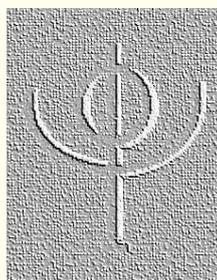


ISSN: 0325-2043



LABORATORIO DE INVESTIGACIONES SENSORIALES (LIS)

Informe XLVII–2014

I N I G E M



CONICET

U B A

Instituto de Inmunología, Genética y Metabolismo
Córdoba 2351, Piso 9, (1121), Buenos Aires
Tel/Fax: 5950-9024
lis@fmed.uba.ar — <http://www.lis.secyt.gov.ar>

Índice

1. Introducción	1
2. Personal	1
3. Proyectos de Investigación	2
3.1. UBACyT CM12, “Pruebas clínicas de análisis de la voz y el habla. Mejoras en la evaluación Audio-Perceptual Evaluación objetiva de la Prosodia”	2
3.1.1. Trabajos terminados	2
3.1.2. Trabajos en desarrollo	2
3.2. Desarrollo de un sistema de conversión de texto en habla para su aplicación en sistemas de telecomunicaciones	3
3.2.1. Trabajos en desarrollo	3
3.3. Desarrollo de un sistema de reconocimiento automático del habla para el español de Argentina	4
3.3.1. Trabajos en desarrollo	4
3.4. Sistema de Información Plenóptica como medio diagnóstico para Lámparas de Hendiduras	5
3.5. CONICET PIP Nro. 5897/06: Análisis de las sensaciones de dulce, agrio y amargo en soluciones puras y mezcladas en medio acuoso y alcohólico	5
3.5.1. Trabajos terminados	6
3.5.2. Trabajos en desarrollo	6
4. Proyectos de investigación conjunta	6
4.1. Proyecto de Cooperación MINCYT-BMBF (AL/12/06) Utilización de Prosodia para el Reconocimiento Automático de Habla	6
4.1.1. Trabajos terminados	6
5. Docencia	8
5.1. Cursos de grado	8
5.2. Cursos de posgrado	8
6. Transferencia Tecnológica	8
6.1. Reconocimiento automático del habla en medios de comunicación	8
7. Intercambio Científico	9
7.1. Visitas al LIS	9
7.2. Estadías en el Exterior de Miembros del LIS	9
8. Propiedad intelectual	9
9. Tesis	9
9.1. Doctorales en curso	9
10. Actividades de Divulgación	12

11. Trabajos que refieren a actividades del LIS	12
11.1. Publicaciones	12
11.2. Presentaciones a congresos	13
12. Publicaciones	13
12.1. Revistas	13
12.2. Congresos	13
12.3. Informes Técnicos	13

1. Introducción

Desde su creación en el año 1968, el LIS publica un informe anual en donde se consignan las publicaciones realizadas, los trabajos en curso, la actividad docente y el intercambio científico.

Los Informes LIS están registrados bajo ISSN 0325-2043 (International Standard Serial Number), a través de Latindex¹, reconocido internacionalmente para la identificación de las publicaciones seriadas. La serie comienza con el Informe I-1968, Laboratorio de Investigaciones Sensoriales, CONICET.

En los informes aparecen siglas que referencian las sedes del LIS, primero en el Hospital Escuela (HE), luego en la Facultad de Medicina (FM) y, actualmente, en el Hospital de Clínicas (HC) de la Universidad de Buenos Aires.

Desde el año 1997, los informes también están disponibles a través del sitio web del laboratorio: <http://www.lis.secyt.gov.ar/>.

2. Personal

Investigadores

- EVIN Diego, Bioingeniero, Dr. en Ciencias de la Computación.
- GUIRAO Miguelina, Prof. Filosofía, Dra. en Psicología Experimental.
- GURLEKIAN Jorge A., Ing. Electrónico, Dr. en Medicina. Responsable del LIS.
- TORRES Humberto, BioIngeniero, Dr. en Ingeniería.

Investigadores que participan en proyectos que se desarrollan en el LIS:

- CALVIÑO Amalia M., Farmacéutica, Dra. en Bioquímica.
- GRAVANO Agustín, Licenciado y Dr. en Ciencias de la Computación.
- VACCARI María Elena, Lic. en Fonoaudiología.

Becarios

- COSSIO MERCADO Christian, Ing. en Informática, Becario CONICET. Tesista de Doctorado UBA.
- MARTINEZ SOLER Miguel, Ing. en Informática, Becario CONICET, Tesista de Doctorado, UBA.

Tesistas

- SIGAL Liliana, Lic. en Fonoaudiología. Tesista de Doctorado UBA.
- TRIPODI Mónica, Lingüista. Tesista de Doctorado UBA.
- UNIVASO Pedro, Ing. Electrónico. Tesista de Doctorado UBA.

