1. **Estudio 1: Efectos de la manipulación acústica del IDS en la atención de bebés en etapa prelingüística**

El Estudio 1 evaluará el impacto del habla dirigida a bebés (IDS, por sus siglas en inglés) en la atención y preferencias de bebés prelingüísticos entre los 3 y 9 meses de edad. Para ello, se emplearán técnicas no invasivas que incluyen el seguimiento ocular (eye-tracking), el análisis automatizado de expresiones faciales y emociones mediante inteligencia artificial, y la estimación de respuestas fisiológicas como el ritmo cardíaco, a partir de cambios en la coloración de la piel registrados en grabaciones de video sincronizadas con los estímulos.

El seguimiento ocular es fundamental, ya que se ha demostrado que los bebés son sensibles desde muy temprana edad a señales visuales provenientes de cuidadores y otras personas [119], y la dirección de la mirada cumple un papel central en el vínculo afectivo y la comunicación visual y verbal temprana con el cuidador principal [118]. Además, al tratarse de una medida no invasiva e indirecta de la atención, que no depende del lenguaje verbal, el seguimiento ocular ha sido ampliamente utilizado para identificar intenciones comunicativas universales en la infancia temprana [118, 119, 171]. Su aplicación en este estudio permitirá vincular las modulaciones acústicas del IDS con patrones atencionales reflejados en la dirección de la mirada.

Para estudiar cómo los bebés distribuyen su atención ante diferentes modulaciones del habla, se requiere registrar con alta precisión parámetros como fijaciones y latencias de fijación mediante el análisis de movimientos sacádicos [172]. Esto es especialmente importante considerando que el comportamiento ocular en la infancia temprana tiende a ser desorganizado y a cambiar rápidamente entre estímulos, debido a la inmadurez atencional típica de esta etapa del desarrollo [173].

Por tanto, es necesario contar con un sistema de seguimiento ocular remoto, de alto rendimiento, que no interfiera con el comportamiento natural de los bebés. El equipo debe tener una tasa de muestreo superior a 100 Hz y una precisión angular menor a 0.5°, características que actualmente solo se alcanzan mediante sistemas basados en cámaras infrarrojas que utilizan la reflexión corneal [171, 172]. Estos permiten estimar con alta precisión la posición tridimensional de ambos ojos y el punto de fijación, y ofrecen seguimiento binocular que tolera movimientos de la cabeza del bebé, reduciendo errores temporales [174]. Además, es imprescindible que el sistema funcione de forma independiente a una pantalla, ya que se empleará durante interacciones sociales en vivo.

**Piloto con cámara web:** A pesar de las ventajas en precisión y control de los sistemas basados en infrarrojos, su uso limita la implementación de estudios transculturales debido a la necesidad de contar con el mismo equipo y software en múltiples ubicaciones, lo que incrementa significativamente los costos y la dependencia de laboratorios especializados.

Por ello, este estudio incluirá un piloto paralelo en el que se recolectarán datos simultáneamente con el rastreador ocular infrarrojo y una cámara web convencional. Aunque estas tecnologías basadas en cámaras web han comenzado a explorarse en investigación, aún no existen estudios validados con bebés, y la precisión que ofrecen no alcanza los estándares de alta calidad científica. Su uso, aunque prometedor, sigue siendo limitado, como lo evidencian estudios en otras áreas y técnicas relacionadas como la pupilometría [175, 176, 177].

Si este piloto demuestra que es posible obtener datos de calidad aceptable mediante cámaras web, se podrá avanzar hacia paradigmas experimentales distribuidos en línea, en los que las familias participen desde sus hogares utilizando únicamente un computador con cámara web. Esta posibilidad, ya sugerida por otros autores [178], permitiría democratizar el acceso a estudios con tecnología de rastreo ocular —históricamente restringida a ciertos contextos institucionales— y ampliar enormemente el alcance y la inclusión geográfica y cultural de este tipo de investigaciones.

**Herramientas de inteligencia artificial para analizar puntos de referencia faciales y emociones, y extraer**

**respuestas fisiológicas (ritmo cardíaco):** Además de ayudar a definir y afinar una herramienta de código abierto para el rastreo ocular, estudio integrará herramientas de inteligencia artificial de código abierto para el análisis automático de expresiones faciales, estimación de emociones y extracción de ritmo cardíaco a partir de video.

