

Batería de Evaluación para Pacientes con Prótesis Auditiva (BEPPA)

Diseño del Corpus de Habla:

Leonor Aronson
Patricia Estiene
FUNDACIÓN ARAUZ

Grabación y Procesamiento Digital:

Diego Milone
César Martínez
Hugo Leonardo Rufiner
FACULTAD DE INGENIERÍA - UNER

Características generales

La batería de evaluación para pacientes con prótesis auditiva (**BEPPA**) fue diseñada especialmente para evaluar la capacidad auditiva de pacientes adultos con audífonos o implantes cocleares. El corpus de habla fue diseñado y ajustado a través de la experiencia cotidiana durante sesiones de calibración de prótesis auditivas con pacientes adultos. Una vez definido el corpus de habla se procedió a la grabación del material en una cámara anecoica y acústicamente aislada. Finalmente, se realizó un post-procesamiento digital de las señales para separar en partes el material, eliminar errores y artefactos, agregar cantidades controladas de distintos ruidos y grabarla en formato audio-CD.

Las características principales de la batería se resumen en la siguiente tabla:

CARACTERÍSTICAS PRINCIPALES DE BEPPA	
Dialecto del Idioma Español	Argentino rioplatense
Cantidad de elocuciones	295
Tipo de elocuciones	<ul style="list-style-type: none"> ▪ Consonantes en contexto vocálico ▪ Transiciones formánticas ▪ Monosílabos ▪ Oraciones cortas de uso cotidiano
Cantidad de hablantes	<ul style="list-style-type: none"> ▪ 1 femenino ▪ 1 masculino
Características de grabación	<ul style="list-style-type: none"> ▪ Resolución: 20 bits. ▪ Frecuencia de muestreo: 48 KHz. ▪ SNR: menor de 50 dB.
Características de uso	<ul style="list-style-type: none"> ▪ Formato: audio-CD. ▪ Resolución: 16 bits. ▪ Frecuencia de muestreo: 44 KHz.
Ruidos adicionados	<ul style="list-style-type: none"> ▪ Tipo: blanco, rosa ▪ Fuente: NOISEX ▪ Niveles: 15, 10, 5 dB
Duración total	11760 segundos

BEPPA fue desarrollada conjuntamente por el Departamento de Implante Coclear y Rehabilitación Auditiva de la Fundación Arauz y el Laboratorio de Cibernética de la Facultad de Ingeniería de la Universidad Nacional de Entre Ríos (FI-UNER).

Diseño del corpus de habla

Sección pendiente destinada a describir el criterio de selección de frases y palabras, fonemas utilizados, frecuencias relativas, criterios de acentuación y pronunciación, etc.

Grabación y procesamiento digital

En esta sección se describen las características principales de los medios empleados para la grabación de la batería, y los procedimientos efectuados sobre las señales hasta obtener su formato final de uso.

- **Cámara anecoica**

La cámara anecoica se encuentra en las afueras de la localidad de Oro Verde, en un edificio de planta horizontal separado del resto de los edificios de la Facultad de Ingeniería de la UNER. Esto facilitó la tarea de aislarla acústicamente, ya que el nivel de ruido aéreo es mínimo y no hay ruido debido a vibraciones de las estructuras. El recinto posee 3,75 x 3,00 m y sus paredes exteriores son dobles. La pared exterior es de ladrillo común de 15 cm de espesor, y la interior de ladrillo cerámico hueco de 12 cm de espesor. La cara interna de la pared se recubrió con una malla metálica conectada a tierra —a manera de jaula de Faraday— para aislar interferencias electromagnéticas. El sistema se complementó con una cámara de aire, lograda con una separación de 5 cm de aire entre la pared y una placa de yeso de 12 mm de espesor. Esta cámara fue sellada herméticamente con masilla de base mineral.