¹Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal. Sitio: <http://www.latindex.unam.mx>

3. Proyectos de Investigación

3.1. UBACyT CM12, “Pruebas clínicas de análisis de la voz y el habla. Mejoras en la evaluación Audio-Perceptual Evaluación objetiva de la Prosodia”

Director: María E. Vaccari (UBA-Fonoaudiología), Co-director: Jorge A Gurlekian, Investigadores: Humberto Torres, Diego Evin, Liliana Sigal.
Período: 2011–2014.

3.1.1. Trabajos terminados

Guía para la segmentación y transcripción fonética en las tecnologías de habla

Gurlekian JA; Torres HM; Evin DA

Resumen:

En este trabajo se define una guía para la segmentación fonética y su transcripción mediante el alfabeto SAMPA (Speech Assessment Methods: Phonetic Alphabet). La transcripción fonética con SAMPA es de uso creciente en las tecnologías de habla en las tareas de preparación de datos acústicos para ser utilizados en el entrenamiento de sistemas de reconocimiento de habla/hablante y en la preparación de unidades de síntesis de habla.

Aplicación de los índices de perturbación integrado y de precisión articulatoria en pacientes con disfonía espasmódica

Sigal L. y Gurlekian J.A.

Resumen:

La disfonía espasmódica es un desorden vocal severo caracterizado por interrupciones involuntarias de la fonación, denominada también distonía focal laríngea. Este trabajo da a conocer resultados de la aplicación del índice de perturbación integrado (IPI), y de precisión articulatoria (IPA), en una población de pacientes de ambos sexos con disfonía espasmódica, y presenta la comparación con un grupo control, sujetos con diagnóstico otorrinolaringológico de cuerdas vocales móviles, sin daños estructurales en la mucosa glótica, ni alteraciones neurológicas que afecten la dinámica cordal. El índice de perturbación agrupa a cuatro parámetros: Jitter, Shimmer, relación Armónico/Ruido y la Amplitud Diferencial del Cepstrum. Los resultados muestran diferencias significativas entre los grupos de casos y controles femeninos y masculinos en los parámetros de índice de perturbación integrado y de grado de precisión articulatoria, por lo que ambos pueden considerarse indicadores apropiados para caracterizar la falta de estabilidad en la función articulatoria que se produce en las cavidades resonanciales de las pacientes con disfonía espasmódica.

3.1.2. Trabajos en desarrollo

Análisis de la percepción de perturbaciones en la frecuencia fundamental

El análisis de la percepción de perturbaciones en la frecuencia fundamental es uno de los trabajos concluidos en este período. Las variaciones involuntarias de la frecuencia fundamental que se producen ciclo a ciclo, y que comúnmente se denomina jitter, es un parámetro de la voz que está directamente asociada a valoraciones como la ronquera. Puede tener orígenes neurológicos, anatómicos, entre otros. En la clínica la valoración del jitter es un estudio perceptual y el resultado depende fuertemente del profesional que lo realiza. El trabajo estuvo

orientado a explorar como percibimos los valores jitter frente a estímulos controlados, midiendo las diferencias inter e intra sujetos evaluadores de los estímulos. Los resultados del análisis de los experimentos perceptuales muestran que la propuesta de utilizar patrones de referencia en el proceso de evaluación favorece la uniformidad de las valoraciones.

3.2. Desarrollo de un sistema de conversión de texto en habla para su aplicación en sistemas de telecomunicaciones

Director: Jorge Gurlekian, Co-directores: Humberto Torres, Diego Evin

3.2.1. Trabajos en desarrollo

Etiquetado automático de clases de palabras

Dada la importancia que han adquirido las clases de palabras, decidimos profundizar nuestros trabajos en esta área. Para ello se propuso trabajar con cinco corpus de texto: Notas extraídas de un diario, los discursos de la Presidenta de la Nación, textos de la enciclopedia en línea Wikipedia, discursos realizados en el Senado de la Nación, y diálogos en programas de radio. Se etiquetaron los tres primeros, y los dos últimos se encuentran en proceso de construcción. Una primera estadística de la distribución de las clases de palabras muestra claras diferencias en los corpora analizados. En el siguiente período se pretende realizar una estimación del nivel de desempeño del etiquetador, comparando su salida con una versión reducida, corregida en forma manual, sobre los tres corpora, y mediante un muestreo aleatorio. En la cuarta parte del informe técnico que se adjunta (Torres, 2014) se describen los trabajos realizados y los resultados obtenidos.