Para la detección de emociones, se utilizarán modelos que estiman estados afectivos (valencia y activación) a partir de grabaciones de video. Existen algoritmos y herramientas de código abierto, principalmente implementadas en Python que emplean redes neuronales convolucionales entrenadas en bases de datos multimodales, y ya han sido validadas en condiciones de laboratorio controladas para inferencia afectiva [X1, X2, X3].

Para la estimación del ritmo cardíaco, se emplearán técnicas de fotopletismografía remota (rPPG), que permiten inferir la frecuencia cardíaca a partir de sutiles cambios en la coloración facial causados por el flujo sanguíneo. Las herramientas más prometedoras en este campo también han sido desarrolladas bajo licencias abiertas y validadas en condiciones con distintas iluminaciones y movimientos de cabeza [X4, X5, X6, X7].

**Estímulos:**

Como estímulos, se usarán audios de madres hablando en IDS, de duración aproximada de 5 segundos. De cada video se realizarán manipulaciones independientes de variables acústicas que tienen efectos importantes en la percepción de las voces. La voz humana tiene características acústicas paralingḯsticas particulares ampliamente estudiadas y salientes, como el tono de voz (qué tan grave o aguda es la voz), que se determina a partir de la frecuencia fundamental (*f*0; CITAS), así como las formantes, que son frecuencias de resonancia características del tracto vocal que aparecen como picos en el espectro sonoro, especialmente relevantes en la producción de vocales, pero que también se asocian con características sexualmente dimórficas como el largo del tracto vocal (Fig. 1).

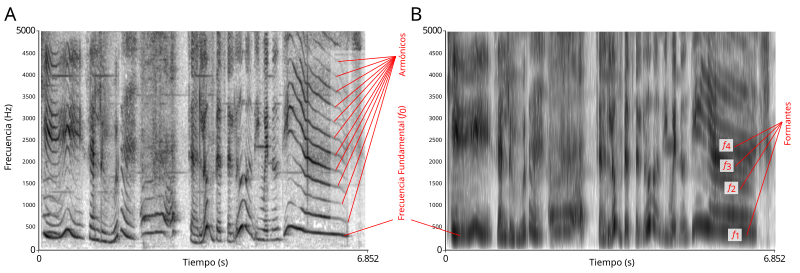


Figura 1. Espectrograma de madre (investigadora del proyecto) hablando en IDS. Los espectrogramas muestran la densidad espectral de potencia (es decir, la energía acústica) distribuida en diferentes frecuencias (eje vertical) a lo largo del tiempo (eje horizontal). Cuanto más oscuro es un punto, mayor es la energía concentrada en esa frecuencia y momento. En este caso, ambos paneles representan la misma grabación, con una duración total de 6.85 segundos, y muestran la energía entre 0 y 5000 Hz. La única diferencia entre los paneles radica en la longitud de la ventana temporal utilizada para calcular la transformada de Fourier en cada paso del análisis espectral. En el panel A, los armónicos (componentes periódicos relacionados con la frecuencia fundamental, f0) se visualizan con mayor claridad, mientras que en el panel B se destacan mejor las formantes (f₁, f₂, f₃, f₄), que reflejan las resonancias determinadas por la forma y longitud del tracto vocal. Los análisis y espectrogramas fueron generados usando Praat (CITA).

A partir de estas características, se manipularán tres variables: el tono promedio de la voz (i.e. qué tan aguda o grave es la voz, en promedio), la variabilidad tonal (que tan “cantada” es la voz), y el largo del tracto vocal (que se relaciona directamente con la percepción de feminidad o masculinidad relativa, y el tamaño corporal). Ya que depende de la dispersión de valores de *f*0, la variabilidad tonal se determina como desviación estándar de la *f*0. (*f*0 SD), mientras que el tono promedio se define como media de la *f*0 (*f*0 mean), y el tracto vocal como la distancia promedio entre las formantes (normalmente conocida como dispersión de las formantes, o *D*f). Cada una de estas variables será manipulada independientemente en 2 niveles (Alto y Bajo), determinadas a partir de estándares acústicos de IDS (CITAS):

* *f*0 SD: Variabilidad tonal (Baja vs Alta) = ± 10 Hz
* *f*0 mean: Tono medio (Baja vs Alta) = ± 20 Hz
* *D*f : Dispersión de formantes (Baja vs Alta) = ± 10 %

