EDITORIAL CSIC

Loquens
13, diciembre 2026, e133
eISSN: 2386-2637
https://doi.org/10.3989/loquens.2026.e133

Artículo

Reconocimiento de la prosodia emocional: creación y validación de un corpus de frases y pseudofrases en español peninsular

Emotional prosody recognition: creation and validation of an emotional corpus of sentences and pseudo-sentences in peninsular Spanish

Lola Terny
Departamento de Metrología y Ciencias del Lenguaje, Universidad de Mons, Bélgica
lola.terny@umons.ac.be, https://orcid.org/0009-0006-4982-5929

Kathy Huet
Departamento de Metrología y Ciencias del Lenguaje, Universidad de Mons, Bélgica
kathy.huet@umons.ac.be, https://orcid.org/0000-0002-7160-7489

Myriam Piccaluga
Departamento de Metrología y Ciencias del Lenguaje, Universidad de Mons, Bélgica
myriam.piccaluga@umons.ac.be, https://orcid.org/0000-0001-6169-5123

Resumen: El reconocimiento de las emociones transmitidas por la prosodia constituye una competencia esencial en la comunicación, pero los recursos experimentales en español siguen siendo limitados. Este estudio tiene como objetivo crear y validar un corpus de estímulos auditivos en español peninsular destinado a evaluar la identificación de emociones a partir de la voz. Una hablante ha producido frases semánticamente neutras y pseudofrases con entonaciones de enfado, miedo, alegría y tristeza. Treinta y tres evaluadores hispanohablantes han participado en un proceso de selección de grabaciones en línea. Los resultados muestran un reconocimiento global superior al azar y variaciones según la emoción representada, sin diferencias significativas entre frases y pseudofrases ni en función del sexo, la edad o la región de los evaluadores. El corpus resultante ofrece un conjunto de estímulos controlados, fiables y válidos, de los cuales se seleccionarán aquellos con índices de identificación superiores al 80% para su utilización en estudios sobre percepción emocional infantil.

Palabras clave: prosodia emocional; reconocimiento de emociones; corpus auditivo; español.

Abstract: Recognizing emotions conveyed by prosody is an essential skill in communication, but experimental resources in Spanish remain limited. This study aimed to create and validate a corpus of auditory stimuli in peninsular Spanish designed to evaluate the identification of emotions based on voice. A speaker produced semantically neutral sentences and pseudo-sentences with intonations of anger, fear, joy, and sadness. Thirty-three Spanish-speaking evaluators took part in the online selection process for the recording. The results showed overall recognition superior to chance and variations depending on the emotion represented, with no significant differences between sentences and pseudo-sentences or based on the gender, age, or region of the participants. The resulting corpus offers a set of controlled, reliable and valid stimuli, from which those with identification rates above 80% will be selected for use in studies on children's emotional perception.

Keywords: emotional prosody; emotion recognition; auditory corpus; Spanish.

Recibido: 09-07-2025 / **Aceptado**: 09-12-2025 / **Publicado**: 21-07-2026

Cómo citar: Terny, L., Huet, K. y Piccaluga, M. (2026). Reconocimiento de la prosodia emocional: creación y validación de un corpus de frases y pseudofrases en español peninsular. *Loquens, 13,* e133. https://doi.org/10.3989/loquens.2026.e133

Información complementaria ↓

1. INTRODUCCIÓN

La competencia emocional se define como la capacidad de identificar, comprender, expresar y gestionar tanto las propias emociones como las de los demás (Mikolajczak, 2020). Como señalan Bänziger *et al.* (2012), los estados emocionales pueden expresarse a través de diversas modalidades, tales como las expresiones faciales, variaciones vocales (prosodia) o gestos. En la mayoría de los casos, estas señales se interpretan de manera simultánea, debido a que la percepción de la emoción es multimodal[1]. En contraste con el reconocimiento de la expresión facial, el estudio de la expresión vocal de las emociones ha recibido una atención relativamente menor en la literatura adulta (Scherer, 2021; Zupan y Eskritt, 2023) y aún más limitada en la literatura infantil (Riddell *et al.*, 2024).

En el ámbito de la investigación sobre el reconocimiento vocal de las emociones se nota un interés general en elementos acústicos no necesarios para la percepción de los sonidos del habla, pero que informan sobre el estado o la intención del hablante. Las especificidades del ritmo y de la entonación, que se denominan habitualmente *prosodia* (Billières, 2014), funcionan como señales complementarias por la percepción del sentido general de las palabras y de las frases. La prosodia abarca los elementos suprasegmentales de la lengua hablada y su primer papel es lingüístico. Puede ser delineada mediante las fluctuaciones de ritmo, entonación y acentuación que se manifiestan en una misma palabra o en una estructura fraseológica. Estas variaciones son específicas de cada lengua y se caracterizan por seguir las reglas prosódicas que la estructuran. En el caso de las lenguas con un sistema acentual, como el inglés o el español, se hace necesario respetar las reglas de acentuación y variar la intensidad de una sílaba a otra dentro de una misma palabra. Este fenómeno, conocido en el ámbito lingüístico como *prosodia lingüística* (PL), se manifiesta en la variación de los patrones de entonación y acentuación en diferentes variedades de una misma lengua. Estas reglas, frecuentemente de carácter inconsciente e implícitas en el idioma nativo, se aplican de manera semejante en la transmisión de emociones. Para referirnos a estas variaciones prosódicas con fines emocionales, se emplea el concepto de *prosodia emocional* (PE). En el caso de que la prosodia de un hablante presente deficiencias, ya sea de índole lingüística o emocional, el hablante nativo será capaz de identificarlas de manera instantánea mediante la aplicación de sus conocimientos previos y la movilización de elementos contextuales asociados al entorno en el que tiene lugar la producción (Bak, 2016). En efecto, la capacidad de procesar de manera adecuada la prosodia emocional constituye un componente esencial de la competencia emocional más amplia del individuo y resulta fundamental para una comunicación efectiva.

Si bien existe un consenso generalizado entre la literatura científica que indica que el reconocimiento de las emociones vocales experimenta un progreso notable durante la transición de la infancia a la edad adulta (véase, por ejemplo, Filippa *et al.*, 2022; Grosbras *et al.*, 2018), y que posteriormente se produce una disminución a partir de cierta edad (Amorim *et al.*, 2021), el desarrollo de la PE parece demorarse más antes de alcanzar el nivel característico de la edad adulta, en contraste con el desarrollo de las expresiones faciales (véase, por ejemplo, Nelson y Russel, 2011; Berman *et al.*, 2016). La maduración de la capacidad de interpretar la PE seguiría una trayectoria de desarrollo específica, distinta del reconocimiento de las emociones faciales, como lo demuestran las diferencias en la decodificación de

[1] Para una comparación de la percepción de la emoción a partir de la expresión facial, la voz y el tacto en adultos, véase Schirmer y Adolphs (2017).

las señales faciales en comparación con las expresiones vocales (Morningstar *et al.*, 2018). En efecto, en adultos, la manifestación de las señales vocales parece ser más específica de la cultura e idioma del individuo (Elfenbein y Ambady, 2002; Cordaro *et al.*, 2018; Laukka y Elfenbein, 2021), a diferencia de las expresiones faciales, que, según ciertas teorías, son más universales (Ekman, 1994; Izard, 1994).

Sin embargo, los estímulos auditivos que se han utilizado en dispositivos metodológicos para investigar la PE son muy variables de un estudio al otro, como lo veremos a continuación. Además, la mayoría de las investigaciones previas se han focalizado en el idioma inglés (véase Tabla 1). No se ha desarrollado material adaptado para el español, y aún menos para una población infantil. Por ello, hemos optado por desarrollar un corpus de frases y pseudofrases emocionales en español peninsular. El presente artículo tiene como objetivo describir la creación y el proceso de validación de un corpus de estímulos auditivos emocionales que servirá de base para estudios posteriores con niños monolingües hispanohablantes y niños bilingües (de 4 a 8 años), con el fin de investigar la trayectoria del desarrollo de la PE en la infancia según distintos perfiles lingüísticos.