Modelización de la prosodia

La asociación del modelo acústico de entonación de Fujisaki con la información lingüística y las mejoras en la obtención de los parámetros del modelo han sido los principales temas de trabajo en el período. En el año 2013 habíamos propuesto introducir información lingüística en la inicialización del método de extracción de parámetros de Mixdorff (Torres et al, 2014), con el objetivo de poder asociar los comandos del modelo con el contenido lingüístico de las oraciones. Si bien se logra un aceptable margen de error en el ajuste de las curvas de entonación, un análisis posterior de los valores de los parámetros de los comandos del modelo reveló que el proceso de optimización de los parámetros diluían las posibilidades de relacionar los comandos con los eventos que le habían dado origen. Bajo la hipótesis de que el problema era la etapa de optimización, en este período propuse un nuevo método de extracción de parámetros (Torres y Gurlekian, 2015). Primero definimos que se espera de los elementos del modelo, relacionando eventos acústicos y lingüísticos. Mantuve las hipótesis para la inicialización ya presentadas, y en todo momento de la optimización posterior de los parámetros se mantiene la asociación entre comandos y eventos lingüísticos. Primero se propuso un método manual de estimación de parámetros, y luego una implementación computacional, con la cual se hicieron pruebas sobre tres corpora, en tres idiomas: alemán, castellano argentino e inglés. Con este nuevo método logramos tener una asociación directa entre comandos del modelo e información lingüística, y además, una mejora superior al 34% en la bondad del ajuste de la curva de F0 con respecto al método estándar. En la primera parte del informe técnico que se adjunta (Torres, 2014) se describen los trabajos realizados y los resultados obtenidos. Continuando con el análisis, se estimaron las distribuciones de los valores de los parámetros

de los comandos del modelo versus clases de palabras que le dieron origen. Se optó por una distribución de valores extremos para aproximar las distribuciones de los datos. Los resultados obtenidos muestran una marcada diferencia entre las clases de palabras, las cuales se pueden agrupar en base a los parámetros de las distribuciones estimadas. Resta realizar pruebas estadísticas para sustentar estos agrupamientos. Esta información podría ser utilizada para mejorar la predicción de los comandos del modelo a partir de un texto de entrada. Además, se espera poder verificar la viabilidad de estos agrupamientos a partir de teorías lingüísticas. En la tercera parte del informe técnico que se adjunta (Torres, 2014) se describen en detalle los trabajos realizados y los resultados obtenidos.

Corpus EMILIA

La tarea de etiquetado del corpus EMILIA aun no fue concluida, restan aun revisar el 2% de las oraciones. Para el próximo período se espera completar el corpus, extrayendo los datos necesarios para actualizar la voz de EMILIA en el sistema de conversión de texto en habla AROMO. Una vez terminado esta etapa, se realizará una prueba perceptual de la calidad del habla generada, en forma similar a la reportada en Gurlekian et al. (2012).

Desde períodos anteriores se viene trabajando en la construcción de corpus orales: corpus para reconocimiento de habla en canal telefónico (Gurlekian et al., 2001a), corpus para el estudio de la prosodia (Gurlekian et al., 2001b), corpus Atlas Multimedia de la Prosodia del Espacio Románico (Gurlekian et al., 2010a, Torres, 2014), corpus para la conversión de texto en habla (Torres, 2012), y corpus para el reconocimiento automático del habla en audios de radio y televisión¹. Esto nos ha llevado desarrollar una serie de definiciones, protocolos y herramientas para la construcción de estos corpus (Gurlekian et al., 2001c, 2003, 2010b). En este período, el conocimiento adquirido en segmentación y etiquetado fonético de corpus orales para el castellano de argentina fue resumido y publicado en la revista Fonoaudiológica (Gurlekian et al., 2014).

3.3. Desarrollo de un sistema de reconocimiento automático del habla para el español de Argentina

Director: Jorge Gurlekian, Co-director: Diego Evin, Humberto Torres

3.3.1. Trabajos en desarrollo

Construcción de un corpus de habla continua para Español de Buenos Aires

Uno de los elementos indispensables para el desarrollo de reconocedores del habla es contar con un gran cuerpo de datos conteniendo muestras del lenguaje que se pretende reconocer, y que pueda ser empleado como fuente de datos para el entrenamiento de los modelos del habla. En esta etapa se inició la construcción de un corpus de habla continua para Español de Buenos Aires. En este sentido, se trabajó en el diseño y la construcción de un corpus de habla, con material obtenido de medios radiales tanto AM como FM de la ciudad de Buenos Aires. El objetivo de este proyecto es conseguir un corpus con al menos 100 horas de habla etiquetadas, en consonancia con el tamaño de proyectos similares para otros idiomas que se pueden encontrar en la literatura. Esta tarea además de la recopilación de datos implica la separación, discriminación y selección del material de acuerdo a su contenido, ya sea habla limpia, habla de canal telefónico, proveniente de pautas publicitarias, música o con diferentes fuentes de ruido de fondo.

