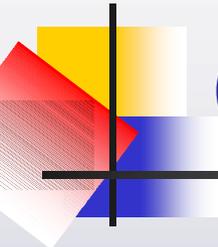


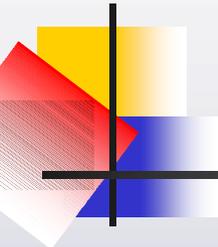
INTRODUCCIÓN A LAS TECNOLOGÍAS DEL HABLA

Presentado por: Luis Fernando D´Haro



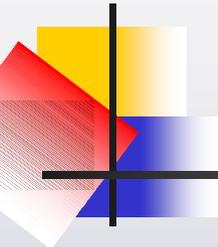
Contenido

- Introducción y perspectiva histórica
- Campos de aplicación
- Retos de las Tecnologías del Habla
- Futuro de las T.H
- Conclusiones



Definición de T.H

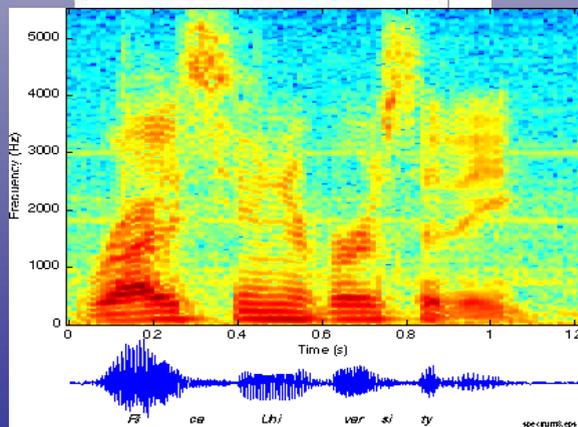
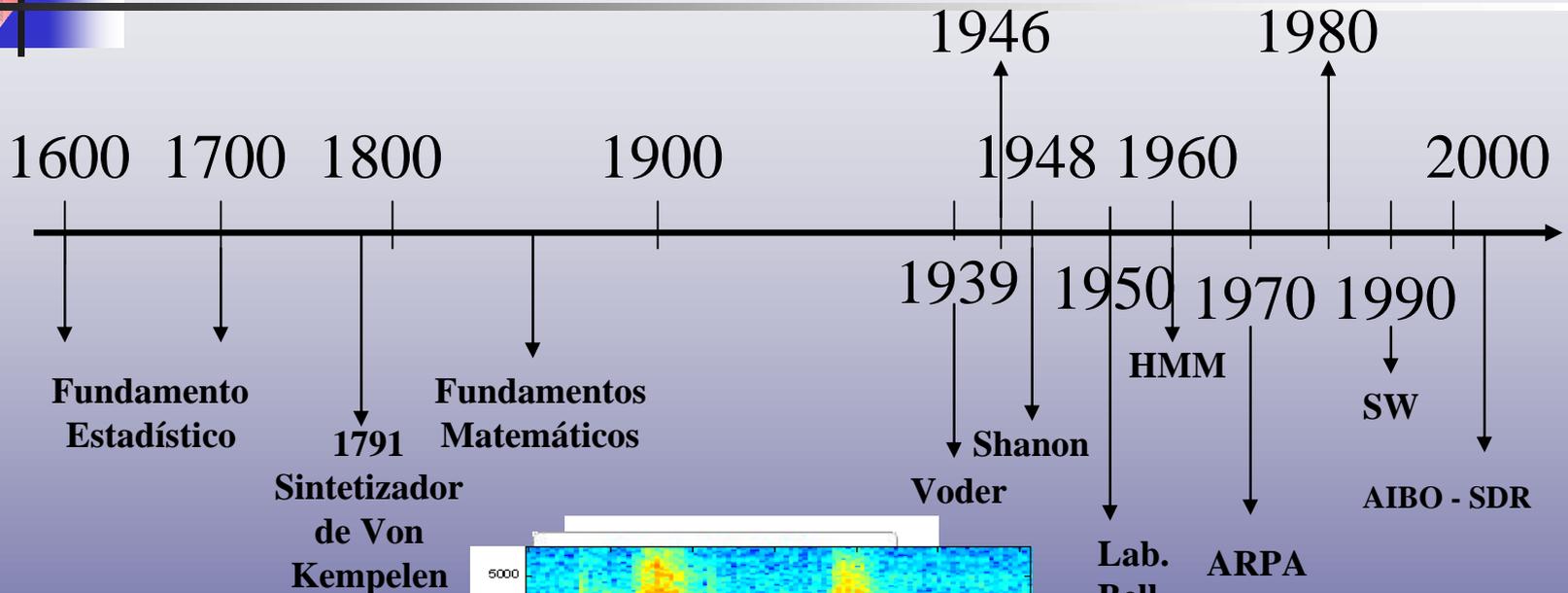
- Procesos necesarios para realizar una interfaz oral, es decir, aquella que nos permita comunicarnos con una máquina o en la comunicación entre personas.



Introducción

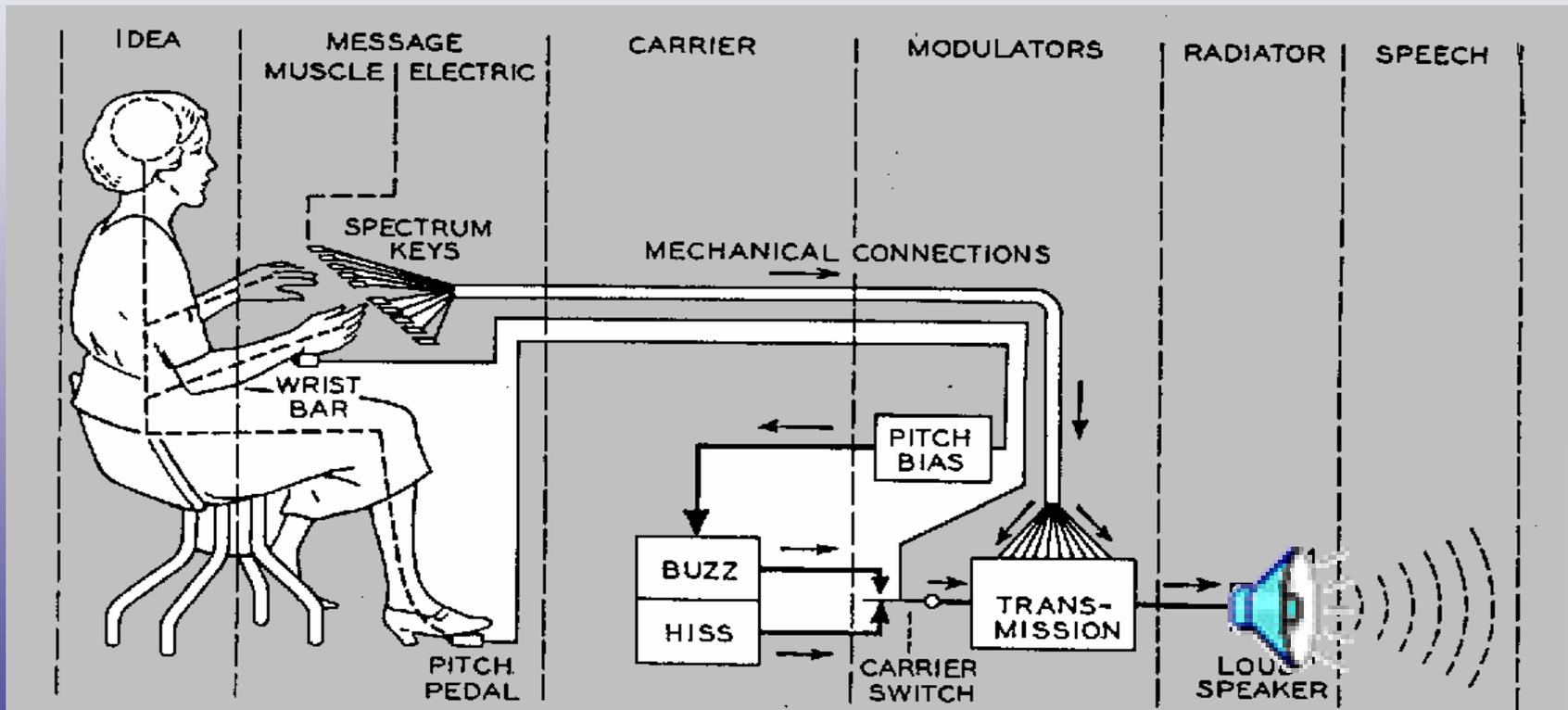
- Necesidad de comunicación hombre-máquina
- Disminución de costos de Tx – Rx
- Ayudas a personas con discapacidad
- Sistemas de autenticación
- Facilitar el intercambio de información

Perspectiva Histórica

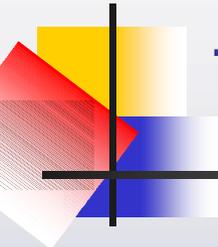


Provided by Sony

Esquema del Voder



Schematic circuit of the voder.