**Diseño y muestra:**

Para el estudio 1, la muestra serán bebés de 3 a 9 meses de edad, prelingüísticos, capaces de sostener su cabeza. Dado que el propósito principal de este estudio es analizar los tiempos de fijación total en los estímulos, el tamaño de muestra fue estimado empíricamente, representando directamente los modelos que se usarán para el análisis de los datos, con base en simulaciones de Monte Carlo realizadas en R. Los detalles y código para reproducir esta simulación, están disponibles en XXXX. Ya que hay tres factores que serán manipulados experimentalmente (*f*0 SD,*f*0 mean, *D*f), cada uno a dos niveles (Bajo vs Alto), esto resulta en ocho manipulaciones experimentales:

* *f*0 SD baja, *f*0 mean baja, *D*f baja
* *f*0 SD baja, *f*0 mean baja, *D*f alta
* *f*0 SD baja, *f*0 mean alta, *D*f baja
* *f*0 SD baja, *f*0 mean alta, *D*f alta
* *f*0 SD, *f*0 mean baja, *D*f baja
* *f*0 SD, *f*0 mean baja, *D*f alta
* *f*0 SD, *f*0 mean alta, *D*f baja
* *f*0 SD, *f*0 mean alta, *D*f alta

Para la estimación de poder, esto fue modelado como un diseño cruzado 2x2x2, con efectos aleatorios para diferentes bebés. Se modelaron efectos intra-sujeto en una población hipotética de 10 000 bebés, medidos como tiempos de fijación sobre 5000 mili-segundos. Los efectos simulados en la población fueron siempre muy conservadores: el efecto principal de *f*0 SD fue modelado como un efecto pequeño ( = 0.15), mientras que los efectos principales de *f*0 mean y *D*f se modelaron como muy pequeños ( ≤ 0.9). Todas las interacciones, aunque modeladas, fueron determinadas como efectos extremadamente pequeños ( ≤ 0.2), ya que no hacen parte de las predicciones puntuales (Fig. 2). El efecto de *f*0 SD se modeló como un efecto más fuerte que los demás, pues para este hay predicciones específicas: la atención de las y los bebés debe aumentar cuando los estímulos tienen alta *f*0 SD (predicción V de CITA). Los efectos principales de 0 mean y *D*f son de menor tamaño pues, aunque no hay predicciones específicas, esperamos que la atención aumente en estímulos con valores altos en comparación con valores bajos, ya que en ambas variables mayores valores se asocian con características más femeninas. Dada la dependencia de las y los bebés de sus mamás para sobrevivir, es de esperar que haya cierta preferencia por características femeninas.