La disminución de la reverberación se logró colocando sobre la placa de yeso tres capas de lana de vidrio de distintas densidades. Cada capa tiene 8 cm de espesor y recubre la totalidad del recinto de grabación. Sus densidades son de 48 gr/cm³, 33 gr/cm³ y 21 gr/cm³, con lo cual se logra un cambio suave de esta propiedad. Esta característica permite disminuir en gran medida la reflexión de las ondas sonoras. El techo es flotante, con recubrimiento similar al realizado a las paredes. El piso, con igual recubrimiento, se completa con una malla de acero de 4.2 mm de diámetro, con figuras cuadradas, fijadas a soportes solidarios de los barrales de la cámara de aire (del piso). La reverberación resultante es menor a 170 ms por debajo de los 400 Hz, menor a 100 ms hasta los 800 Hz y menor a 20 ms para frecuencias superiores [AGM+01].

- **Micrófono**

Los registros se realizaron con un micrófono Shure SM58, ubicado a unos 15 cm de la boca del locutor. El SM58 es un micrófono dinámico (bobina móvil) unidireccional, diseñado especialmente para registros de voces. Posee un filtro esférico incorporado que reduce al mínimo los ruidos causados por el viento y por el aliento. La dispersión polar de cardioide aísla la fuente sonora principal a la vez que reduce al mínimo los ruidos de fondo. El micrófono tiene una respuesta en frecuencia (50 Hz a 15 kHz) ajustada para la captación de voces, con aumento de frecuencias medias y atenuación progresiva de frecuencias bajas [SI02].

- **Digitalización**

Las voces se digitalizaron con una placa de adquisición de sonidos profesional Turtle Beach Multisound FIJI. Los registros se realizaron con una resolución de 20 bits, a una frecuencia de muestreo de 48 kHz. El conversor de entrada está basado en tecnología de modulación sigma-delta, con un sobremuestreo de 128 veces. La relación señal/ruido provista es mejor a 97 dB y la distorsión armónica total (THD) menor al 0.005%. Esta

placa posee una respuesta en frecuencia máximamente plana entre 10 Hz y 22 kHz (± 1 dB), y una respuesta de fase con ± 0.1 grados. La distorsión de intermodulación (IMD) es menor a 0.01% [TB97].

- *Procesamiento digital*

Una vez registrado cada hablante, se separó el material según los grupos que se indican en el Apéndice A. Cada conjunto de elocuciones fue revisado para detectar posibles repeticiones del locutor y eliminarlas. Se insertaron y/o eliminaron silencios donde fue necesario, a fin de mantener aproximadamente constante la separación temporal entre emisiones. Finalmente, se normalizaron las amplitudes de pico de todas las emisiones a una potencia de -3 dB para aprovechar el rango dinámico al momento de combinar habla limpia con ruido y evitar además la pérdida de información en la grabación a CD, ya que este proceso reduce la resolución a 16 bits.

A cada grupo de elocuciones se le sumaron cantidades controladas de ruidos de la base de datos NOISEX [VS93]. El ruido blanco de esta base de datos tiene asegurada igual energía en cada banda de frecuencia y fue digitalizado a partir de un generador de alta calidad (Wandel & Goltermann). El mismo equipo fue utilizado para generar el ruido rosa, asegurando igual energía cada 1/3 de octava. Estos registros se realizaron con una resolución de 16 bits y una frecuencia de muestreo de 19980 Hz.

Antes de realizar la mezcla del ruido y la voz, se remuestrearon ambas señales a 44100 Hz y se redujo la resolución de la señal de voz limpia a 16 bits [RS78, OS89]. A partir de la energía total de cada tramo de voz y de la energía del ruido correspondiente, se calculó el coeficiente de amplificación para el ruido según [DPH93]: $A^2 = 10^{-(SNR/10)}$ E_{voz} / E_{ruido} , donde E_{voz} es la energía de la señal de voz, y E_{ruido} la energía del ruido. Finalmente se promedió punto a punto ambas señales de forma de obtener el archivo mezclado para cada tipo y nivel de ruido.

- *Grabación en formato audio-CD*

Las elocuciones se grabaron en formato audio-CD a una resolución de 16-bits y frecuencia de muestreo de 44100 Hz. Para aumentar la compatibilidad del CD grabado se utilizaron dos canales (estereo), copiando la misma señal en ambos. La lista de correspondencias entre el texto de las elocuciones y la pista de cada CD se detalla en el Apéndice B.