Tecnologías Básicas

- ✓ Codificación de Voz
- ✓ Síntesis de voz
- ✓ Reconocimiento automático del habla
- ✓ Verificación o identificación de locutor

- ✓ Identificación de idiomas

Tipos de Aplicaciones

- Locales:
 - Discapacitados: lectura, dictado, procesos ágiles.
 - Aprendizaje de idiomas
 - Generación de reportes, trabajos, informes.
 - Aplicaciones multimedia



Tipos de Aplicaciones

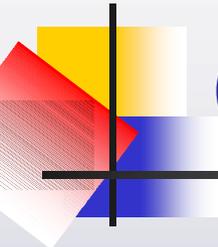


- Respuestas vocales interactivas: Reemplazo al DTMF
- Difusión de Información: Noticias, info. Bursátil.
- Captura de Información: Votaciones, compras por teléfono
- Sistemas de Dialogo: Iniciativa del sistema, usuario o mixta.

Tipos de Aplicaciones

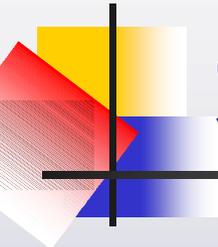
- ✿ Automatización de sistemas telefónicos: correo electrónico, mensajería integrada.
- ✿ Marcación vocal (manos libres)
- ✿ Directorio público
- ✿ Identificación de abonado
- ✿ Despertador
- ✿ Cobros revertidos





Codificación de Voz

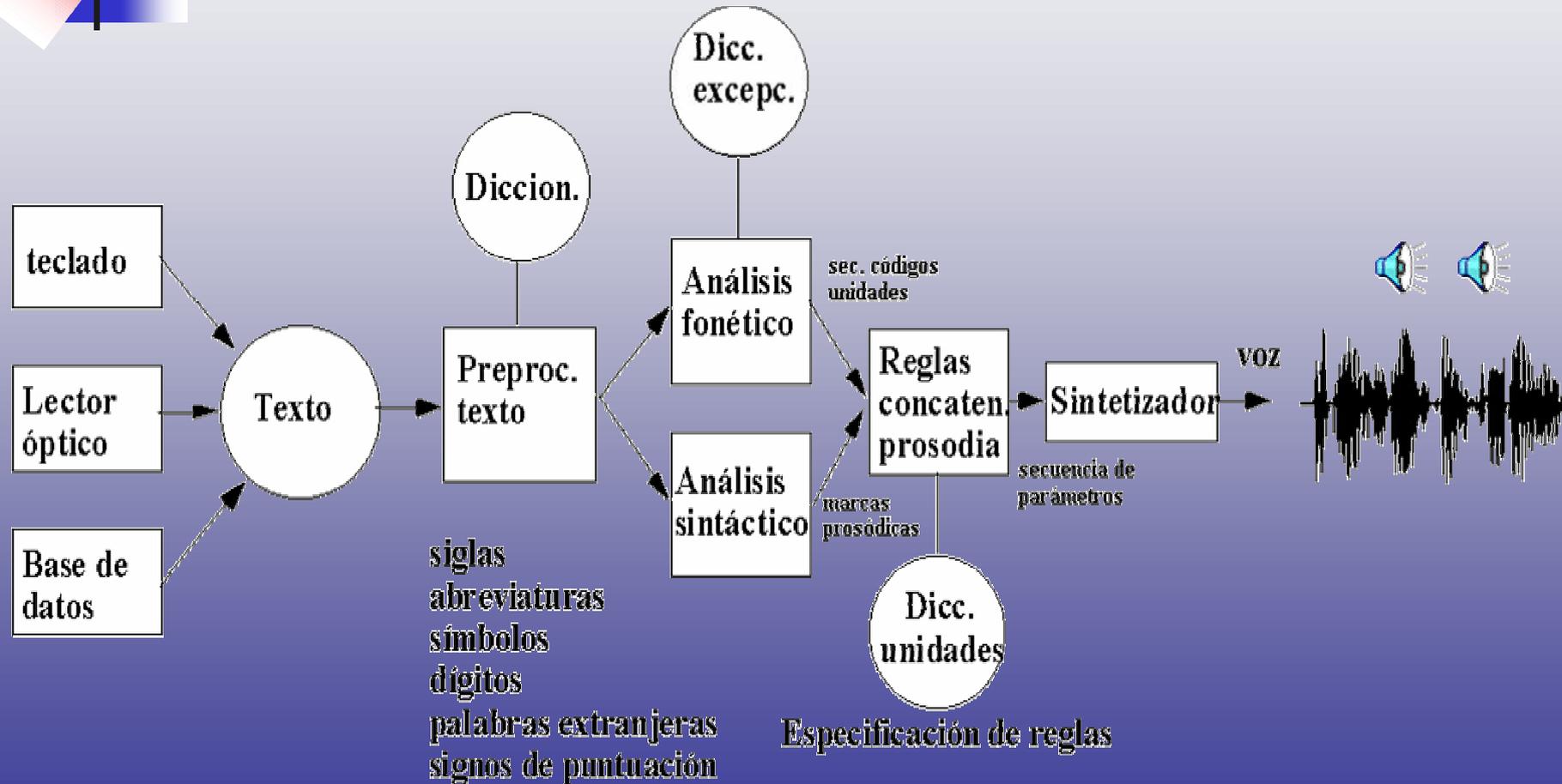
- Finalidad: Compresión de la información de la señal de voz para Tx o Almacenar economizando BW del canal o del soporte de almacenamiento
- Conversión Análoga – Digital
- Factor de compresión: Frecuencia de muestreo (Banda), Número de Bits, Canales.
- Codificación: Ley A, Ley μ , PCM, ADPCM, MP3, LPC, CELP.
- Conversión Digital - Análoga