Construcción de sistemas de referencia para el reconocimiento de habla continua e identificación de palabras claves

El objetivo de esta tarea consiste en implementar sistemas de referencia para el reconocimiento de habla continua e identificación de palabras claves que puedan ser empleados en aplicaciones prácticas, de acuerdo al estado del arte del área. Empleando el material que se fue recopilando de medios radiales tanto AM como FM de la ciudad de Buenos Aires, se comenzaron a construir sistemas de reconocimiento automático del habla de gran vocabulario, y se diseñaron alternativas para el desarrollo de un primer sistema para la detección de palabras clave (*keyword spotting*).

Estudio de información prosódica relacionada con la identificación de palabras clave

La información prosódica del habla señala varios aspectos tanto lingüísticos como paralingüísticos que los seres humanos empleamos para comunicarnos. El objetivo de este tema de investigación es determinar las pistas informativas que se pueden encontrar en este canal y que puedan servir para facilitar la detección de palabras que resultan de mayor relevancia para el hablante o el oyente. Se trabajó de manera colaborativa con investigadores de la Beuth Hochschule für Technik en el estudio de la manifestación acústica de prominencia y foco.

3.4. Sistema de Información Plenóptica como medio diagnóstico para Lámparas de Hendiduras

Director: Bartolomé Drozdowicz, Codirector: Adrián Carlos Salvatelli

Investigadores: Gustavo Bizai, Alejandro Hadad, Diego Evin, Rodrigo Torres

Resumen:

En oftalmología, las imágenes resultan fundamentales tanto para establecer diagnósticos como para evaluar la evolución y respuesta terapéutica. La biomicroscopía del fondo de ojos es un estudio sencillo que utiliza una lámpara de hendidura, equipo disponible en la gran mayoría de los consultorios oftalmológicos, para la detección de maculopatías. Una de las principales causas de ceguera legal a nivel mundial es la maculopatía relacionada con la edad y la retinopatía diabética. El presente proyecto propone brindarle valor diagnóstico agregado a las lámparas de hendidura del mercado, a través de la obtención de una imagen 3D del fondo de ojo, en particular la mácula, mediante la captura de imágenes de campos de luz para fotografía plenóptica. Esta novedosa técnica permite la obtención de una serie de imágenes con múltiples focos y ángulos de visión (paralaje) a partir de una única captura. Contando con la información 3D agregada, se espera mejorar la sensibilidad y especificidad en el diagnóstico de enfermedades maculares, facilitando su detección precoz y mejorando la calidad de vida de las personas.

3.5. CONICET PIP Nro. 5897/06: Análisis de las sensaciones de dulce, agrio y amargo en soluciones puras y mezcladas en medio acuoso y alcohólico

Dirección: Miguelina Guirao, Codirección: Amalia Mirta Calviño

3.5.1. Trabajos terminados

Efecto del etanol en el gusto de la sacarosa y del ácido cítrico

Guirao M. y Evin D.A.

Resumen:

En trabajos anteriores se ha observado que el etanol modifica el dulce de la sacarosa y el agrio del ácido cítrico y que el efecto es en algún modo diferente para uno y otro gusto. En este trabajo se examinan esas posibles diferencias teniendo en cuenta no solo los efectos en la intensidad sino también en la cualidad y en la persistencia, es decir los cambios que se perciben desde el comienzo al final de la sensación gustativa

3.5.2. Trabajos en desarrollo

Efecto del etanol en el gusto de la cafeína

Resumen:

En este trabajo se estudia la influencia del etanol en el gusto de la cafeína. Las muestras consisten de siete concentraciones de (3, 6, 12, 25, 35, 55 y 80 mM) de cafeína disueltas en agua y de dos niveles (8% and 15%) de etanol. Las soluciones se preparan diluidas en medio acuoso con y sin etanol y se presentan en vasos plásticos en forma combinada y aleatoria. Los panelistas prueban la solución, la mantienen en la boca durante varios segundos y la expulsan a un contenedor. Para evaluar el efecto en la intensidad se aplican dos métodos psicofísicos diferentes, comparación por pares y estimación de la magnitud, y para evaluar la duración del gusto se aplica el método Tiempo-Intensidad.

4. Proyectos de investigación conjunta

4.1. Proyecto de Cooperación MINCYT-BMBF (AL/12/06) Utilización de Prosodia para el Reconocimiento Automático de Habla

Proyecto Bilateral Cooperation in Education and Research, Ministerio de Ciencia y Técnica (Argentina) - International Bureau of the Federal Ministry of Education and Research (Alemania)

Director Argentino: Jorge Gurlekian, Director Alemán: Hansörj Mixdorff (Beuth-Hochschule Für Technik Berlin)

Investigadores: Humberto Torres, Diego Evin, Hartmut Pfstzinger (Universidad de Kiel)

4.1.1. Trabajos terminados

Linguistically motivated parameter estimation methods for a superpositional intonation model

Torres H.M., Gurlekian J.A., Mixdorff H. and Pfstzinger H.

