

	Procesamiento Digital de la Información GITT-Grado en Ingeniería en Tecnologías de las Telecomunicaciones Departamento de Teoría de la Señal, Sistemas Telemáticos y Computación Universidad Rey Juan Carlos - Curso 2023/24
--	---

Laboratorio de Habla 3

Construcción de un Vocoder

Objetivos de Aprendizaje

1. Ser capaz de manipular y visualizar señales de voz en Julia.
2. Ser capaz de realizar un análisis LPC de la señal de voz en Julia.
3. Ser capaz de reconstruir señales usando filtros de reconstrucción.

Consejos para la Solución

1. Utiliza las convenciones de Julia para escribir código. Puedes encontrarlas en:

https://web.mit.edu/julia_v0.6.2/julia/share/doc/julia/html/en/manual/style-guide.html

2. Si copias el texto de los enunciados como parte (comentada) de un fichero de Julia podrás ejecutarlo y obtener un fichero que entregar.
3. Mi consejo es que intentes “crear una historia” que haga amena la lectura del entregable.
4. Recuerda que en la puntuación de entregables cuenta tanto el contenido como la presentación.
5. Esta práctica se puede hacer con Julia básico o usando paquetes de procesamiento de señal, entre los cuales se encuentran **SampledSignals.jl**, **PortAudioStream.jl**, **SignalAnalysis.jl**, **DSP.jl** ó **AuditoryStimuli.jl**, además de **WAV.jl**. Estos paquetes permiten hacer cálculos mucho más potentes basado en el concepto de flujos de datos de una fuente y de bloques de procesamiento, pero implican una mayor complejidad conceptual y conocimientos operativos, que se alcanzarán empleando más tiempo para su estudio.
6. Un paquete que merece la pena investigar es **Unitful.jl** que pone unidades del SI en todas las cantidades. Su propósito es el mismo que el de aquellas: que las señales y constantes que construimos tenga un sentido físico.

Introducción: la Vocodificación LPC

Para entender la codificación por predicción lineal (“Linear Predictive Coding”, LPC en inglés) se puede consultar los apuntes del tema 4.3 (p. 30 y ss.) o cualquier libro de procesamiento digital de audio avanzado como el que aparece debajo. Para una introducción somera, se puede leer el artículo de la Wikipedia:



Figura 1: Diagrama de un analizador (izquierda) y sintetizador (derecha) de voz LPC

La figura 1 muestra un diagrama de bloques de un vocoder (“voice encoder-decoder”, en inglés). Como existe una correlación grande entre muestras de voz, la ganancia predictiva entre muestras es muy alta y

	Procesamiento Digital de la Información GITT-Grado en Ingeniería en Tecnologías de las Telecomunicaciones Departamento de Teoría de la Señal, Sistemas Telemáticos y Computación Universidad Rey Juan Carlos - Curso 2023/24
--	---

se pueden usar técnicas de codificación lineal para mejorar la tasa binaria de “representación” de la señal, tanto para su transmisión como para su almacenamiento.

La teoría de predicción lineal para voz es muy sencilla [1]: sea una señal de voz muestreada $s[n]$ cada una de cuyas muestras queremos predecir mediante p muestras anteriores:

$$\hat{s}[n] = \sum_{k=1}^p a_k s[n-k]$$

En donde el parámetro p es el orden del filtro de predicción. El error cometido en la estimación es entonces:

$$e[n] = s[n] - \hat{s}[n] = s[n] - \sum_{k=1}^p a_k s[n-k]$$

Tomando la transformada Z del error y calculando la función de transferencia del *filtro análisis o predicción*:

$$E(z) = Z\{e[n]\} = Z\{s[n] - \sum_{k=1}^p a_k s[n-k]\} = S(z) \left(1 - \sum_{k=1}^p a_k z^{-k} \right)$$

$$H^{-1}(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p a_k z^{-k}$$

Tomando la función de transferencia inversa construimos el *filtro de síntesis o reconstrucción*:

$$H(z) = \frac{1}{H^{-1}(z)} = \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}}$$

Esta solución, de momento, tiene como parámetros libres los coeficientes del filtro $\{a_k\}$. Para estimarlos (temas 2 y 3) usamos el criterio de minimizar el error de predicción para cada trama de análisis:

$E = \sum_n e^2[n]$. Por lo que derivamos el error en los coeficientes $\frac{\partial E}{\partial a_k} = 0$ para $k = 1, \dots, p$

$$\frac{\partial E}{\partial a_i} = 0 = \sum_{n=-\infty}^{\infty} s[n]s[n-i] - \sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} s[n-k]s[n-i]$$

Y definiendo la autocorrelación de la señal como $R[k] = \sum_{n=-\infty}^{\infty} s[n]s[n-k]$ tenemos el sistema de ecuaciones:

$$0 = R[i] - \sum_{k=1}^p a_k \sum_{n=-\infty}^{\infty} R[k-i] \quad j = 1, \dots, p$$

Sistema de ecuaciones que expresado en forma matricial queda:

$$\begin{bmatrix} R[0] & R[1] & \dots & R[p-1] \\ R[1] & R[0] & \dots & R[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ R[p-1] & R[p-2] & \dots & R[0] \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R[1] \\ R[2] \\ \vdots \\ R[p] \end{bmatrix}$$

O escrito en forma compacta $R \cdot a = r$. Dado que la matrix R es Toeplitz, hay diversos modos de resolver el sistema, pero para órdenes p pequeños la inversión es un método factible.