Al abordar el estudio de las expresiones emocionales vocales, resulta imperativo especificar con precisión el tipo de estímulos empleados. De hecho, se ha observado la creación y utilización de bases de datos o grabaciones de audio que contienen estímulos de diversa naturaleza. Estos estímulos abarcan desde vocalizaciones, tales como gritos, llantos y suspiros (Lima *et al.*, 2013), hasta palabras, interjecciones (Belin *et al.*, 2008) o pseudofrases (Bänziger *et al.*, 2012) y frases (Castro y Lima, 2010) con un contenido semántico emocionalmente neutro. Por tanto, se podrían clasificar los estímulos emocionales utilizados en tres grandes categorías. En primer lugar, se encuentran los estímulos lingüísticos, como las *frases* y las *palabras*, que presentan variaciones prosódicas (PL) propias del sistema lingüístico al que pertenecen. En segundo lugar, los estímulos denominados *expresiones no verbales* o *vocalizaciones* —como gritos o suspiros— que carecen de contenido lingüístico y no se asocian a ninguna lengua en particular. Finalmente, las *pseudofrases* que constituyen una categoría intermedia, aunque no tienen significado ni referente semántico, respetan las combinaciones fonotácticas propias de la lengua en la que han sido construidas y las reglas prosódicas (por ejemplo, reglas de acentuación del español).

En los corpus anteriores que se han creado, todos estos tipos de estímulos se producen con una intencionalidad emocional, es decir, presentan variaciones prosódicas diseñadas para expresar una emoción específica (PE). Algunos corpus recurren a actores, mientras que otros no lo hacen. En la escasa literatura existente sobre la identificación de emociones a partir de la prosodia en niños, las características de los estímulos empleados y desarrollados presentan una notable variabilidad (véase Tabla 1).

En el contexto de los estudios presentados en la Tabla 1, el desarrollo de la identificación de la prosodia emocional exhibe variaciones en función de la emoción asociada (por ejemplo, ciertas emociones presentan una identificación más precisa que otras), del tipo de estímulo (el caso de las vocalizaciones no verbales que se identifican con mayor facilidad en comparación con las frases o pseudofrases) y del idioma en los audios. En lo que respecta a esta última variable, se ha demostrado que, entre los tres y los cinco años, los niños poseen la capacidad de identificar las emociones transmitidas por las características prosódicas de la producción de la señal vocal de la misma manera en su lengua materna y en lenguas con las que no están familiarizados (Ma *et al.*, 2022), sugiriendo una habilidad universal. De acuerdo con los datos proporcionados por Chronaki *et al.* (2018), a partir de los ocho años de edad se evidenciaría una ventaja del grupo lingüístico, en el sentido de que los niños exhiben mejores resultados en su lengua materna (inglés) que en lenguas desconocidas para ellos (español, árabe y chino mandarín). No obstante, los enfoques metodológicos y los tipos de estímulos utilizados difieren entre ambos estudios, lo que dificulta su comparación. Además, no se han identificado otras investigaciones que hayan explorado esta habilidad desde una perspectiva interlingüística en población infantil, con la excepción del estudio de Roseberry-McKibbin y Brice (1999) realizado con niños bilingües hispanoamericanos. Por lo tanto, la prosodia constituye un área de estudio que se ha caracterizado por una escasa investigación desde una perspectiva de desarrollo, en comparación con otros aspectos del lenguaje. Desde una perspectiva interlingüística, se han realizado algunos estudios sobre el desarrollo de la prosodia lingüística, como los de Filipe *et al.* (2017) y Peppé *et al.* (2010), quienes desarrollaron y validaron el test PEPS-C en cinco lenguas europeas; sin embargo, el desarrollo de la prosodia emocional ha sido escasamente abordado.

Tabla 1: Características de los estímulos auditivos empleados en otros estudios con niños

Autores (año)	Tipo de estímulo	Idioma(s) del estímulo	Ejemplo (si lo hay)	Actores (sí/no)	Base de datos o creación original
Amorim *et al.* (2021)	vocalizaciones no verbales	portugués	vocalizaciones breves sin contenido verbal	no	creación original
Charpentier *et al.* (2018)	vocal /a/	francés	/a/	no	creación original
Chronaki *et al.* (2015)	interjecciones	francés	ah	sí	Maurage *et al.* 2007
*Chronaki *et al.* (2018)	**pseudofrases**	**inglés, árabe, español, chino**	**I nestered the flugs**	**no**	Pell *et al.*, 2009a; Pell *et al.*, 2009b; Liu y Pell, 2012
Correia *et al.* (2019)	frases semánticamente neutras	portugués	O futebol e um desporto	no	Castro y Lima, 2010
Doherty *et al.* (1999)	frases semánticamente neutras	inglés	The dog jumped into the back seat	no	creación original
Filippa *et al.* (2022)	pseudofrases vocal /a/	francés	Nekal ibam soud molen!; koun se mina lod belam?	sí	Geneva Multimodal Emotion Portrayals
Fujisawa y Shinohara (2011)	palabras semánticamente neutras	japonés	Suzuki-san (nombre propio)	sí	creación original
Grosbras *et al.* (2018)	vocalizaciones no verbales	inglés	ah	sí	Montreal Affective Voices
Guidetti (1991)	pseudofrases	lenguas indoeuropeas	Zi göt laich schong kil goster et hat sandik pronn ju ventsi	sí	Scherer y Wallbott, 1989
Leppänen y Hietanen (2001)	palabras semánticamente neutras	finés	Sarah (nombre propio)	sí	Leinonen *et al.*, 1997
*Ma *et al.* (2022)	**frases semánticamente neutras**	**inglés, francés, chino, español**	**I see a rug on the floor; This is a garbage can**	**no**	**creación original**
Matsumoto y Kishimoto (1983)	palabras semánticamente neutras	inglés, japonés	English alphabet and Japanese syllabary	sí	creación original
McCluskey *et al.* (1975)	frases filtrradas de paso bajo	/	/	sí	creación original
McCluskey y Albas (1981)	frases filtradas de paso bajo	/	/	no	creación original
Morningstar *et al.* (2019)	frases semánticamente neutras	inglés	Why did you do that?; I was just trying to be nice; I didn't know about it; You shouldn't have done that	sí	Morningstar *et al.*, 2017
Morningstar *et al.* (2024)	frases semánticamente neutras	inglés	I didn't know about it	sí	Morningstar *et al.*, 2017
*Morton y Trehub (2001) Experimento 2	**frases semánticamente neutras**	**inglés, italiano**	***No hay ejemplos de las frases en italiano***	**no**	**creación original**
*Morton y Trehub (2001) Experimento 3	**filtro de paso bajo**	**/**	**/**	**no**	**creación original**
Nelson y Russell (2011)	frases semánticamente neutras	inglés	I felt this feeling before; it was just a few days ago	sí	creación original
Autores (año)	Tipo de estímulo	Idioma(s) del estímulo	Ejemplo (si lo hay)	Actores (sí/no)	Base de datos o creación original
Neves *et al.* (2021)	frases semánticamente neutras; pseudofrases; vocalizaciones no verbales	portugués	Esta mesa é de madeira; Esta dêpa é de faneira; vocalizations	no	Lima *et al.*, 2013; Castro y Lima, 2010
Quam y Swingley (2012)	habla espontánea	inglés	*Habla espontánea*	no	creación original

*Roseberry-McKibbin y Brice (1999)	**frases semánticamente neutras**	**inglés, mexicano**	**Will you bring me the ball?**	**no**	**creación original**
Sauter *et al.* (2013)	vocalizaciones no verbales; palabras neutras	inglés	risas, suspiros, gruñidos; cifras (“uno”, “dos”...)	sí	Sauter, 2006; Sauter *et al.*, 2010
Śmiecińska (2016)	palabras semánticamente neutras	polaco	marchewka (“zanahoria” en polaco)	no	creación original
Zupan (2015)	frases semánticamente neutras	inglés	I’m going out of the room now and I’ll be back later	no	Diagnostic Analysis of Nonverbal Accuracy (DANVA-2)
*Tres estudios en lenguas maternas y desconocidas (Ma *et al.*, 2022; Chronaki *et al.*, 2018; Morton y Trehub, 2001) y una investigación con niños bilingües (Roseberry-McKibbin y Brice, 1999)					

Nuestro corpus se ha elaborado basándose en el estudio de Castro y Lima (2010), cuyo objetivo era crear una base de datos de frases y pseudofrases en portugués para investigar la prosodia emocional en adultos en su lengua materna y permitir comparaciones interlingüísticas. Compartimos, por tanto, sus conceptos de *complejidad lingüística* y *especificidad de la lengua*, que desarrollamos a continuación. En el ámbito del reconocimiento vocal de las emociones, se han formulado pocas hipótesis sobre el efecto del tipo de estímulo, denominado *complejidad lingüística* por Castro y Lima (2010). Esta hipótesis sostiene que un estímulo en el que la señal ha sido filtrada mediante un paso bajo es simple, mientras que una frase se considera compleja, en la medida en que su contenido semántico requiere un mayor procesamiento por parte del oyente. Si imaginamos un continuo que va de lo simple a lo complejo, en términos de procesamiento semántico, situaríamos en el extremo más simple los ruidos (por ejemplo, vocalizaciones como gritos), seguidos por interjecciones, palabras, pseudofrases y, finalmente, frases completas. Sin embargo, la naturaleza del estímulo no ha sido objeto de una investigación exhaustiva. En cuanto a la *especificidad lingüística*, se observa una situación similar: los criterios prosódicos específicos de cada lengua no se han considerado sistemáticamente como variables influyentes en el reconocimiento vocal de las emociones. Además, son escasos los estudios que adoptan un enfoque psicolingüístico y evolutivo de la prosodia emocional. Cabe añadir que la creación de los audios suele carecer de justificaciones metodológicas: con frecuencia, los autores no explicitan por qué han privilegiado un tipo de estímulo sobre otro.

A diferencia del trabajo de Castro y Lima (2010), el presente material ha sido específicamente diseñado para su aplicación en población infantil. Se ofrece así un conjunto restringido de estímulos auditivos destinado a futuras investigaciones sobre la trayectoria evolutiva de la prosodia emocional en la lengua materna, tanto desde una perspectiva interlingüística como en contextos bilingües.

2. MÉTODO

2.1. Creación y grabación

Hasta donde alcanza nuestro conocimiento, no se ha desarrollado ningún corpus orientado al estudio de la prosodia emocional en niños hispanohablantes de entre 4 y 8 años. Tomando como referencia el trabajo de Castro y Lima (2010), elaboramos dos tipos de estímulos adaptados con el objetivo de focalizar la atención en la prosodia, evitando tanto el procesamiento semántico como el sintáctico. Por una parte, diseñamos frases con contenido semántico emocionalmente neutro; por otra, generamos pseudofrases derivadas de estas. En la Tabla 2 se exponen las características lingüísticas de todas las frases y pseudofrases que se han construido y empleado para el corpus final. Según los autores, esta doble tipología de estímulos permite examinar el papel de los procesos lingüísticos generales frente a los procesos específicos implicados en la prosodia emocional.

2.1.1. Material

La creación de estímulos de tipo *frase* persigue dos objetivos principales:

1. Investigar el reconocimiento de emociones a partir de la prosodia emocional, así como las variaciones prosódicas asociadas a la expresión de cada emoción en español
2. Analizar un posible efecto de la lengua —y en particular de las restricciones prosódicas propias del español— en el reconocimiento de emociones (lo cual no sería posible utilizando únicamente interjecciones o vocalizaciones no verbales)

La necesidad de generar estímulos de tipo *pseudofrase* se justifica por las razones siguientes:

1. Neutralizar cualquier sesgo o interferencia relacionada con el contenido semántico
2. Contrarrestar la posible ausencia de neutralidad emocional en las frases (una frase aparentemente neutra puede adquirir una connotación positiva o negativa en función de la experiencia del oyente)

Las frases se han creado controlando varias variables en relación con los niños. En cuanto a los sustantivos y los verbos, nos basamos en las imágenes del vocabulario receptivo de los niños de cuatro años consultando la tarea de denominación de la prueba PPVT-5 (PeaBody versión española). Después, comprobamos la frecuencia léxica de sustantivos y verbos utilizando la base de datos EsPal[2]. Las frases fueron leídas y comprobadas por dos logopedas del departamento, que confirmaron que correspondían a frases comprensibles y utilizables por niños a partir de cuatro años. Son frases simples con vocabulario de alta frecuencia que va de 0.50 ("escoba") a 2.02 ("negro")[3]. También se comprobó el número de sílabas (mínimo cinco, máximo diez). Las pseudofrases se obtuvieron a partir de frases preconstruidas, respetando las combinaciones fonotácticas permitidas en la lengua española.

Tabla 2: Características lingüísticas de las frases y de las pseudofrases del corpus (entre paréntesis, el número de sílabas; debajo, la transcripción fonológica; debajo, la combinación fonotáctica: C = consonante, V = vocal)

Frases *(número de silabas)* *CV | VVC combinación fonotáctica*	**Pseudofrases** *(número de silabas)* *CV | VVC combinación fonotáctica*
El gato come carne (7**)** /el ɣato kome karne/ VC | CVCV | CVCV | CVCCV	**El tago come parne (**7**)** /el taɣo kome parne/ VC |CVCV| CVCV |CVCCV
El helicóptero es negro (9**)** /el elikoptero es neɣro/ VC | VCVCVCCVCV | VC | CVCCV	**El ilekáptore es nagro (**9**)** /el ilekaptore es naɣro/ VC |VCVCVCCVCV| VC| CVCCV
Toca el tambor (5**)** /toka el tambor/ CVCV | VC | CVCCVC	**Teka el pambor (**5**)** /teka el pambor/ CVCV|VC|CVCCVC
La escoba está en la cocina (11**)** /la eskoβa esta en la coθina/ CV | VCCVCV | VCCV | VC | CV | CVCVCV	**La esposta está en la cinoca (**11**)** /la espota esta en la θinoka/ CV | VCCVCV| VCCV| VC| CV CVCVCV
El barco es verde (6**)** /el barko es berde/ VC | CVCCV | VC | CVCCV	**El borca es varde (**6**)** /el borka es barde/ VC |CVCCV| VC |CVCCV
La chica vive en un castillo (10**)** /la tʃika vive en un kastiʝo/ CV | CCVCV | CVCV | VC| VC | CVCCVCV	**La chipa vove en un postilla (**10**)** /la tʃipa vove en un postiʝa/ CV| CCVCV| CVCV |VC| VC CVCCVCV

[2] https://www.bcbl.eu/databases/espal/

[3] Se basa en el log_frq de EsPal, que tiene como valor mínimo 0.0014, como valor máximo 4.8523 y como valor medio 0.1393.

Frases *(número de silabas)* *CV | VVC combinación fonotáctica*	**Pseudofrases** *(número de silabas)* *CV | VVC combinación fonotáctica*
El plátano es amarillo (9**)** /el platano es amari̯o/ VC | CCVCVCV | VC | VCVCVCV	**El plánato es aramillo (**9**)** /el planato es aramijo/ VC| CCVCVCV| VC | VCVCVCV
La lámpara está en la mesa (10**)** /la lampara esta en la mesa/ CV | CVCCVCV | VCCV | VC | VC | CVCV	**La támpara esta en la desa (**10**)** /la tampara esta en la desa/ CV | CVCCVCV | VCCV | VC | CV| CVCV

2.1.2. Grabación

Las grabaciones se llevaron a cabo en una cámara anecoica de la Universidad de Mons (Bélgica), en el Departamento de Metrología y Ciencias del Lenguaje (Facultad de Psicología y Ciencias de la Educación), utilizando un micrófono Zoom H5 Handly Recorder Stereo (4 pistas). La digitalización se realizó a una frecuencia de muestreo de 44 kHz y una resolución de 16 bits. Una locutora bilingüe francés-español realizó las grabaciones emoción por emoción, grabándose a sí misma en múltiples ocasiones en un mismo audio y variando las características prosódicas con el fin de inducir cuatro emociones: la alegría, la tristeza, el enfado y el miedo. Para cada emoción, se efectuó una grabación larga, con múltiples repeticiones de cada una de las frases y de las pseudofrases de la Tabla 2. Cada audio inicial tenía una duración aproximada de diez minutos por emoción, debido a los varios intentos de la locutora. Posteriormente, se recurrió al uso del programa Audacity (Audacity 3.7.1) para la edición de los estímulos. En cada audio de diez minutos se segmentaron las producciones emocionales y se guardaron una a una las frases y pseudofrases producidas en formato.WAV. Se eliminaron algunas producciones, principalmente por motivos técnicos, como ruido del micrófono o pasajes inaudibles. Una vez recortados y guardados los estímulos emocionales, el informático del laboratorio implementó un sistema en línea que permite realizar una tarea de identificación forzada y que se detallará a continuación. En total, se presentaron 136 estímulos a los evaluadores, cuya distribución según la emoción (alegría, tristeza, enfado, miedo) y el tipo de estímulo (frase, pseudofrase) se detalla en la Tabla 3.

Los estímulos fueron grabados por una única locutora bilingüe francés-español de 29 años, por varias razones. En primer lugar, cabe señalar que está previsto desarrollar, en un estudio futuro, un corpus equivalente en lengua francesa. Por ello, se optó por una locutora bilingüe, lo que facilitará la comparación interlingüística utilizando un mismo perfil vocal. En segundo lugar, en una tarea de identificación de emociones entran en juego múltiples variables y el uso de diferentes locutores habría supuesto la incorporación de una variable adicional que no sería de mayor interés para la futura investigación. Finalmente, por razones técnicas: era más fácil volver a grabar los estímulos en caso de dificultad, ya que la locutora forma parte del departamento de Metrología y Ciencias del Lenguaje. Al finalizar las pruebas, se preguntó a los evaluadores si percibieron que el español no era la lengua materna de la locutora; todos respondieron de forma unánime que no detectaron ningún acento.

Tabla 3: Número de audios (*N*) en cada tipo de estímulos
para la etapa de validación por los evaluadores

Emoción / Condición	Alegría	Tristeza	Enfado	Miedo	Total
Frases	24	23	16	15	78
Pseudofrases	16	13	16	13	58
Total	40	36	32	28	136

2.2. Validación

2.2.1. Participantes

Un total de treinta y tres evaluadores hispanohablantes, de edades y procedencias geográficas diversas (véase Tabla 4), participaron en el estudio mediante una plataforma en línea. Se ha utilizado una muestra ocasional (también llamada "de conveniencia"), dado que los participantes fueron reclutados de manera voluntaria a través de una convocatoria en línea dirigida a hablantes nativos de español. Cada participante proporcionó información sociodemográfica básica, incluyendo fecha de nacimiento, nacionalidad, país y ciudad de residencia, lenguas habladas y en contacto, así como la presencia o ausencia de formación musical. Nueve evaluadores eran originarios de distintos países de América Latina: Argentina (*n* = 3), Chile (*n* = 5) y Ecuador (*n* = 1). Los veinticuatro restantes procedían de diferentes comunidades autónomas de España: Galicia (*n* = 12), Castilla y León (*n* = 1), Comunidad Valenciana (*n* = 1), Comunidad de Madrid (*n* = 5), Cataluña (*n* = 1), Andalucía (*n* = 3) y otra comunidad no especificada (*n* = 1). La muestra no se limita a los locutores del español originarios de la península ibérica, sino que también se han incluido evaluadores de América Latina. En efecto, la heterogeneidad de la muestra, tanto en términos etarios como geográficos, permite considerar que un alto porcentaje de reconocimiento emocional para un determinado audio no se debe a factores contextuales o individuales específicos, sino que refleja una expresión prosódica eficaz y suficientemente clara de la emoción objetivo. Así, puede inferirse que dicho audio constituye un ejemplar representativo de la emoción que se procura transmitir.

Tabla 4: Características descriptivas de la muestra según la edad y el origen geográfico

			Edad (en años)			
Origen geográfico	**Sexo**	**Número**	**Media**	**Desviación estándar**	**Mínima**	**Máxima**
España	Mujeres	12	31.1	5.28	21	39
	Hombres	12	32.5	6.56	23	47
América latina	Mujeres	3	35.3	10.41	27	47
	Hombres	6	29.0	4.82	23	34

2.2.2. Procedimiento

Se les proporcionan a los participantes las instrucciones lo más detalladas posible, dado que la tarea se realiza de forma remota. En caso de surgir alguna dificultad, se indica que pueden contactar con la investigadora principal. Se informa, en primer lugar, de que la actividad tiene una duración estimada de entre 20 y 30 minutos y se divide en dos partes: la primera, compuesta por frases, y la segunda, por pseudofrases. Ambas pueden completarse en momentos distintos o en una sola sesión. En total, se presentan 136 estímulos vocales: 78 en la primera parte y 58 en la segunda. Seguidamente, se pide a cada evaluador que dé su consentimiento informado antes de comenzar la tarea (haciendo clic en una casilla). Se le informa de que, una vez procesados los datos, la información recogida será anonimizada[4]. Solo el equipo de investigación encargado de este proyecto tendrá acceso a la información indicada en los datos de identificación.

A continuación, se recomienda a los evaluadores que utilicen auriculares y se encuentren en una habitación tranquila, con el fin de garantizar las mejores condiciones para la realización de las tareas. Se les indica a los evaluadores que deben responder de la manera más espontánea posible, basándose

[4] Se le recuerda que la UMONS cumple con el Reglamento General de Protección de Datos (RGPD) y concede gran importancia a la protección de sus datos personales. Se le aconseja que se ponga en contacto con el Responsable de Protección de Datos en dpo@umons.ac.be si tiene alguna duda sobre la protección de sus datos por parte de la UMONS.

en su primera impresión. Asimismo, se les informa que pueden reproducir los audios tantas veces como necesiten y que no existe límite de tiempo para completar la tarea. Finalmente, se informa a los evaluadores que la situación en la que participan no es natural, ya que se trata de un experimento con fines de investigación. Por ello, es normal que en algunos momentos les resulte difícil responder o que se sientan confundidos ante uno o varios audios. En caso de que la confusión sea excesiva, se les recomienda hacer una pausa. De lo contrario, deben continuar respondiendo de la forma más espontánea posible, basándose en su primera impresión.

La tarea se compone de dos partes: la primera consiste en identificar la emoción vinculada al audio. Cada evaluador debe elegir entre alegría, tristeza, enfado, miedo u otra. La alternativa "otro" se añadió en caso de que ninguna de las proposiciones correspondiera a lo que el evaluador quería responder (y para no elegir una de las emociones por defecto). Previamente al inicio de la tarea, se utilizan dos audios ejemplos, con el propósito de asegurar una comprensión clara y completa de los procedimientos a seguir por parte de cada evaluador, así como para facilitar su familiarización con la tarea. Tras la elección forzada, el evaluador debía situarse en 4 escalas según su elección:

- una escala de certeza (¿cuán seguro está de la respuesta dada, en una escala de 0 a 10?);
- una escala de valencia (¿es la emoción positiva o negativa, en una escala de −5 a +5?);
- una escala de intensidad (¿es intensa la emoción, en una escala de 0 a 10?);
- una escala de calidad de audio (¿es el audio de buena calidad, en una escala de 0 a 10?).

A continuación, los estímulos fueron presentados en un orden pseudoaleatorio previamente establecido, idéntico para todos los participantes.

3. RESULTADOS Y DISCUSIÓN

Conviene recordar que el objetivo principal de este estudio es seleccionar los estímulos auditivos con los mayores índices de reconocimiento entre los presentados a los evaluadores, con el fin de conformar un corpus validado que pueda ser utilizado en investigaciones futuras con niños hispanohablantes, tanto monolingües como bilingües[5]. Antes de presentar la selección final de estímulos, se expondrán algunos resultados relacionados con los porcentajes de identificación en función de la emoción transmitida y del tipo de estímulo. Asimismo, se analizará el posible efecto de variables como el sexo, la edad y la región geográfica de los evaluadores. Todos los análisis estadísticos se realizaron con *Jamovi* (2.4.8)[6]. Es importante señalar que algunas respuestas están ausentes (n = 198), ya que no era obligatorio seleccionar una emoción para avanzar al siguiente audio. En algunos casos, los participantes olvidaron hacer clic, por lo que la respuesta no se registró. Además, uno de los evaluadores no completó la segunda parte de la tarea (pseudofrases). En total, se esperaba un conjunto de 4488 respuestas (33 evaluadores*136 estímulos), pero se obtuvo un 4.4% menos debido a las omisiones mencionadas ((198/4488)*100 = 4.41%). Cabe añadir que tres evaluadores comenzaron por la segunda tarea (pseudofrases), mientras que los demás iniciaron con la primera (frases). La mayoría de los participantes completó ambas tareas en dos días diferentes.

3.1. Reconocimiento de las emociones

En primer lugar, se presenta el porcentaje de reconocimiento obtenido para cada emoción considerando el total de evaluadores (véase Figura 1). Todas las emociones fueron identificadas por encima del nivel

[5] Si desea acceder al corpus con todos los audios, póngase en contacto con el primer autor del artículo.

[6] *Jamovi* (https://www.jamovi.org/) es una interfaz gráfica para R con apariencia de hoja de cálculo. Ejecuta código de R de forma interna al seleccionar opciones de análisis, sin que el usuario deba interactuar con el código directamente.

de azar, siendo la tristeza la emoción con mayor porcentaje de aciertos (83.8%) y el enfado la que obtuvo el valor más bajo (52.8%). Como se observa en la Tabla 5, algunos estímulos correspondientes a las emociones de tristeza y miedo alcanzaron un 100% de reconocimiento por parte de ciertos evaluadores. El hecho de que el enfado sea la emoción con el porcentaje más bajo de reconocimiento es sorprendente, ya que en la literatura científica sobre población adulta esta emoción es una de las que mejor se identifican a nivel vocal (Chronaki *et al.*, 2018; Filippa *et al.*, 2022). Sin embargo, el estudio de Chronaki *et al.* (2018) se realizó con participantes que tienen el inglés como lengua materna, y el de Filippa *et al.* (2022) con participantes que tienen el francés como lengua materna. Podemos suponer que habría diferencias en el reconocimiento vocal de las emociones según el idioma considerado en el estudio y la tarea. Por otro lado, la tristeza alcanza el mayor porcentaje de reconocimiento con un 83.8%. Es una emoción que ha demostrado ser bien reconocida tanto por los niños como por los adultos (Amorim *et al.*, 2021) a nivel vocal. En lo que concierne al miedo, se ha reconocido con un 76.2% a nivel global y es una emoción menos reconocida, a menudo confundida con la tristeza (Pell *et al.* 2009a). Por último, la alegría tiende a ser mejor reconocida (77.4%) en este tipo de dispositivo de elección forzada porque es la única emoción positiva y su valencia emocional es, por tanto, opuesta a la tristeza, al miedo y al enfado.

Tabla 5: Porcentaje de reconocimiento según la emoción vinculada y el tipo de estímulo auditivo

Condición	Emoción vinculada	Media (%)	Desviación estandar (%)	Minima (%)	Maxima (%)
Frase	Enfado	51.5	25.43	3.03	97.0
	Alegria	79.3	14.34	39.39	97.0
	Miedo	77.4	13.40	51.52	100.0
	Tristeza	82.6	10.97	63.64	97.0
Pseudofrase	Enfado	54.2	22.29	21.21	90.9
	Alegria	74.6	17.03	30.30	93.9
	Miedo	74.8	12.15	48.48	87.9
	Tristeza	86	3.40	78.79	90.9
Total	Enfado	52.8	23.56	3.03	97.0
	Alegria	77.4	15.44	30.30	97.0
	Miedo	76.2	12.66	48.48	100.0
	Tristeza	83.8	9.08	63.64	97.0

Dado que no se cumplían las condiciones de normalidad y homogeneidad de los datos, se realizó una prueba de Kruskal–Wallis para examinar el efecto de la emoción sobre los porcentajes de identificación, mostrando un efecto global significativo con un tamaño de efecto grande ($\chi^2(3) = 36.9$, $p < .001$, $\varepsilon^2 = .27$). Las comparaciones post hoc (pruebas de Wilcoxon) revelaron que los porcentajes de identificación para la emoción de enfado diferían significativamente de los de alegría ($Z = 6.19$, $p < .001$, $r = .73$, efecto grande), miedo ($Z = 5.43$, $p < .001$, $r = .70$, efecto grande) y tristeza ($Z = 7.64$, $p < .001$, $r = 0.93$, efecto muy grande). Esto indica que el enfado se reconoce significativamente peor que las demás emociones. También se observó una diferencia moderada entre miedo y tristeza ($Z = 3.95$, $p = .027$, $r = .49$, efecto moderado). No se encontraron diferencias significativas entre las demás emociones.

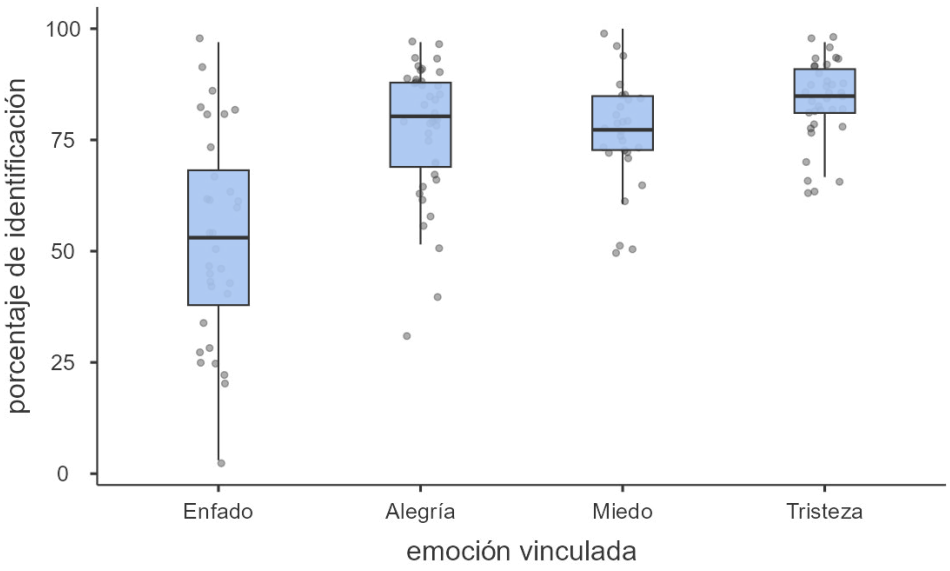

Figura 1: Reconocimiento de las emociones (%) considerando todos los tipos de estímulos.

Después, se comparó el porcentaje de reconocimiento de los estímulos de tipo *frase* con el de las *pseudofrases*, y se realizó la prueba U de Mann-Whitney[7] para comparar los dos tipos de estímulos en función de las respuestas correctas de los evaluadores. No se encontraron diferencias significativas en la identificación de las emociones en los estímulos de tipo *frase* en comparación con los estímulos de tipo *pseudofrase*. Se podría haber esperado mejores resultados con las pseudofrases porque, al tratarse de la segunda tarea, es posible que los evaluadores se acostumbraran a la tarea y a la voz del hablante. Este resultado coincide con el estudio de Castro y Lima (2010), que solo observaron un efecto del tiempo de reacción, que no se ha medido en nuestro dispositivo. En otros estudios, se ha observado un efecto del tipo de estímulo cuando se comparan las vocalizaciones no verbales y los estímulos de tipo *frase* o *pseudofrase*. En efecto, las vocalizaciones no verbales se identifican mejor tanto en niños (Neves *et al.*, 2021) como en adultos (Pell *et al.*, 2015).

Para profundizar en este aspecto, la Figura 2 muestra la distribución de los porcentajes de respuestas correctas en el reconocimiento de emociones, en función de la emoción objetivo y del tipo de estímulo (frases a la izquierda y pseudofrases a la derecha). En los estímulos de tipo *frase* (lado izquierdo de la Figura 2) se observa una mayor dispersión en las respuestas asociadas al enfado (*DE* = 25.43; rango = 3.03–97.0) y al miedo (*DE* = 13.40; rango = 51.52–100.0). La alegría mostró una variabilidad intermedia (*DE* = 14.34; rango = 39.39–97.0), mientras que la tristeza presentó las respuestas más homogéneas (*DE* = 10.97; rango = 63.64–97.0). Este patrón sugiere que el contenido semántico podría influir en la interpretación emocional. En los estímulos de tipo *pseudofrase*, las distribuciones aparecen más concentradas, con una variabilidad menor en las respuestas asociadas al miedo (*DE* = 12.15) y a la tristeza (*DE* = 3.40), y más heterogéneas para el enfado (*DE* = 22.29) y la alegría (*DE* = 17.03). Esto podría reflejar una mayor dependencia de las claves prosódicas en ausencia de contenido léxico. En conjunto, la tristeza emergió como la emoción mejor reconocida y con menor dispersión, mientras que el enfado fue la emoción menos reconocida y más variable, tanto en las frases como en las pseudofrases. Finalmente, aunque en algunos casos la desviación estándar no es muy elevada, se observa una amplia diferencia entre los valores mínimo y máximo, lo que confirma la heterogeneidad interindividual en las respuestas. En la selección final, solo se han elegido estímulos del lado derecho de la barra negra (alcanzado un porcentaje de reconocimiento superior al 80%).

[7] La prueba U de Mann-Whitney es una técnica estadística no paramétrica utilizada cuando los datos no cumplen los supuestos requeridos por las pruebas paramétricas.

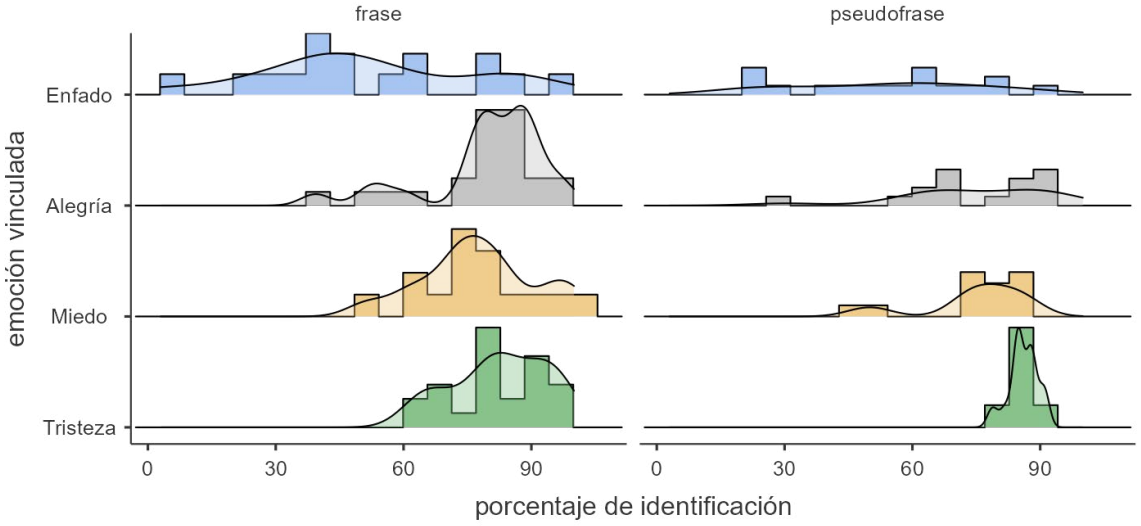

Figura 2: Curvas de densidad de las respuestas (en %) superpuestas a los histogramas que representan la distribución de los porcentajes de identificación por emoción *(enfado, alegría, miedo y tristeza)* y por tipo de estímulo *(frase* y *pseudofrase).*

3.2. Origen geográfico, sexo y edad

Con el propósito de garantizar que la región geográfica no ejercía influencia alguna en la identificación de las emociones, se ha llevado a cabo una comparación entre los evaluadores según su origen geográfico. En primer lugar, se han creado dos grupos: uno con los evaluadores originarios de España (*n* = 24) y otro con los originarios de América Latina (*n* = 9). Para comparar las medianas de estos dos grupos independientes, que no son equilibrados y cuya distribución de los datos no sigue una distribución normal, se realizó una prueba U de Mann-Whitney. No se ha demostrado una diferencia significativa entre estos dos grupos en relación con el porcentaje de respuestas correctas. En segundo lugar, dentro de los participantes españoles (*n* = 24), se observa que el 50% reside en la Comunidad Autónoma de Galicia (*n* = 12). Por consiguiente, se determinó la necesidad de corroborar que los resultados obtenidos no diferían de los de los españoles originarios de otras comunidades autónomas distintas a Galicia. De nuevo, se han creado dos grupos: uno con los evaluadores originarios de la Comunidad Autónoma de Galicia (*n* = 12) y otro con los evaluadores originarios de otra Comunidad Autónoma (*n* = 12), seleccionando únicamente los participantes originarios de España para el análisis. Se realizó una prueba U de Mann-Whitney para observar si los evaluadores originarios de Galicia tenían porcentajes de identificación correcta que difieren significativamente de los evaluadores originarios de otra Comunidad Autónoma. Se procedió a este análisis a nivel global (porcentaje global con los dos tipos de estímulo y las cuatro emociones) y emoción por emoción. Los resultados no revelaron diferencias significativas en el estudio de los dos grupos, tanto en términos de porcentaje global como el porcentaje según la emoción vinculada (alegría, tristeza, miedo o enfado). Se replicó el mismo procedimiento estadístico para comprobar si había diferencias entre hombres y mujeres en cuanto al porcentaje total, al porcentaje por emoción y al porcentaje por tipo de estímulo. Se conformaron dos grupos (hombres: *n* = 18; mujeres: *n* = 16) y se compararon mediante la prueba U de Mann-Whitney. Los resultados (véase Tabla 6 para los valores detallados) no mostraron diferencias significativas ni a nivel general, ni en función de la emoción evaluada o el tipo de estímulo presentado. En consecuencia, no se observaron diferencias significativas entre hombres y mujeres, ni entre participantes originarios de España o América Latina y aquellos de Galicia u otras comunidades, en la tarea de reconocimiento emocional. Asimismo, el porcentaje de reconocimiento no varió de forma significativa según la emoción expresada o el tipo de estímulo.

Tabla 6: Resultados detallados de los análisis estadísticos no significativos

	Origen geográfico (España: *n* = 24, América latina: *n* = 9)		**Comunidad Autónoma** (Galicia: *n* = 12, otras: *n* = 12)		**Sexo** (hombre: *n* = 18; mujer: *n* = 15)	
% total	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*
	76	.131	71	.977	121	.625
% por emoción[1]	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*
alegría	111.5	.906	58.5	.451	106	.293
tristeza	68.5	.062	61.5	.559	108	.326
miedo	86	.262	70	.931	129	.842
enfado	79	.163	71.5	1.000	119	.561
% tipo de estímulo[1]	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*	U de Mann-Whitney	*p*
frase	83	.217	61.5	.563	113	.437
pseudofrase	106.5	.754	62.5	.602	113	.426

[1] Los resultados por emoción contienen todos los tipos de estímulos juntos (frases pseudofrases) y los resultados por tipos de estímulos contienen todas las emociones juntas (alegría, tristeza, miedo y enfado).

Finalmente, se calcularon correlaciones de Bravais-Pearson entre la edad de los evaluadores y los distintos porcentajes de reconocimiento (globales, por emoción y según el tipo de estímulo). Todas estas correlaciones resultaron no significativas y presentaron coeficientes bajos, lo que indica una ausencia de asociación entre ambas variables sugiriendo que los participantes, independientemente de su edad, obtuvieron resultados comparables en la tarea. Por lo tanto, dado que no se ha evidenciado ningún efecto de la procedencia geográfica ni del sexo, y tampoco se ha encontrado relación alguna entre los resultados y la edad, la selección de los estímulos no debe tener en cuenta estas variables.

3.3. Selección final

En esta sección se presenta la selección final del corpus de estímulos auditivos (véase Tabla 7). Tal como se señaló previamente, las variables sociodemográficas de los evaluadores, edad, región geográfica y sexo, no mostraron efectos significativos en el reconocimiento de los estímulos emocionales. Por tanto, dichas variables no fueron consideradas en el proceso de selección final, dado que no parecieron influir en las elecciones de los participantes. Con el objetivo de seleccionar las ocho frases y las ocho pseudofrases que conformarán la base del futuro estudio con niños, se calculó el porcentaje de reconocimiento de cada uno de los 136 audios por parte de los treinta y tres evaluadores. A continuación, se clasificaron todos los estímulos según su nivel de reconocimiento, seleccionando aquellos cuya tasa de identificación superaba el 80%. En esta primera fase, se identificaron 36 frases (enfado: *n* = 4; tristeza: *n* = 15; alegría: *n* = 13; miedo: *n* = 6) y 26 pseudofrases (enfado: *n* = 3; tristeza: *n* = 12; alegría: *n* = 7; miedo: *n* = 4) que cumplían este criterio. Además, se comprobó que el nivel de certidumbre medio fuera superior a siete, que la intensidad media fuera de al menos cinco sobre diez y que la calidad media estuviera evaluada entre ocho y diez. En una segunda fase, se seleccionaron ocho frases y ocho pseudofrases (dos por emoción en cada caso), procurando maximizar el porcentaje de reconocimiento y garantizar que no hubiera repeticiones dentro del conjunto de estímulos. Dado el número limitado de ejemplos disponibles para algunas emociones (por ejemplo, solo cuatro frases de enfado), la selección se realizó de manera prioritaria para estos casos. Como resultado, el conjunto final incluye dieciséis estímulos: para cada tipo de material (frase y pseudofrase), se dispone de dos ejemplos correspondientes a cada una de las cuatro emociones (alegría, tristeza, miedo y enfado). La Tabla 6 presenta los estímulos seleccionados, junto con el número de evaluadores que identificaron correctamente la emoción representada, el porcentaje total de reconocimiento, así como la media de las puntuaciones relativas a la intensidad, la certidumbre y la calidad del audio.

Tabla 7: Estímulos auditivos seleccionados para el corpus final

Código estímulo		Condición	Emoción	Número de evaluadores	Total (%)	Intensidad media (/10)	Certidumbre media (/10)	Calidad media (/10)
62	E	Frase	Miedo	27	81.82	6.64	7.52	9.12
58	G	Frase	Enfado	28	84.85	5.85	8	8.91
54	D	Frase	Tristeza	30	90.91	7.03	8.48	9.09
32	A	Frase	Tristeza	32	96.97	7	8.55	9.12
44	C	Frase	Alegría	32	96.97	6.67	7.88	9.39
56	F	Frase	Miedo	32	96.97	7.36	8	9.15
59	B	Frase	Alegría	32	96.97	7.55	8.27	9.24
70	H	Frase	Enfado	32	96.97	6.73	8.15	8.73
52	C	Pseudofrase	Enfado	27	81.82	7	7.72	8.16
28	E	Pseudofrase	Tristeza	28	84.85	6.56	7.91	8.94
34	G	Pseudofrase	Miedo	28	84.85	8.19	8.16	8.69
09	A	Pseudofrase	Miedo	29	87.88	7.63	8.31	8.88
19	B	Pseudofrase	Tristeza	29	87.88	6.69	8.13	8.66
04	H	Pseudofrase	Enfado	30	90.91	7.81	8.44	8.34
54	F	Pseudofrase	Alegría	30	90.91	6.5	7.56	8.53
14	D	Pseudofrase	Alegría	31	93.94	7.16	8.38	9

Por último, en lo que respecta a la fiabilidad global de las evaluaciones relativas a los dieciséis estímulos auditivos, se observa un alto grado de consistencia entre los evaluadores, con un coeficiente alfa de Cronbach de 0.91 calculado a partir de las escalas de certidumbre. Asimismo, se evidencia una elevada coherencia interna en ambos tipos de estímulos, con un alfa de 0.85 para las frases y de 0.84 para las pseudofrases.

4. CONCLUSIÓN

El objetivo de este estudio fue la creación y validación de un corpus de estímulos auditivos emocionales en lengua española, con vistas a su utilización posterior en una población infantil. Dado que no existía un material de este tipo, se optó por crear estímulos propios en forma de frases y pseudofrases, justificando los criterios de selección y las variables controladas en función de la población destinataria. Estas frases y pseudofrases fueron grabadas por una locutora bilingüe francés-español que expresó las cuatro emociones objetivo (enfado, alegría, miedo y tristeza), y posteriormente fueron sometidas a un procedimiento de validación mediante la participación de hablantes hispanohablantes —hombres y mujeres— procedentes de distintas regiones del mundo hispánico, con identidades socio-lingüísticas diversas y con edades variadas. En total, se evaluaron 136 estímulos en función de la emoción expresada, su intensidad y su valencia. Asimismo, se solicitó a los evaluadores que indicaran el grado de certeza de su respuesta respecto a la emoción percibida. Los análisis estadísticos no revelaron ningún efecto significativo del sexo o de la procedencia geográfica y ningún vínculo con la edad de los participantes sobre las valoraciones recogidas. El porcentaje de reconocimiento correcto varió significativamente en función de la emoción expresada, pero no en función del tipo de estímulo (frase o pseudofrase). Finalmente, se seleccionaron dieciséis estímulos: ocho frases y ocho pseudofrases, distribuidas equitativamente según la emoción expresada, en base al porcentaje de reconocimiento correcto (igual o superior al 80%) y procurando además garantizar una variedad en términos de construcción sintáctica, léxico y número de sílabas. Estos dieciséis estímulos constituyen una nueva tarea estandarizada de reconocimiento de emociones en español, adaptada al público infantil.

Para futuras investigaciones se prevé investigar el reconocimiento de las emociones a partir de variaciones vocales en niños con perfiles lingüísticos variados y con una perspectiva interlingüística, es

decir, con audios en el idioma materna del niño y en un idioma que desconoce. Esto es especialmente relevante dado que se ha demostrado que la competencia de los niños en tareas prosódicas, tanto lingüísticas como emocionales, se desarrolla a ritmos diferentes según su lengua materna (Filipe *et al.*, 2017; Peppe *et al.*, 2010). Con este propósito, también se creará un corpus equivalente en lengua francesa, utilizando el mismo procedimiento. Este material está destinado a estudios futuros con niños monolingües francófonos, monolingües hispanohablantes y bilingües francés-español o español-francés, con el objetivo de documentar la trayectoria del desarrollo de la identificación emocional basada en la prosodia en niños con distintos perfiles lingüísticos. Dado que la literatura demostró una diferencia entre vocalizaciones y frases/pseudofrases, se integraría también este tipo de estímulo en los futuros estudios con los niños de distintos perfiles lingüísticos.

Información complementaria ↑

Fuentes de financiación

No procede.

Material suplementario

No procede.

Disponibilidad de datos

No procede.

Agradecimientos

Queremos expresar nuestro agradecimiento a todas las personas evaluadoras que participaron en el estudio y que dedicaron su tiempo y atención para que la creación del corpus fuera posible. Su colaboración ha sido fundamental para el desarrollo de esta investigación.

Declaración de contribución de autoría

Lola Terny: Conceptualización, Análisis formal, Adquisición de fondos, Investigación, Metodología, Administración del proyecto, Supervisión, Validación, Visualización, Redacción –borrador original, Redacción– revisión y edición.

Kathy Huet: Conceptualización, Gestión de datos, Recursos, Software, Análisis formal, Investigación, Metodología, Redacción – revisión y edición.

Myriam Piccaluga: Conceptualización, Gestión de datos, Recursos, Software, Análisis formal, Investigación, Metodología, Redacción – revisión y edición.

Declaración de conflicto de intereses

Los/as autores/as de este artículo declaran no tener conflictos de intereses financieros, profesionales o personales que pudieran haber influido de manera inapropiada en este trabajo.

Declaración de uso de Inteligencia Artificial

No procede.

REFERENCIAS

Amorim, M., Anikin, A., Mendes, A. J., Lima, C. F., Kotz, S. A. y Pinheiro, A. P. (2021). Changes in vocal emotion recognition across the life span. *Emotion*, *21*(2), 315-325. https://doi.org/10.1037/emo0000692

Bänziger, T., Mortillaro, M. y Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion*, *12*(5), 1161-1179. https://doi.org/10.1037/a0025827

Bąk, H. (2016). The State of Emotional Prosody Research—A Meta-Analysis. En H. Bąk (Ed.), *Emotional Prosody Processing for Non-Native English Speakers* (pp. 79-115). Springer International Publishing. https://doi.org/10.1007/978-3-319-44042-2_5

Belin, P., Fillion-Bilodeau, S. y Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531-539. https://doi.org/10.3758/BRM.40.2.531

Berman, J. M. J., Chambers, C. G. y Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology, 142*, 391-399. https://doi.org/10.1016/j.jecp.2015.09.014

Billières, M., (2014), *Au son du fle - Michel Billières*. https://www.verbotonale-phonetique.com/phonetique-didactique/

Castro, S. L. y Lima, C. F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods, 42*(1), 74-81. https://doi.org/10.3758/BRM.42.1.74

Charpentier, J., Kovarski, K., Roux, S., Houy-Durand, E., Saby, A., Bonnet-Brilhault, F., Latinus, M. y Gomot, M. (2018). Brain mechanisms involved in angry prosody change detection in school-age children and adults, revealed by electrophysiology. *Cognitive, Affective & Behavioral Neuroscience*, *18*(4), 748763. https://doi.org/10.3758/s13415-018-0602-8

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P. y Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology*, *33*(2), 218236. https://doi.org/10.1111/bjdp.12075

Chronaki, G., Wigelsworth, M., Pell, M. D. y Kotz, S. A. (2018). The development of cross-cultural recognition of vocal emotion during childhood and adolescence. *Scientific Reports*, *8*(1), 1-17. https://doi.org/10.1038/s41598-018-26889-1

Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N. y McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion, 18*(1), 75-93. https://doi.org/10.1037/emo0000302

Correia, A. I., Branco, P., Martins, M., Reis, A. M., Martins, N., Castro, S. L. y Lima, C. F. (2019). Resting-state connectivity reveals a role for sensorimotor systems in vocal emotional processing in children. *NeuroImage*, *201*. https://doi.org/10.1016/j.neuroimage.2019.116052

Doherty, C. P., Fitzsimons, M., Asenbauer, B. y Staunton, H. (1999). Discrimination of prosody and music by normal children. *European Journal of Neurology*, *6*(2), 221226. https://doi.org/10.1111/j.1468-1331.1999.tb00016.x

Ekman, P. (1994). Strong Evidence for Universals in Facial Expressions: A Reply to Russell's Mistaken Critique. *Psychological Bulletin, 115*(2), 268-287. https://doi.org/10.1037/0033-2909.115.2.268

Elfenbein, H. A. y Ambady, N. (2002). Is there an in-group advantage in emotion recognition? *Psychological Bulletin, 128*(2), 243-249. https://doi.org/10.1037/0033-2909.128.2.243

Filipe, M. G., Peppé, S., Frota, S. y Vicente, S. G. (2017). Prosodic development in European Portuguese from childhood to adulthood. *Applied Psycholinguistics, 38*(5), 1045-1070. https://doi.org/10.1017/S0142716417000030

Filippa, M., Lima, D., Grandjean, A., Labbé, C., Coll, S. Y., Gentaz, E. y Grandjean, D. M. (2022). Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. *Scientific Reports, 12*(1), 1-8. https://doi.org/10.1038/s41598-022-21554-0

Fujisawa, T. X. y Shinohara, K. (2011). Sex differences in the recognition of emotional prosody in late childhood and adolescence. *Journal of Physiological Sciences*, *61*(5), 429435. https://doi.org/10.1007/s12576-011-0156-9

Grosbras, M.-H., Ross, P. D. y Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. *Scientific Reports, 8*(1), 14791. https://doi.org/10.1038/s41598-018-32868-3

Guidetti, M. (1991). L'expression vocale des émotions : Approche interculturelle et développementale = Vocal expression of emotions : A crosscultural and developmental approach. *L'Année Psychologique*, *91*(3), 383396. https://doi.org/10.3406/psy.1991.29473

Izard, C. E. (1994). Innate and Universal Facial Expression: Evidence from Developmental and Cross-Cultural Research. *Psychological Bulletin, 115*, 288-299. https://doi.org/10.1037/0033-2909.115.2.288

Laukka, P. y Elfenbein, H. A. (2021). Cross-Cultural Emotion Recognition and In-Group Advantage in Vocal Expression: A Meta-Analysis. *Emotion Review, 13*(1), 3-11. https://doi.org/10.1177/1754073919897295

Leinonen, L., Hiltunen, T., Linnankoski, I. y Laakso, M. L. (1997). Expression of emotional–motivational connotations with a one-word utterance. *The Journal of the Acoustical society of America*, *102*(3), 1853-1863. https://doi.org/10.1121/1.420109

Leppänen, J. M. y Hietanen, J. K. (2001). Emotion recognition and social adjustment in school-aged girls and boys. *Scandinavian Journal of Psychology*, *42*(5), 429435. https://doi.org/10.1111/1467-9450.00255

Lima, C. F., Castro, S. L. y Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods, 45*(4), 1234-1245. https://doi.org/10.3758/s13428-013-0324-3

Liu, P. y Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior research methods*, *44*(4), 1042-1051. https://doi.org/10.3758/s13428-012-0203-3

Ma, W., Zhou, P. y Thompson, W. F. (2022). Children's decoding of emotional prosody in four languages. *Emotion, 22*(1), 198-212. https://doi.org/10.1037/emo0001054

Matsumoto, D. y Kishimoto, H. (1983). Developmental characteristics in judgments of emotion from nonverbal vocal cues. *International Journal of Intercultural Relations*, *7*(4), 415424. https://doi.org/10.1016/0147-1767(83)90047-0

Maurage, P., Joassin, F., Philippot, P. y Campanella, S. (2007). A validated battery of vocal emotional expressions. *Neuropsychological Trends*, 2(1), 63-74.

McCluskey, K. W. (1975). Cross-cultural differences in the perception of the emotional content of speech: A study of the development of sensitivity in Canadian and Mexican children. *Developmental Psychology*, *11*(5), 551555. https://doi.org/10.1037/0012-1649.11.5.551

McCluskey, K. W. y Albas, D. C. (1981). Perception of the emotional content of speech by Canadian and Mexican children, adolescents, and adults. *International Journal of Psychology*, *16*(14), 119132. https://doi.org/10.1080/00207598108247409

Mikolajczak, M. (2020). *Les compétences émotionnelles*. Dunod.

Morningstar M, Huang S. y Dirks MA (2017). Vocal cues underlying youth and adult portrayals of socio-emotional expressions. *Journal of Nonverbal Behavior*, 1–29. https://doi.org/10.1080/15374416.2017.1350963

Morningstar, M., Nelson, E. E. y Dirks, M. A. (2018). Maturation of vocal emotion recognition: Insights from the developmental and neuroimaging literature. *Neuroscience & Biobehavioral Reviews, 90*, 221-230. https://doi.org/10.1016/j.neubiorev.2018.04.019

Morningstar, M., Dirks, M. A., Rappaport, B. I., Pine, D. S. y Nelson, E. E. (2019). Associations between anxious and depressive symptoms and the recognition of vocal socioemotional expressions in youth. *Journal of Clinical Child & Adolescent Psychology*, *48*(3), 491-500. https://doi.org/10.1080/15374416.2017.1350963

Morningstar, M., Hughes, C., French, R. C., Grannis, C., Mattson, W. I. y Nelson, E. E. (2024). Functional connectivity during facial and vocal emotion recognition: Preliminary evidence for dissociations in developmental change by nonverbal modality. *Neuropsychologia*, *202*, 108946. https://doi.org/10.1016/j.neuropsychologia.2024.108946

Morton, J. B. y Trehub, S. E. (2001). Children's Understanding of Emotion in Speech. *Child Development*, *72*(3), 834843. https://doi.org/10.1111/1467-8624.00318

Nelson, N. L. y Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*(1), 52-61. https://doi.org/10.1016/j.jecp.2011.03.014

Neves, L., Martins, M., Correia, A. I., Castro, S. L. y Lima, C. F. (2021). Associations between vocal emotion recognition and socio-emotional adjustment in children. *Royal Society Open Science*, *8*(11), 1-16. https://doi.org/10.1098/rsos.211412

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A. y Kotz, S. A. (2009a). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, *37*(4), 417-435. https://doi.org/10.1016/j.wocn.2009.07.005

Pell, M. D., Monetta, L., Paulmann, S. y Kotz, S. A. (2009b). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120. https://doi.org/10.1007/s10919-008-0065-7

Pell, M. D., Rothermich, K., Liu, P., Paulmann, S., Sethi, S. y Rigoulot, S. (2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological Psychology, 111*, 14-25. https://doi.org/10.1016/j.biopsycho.2015.08.008

Peppé, S. J. E., Martínez-Castilla, P., Coene, M., Hesling, I., Moen, I. y Gibbon, F. (2010). Assessing prosodic skills in five European languages: Cross-linguistic differences in typical and atypical populations. *International Journal of Speech-Language Pathology, 12*(1), 1-7. https://doi.org/10.3109/17549500903093731

Quam, C. y Swingley, D. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236250. https://doi.org/10.1111/j.1467-8624.2011.01700

Riddell, C., Nikolić, M., Dusseldorp, E. y Kret, M. E. (2024). Age-related changes in emotion recognition across childhood : A meta-analytic review. *Psychological Bulletin*, *150*(9), 10941117. https://doi.org/10.1037/bul0000442

Roseberry-McKibbin, C. y Brice, A. (1999). The perception of vocal cues of emotion by Spanish-speaking limited english proficient children. *Communication Disorders Quarterly, 20*(2), 19-24. https://doi.org/10.1177/152574019902000203

Sauter, D. A. (2006). *An investigation into vocal expressions of emotions: The roles of valence, culture, and acoustic factors* (Unpublished PhD thesis). University College London, London.

Sauter, D. A., Eisner, F., Calder, A. J. y Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly journal of experimental psychology*, *63*(11), 2251-2272. https://doi.org/10.1080/17470211003721642

Sauter, D. A., Panattoni, C. y Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, *31*(1), 97-113. https://doi.org/10.1111/j.2044-835X.2012.02081.x

Scherer, K. R., Wallbott, H. G. (1989) Development of a standardized vocal emotion recognition test, unpublished document

Scherer, K. R. (2021). Comment: Advances in Studying the Vocal Expression of Emotion: Current Contributions and Further Options. *Emotion Review, 13*(1). 57-59. https://doi.org/10.1177/1754073920949671

Schirmer, A. y Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences, 21*(3), 216-228. https://doi.org/10.1016/j.tics.2017.01.001

Śmiecińska, J. (2016). Emotional and linguistic prosody development in Polish children: Three different paths. *Poznan Studies in Contemporary Linguistics*, *52*(3), 519554. https://doi.org/10.1515/psicl-2016-0020

Zupan, B. y Eskritt, M. (2023). Facial and Vocal Emotion Recognition in Adolescence: A Systematic Review. *Adolescent Research Review, 9*(2), 1-25. https://doi.org/10.1007/s40894-023-00219-7

Zupan, B. (2015). Recognition of High and Low Intensity Facial and Vocal Expressions of Emotion by Children and Adults. *Journal of Social Sciences and Humanities*, *1*(4), 332-344. http://files.aiscience.org/journal/article/html/70320049.html