Abstract:

This paper proposes two novel approaches for parameter estimation of a superpositional intonation model. These approaches present linguistic and para-linguistic assumptions for initializing a pre-existing standard method. In addition all restrictions on the configuration of commands were eliminated. The proposed linguistic hypotheses can be based on either pitch accents or lexical stress, which give rise to two different estimation methods. These two

hypotheses were validated by comparison of the estimation performance relative to two standard methods, one manual and one automatic. The results of the experiments for German, English and Spanish corpora show that the proposed methods outperform the standard ones.

Avances en la estimación y el análisis del modelo de entonación de Fujisaki

Torres, H.M.

Resumen:

En este informe resumo las actividades realizadas en el año 2014 en temas relacionados al modelo de entonación de Fujisaki. En el año 2013 habíamos propuesto introducir información lingüística en la inicialización del método de extracción de parámetros de Mixdorff (Torres et al, 2014), con el objetivo de poder asociar los comandos del modelo con el contenido lingüístico de las oraciones. Si bien se logra un aceptable margen de error en el ajuste de las curvas de entonación, un análisis posterior de los valores de los parámetros de los comandos del modelo reveló que el proceso de optimización de los parámetros diluían las posibilidades de relacionar sus valores con los eventos que le habían dado origen. Por lo cual en este período, propuse un nuevo método de extracción de parámetros (Torres y Gurlekian, 2015). Primero definí que se espera de los elementos del modelo, relacionando eventos acústicos y lingüísticos. Se mantienen las hipótesis para la inicialización del ya presentadas, y en todo momento de la optimización posterior de los parámetros se mantiene la asociación entre comandos y eventos lingüísticos. Primero se propuso un método manual de estimación de parámetros, y luego una implementación computacional, con la cual se hicieron pruebas sobre tres corpora, en tres idiomas: alemán, castellano argentino e inglés. Con este nuevo método logramos tener una asociación directa entre comandos del modelo e información lingüística, y además, una mejora superior al 34 % en la bondad del ajuste de la curva de F0 con respecto al método estándar. El método propuesto y los resultados obtenidos con este se presentan en la Parte I de este informe.

Luego decidimos analizar los valores de los parámetros extraídos. Primero se estudia la relación entre foco y prominencia versus valores de los parámetros de los comandos de acento. Para ello se diseñó, grabó y etiquetó un corpus de datos oral en castellano de argentina y alemán, con oraciones declarativas e interrogativas, sin foco y con foco en determinadas estructuras de las frases. Luego se realizó una evaluación perceptual de foco y de prominencias en las sílabas. Finalmente se exploraron posibles relaciones entre comandos y sílabas más prominentes, y comandos con parámetros de mayor valor y su relación con el foco de la frase. Nuestras pruebas muestran una correlación entre prominencia y los valores de los parámetros de los comandos del modelo de Fujisaki. Sin embargo, no se encontró una relación entre foco y los comandos. En Mixdorff et al. (2015) se presenta una detallada descripción del trabajo realizado para el alemán. En la Parte II de este informe se describen los trabajos realizados y resultados obtenidos para el castellano de la argentina.

Continuando con el análisis, se estimaron las distribuciones de los valores de los parámetros de los comandos del modelo versus clases de palabras que le dieron origen. Se optó por una distribución de valores extremos para aproximar las distribuciones de los datos. Los resultados obtenidos muestran una marcada diferencia entre las clases de palabras, las cuales se pueden agrupar, en base a los parámetros de las distribuciones estimadas. Restan realizar pruebas estadísticas para sustentar estos agrupamientos. Esta información podría ser utilizada para mejorar la predicción de los comandos del modelo a partir de un texto de entrada. Además, se espera poder verificar la viabilidad de estos agrupamientos a partir de teorías lingüísticas. Estos trabajos se describen en la Parte III de este informe.

Por último, se trabajó con el etiquetado de texto con las clases de palabras. Dada la importancia que han adquirido las clases de palabras, creemos que es necesario profundizar en esta área. Para ello se propuso trabajar con cinco corpus de texto: Notas extraídas de un diario, los discursos de la Presidenta de la Nación, textos de la enciclopedia en línea Wikipedia, discursos realizados en el Senado de la Nación, y diálogos en programas de radio. Se etiquetaron los tres primeros, y los dos últimos se encuentran en proceso de construcción. Una primera estadística de la distribución de las clases de palabras en los corpora analizados muestran claras diferencias. Se pretende realizar una estimación del nivel de desempeño del etiquetador, comparando su salida con una versión reducida, corregida en forma manual, sobre los tres corpora, y mediante un muestreo aleatorio. Estos trabajos se detallan en la Parte IV de este informe.

5. Docencia

5.1. Cursos de grado

Dr. Humberto M. Torres

Profesor Adjunto de la cátedra *Señales y Sistemas*, Facultad de Ingeniería UBA. Desde el 26 de Septiembre de 2011.

Dr. Jorge A Gurlekian

Dictado del Seminario *Laboratorio de Voz*. Facultad de Medicina. Área Fonoaudiología. Desde el 2001.

Ing. Christian Cossio Mercado

Ayudante de primera regular, en la materia *Paradigmas de Lenguajes de Programación*, Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA. Desde marzo de 2014.

5.2. Cursos de posgrado

Funciones neurosensoriales del sistema gustativo

Docente: Dra. Miguelina Guirao

Curso brindados para la Carrera de Especialistas en ORL, Facultad de Medicina, UBA. Asociación Médica Argentina. 6 de Agosto 2014.

6. Transferencia Tecnológica

6.1. Reconocimiento automático del habla en medios de comunicación

Parte de los objetivos de investigación y desarrollo se llevaron adelante asociados y en el marco de un programa de vinculación entre nuestro laboratorio y la empresa Global News Media, formalizada a través de un Servicio Tecnológico de Alto Nivel (STAN). Esta empresa realiza monitoreo de contenidos en radio y televisión con fines de marketing, de manera manual, y estaba interesada en automatizar parte de ese proceso.

Los trabajos tienen como objetivo la construcción de un sistema de reconocimiento automático del habla para desgrabar e indexar los audios de radio y videos de televisión. En este período se ha continuado con la creación del corpus de datos para el entrenamiento del sistema, el cual se espera concluir en el siguiente período. La etapa crítica de esta tarea es la transcripción y etiquetado manual de los diálogos en radio y televisión. Cabe resaltar que este será el primer corpus de este tipo del castellano hablado en Argentina. Además, se espera que a raíz de esta interacción permitan abrir nuevas líneas de investigación y se generen en forma conjunta nuevos proyectos de investigación y desarrollo.

7. Intercambio Científico

7.1. Visitas al LIS

Las estudiantes Christiane Hagedorn y Anna Lehman, de la Universidad Beuth Hochschule de Berlín, concurren al LIS en el mes de Febrero de 2014. Aquí tuvieron un entrenamiento fonético en español, y trabajaron en el desarrollo de una plataforma de grabación multimedia dentro del proyecto bilateral BMBF-MINCYT.

7.2. Estadías en el Exterior de Miembros del LIS

Ing. Christian Cossio Mercado e Ing. Miguel Martinez Soler, Berlín (Alemania), 9 de octubre de 2014 al 23 de febrero de 2015

Los dos becarios y tesistas del LIS, viajaron a Alemania como parte del proyecto bilateral BMBF-MINCYT, desde el mes de Octubre hasta Febrero de 2015. Allí realizaron tareas de investigación bajo la dirección del Dr. Hanjörg Mixdoff (Beuth Hochschule für Technik). Además, realizaron una estadía doctoral en la Universidad Tecnológica de Berlin (TU-Berlin) para cursar materias vinculadas a la especialidad, bajo la coordinación del Dr. Sebastian Möller (TU-Berlin).

8. Propiedad intelectual

Se iniciaron los trámites para la presentación una solicitud de propiedad intelectual sobre un software desarrollado para la Medición y prevención de daños auditivos producidos por ruidos ambientales.

9. Tesis

9.1. Doctorales en curso

Evaluación Automática de Calidad del Habla Artificial

Tesista: Christian Cossio Mercado

Directores: Dr. José Castaño (FCEyN-UBA) y Dr. Jorge Gurlekian

Consejero de Estudios: Dr. Agustín Gravano (FCEyN-UBA)

Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales

Resumen:

Un sistema de conversión de texto a habla (o sistema TTS, por el acrónimo del inglés *Text-To-Speech*) se encarga de transformar un texto de entrada en una secuencia de sonidos equivalente a la que produciría una persona al leerlo en voz alta. Es decir, dadas las palabras escritas, debe encargarse de pronunciarlas de manera: *inteligible*, que se comprendan todas las palabras dichas; *natural*, que sea similar a como lo diría un ser humano; y *expresiva*, que acompañe, refuerce y sea consistente con el mensaje comunicado.

En la mayoría de los sistemas actuales se alcanza una inteligibilidad próxima a la del habla humana. Sin embargo, un problema hasta ahora no resuelto satisfactoriamente es el de la naturalidad y expresividad del habla sintetizada. Para que un usuario preste atención al habla artificial durante un tiempo prolongado, es necesario que esta sea lo más natural posible dado que, de lo contrario, perderá la concentración, se sentirá cansado y perderá gran parte de la información, aun cuando la voz sea inteligible. Por otra parte, especialmente en sistemas de diálogo, la expresividad será necesaria por cuanto se desea transmitir un mensaje con una intención específica, demostrar un estado de ánimo determinado o indicar un tipo de relación con el oyente (e.g., de amistad o distante).

La falta de naturalidad y expresividad en los sistemas de diálogo automático ha restringido su empleo a ámbitos específicos como, por ejemplo, lectura de mensajes escritos, acceso por voz a información de servicios, y solicitud de reservas, aunque, por caso, no es aceptable para lectura de libros o diálogos extensos.

A continuación se listan las principales hipótesis de trabajo de la tesis:

- Es posible mejorar los métodos de evaluación automática de habla incorporando criterios utilizados por los seres humanos en el procesamiento del habla natural.
- Al obtener atributos propios del habla natural que están relacionados con percepciones de calidad buenas y malas, estos se pueden usar como referencia para evaluar habla sintetizada y predecir su calidad.
- Es posible mejorar la predicción de la evaluación de sistemas TTS al integrar en un único modelo atributos objetivos del habla, así como otros originados en el análisis de textos y acústicos, y el estudio de concordancia entre los resultados de estos dos análisis.
- El posible desarrollar métricas objetivas asociadas a la agradabilidad de un sistema de TTS y utilizarlas para mejorar la predicción de la calidad de un sistema de conversión de texto a habla.

Esta tesis en desarrollo busca realizar un **modelo integral para evaluación automática de la calidad del habla artificial**, aprovechando las ventajas de los modelos multidimensionales, en el cual se incluirán características originadas en el estudio de la percepción de habla natural y del habla sintetizada, técnicas de reconocimiento automático de habla y descripciones de las Neurociencias.

El trabajo tiene como aporte principal la integración en un mismo modelo de métricas originadas en la percepción humana con otras basadas en atributos físicos acústicos del habla. Adicionalmente, se tendrá especial consideración para realizar una evaluación que permita su utilización en un esquema de Aprendizaje Automático.

Reconocimiento forense de hablantes mediante el uso de información de alto nivel y metadatos

Tesista: Miguel Martínez Soler

Directores: Dr. Jorge A. Gurlekian (CONICET) y Dr. Agustín Gravano (FCEyN-UBA)

Consejero de Estudios: Dr. Diego Garbervetsky (FCEyN-UBA)

Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales

Resumen:

En la tarea de reconocimiento forense de hablantes se intenta determinar si un individuo conocido ha producido la voz registrada en una grabación, realizando comparaciones a diferentes niveles (acústico, lingüístico, prosódico, etc.). Las soluciones propuestas para resolver este problema incluyen la evaluación perceptual realizada por un panel de oyentes experimentados, el tratamiento y modelado de señales por sistemas totalmente automáticos y el uso de sistemas semi-automáticos, normalmente basados en la extracción de rasgos acústicos bajo la dirección del usuario.

En la última década el uso de enfoques bayesianos cobró popularidad en el área. Estos enfoques permiten calcular una probabilidad a posteriori de una hipótesis (dada la evidencia) tomando en cuenta su probabilidad a priori y una evaluación de la evidencia (dada la hipótesis). La principal ventaja que tiene esta forma de encarar el problema es que los sistemas que utilizan este enfoque emiten resultados fácilmente interpretables por humanos.

Las fuentes de información disponibles en el habla se pueden organizar en diferentes niveles de abstracción que van desde el nivel más bajo, que contiene la información acústica, hasta el nivel más alto, que describe la forma en que se usa el lenguaje. En medio de estos extremos encontramos también la información fonológica y prosódica, entre otras. Las diferencias dialectales pueden encontrarse en cualquiera de estos niveles, siendo en nuestro país muy notorias las diferencias que se encuentran a nivel fonológico, prosódico y léxico.

Además de la información contenida en la señal de habla en sí, se ha demostrado en trabajos anteriores que, utilizando un enfoque bayesiano, se puede mejorar la confianza del reconocimiento de hablantes mediante el uso de metadatos, como por ej. las duraciones de los segmentos de audio en estudio, o el tipo de canal (directo, telefónico convencional, telefónico celular, etc.).

En un trabajo de 2008, Ferrer et. al. [27] obtuvieron una mejora notable en una tarea de reconocimiento automático de hablantes mediante una estimación del grado de certeza de que el hablante sea nativo. Este resultado sugiere que el uso de información dialectal podría aportar mejoras a la tarea de reconocimiento de hablantes. Dado que se encuentran a disposición dos corpus de datos de habla segmentados por regiones de la República Argentina, se propone explotar el uso de información de alto nivel en la señal de habla para la tarea de reconocimiento forense de hablantes, tomando en cuenta las diferencias que se puedan encontrar en las diferentes regiones de nuestro país. Se estudiará también el uso de metadatos y su impacto en la confianza de las estimaciones. La evaluación del trabajo será realizada mediante la comparación con sistemas de reconocimiento de hablantes automáticos y semi-automáticos implementados mediante técnicas del estado del arte.

Reconocimiento automático de hablantes empleando información de largo plazo

Tesista: Pedro Univaso

Director: Dr. Jorge A. Gurlekian

Universidad de Buenos Aires, Facultad de Ingeniería.