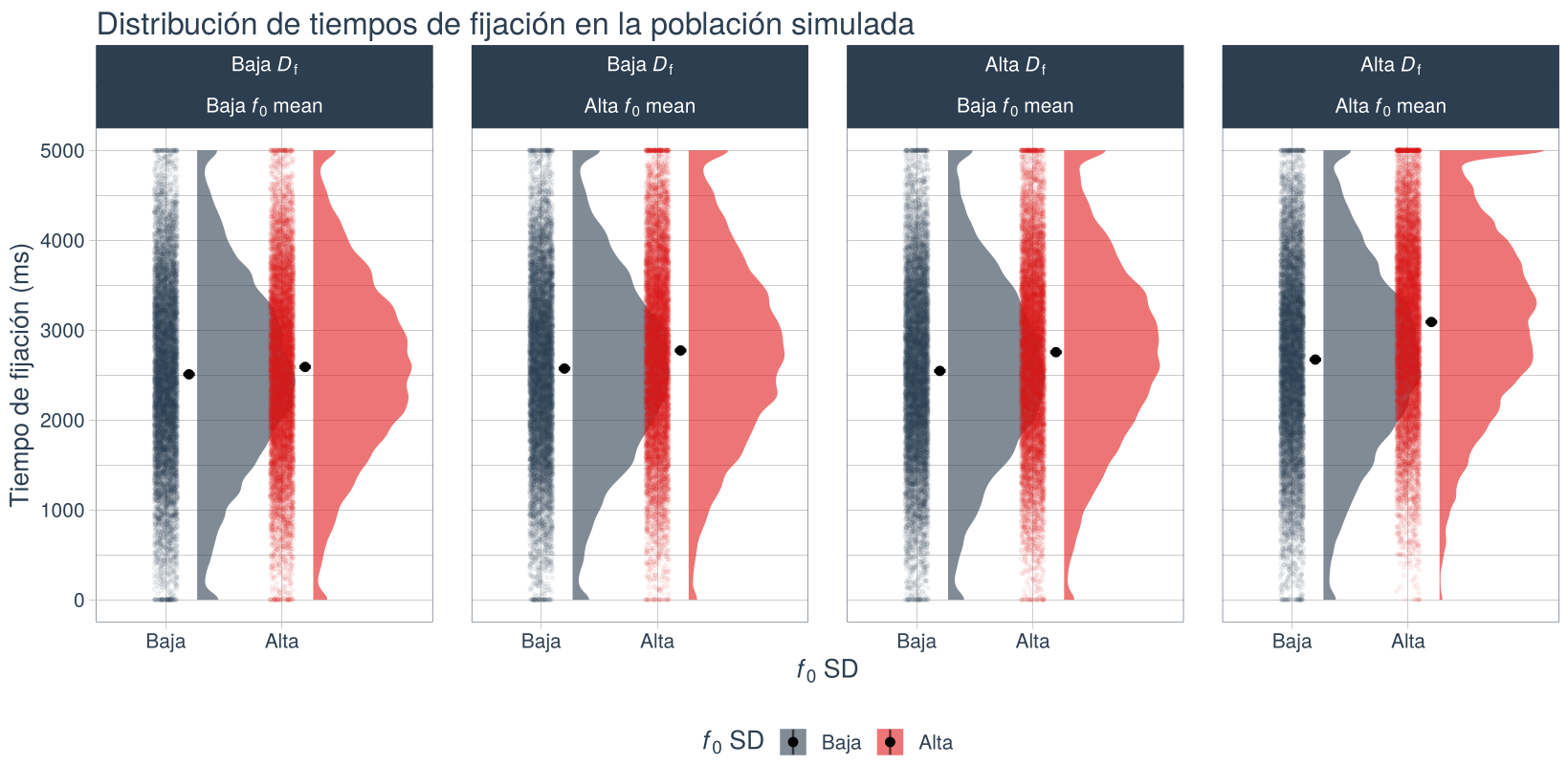


Figura 2. Distribución de los tiempos de fijación en función de f0 SD (eje x), dividida por f0 mean y Df. Cada panel muestra la distribución de densidad, los puntos individuales (dispersados), y los promedios por condición con sus barras de error. Los tiempos de fijación aumentan con mayor f0 SD, y los efectos se ven modulados por las demás características acústicas.

Una vez modelada la población hipotética de 10 000 bebés con estas características, simulamos muestras aleatorias de diferentes tamaños, y ajustamos el modelo lineal mixto con los efectos principales de *f*0 SD, *f*0 mean, y *D*f así como sus interacciones como efectos fijos, y efectos aleatorios para cada bebé. Específicamente, a partir de la población simulada (N = 10 000), extraemos 1 000 muestras aleatorias de tamaño n = 10, luego otras 1 000 de tamaño n = 20, y así sucesivamente hasta obtener 1 000 muestras de tamaño n = 200. Para cada una de las 20 000 muestras resultanes, ajustamos el mismo modelo lineal mixto, examinamos la distribución de los valores *p*, y estimamos la probabilidad de detectar un efecto estadísticamente significativo para cada término del modelo (Fig. 3).