Referencias

[AGM+01] L. Aronson, L. Gamero, C. Martínez, D. Milone, L. Di Persia, H. Rufiner, D. Tochetto, H. Torres, M. Torres, "*Sistema de reconocimiento automático del habla*". Informe final del Proyecto de Investigación y Desarrollo PID 6036, FI-UNER, 2001.

[SI02] Shure Inc., "*Model SM58 User Guide*",
<http://www.shure.com/microphones/models/sm58.asp>, 2002.

[TB97] Turtle Beach, "*Getting Started Guide*",
http://www.turtlebeach.com/site/kb_ftp/540ftp.asp, 1997.

[VS93] A. Varga, H. Steeneken, "*Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems*", *Speech Communication*, 12(3):247--251, 1993.

[RS78] L. R. Rabiner, R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.

[OS89] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Inc., 1989.

[DPH93] J. R. Deller, J. G. Proakis, J. H. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing, 1993.

Apéndice A: listado de elocuciones por categoría.

**consonantes en contexto
vocalico**

asa
ama
ala
ana
afa
aja
apa
aba
ada
aca
ata
aga
arra

**transiciones
formánticas 01**

aéreo
aire
ahora
auto
real
veinte
león
deuda
piano
hielo
ciudad
almohada
cohete
boina
suave
pueblo

**transiciones
formánticas 02**

huida
saeta
caimán
aorta
ausente
fea
seis
reo
reuma
diario
hielo
diurno
proa
roedor
oigo
cuatro
suelo
buitre

**transiciones
formánticas 03**

paella
baile
ahogo
cauto
leal
reina
peón
seudo
viaje
miedo
boato
poeta
troica
cuadro
nuevo
huida

**transiciones
formánticas 04**

maestra
braile
mahoma
paula
leandro
viene
meollo
rehúsa
diálogo
quiero
camión
croacia
bohemio
moira
suave
huevo
cuidar

**transiciones
formánticas 05**

maestro
aislado
caos
audaz
peaje
aceite
leopardo
neurona
diablo
cierre
trunfo
boa
enmohecido
guante
abuela
ruido

**transiciones
formánticas 06**

aeropuerto
zaino
zanahoria
austral
pelea
treinta
trofeo
europa
sabia
cierre
silencio
diurético
loa
poema
mohín
guapo
juego
suizo

monosílabos 01

ven
con
flor
pul
ley
gran
tren
paz
crin
voz
tres
muy

monosílabos 02

ron
san
don
gol
tul
vil
fin
cruz
fray
flan
pan
pez

monosílabos 03

sal
doy
sol
hoy
luz
gas
red
sed
as
fe
mal
res

monosílabos 04

tal
dos
van
mes
can
fax
bol
tres
voy
rey
mil
tos

monosílabos 05

soy
re
tras
gin
rock
fui
cual
miel
yo
sus
clan
chal

monosílabos 06

pus
tan
que
do
sin
los
sor
bar
clon
gel
bis
vas

oraciones 01

sesenta gramos
te gustó?
quién vino?
me queda grande
sonó el teléfono?
dame la sal
prestame el lápiz
qué porquería!
estoy impresionado
ya volvió?
acordate de llamar
todavía no me llamó

oraciones 02

dale la sal
setenta gramos
el piso está mojado
qué interesante!
te divertiste?
está muy claro
qué barato!
Dejate de molestar!
andá a dormir
comprá dos kilos
cuántas veces?
Estás arrepentido?

oraciones 03

a qué hora es?
cuánto pagaste?
ahora no puedo
cuándo vas a venir?
quién habla?
cuando vos quieras
por favor avisame
me parece que no
te llamo más tarde
dónde fuiste?
volvé temprano
cómo te fue?

oraciones 04

avisame cuándo volvés
no me gusta
no tengo ganas
vové temprano
qué pasó?
feliz cumpleaños!
el fuego se apagó
me parece que sí
muchas gracias
te felicito
todavía no llegó
qué lástima!

oraciones 05

no puede ser
cuidate mucho
vos estás loca / o!
Estoy muy cansada / o
no me importa
no te escucho
hablá más fuerte!
no me molestes!
Dejame tranquilo / a
Mañana voy a verte
el juego se acabó
vení cuando quieras

oraciones 06

llueve?
qué hora es?
no me gustó
no le gustó
no te gustó
quién llamó?
qué asco!
no me des
no le des
no les des

oraciones 07

qué escándalo!
hablá más suave
ahora no puedo
luis pisa el pasto
el suelo está sucio
me prestás ese libro?
andá a descansar
vení más tarde
ya me cansé
entonces, qué hacemos?

oraciones 08

está lloviendo
me voy a dormir
la fruta está fresca
quiero té de tilo
decile que no
viajaste bien?
falta mucho?
Me duele la garganta
la policía apresó al ladrón
dame una taza de té

oraciones 09

tiene filo
estoy aburrida
siento frío
quiero vino
esa película ya la ví
qué pensás hacer?
no estoy de acuerdo
pablo pintó su casa
en bariloche está nevando
elena espera un bebé

oraciones 10

ayer viajaste en subte?
el tránsito está pesado
necesito hablar con vos
la alfombra está manchada
el chocolate me cae mal
la heladera está vacía
el bebé tiene hambre
terminamos a las siete
cuánto tengo que esperar?
caminamos diez kilómetros

SEGUNDA PARTE

oraciones 01

ayer no pude dormir
 quién vino a visitarte?
 cuántos años pasaron?
 nicolás cumplió diez años
 yo no te engañé
 hoy se fue temprano
 la casa está en venta
 no tengo idea
 quién pidió otra porción
 por qué gritaste?

oraciones 02

la pared está rota
 el cuaderno está completo
 el teléfono sonó una vez
 cuándo llamó el plomero?
 el viernes fuimos a cenar
 el auto está en el taller
 el reloj es de oro
 fernando estudia mucho
 quién salió a esta hora?
 la radio no funciona

oraciones 03

no recuerdo dónde lo puse
 las cartas se perdieron
 no te puedo contar
 cuántos platos necesitás?
 quiénes estaban reunidos?
 la pintura está seca
 por qué no volvés mañana?
 el jarrón es muy antiguo
 la conferencia terminó
 el paquete es muy pesado

oraciones 04

cuánto te costó?
 no me gustó?
 dame la sal
 tené paciencia
 ya terminaste?
 no te puedo creer?
 fuiste a misa?
 no entendí nada
 dame sesenta gramos
 ya me cansé

oraciones 05

otra vez?
 quién te pegó?
 más tarde te llamo
 ese vestido te queda mal
 no me interesa
 que te pareció?
 tené paciencia
 mañana te llamo
 prepará tu ropa
 tengo mucho hambre

oraciones 06

por fin llegaste
 cómo te llamas?
 me quedé dormido
 qué suerte!
 decile que no
 te vas a acordar
 hablá más suave
 es interesante
 andá a descansar
 a qué hora empieza

oraciones 07

quién te lo regaló
 es un desastre
 se terminó
 no tengo tiempo
 hay alguna señal
 no es verdad
 dónde está el mapa
 me olvidé
 cortaron el agua?
 dame otra porción

oraciones 08

preparaste la cena
 ya no aguanto más
 se descompuso el tren
 las flores están secas
 el agua está muy fría?
 no me gusta manejar
 estás de vacaciones?
 hay olor a gas
 está nevando
 qué musica hermosa!

oraciones 09

demasiado pesado
 tomaste el remedio
 los niños se durmieron
 qué problema, no?
 el accidente fue terrible
 amaneció nublado
 la autopista estaba vacía
 cerraron el museo
 la carne está cruda
 recordás el número?

transiciones 01

acaecer
 amaina
 vaho
 astronauta
 teatral
 deleitar
 teórico
 terapeuta
 cristian
 cien
 rioja
 triunfar
 loable
 poético
 coincide
 evaluar
 muestra
 cuida

transiciones 02

ahora
 decae
 aire
 gaona
 ausente
 leal
 reir
 rejoy
 tiara
 siete
 diodo
 ciudad
 proa
 cohete
 coima
 cuatro
 luego
 luis

transiciones 03

traer
 amainar
 ahogo
 aula
 marea
 deleite
 teoría
 neutral
 diagonal
 ciento
 violeta
 miura
 loada
 roer
 oir
 cuándo
 rueda
 suicida

transiciones 04

caer
 caiga
 paola
 flauta
 peleado
 deleite
 rejoy
 neurólogo
 liviano
 siempre
 piojo
 viuda
 coalición
 roer
 boicot
 actuar
 tuerto
 ruina.