Diagnostico diferencial de pacientes con movimientos anormales laringeos, complementacion entre el diagnostico neurológico y los resultados que brinda el abordaje otorrino-fonoaudiologico

Tesista: Liliana Sigal

Universidad de Buenos Aires, Facultad de Medicina

Director : Dr. Jorge A. Gurlekian.

Consejero : Profesor Dr Federico Micheli

Resumen:

El presente trabajo se desarrollará en el Laboratorio de Investigaciones Sensoriales, Conicet, la División Otorrinolaringología, el Departamento de Movimientos Anormales y el Departamento de Neurofisiología del Hospital de Clínicas José de San Martín. Tiene por objeto el estudio de la distonía laríngea, alteración que afecta el movimiento de las cuerdas vocales. Los síntomas son generalmente graduables desde una inestabilidad moderada a cortes incontrolables en la voz y un creciente esfuerzo que repercute en la inteligibilidad del habla.

10. Actividades de Divulgación

- Cossio-Mercado, Christian. “Descubriendo el lado oculto de nuestra voz”, Programa Exactas va a la Escuela 2014, Escuela Instituto Sagrado Corazón, CABA. Charla sobre análisis automático de habla y su utilización en el desarrollo de sistemas de diálogo orientados a predecir el estado físico y emocional de los usuarios, y cómo aprovechar esa información para mejorar la interacción con ellos. 6 de Mayo de 2014
- Cossio-Mercado, Christian. “Percepción Computacional: nuestros sentidos en una computadora”, Semana de la Computación 2014, Facultad de Ciencias Exactas y Naturales, UBA. Charla sobre la percepción humana y su funcionamiento, disfuncionamiento y modelización de los sentidos con una computadora. 19 de Junio de 2014.
- Miguelina Guirao. Participación en la nota “Los otros cinco sentidos”. Clarin, Revista VIVA, 23 de Noviembre de 2014.

11. Trabajos que refieren a actividades del LIS

11.1. Publicaciones

- Piñeda, M. A., & Jacó- Vilela, A. M. (2014). “Ciencia psicológica y profesionalización en Argentina y Brasil: 1930–1980”. *Universitas Psychologica*, vol 13(5), pp. 2015–2033, 2014. <http://dx.doi.org/10.11144/Javeriana.upsy13-5.cppa>.

11.2. Presentaciones a congresos

- Piñeda, Maria Andrea
Tema: Miguelina Guirao y los estudios interdisciplinarios de la percepción en el Laboratorio de Investigaciones Sensoriales
XV Encuentro Argentino de Historia de la Psiquiatría, Psicología y Psicoanálisis. Institución organizadora: Facultad de Psicología de la Universidad Nacional de La Plata. La Plata 2014.
- Piñeda, Maria Andrea y Scherman, P.
Tema: Los estudios experimentales de la percepción en Argentina a mediados del siglo XX. La UNSL, el CIAL y el LIS XI. Encontro Clio Psyché Discursos e Praticas na História da Psicologia; Rio de Janeiro 2014.

12. Publicaciones

12.1. Revistas

- Gurlekian, J.A., H. Torres y D. A. Evin, “Guía para la segmentación y transcripción fonética para las tecnologías del habla”, *Fonoaudiológica*, Vol. 61 (2), pp. 24–47. Diciembre de 2014.
- Sigal, L., y Gurlekian J.A., “Aplicación de los índices de perturbación integrado y de precisión articuladora en pacientes con disfonía espasmódica”. *Revista de Investigación en Logopedia*, Madrid, 2 (2014), pp.132–150. Disponible en <http://revistalogopedia.uclm.es>.
- Torres H.M., Gurlekian J.A. and Mixdorff H., “Linguistically motivated parameter estimation methods for a superpositional intonation model”. *EURASIP Journal on Audio, Speech, and Music Processing*, Vol. 2014, Issue 1. Disponible en <http://asmp.eurasipjournals.springeropen.com/articles/10.1186/s13636-014-0028-3>.
- Guirao, M., y Evin, D. A, “Efecto del etanol en el gusto de la sacarosa y del ácido cítrico”. Aceptado para su publicación en *Revista de La Alimentación Latinoamericana*.

12.2. Congresos

- Drozdowicz, B., Salvatelli, A., Bizai, G., Hadad, A., Evin, D. y Torres, R., “Análisis y Propuesta de Trabajo para la Reconstrucción Tridimensional de Imágenes Maculares Obtenidas Mediante Lámparas de Hendidura”. En *Proceedings del 6to Congreso Argentino de Informática y Salud*, Rosario, 2014.

12.3. Informes Técnicos

- Torres, H.M., “Avances en la estimación y análisis del modelo de entonación de Fujisaki”. Laboratorio de Investigaciones Sensoriales, LIS, INIGEM, CONICET-UBA. Diciembre de 2014.