la intel.ligència artificial. El curs comprèn el desenvolupament de pràctiques sobre l'entorn del llenguatge PADD. Es van manipular estructures d'arbres per a la representació de les xarxes i de les probabilitats, i per llur càlcul partint de dades reals, a fi de representar una base de coneixement i interferència probabilística. Va comprendre unes vint hores de teoría i d'un període de pràctiques d'unes tres setmanes de duració.

5.- Conclusió.

La col.laboració professors-estudiants que ha causat aquesta breu article ha estat una agradable experiència, guiada per la motivació i la il.lusió d'apendre per part de tots. El curs va anar molt be i tothom ha après tot divertint-se. Animem, doncs, a altres grups a fer el mateix !.

REFERENCIES.

[1] M. Bertran, ' On a Formal Definition and Application of Dimensional design' *Software-Practice and Experience*, Vol.18(11), pp 1029-1045 (november 1988).

[2] A. Papoulis, M. Bertran, *Sistemas y Circuitos*, (Chapter 8) Marcombo, Barcelona, 1989.

[3] M. Bertran, J. Forga, F. Oller, J.A. Frau, ' PADDS: an envirement for the design of concurrent systems by simulation', *Proc. II Jornades sobre concurrència*, Universitat de les illes Balears, setembre1990, 71-83.

[4] M. Bertran, J. Forga, F. Oller, J. Viaplana, F. Alvarez-Cuevas, M. Porta, ' Integrated simulation and design of communication systems in a PADD envoirement', *IEEE Communications society Computer Aided Modeling and Design of Communications Links and Networks CAMAD'92*, Montebello, Quebec, Canada, 29 september- 2 october 1992.

[5] M. Bertran, ' PADD: a Schema Notation Integrating Parallelism and Abstraction', Report, ETSE Telecom. (UPC), Barcelona, Autumn 1989. Also in *Proc. IEEE Com. soc. CAMAD'92*, Montebello, quebec, Canada, 1992.

Actividades en Tratamiento de Voz del Grupo de Procesado de Señal

J. Hernando (*)

Departamento de Teoría de la Señal y Comunicaciones

El Grupo de Procesado de Señal (GPS) desarrolla sus actividades de investigación y desarrollo en el Departamento de Teoría de la Señal y Comunicaciones (TSC) de la Universidad Politécnica de Cataluña (UPC). Dentro del grupo tres subgrupos comparten la misma infraestructura de laboratorios de investigación y la administración económica, cada uno de ellos centrado en temas de:

- Procesado de voz,
- Procesado de imagen y
- Procesado digital de señal en comunicaciones, radar y sonar

Las actividades del subgrupo de Procesado o Tratamiento de Voz abarcan todas las áreas de investigación relacionadas con la transmisión de la señal de voz y la comunicación oral hombre-máquina. Seguidamente se describirán las líneas principales en cada área.

Codificación de la señal de voz

Se trabaja en codificación de voz a **velocidades de transmisión media y baja**. Para minimizar la pérdida de calidad a estas velocidades se han desarrollado sistemas de cuantificación vectorial adaptativa de una o varias etapas que se han aplicado con éxito a varios codificadores de voz desde 16 kbps a 4,8 kbps.

En codificación de **audio de banda ancha** se investiga en los tres niveles de calidad que suelen distinguirse: en el primer nivel, 7 kHz, se exige una calidad conversacional; en el segundo nivel, de 7 a 15 kHz, se exige un buen sonido musical; y en el tercer nivel, 20 kHz se requiere una calidad de alta fidelidad para la señal musical. Cada uno de estos niveles plantea problemas distintos que se han de atacar con soluciones específicas. Por ejemplo, en el tercer nivel se pretende alcanzar una calidad de CD en el rango de 96-128 kbps (¡700 kbps en los sistemas CD actuales!). Para ello se requiere un aprovechamiento exhaustivo de las características psicoacústicas de enmascaramiento espectral del sistema auditivo humano, las bandas críticas y los umbrales de distorsión audibles. Se están analizando soluciones basadas en nuevas transformaciones tiempo-frecuencia.

Reconocimiento del habla

Para robustecer el sistema de reconocimiento frente al ruido ambiente se estudian **nuevas representaciones** de la señal de voz que sean resistentes al mismo desde dos enfoques distintos: uno desde el punto de vista de procesado de señal y otro que trata de emular la capacidad auditiva humana. Con respecto al primer enfoque, se trabaja en la estimación del espectro analítico o la envolvente espectral en lugar del propio espectro, la descomposición en subespacio de señal y subespacio de ruido, etc. En cuanto al segundo enfoque, se consideran la sensibilidad logarítmica en frecuencia y en intensidad del oído y también el efecto de las bandas críticas. Con el mismo propósito se estudian representaciones dinámicas o **filtradas** del habla y **medidas de distancia** entre vectores de parámetros.

Se está desarrollando un sistema de **reconocimiento de habla continua** para el español que utiliza la **semisílaba** como unidad de reconocimiento y los **modelos ocultos de Markov** para describir de forma probabilística los características del habla. Se ha elegido la semisílaba como unidad de reconocimiento debido a la estructura silábica del español y a que existe un número relativamente reducido de ellas: menos de 750 en español. En cuanto a los modelos ocultos de Markov son los que en la actualidad proporcionan mejores prestaciones en los sistemas en desarrollo. A este sistema se le ha bautizado con el nombre de RAMSES (Reconocimiento Automático Mediante SEmiSílabas). En el problema

del reconocimiento de cadenas de números del 0 al 999, la tasa de reconocimiento de RAMSES ronda el 96 %.

Se investiga también en el problema de la detección de un conjunto dado de palabras en un discurso continuo (***word spotting***). En este campo se trabaja en algoritmos eficientes de búsqueda y en el modelado mediante modelos ocultos de Markov de las palabras de dicho conjunto, palabras clave, y de las que no están incluidas en él y los sonidos extraños.

Además, se estudian sistemas híbridos de reconocimiento de palabras aisladas basados en la integración de **redes neuronales** en el contexto de los modelos ocultos de Markov.

Reconocimiento del locutor

Se está iniciando la investigación en el área de reconocimiento del locutor, tanto en su versión de **identificación**, cuando el hablante no declara su identidad, como en la de **verificación**, cuando se ha de decidir si el hablante falsea o no su identidad. Para ello se están utilizando modelos autorregresivos vectoriales de la evolución temporal de los parámetros de la señal de voz.

Síntesis del habla

Se está desarrollando un sistema de **conversión texto-voz** mediante la **concatenación** de difonemas y trifonemas en español y catalán. En particular, se investiga la normalización de las unidades extraídas de un texto leído y los tipos de patrones de entonación representativos.

Mejora de la señal de voz

Una posibilidad de mejora de señal de voz estudiada es la **cancelación adaptativa de ruido**, que puede utilizarse cuando se dispone de una o varias referencias de ruido obtenidas a partir de las fuentes externas correspondientes mediante micrófonos auxiliares. En particular, se ha aplicado esta técnica a la cancelación de ruido en el interior de un automóvil tomando las referencias en el motor y las ruedas del mismo. Para esta aplicación, se propone utilizar filtros IIR adaptativos autoestabilizados y rápidos para identificar los canales de transmisión de la señal y añadir un grado de libertad (un filtro adicional) que permita la inversión exacta de los caminos de fase no mínima.

Se está considerando además la utilización de **análisis estadísticos de orden superior** en algunos algoritmos clásicos de mejora de la voz basados en un análisis de segundo orden, teniendo en cuenta la habilidad de los análisis de orden superior para distinguir entre procesos gaussianos y no gaussianos o entre procesos con pdf simétrica o asimétrica. Se ha aplicado este tipo de análisis al filtrado de Wiener iterativo. Se espera aplicar estas técnicas al reconocimiento robusto del habla.

Análisis de la señal de voz

La **estimación del tono**, la frecuencia fundamental de la señal de voz, en los sonidos sonoros y la **detección de sonoridad**, es decir, la decisión de si un segmento de señal de voz es sonoro o sordo, son importantes en muchas aplicaciones de procesado del habla (codificación, reconocimiento del locutor, síntesis, ayuda a discapacitados, etc.), pero tienen difícil solución debido a que la periodicidad de la señal de voz no es exacta. En particular, se ha trabajado en el caso de que la señal de voz está contaminada por ruido: se han desarrollado sistemas robustos basados en el espectro analítico y en estadísticas de orden superior y se ha realizado un sistema robusto de seguimiento de los valores proporcionados por un estimador clásico.

En los sistemas de codificación y reconocimiento, una buena detección de **voz/silencio** es fundamental. El objetivo es desarrollar un detector con el mínimo retardo que sea además robusto al ruido ambiental.

Bases de Datos

Junto con otros cinco grupos españoles, el grupo participa en el diseño y recolección de la base de datos en español ALBAYZIN para reconocimiento del habla. El grupo es el encargado de la coordinación del proyecto y además participa en el desarrollo de la base de datos fonética. El grupo también participa en la elaboración de otras bases de datos dentro del marco de proyectos europeos.

Aplicaciones en DSP´s

Se han realizado varios sistemas de **codificación de voz**, cubriendo todo el rango de las principales velocidades de transmisión: LPC-10 a 2,4 kbps, CELP a 4,8 kbps, Codificador Multipulso a 9,6 kbps, Codificador Subbanda a 9,6 y 12 kbps, Codificador Adaptativo a 9,6 kbps y CVSD a 16 y 32 kbps.

Se han desarrollado dos sistemas de **reconocimiento de habla** en tiempo real sobre DSP's. El primero es un reconocedor de palabras aisladas mediante modelos ocultos de Markov que es capaz de reconocer más de 100 palabras en tiempo real y ha sido aplicado al «juego de los barcos» en un PC. El segundo es un reconocedor de palabras clave (*word spotting*) también mediante modelos ocultos de Markov con un vocabulario de hasta 100 palabras clave y varios modelos de palabras fuera del vocabulario y sonidos extraños según la aplicación. Este último sistema se ha aplicado con éxito a detectar en un discurso continuo los comandos usados en IBERCOM. Esta aplicación, conocida como TELEMACO, trabaja en un entorno de PC y reconoce 25 palabras clave utilizando un modelo para las palabras fuera del vocabulario y los sonidos extraños. Ahora se está trabajando en el desarrollo de una versión simplificada de RAMSES sobre DSP para reconocer los número de cero a un millón en tiempo real.

(*) Información dada por todos los profesores del grupo.