Use los módulos a los que está acostumbrad@ en esta práctica: por ejemplo WAV.jl, DSP.jl, etc.

	Procesamiento Digital de la Información GITT-Grado en Ingeniería en Tecnologías de las Telecomunicaciones Departamento de Teoría de la Señal, Sistemas Telemáticos y Computación Universidad Rey Juan Carlos - Curso 2023/24
--	---

Referencias

[1] Deller, John R Jr, John H L Hansen, and John G Proakis. 2000. *Discrete-Time Processing of Speech Signals*. Wiley-IEEE Press.

[2] Linear Predictive Coding. https://en.wikipedia.org/wiki/Linear_predictive_coding

Ejercicio 1: Análisis Exploratorio de la Señal

Objetivo

Determinar si la señal proporcionada puede ser sujeta a codificación LPC efectiva.

Contenido

1. Lea y reproduzca la señal de voz suministrada con este enunciado.
2. Represente el sonido en el dominio del tiempo.
3. Represente la energía localizada y la tasa de cruces por cero como función del tiempo, entramando la señal con diferentes tamaños de ventana y usando las primitivas desarrolladas en la práctica 2.

Ejercicio 2: Análisis de una trama

Objetivo:

Desarrollar el código de análisis-síntesis.

Contenido

1. Seleccione y visualice en el tiempo y la frecuencia una trama sonora de la señal.
2. Obtenga la función de autocorrelación de la trama usando una función `xcorr` que reciba una trama y un orden y obtenga el vector de autocorrelación de ese orden.
3. Genere la matriz R usando la biblioteca `ToeplitzMatrices.jl`
4. Obtenga los coeficientes de predicción lineal usando el sistema anterior.

Ejercicio 3: Síntesis de una trama

Objetivo

Depurar el código de análisis-síntesis.

Contenido

	Procesamiento Digital de la Información GITT-Grado en Ingeniería en Tecnologías de las Telecomunicaciones Departamento de Teoría de la Señal, Sistemas Telemáticos y Computación Universidad Rey Juan Carlos - Curso 2023/24
--	---

1. Obtenga, para la trama dada, el error de predicción como una señal, filtrando la trama a través del filtro FIR inverso.
2. Sintetice la trama de voz analizada mediante el filtro de síntesis calculado en el apartado 1.4
3. Inspeccione, auditivamente, en el tiempo y en la frecuencia las tramas de voz original y sintetizada. ¿Son correctos los coeficientes de análisis-síntesis?

Ejercicio 4: Síntesis LPC

Objetivo:

Realizar el análisis-síntesis LPC de una elocución de voz.

Contenido

En este ejercicio vamos a obtener el resto de los parámetros necesarios para codificar, transmitir y reconstruir la señal de voz con efectividad.

1. Identificar si cada trama de voz es sonora o sorda usando tanto la tasa de cruces por cero como la energía localizada.
2. Calcular la frecuencia fundamental de las tramas sonoras, si puede, usando la FFT.
3. De nuevo, calcule la frecuencia fundamental pero usando en este caso la autocorrelación. ¿En qué propiedad fundamental de la teoría de señales se basa este método? Para responder, reflexione acerca de lo que es la función de autocorrelación desarrollada más arriba.
4. Encuentre una trama sorda y una trama sonora de su señal de voz como en 1.3 y calcule y visualice sus autocorrelaciones como en 2.2. Para la trama sonora, estime el tono fundamental como en 4.1
5. Halle los espectros de las tramas seleccionadas y represéntelos para explorarlos.
6. Construya una función que reciba una trama y devuelva todos los parámetros de análisis necesarios para re-sintetizar la trama.
7. Construya una función que reciba todos los parámetros de análisis y reconstruya una trama de voz en el tiempo.
8. Para toda la señal de voz $s[n]$ diseñe e implemente un esquema de enventanado con ventanas de longitud N que se solapen un 50% de la longitud. Ayúdese de alguna referencia técnica si no entiende qué tiene que hacer. También, decida qué orden de filtro LPC va a usar...
9. Sobre el diseño de enventanado anterior, cuando pase por una trama, analicela y sintetícela para construir una señal aproximada $\hat{s}[n]$.
10. Compare $s[n]$ y su aproximación $\hat{s}[n]$ visual y auditivamente. Investigue diferentes valores del orden del filtro. ¿Puede observar algún efecto en la señal reconstruida?

	Procesamiento Digital de la Información GITT-Grado en Ingeniería en Tecnologías de las Telecomunicaciones Departamento de Teoría de la Señal, Sistemas Telemáticos y Computación Universidad Rey Juan Carlos - Curso 2023/24
--	---

Ejercicio 5: Reflexión

Es deseable que cualquier actividad que curse conlleve un aprendizaje. Este aprendizaje será más profundo y satisfactorio si usted es capaz de darse cuenta de él y enunciarlo.

Objetivo

Aprender a descubrir y ser consciente de los resultados de aprendizaje propios.

Actividad

Escriba un párrafo listando qué conocimientos y destrezas que haya adquirido en esta práctica.