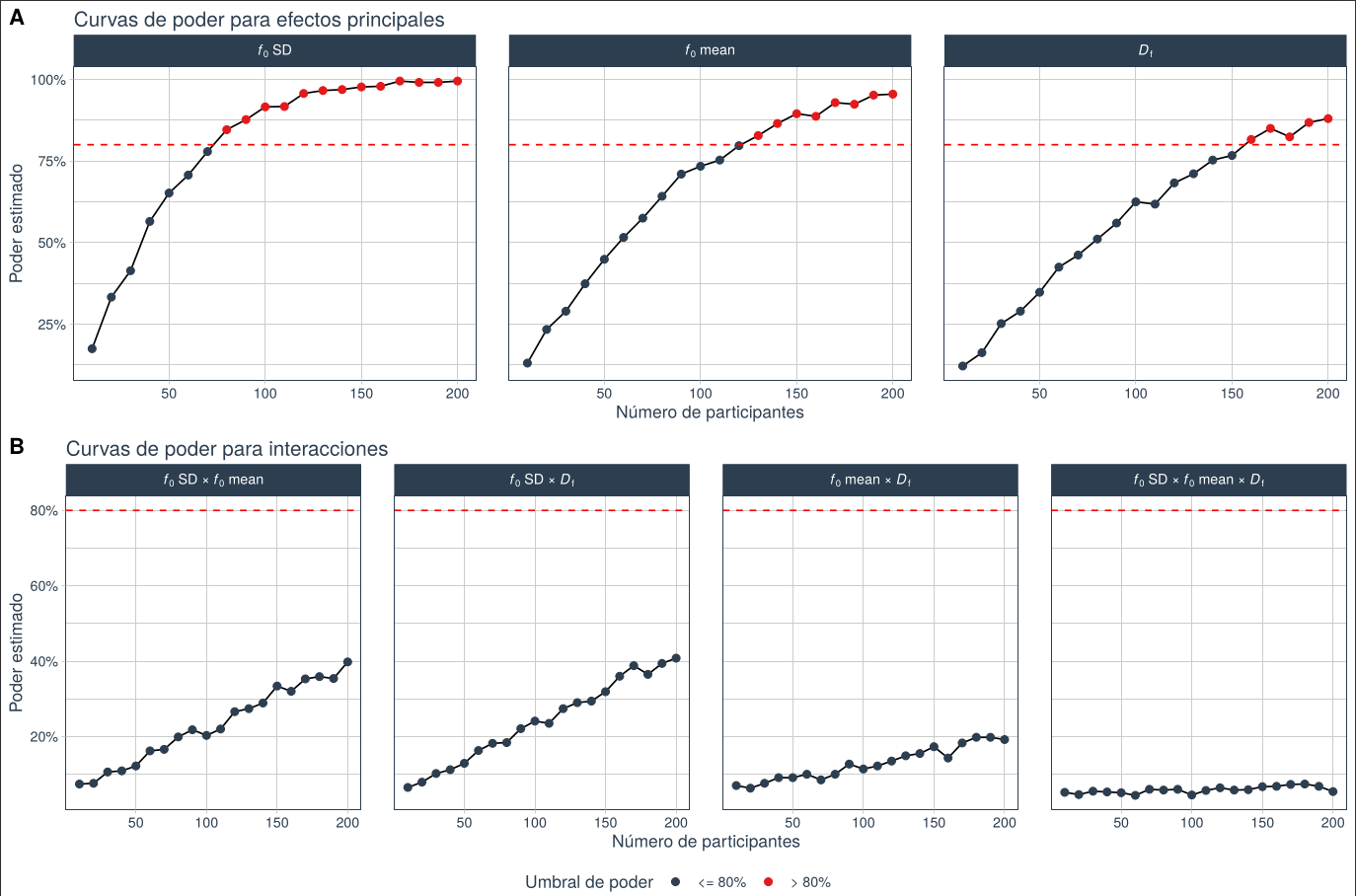


Figura 3. Curvas de poder para detectar efectos principales e interacciones. Los paneles muestran el poder estadístico como función del tamaño muestral para cada efecto fijo (A) e interacción (B). Los puntos indican las estimaciones de poder basadas en simulaciones para cada tamaño muestral, con líneas punteadas rojas que marcan el umbral del 80 %. La mayoría de los efectos principales alcanzan un alto poder con menos de 160 participantes, mientras que las interacciones permanecen con bajo poder en este rango, reflejando sus tamaños de efecto más pequeños.

Esta simulación mostró que, con un tamaño de muestra de 160 bebés, el poder supera el 80 % para los efectos principales. De hecho, la probabilidad de detección para la variable principal (*f*0 SD) llegaría a aproximadamente el 98% (Fig. 4).



Figura 4. Distribución de los valores p en las simulaciones con n = 160. Esta figura muestra la distribución de los valores p para cada efecto, basada en 1000 simulaciones utilizando el tamaño muestral final recomendado. Los efectos principales muestran una fuerte asimetría hacia valores p bajos, consistente con alto poder. Las interacciones producen distribuciones más uniformes, lo que indica una sensibilidad limitada para detectar estos efectos con este tamaño muestral.

