metodologia d'investigació i desenvolupament ha estat l'adient, tot i que si s'hagués decidit la plataforma de desenvolupament abans de cercar biblioteques de captura de so, possiblement no caldria haver revisat cap informació d'OpenAL, guanyant algunes hores de treball. Tot i el canvi realitzat en la part troncal del treball, crec que era impossible adonar-me'n de manera prèvia a l'execució del mateix, pel que no considero que aquest sigui un error metodològic. De fet, considero que la capacitat d'haver evitat el bloqueig total del treball optant per una segona via és una solució correcta al problema trobat.

La planificació del projecte s'ha acomplert correctament, completant totes les fites i lliurant els entregables tal com es va establir al cronograma. Cal dir que es van unir dues entregues (4 i 5) perquè es va estimar que caldria dedicar més esforços a la primera, quan realment va ser trivial.

Finalment, cal dir que l'objectiu de desenvolupar una eina informàtica que generi claus a partir de la veu s'ha acomplert, tot i que aquesta podria ser més vistosa o ser disponible en línia. Aquests dos aspectes podrien ser susceptibles de ser treballats en el futur.

4. Glossari

- Hz: Hertz, unitat de mesura de freqüències, que indica les vegades que succeeix un fenòmen per segon. 44100Hz = 44100 vegades per segon.
- OpenAL: Biblioteca d'accés a dispositius de so. No s'empra a l'implementació final.
- Bit: Unitàt bàsica d'informació, representada per un 0 o un 1.
- FFT: Transformada ràpida de Fourier.
- MFCC: Mel Frequency Cepstral Coefficients, algorisme per obtenir la pressió sonora relativa de diverses bandes freqüències.
- Centroid: Punt central de l'espectre de so. No s'empra a l'implementació final.

5. Bibliografia

1. Fabian Monrose, Michael K. Reiter, Qi Li, Susanne Wetzel. *"Cryptographic Key Generation from Voice".* 2001, conferència IEEE.
2. L. Paola Garcia-Perera, J. Carlos Mex-Perera, Juan A. Nolazco-Flores. *"Multi-speaker voice cryptographic key generation*". 2005, conferència IEEE.
3. Yao-Jen Chang, Wende Zhang, Tsuhan Chen. "*Biometrics-based cryptographic key generation*". 2004, conferència IEEE.
4. Brent Carrara, Charlisle Adams. "*You are the Key: Generating Cryptographic Keys from Voice Biometrics*". 2010 Eighth Annual International Conference on Privacy, Security and Trust.
5. K. Inthavisas, D. Lopresti. "*Speech Cryptographic Key Regeneration based on Password*". 2011, conferència IEEE.
6. http://www.wikijava.org/wiki/The_Fast_Fourier_Transform_in_Java_%28part_1%29
7. https://www.ee.columbia.edu/~ronw/code/MEAPsoft/doc/html/FFT_8java-source.html
8. https://sites.google.com/site/piotrwendykier/software/jtransforms
9. http://www.fftw.org/download.html

6. Annexos

Alhora que la memòria, s'entreguen els documents d'entrega de les diferents fites, així com el codi font de l'eina informàtica desenvolupada.