Síntesis de Voz

- Finalidad: Réplica sintética de una señal de voz para Tx un mensaje desde una máquina.
- Campos: Suministrar información variable o elevada.
- Beneficio: Sistemas amigables
- Bajo almacenamiento (reducción de más de 200 veces si ADPCM)

Esquema de un CTV



Síntesis de Voz



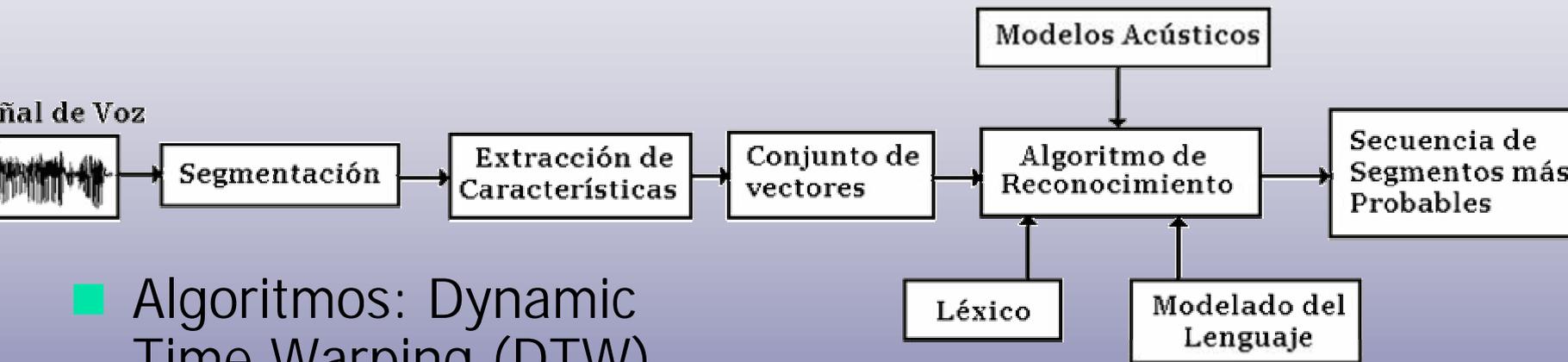
- Tipos: Articulatorios, Paramétrico o por Formantes, Concatenación de onda, Visual.
- Unidades: Fonemas, difonemas, trifonemas, demisilabas, silabas, palabras, morfemas.
- Conocimientos: Fonéticos, Ortográficos, Fonológicos, TDS.
- Ventaja del español: Correspondencia Fonema – Grafema
Reglas de transcripción

Reconocimiento de Voz

- Finalidad: Identificación de las palabras y estructuras lingüísticas complejas que forman el lenguaje hablado.
- Tipos: Dependiente/Independiente de Locutor
- Habla aislada/Habla continua
- Dominio restringido/abierto (flexible?)



Reconocimiento de Voz



- Algoritmos: Dynamic Time Warping (DTW)

- Hidden Markov Models: Discretos, semicontinuos y continuos.

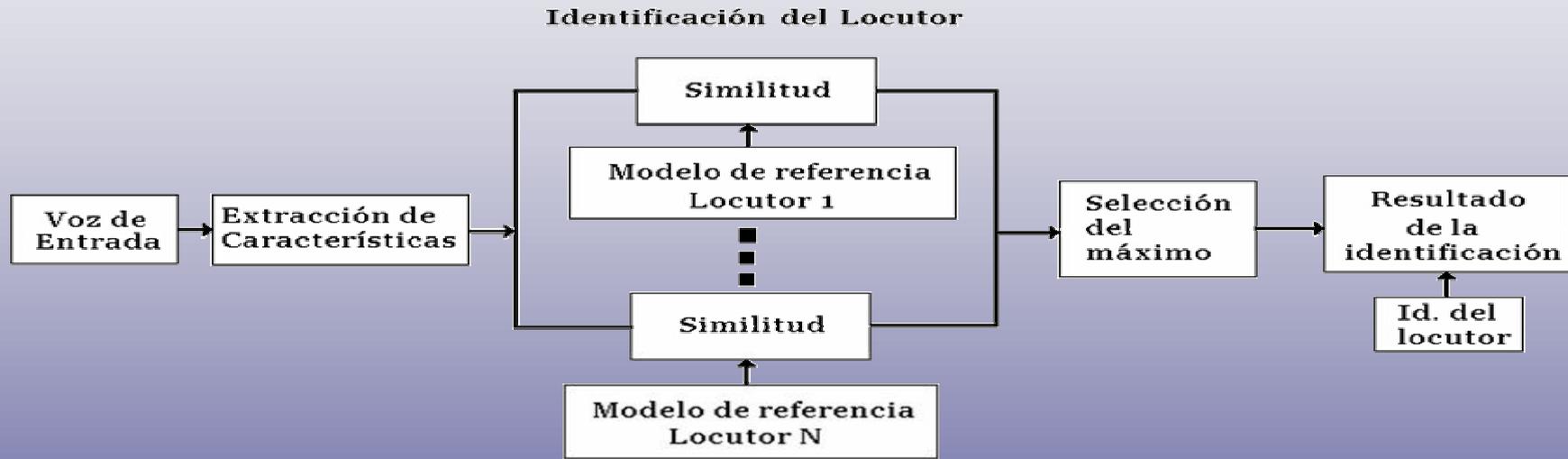
- Modelos de Lenguaje: N-gramas.

Identificación del Locutor

- Identificación del Locutor:
Determinar que usuario registrado es el que habla.
- Verificación del Locutor:
Aceptación o rechazo de una identificación dada.
- Aplicaciones: Control de acceso a servicios de voz, compras, bases de datos, correo de voz, control de seguridad.
- Problemas: Grabaciones. Tasas de aceptación falsas, tasa de rechazo verdadero.

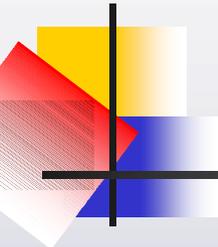


Identificación del Locutor



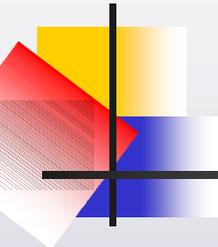
Verificación del Locutor





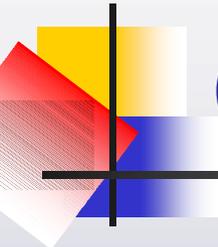
Dificultades de las T.H

- Variabilidad acústica de la voz y de los sonidos del habla.
- Ruidos externos y de las personas.
- Parámetros insuficientes para el modelado?
- Bases de datos
- Reducción del tiempo de procesamiento Vs Algoritmos y Tamaño del vocabulario y modelos.
- Idiomas.



Futuro y Retos de las T.H

- Compresión:
Disminuir la tasa sin perder naturalidad.
Bajo coste computacional (RT)
- Reconocimiento:
Independencia de locutor, vocabulario (tarea), idioma.
- Síntesis:
Naturalidad
Emotividad
- Identificación de Locutor:
Engaños con grabaciones



Conclusiones

- Diversos campos de aplicación
- Amplias posibilidades de desarrollo
- Explotación – Limitación
- Desarrollos con IA y DSP (SW, HW)
- Trabajo multidisciplinar
- Español: Facilidad de reglas, pocos desarrollos, comunidad grande