Apéndice B: listado de las pistas grabadas en cada CD

La batería se dividió en dos partes. Los cuatro CDs correspondientes a la primera parte tienen la siguiente disposición de pistas:

1. Oraciones de uso cotidiano 01
2. Oraciones de uso cotidiano 02
3. Oraciones de uso cotidiano 03
4. Oraciones de uso cotidiano 04
5. Oraciones de uso cotidiano 05
6. Oraciones de uso cotidiano 06
7. Oraciones de uso cotidiano 07
8. Oraciones de uso cotidiano 08
9. Oraciones de uso cotidiano 09
10. Oraciones de uso cotidiano 10
11. Transiciones formánticas 01
12. Transiciones formánticas 02
13. Transiciones formánticas 03
14. Transiciones formánticas 04
15. Transiciones formánticas 05
16. Transiciones formánticas 06
17. Monosílabos 01
18. Monosílabos 02
19. Monosílabos 03
20. Monosílabos 04
21. Monosílabos 05
22. Monosílabos 06
23. Consonantes en contexto vocálico
24. Transiciones formánticas /io/ extraídas

Las primeras 24 pistas (como las enunciadas) corresponden a voz limpia. El resto de pistas en cada CD repiten las categorías anteriores agregando ruido de acuerdo a la siguiente disposición:

- 25 a 48: voz + ruido con SNR 5 dB.
- 49 a 72: voz + ruido con SNR 10 dB.
- 73 a 96: voz + ruido con SNR 15 dB.

Según la organización expuesta, los cuatro CDs de la primera parte presentan el siguiente contenido:

- CD 1: hablante femenino con ruido blanco.*
- CD 2: hablante femenino con ruido rosa.*
- CD 3: hablante masculino con ruido blanco.*
- CD 4: hablante masculino con ruido rosa.*

La segunda parte de la batería también consta de cuatro CDs, con la siguiente disposición de pistas:

1. Consonantes en contexto vocálico, ordenada
2. Consonantes en contexto vocálico, mezcla 1
3. Consonantes en contexto vocálico, mezcla 2
4. Consonantes en contexto vocálico, mezcla 3
5. Consonantes en contexto vocálico, mezcla 4
6. Consonantes en contexto vocálico, mezcla 5
7. Oraciones de uso cotidiano 01
8. Oraciones de uso cotidiano 02
9. Oraciones de uso cotidiano 03
10. Oraciones de uso cotidiano 04
11. Oraciones de uso cotidiano 05
12. Oraciones de uso cotidiano 06
13. Oraciones de uso cotidiano 07
14. Oraciones de uso cotidiano 08
15. Oraciones de uso cotidiano 09
16. Transiciones formánticas 01
17. Transiciones formánticas 02
18. Transiciones formánticas 03
19. Transiciones formánticas 04

Como en el caso anterior, las primeras 19 pistas corresponden a voz limpia. El resto de pistas en cada CD repiten las categorías anteriores agregando ruido de acuerdo a la siguiente disposición:

- 20 a 38: voz + ruido con SNR 15 dB.
- 39 a 57: voz + ruido con SNR 10 dB.
- 58 a 76: voz + ruido con SNR 5 dB.
- 77 a 95: voz + ruido con SNR 0 dB.

Según esta organización, los cuatro CDs de la segunda parte de la batería presentan el siguiente contenido:

- CD 1: hablante femenino con ruido blanco.*
- CD 2: hablante femenino con ruido rosa.*
- CD 3: hablante masculino con ruido blanco.*
- CD 4: hablante masculino con ruido rosa.*