

8 CONCLUSIONES Y TRABAJO FUTURO

El presente proyecto se ha centrado en la investigación y mejora de las técnicas de Verificación Automática de Locutor más utilizadas en los últimos años, analizando y contrastando las mejoras que se han ido incorporando durante ese tiempo, con el objetivo de llegar a obtener resultados altamente competitivos al nivel del estado del arte actual.

En una primera parte durante las simulaciones, se ha experimentado con GMM-UBM todas las posibilidades existentes con respecto a los parámetros de configuración para fijar los valores más óptimos y utilizarlos en el resto de las técnicas para comparar en igualdad de condiciones. Destacar que la normalización de puntuaciones mediante T-Norm ha mejorado notablemente los resultados, consagrándose como la mejor opción de normalización para la base del NIST de 2005. A raíz de los resultados en SAD, se puede considerar especialmente llamativa la diferencia de rendimiento en función del parámetro α_{SAD} , acontecimiento que no preveíamos debido a que 2.5 minutos se considera una cantidad de tiempo bastante elevada como para que dicho parámetro influya con tanta relevancia. Finalmente, con respecto a todas las opciones barajadas y experimentadas en GMM, resaltar la enorme diferencia de eficiencia al trabajar con 10 segundos en lugar de 2.5 minutos. Esto se presenta como una línea de trabajo actual y futura de enorme importancia, puesto que es muy complicado en una aplicación real disponer de 2.5 minutos para un usuario. De ahí que uno de los objetivos primordiales de los investigadores actuales en Verificación Automática de Locutor es la de acercar lo máximo posible los resultados correspondientes a grabaciones de corta duración a los de larga duración.

Con SVM no se ha mejorado los resultados de GMM-UBM en la proporción que se esperaba, a pesar del hecho de que el concepto de máquina de vectores se introdujo en el campo de la Verificación Automática de Locutor con el simple propósito de mejorar las prestaciones de GMM, especialmente en relación al minDCF. El hecho de que la técnica de NAP, cuyo procedimiento de entrenamiento y evaluación es idéntico al de SVM obtenga una mejora significativa justifica el hecho de que para las máquinas de vectores supere el rendimiento de la tradicional técnica generativa de GMM es necesario que la cantidad de datos también sea superior.

La mejora sustancial con NAP y FA resalta uno de los teoremas más importantes de este proyecto: pequeños incrementos en la cantidad de datos pueden provocar enormes mejoras en el rendimiento de los sistemas siempre y cuando estos datos se manejen de manera eficiente. Además estas dos técnicas ponen de manifiesto las principales carencias de GMM-UBM y SVM que consisten en que no se intentan contrarrestar el grave inconveniente de la variabilidad de intercesión. Todas las líneas de investigación actuales, entre las que cabe destacar principalmente a JFA (Joint Factor Analysis) y i-vector tratan, al igual que NAP y FA, de modelar correctamente la compensación de canal, aunque las dos primeras detectan las direcciones de máxima variabilidad de canal en un espacio de características de muy altas dimensiones

(mientras que NAP y FA lo hacen en un subespacio de bajas dimensiones). En concreto, la Universidad de Swansea se encuentra actualmente finalizando la implementación de las técnicas de JFA y i-vector en ALIZE, y si todo marcha tras lo previsto, las utilizará en las evaluaciones del NIST de 2012 en el próximo mes de Octubre. Esto lleva a que la continuación natural de este proyecto podría consistir perfectamente en un análisis e implementación de estas dos técnicas, comparando los resultados con los obtenidos en este proyecto y constatar si se producen notables mejoras. Dichas técnicas son muy parecidas en estructura a FA, con la salvedad de que incluyen nuevas componentes en el modelado del subespacio vectorial que caracteriza la variabilidad de canal. Este proyecto podría ser una referencia ideal para adquirir toda la base necesaria para afrontar esta nueva tarea.

De lo explicado en este proyecto, también se concluye que aunque existen multitud de técnicas usadas para verificación, en cuanto al número de posibilidades para la puntuación, al final son únicamente dos las opciones que se barajan actualmente: GMM (generativa) y SVM (discriminativa).

Una de las principales ventajas de este proyecto es que aunque éste se encuadra dentro del reconocimiento de voz, los algoritmos empleados para las distintas técnicas pueden ser utilizados perfectamente en cualquier campo del reconocimiento biométrico. De hecho, dentro de este proyecto, el tratamiento digital de voz termina una vez llevada a cabo la parametrización de la voz obteniendo los coeficientes MFCC, y a partir de ahí todo el procedimiento empleado es puramente estadístico e independiente de la ciencia biométrica en la que nos estemos moviendo. Además de utilizarse en otros campos del reconocimiento de voz como es la identificación idiomática o reconocimiento de habla, SVM y GMM son también ampliamente tenidos en cuenta en otras técnicas biométricas como el reconocimiento facial. Otra alternativa bastante interesante a este proyecto podría resultar en la implementación de todas estas técnicas en otra área biométrica y comprobar que efectivamente funcionan perfectamente.

Por último, me gustaría concluir comentando que en la vida real, generalmente no nos vamos a encontrar los datos tan claros y “limpios” (de ruido) cómo nos lo presenta el NIST. Y además también tenemos que tener en cuenta que en aplicaciones reales los usuarios intentan engañar, usurpar a otros usuarios grabando su voz por ejemplo, lo cual se traduce en un problema mucho más complejo que lo abordado en este proyecto. Toda esta problemática, conocida actualmente como *‘Replay Attack’*, tiene en cuenta mucho de los principios establecidos en este proyecto, aunque la complejidad es mayor porque tiene que detectar por ejemplo si una persona está hablando en tiempo real o se trata de una grabación. También se podría considerar este tema como una interesante línea de continuación para el proyecto.