**Procedimiento:**

Cada bebé participará en una única sesión presencial, acompañado por su madre, padre o guardián legal, quien firmará un consentimiento informado antes de iniciar. Durante la sesión, el adulto sostendrá al bebé en sus piernas, de modo que este quede orientado hacia una pantalla ubicada al frente.

En la pantalla se presentarán clips de video de aproximadamente 5 segundos de duración, correspondientes a estímulos de habla dirigida a bebés (IDS) con distintas manipulaciones acústicas previamente descritas. Los estímulos se presentarán en bloques, con pausas breves entre ellos para mantener la atención del bebé y evitar fatiga.

Durante la sesión, se registrarán los movimientos oculares del bebé mediante un sistema remoto de rastreo ocular por infrarrojos, y se grabarán en video sus reacciones faciales. Estas grabaciones se analizarán posteriormente mediante herramientas de inteligencia artificial para estimar, de forma no invasiva, indicadores afectivos (como valencia y activación emocional) y fisiológicos (como el ritmo cardíaco), a partir de cambios sutiles en la expresión facial y la coloración de la piel.

La sesión tendrá una duración aproximada de 30 minutos. La participación se interrumpirá de inmediato si se detectan señales de malestar o si el bebé se duerme. Al finalizar, se entregará un subsidio de transporte a la persona acompañante.

1. **Estudio 2: Efectos del IDS en habilidades lingüísticas y musicales posteriores**

XXXX

Participantes y Muestra:

El tamaño de muestra para el Estudio 2 se determinó mediante análisis de poder estadístico que combinan dos enfoques complementarios: la detección de efectos significativos y la posibilidad de concluir equivalencia con cero. Esta estrategia permite maximizar la sensibilidad del estudio tanto para identificar asociaciones relevantes como para descartar efectos triviales con base empírica.

En primer lugar, se estimó analíticamente el poder para detectar una correlación poblacional de al menos ρ = 0.30 entre características acústicas del IDS materno (*f*0 SD,*f*0 mean, *D*f) y medidas del desarrollo lingüístico y musical. Teniendo en cuenta que no hay estudios previos que hayan medido este efecto, y para producir información confiable que supere las limitaciones tradicionales de los estudios correlacionales y exploratorios, se definió un tamaño de muestra mínimo que pemitiera detectar un umbral razonable para un efecto relevante en el contexto del desarrollo infantil temprano (CITAR LAKENS).

En segundo lugar, se realizó un análisis de equivalencia estadística para determinar si una correlación observada puede considerarse estadísticamente indistinguible de cero, dentro de un margen predefinido de ±0.15. Este margen fue definido como umbral práctico para un efecto trivial, con base en estándares comunes para interpretar correlaciones pequeñas CITAR LAKENS).

Ambas pruebas se implementaron utilizando funciones analíticas del paquete TOSTER (CITA), y se complementaron con simulaciones Monte Carlo para estimar la probabilidad de obtener resultados concluyentes (efecto positivo, equivalencia o resultado inconcluso) bajo distintos valores poblacionales. Los resultados mostraron que una muestra de aproximadamente 480 participantes permite alcanzar al menos un 90 % de poder para ambas pruebas:

* Detectar una correlación positiva verdadera de ρ = 0.30 con ≥ 90 % de probabilidad.
* Concluir equivalencia con cero si la correlación observada cae completamente dentro del margen de ±0.15 (con IC del 90 %), también con ≥ 90 % de probabilidad.

Además, este enfoque permite evitar conclusiones erróneas ante correlaciones pequeñas pero estadísticamente significativas que, sin una prueba de equivalencia, podrían interpretarse erróneamente como efectos relevantes. Así, el análisis dual fortalece la capacidad inferencial del estudio, aportando mayor solidez a las conclusiones sobre la presencia o ausencia de efectos del IDS en el desarrollo infantil.

Instrumentos:

Las habilidades lingüísticas se evaluarán mediante pruebas adaptadas culturalmente que midan vocabulario receptivo y expresivo, comprensión verbal y producción oral. Las habilidades musicales se medirán a través de tareas estructuradas de percepción rítmica, discriminación melódica y reproducción de patrones musicales, seleccionadas o adaptadas con base en su validez para población preescolar.

Procedimiento: