UNIVERSITAT DE BARCELONA

Neural encoding of speech in newborns: voice pitch, formant structure and the effects of prenatal acoustic environment

Sonia Arenillas-Alcón

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (**www.tdx.cat**) i a través del Dipòsit Digital de la UB (**diposit.ub.edu**) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (**www.tdx.cat**) y a través del Repositorio Digital de la UB (**diposit.ub.edu**) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (**www.tdx.cat**) service and by the UB Digital Repository (**diposit.ub.edu**) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

***Brainlab* – Cognitive Neuroscience Research Group**

Departament de Psicologia Clínica i Psicobiologia (secció Psicobiologia)

Facultat de Psicologia

Universitat de Barcelona

Neural encoding of speech in newborns: voice pitch, formant structure and the effects of prenatal acoustic environment

Thesis presented by

Sonia Arenillas-Alcón

To obtain the

Grau de Doctor per la Universitat de Barcelona

In accordance with the requirements for the

International PhD Diploma

Programa de Doctorat en Biomedicina

Sonia Arenillas Alcón

Carles Escera

Jordi Costa Faidella

Supervised by **Dr. Carles Escera** and **Dr. Jordi Costa-Faidella**

Barcelona, July 2023

AGRADECIMIENTOS – ACKNOWLEDGEMENTS

Nunca he escrito unos agradecimientos, pero tampoco había escrito nunca antes una tesis y aquí está, así que espero que esto también sepa manejarlo. Aunque en general soy muy abierta con cualquier tema, no suelo expresar mucho cuánto me importan todas las personas que me rodean, así que ésta es la mejor oportunidad que tendré para agradecer a todos los que me han ayudado a que pueda llegar a este punto; permitidme que me ponga sentimental. Además, es la única parte que sé seguro que os leeréis todos los que no pertenezcáis a mi tribunal y, por lo tanto, estéis obligados a leeros todo lo demás. ¡Espero no decepcionaros!

En primer lugar, gracias a mi madre y mi padre. Vivir lejos de las personas que quieres es difícil, pero ellos siempre me han animado a no conformarme con lo que es fácil y a trabajar duro, me han apoyado en cualquiera de mis objetivos aunque implicasen estar lejos de casa, y siempre siempre siempre han confiado en que era capaz de hacer cualquier cosa aunque yo pensase que no. Os lo debo todo. Gracias a mi hermana, la única que tengo pero que hace por siete: somos como el día y la noche, pero ninguna de las dos puede ser sin la otra. Gracias por siempre ser valiente con lo que quieres e inspirarme para hacer yo lo mismo, gracias por ser tan generosa y tan fuerte, no sabes lo orgullosa que estoy de ti.

Many thanks to my significant other. Thanks for your infinite support, your unlimited patience and your kindness. Thank you for opening the doors of your world to me, for growing up with me every day and always helping to look on the bright side. Danke for never giving up, and never letting me give up. Thank you for always being willing to exchange the food when I don't like what I ordered, and for getting out of bed even with a smile to turn off the light. I'm looking forward to starting together the next adventure, and sharing countless Kaffee und Kuchen. Dankeschön für alles, ich liebe dich ganz ganz dolle.

Después, quiero agradecer a todos y cada uno de los que formáis el Brainlab. De verdad, habéis sido mi familia aquí. Es difícil encontrar gente tan inteligente, pero que encima eso se dé combinado con personas que son tan extraordinarias, buenas y a la vez tan humildes, parece que me haya tocado la lotería por poder compartir estos años con vosotros. No sé ni por dónde empezar a daros las gracias a todos.

Lo primero de todo, infinitas gracias al Dr. Carles Escera y Dr. Jordi Costa-Faidella, gracias por ser los mejores directores y jefes del mundo (¡y no es peloteo, que ya me habéis firmado la tesis!). Gracias Carles por ser el capitán del barco, pero hacernos sentir a todos como si fuéramos el primer oficial. Gracias por haberme dado la oportunidad de venir aquí y aprender a tu lado.

Gracias por haber confiado en mí y en que podía hacerlo, por dedicarnos siempre tu tiempo como si fuéramos lo más importante de la agenda, por transmitirme tu pasión por lo que haces, por hacerme crecer en todos los aspectos. Gracias también a ti, Jordi, por concederme el privilegio de ser una de tus primeras supervisadas, por enseñarme a perseguir siempre la perfección y a la vez admitir abiertamente cuando no se consigue. Por enseñarme que no puedo saber de todo, pero que eso no me hace menos válida; por ser tan genio y a la vez capaz de transmitir lo que sabes de forma tan sencilla. Gracias por tus acertadas ~~e infinitas~~ correcciones que han hecho mejorar mi trabajo día tras día. Gracias Jose por convertir absolutamente todo lo que tocas en una maravilla: de mayor quiero ser como tú, ni que sea con la mitad de tu talento. Gracias Iria por demostrarme que se puede tener tiempo para todo: ser una gran madre y una gran científica. Gracias MJ y Judith por ponerlo tan fácil y ayudarnos tanto cuando nos llega la otra parte del doctorado: la docencia. Doctor Via, gracias por ser tan cercano siempre, y por alegrar los días con los chistes ~~tan malos~~ que todos apreciamos. Gracias Imma por ser como una segunda mamá y siempre gritar por el pasillo "¿y cómo está la sevillana?". Gracias a Paco, mi malagueño favorito, aunque me haya abandonado por pastos más ~~andaluces~~ verdes; gracias por enseñarme Matlab con 'tologordo' y por haberme explicado física básica con paciencia cuando no sabía ni lo que era una frecuencia. Gracias a todo el Institut: Marta Turró, Cristina, y Marta Rubio, gracias por salvarme la vida tantísimas veces con el papeleo, por ponérnoslo fácil, y gracias por siempre ser la alegría cuando os unís a desayunar.

Sobre los Brainlabers que conocí cuando aún caminaban por la senda de los predocs (aunque algunos de ellos ya hayan pasado al rango de doctores), sólo tengo palabras bonitas. Iré por despachos, y así espero no dejarme a nadie. Como no podía ser de otra manera, primero, gracias a mi Tere. Gracias por acogerme con cariño desde el minuto uno, gracias por enseñarme absolutamente todo lo que sé y por guiarme durante este proceso como lo has hecho, gracias por dejarme conocerte. Me harían falta libros para poder hacerte justicia: eres una persona maravillosa por dentro y por fuera, siempre dispuesta a dar el 200% en todo lo que haces y que aun así te parezca poco. Lo prometo: nos vemos en Ibiza. Gracias Natàlia por ser mi hada madrina, saber siempre de todo y cuidar tantísimo de mí en todos los aspectos. Te admiro muchísimo, aunque el primer día pensase que me odiabas. Gracias por obligarme a parar cuando me hacía falta, y por motivarme a seguir cuando me faltaban ganas. Gracias por enseñarme catalán con paciencia, acompañarme durante mis meses fuera, y por hacer que me sintiese más cerca de todos. Et trobaré a faltar molt. Gracias Mondejillar por haberte puesto en mis manos para supervisarte, y por haber demostrado desde el primer momento que merecías estar donde estás. Gracias por tener un humor más negro que el mío, y por ser una persona tan generosa y

con tantas cualidades pero a la vez tan sencilla. Alex, respira hondo y confía en ti en este camino que empiezas ahora, que esto es una carrera de fondo. Gracias a las gemelas secsis de las cuevas: Sam y Raquel. Gracias Sam por ser tan auténtica y tan artista, y por dedicarme la primera edición de tu primer disco: cuando seas mundialmente famosa podré venderlo y hacerme rica. Gracias Raquel por ser esa mezcla tan extraordinaria de persona dulce y ácida, y por alegrarte casi más que yo con mis progresos en la tesis. Gracias Fran, por dejarme tu mesa cuando aún ni me conocías, y por ser mi piedra de toque en Pittsburgh. Gracias por enseñarme lo mejor de los USA: siempre recordaré mi primer arcade, mi primer pumpkin carving y mi primera barbacoa con s'mores. Gracias gracias gracias. Gracias Vittoria, por ser siempre tan risueña y amable incluso cuando iba a molestarte a tu despacho mientras estabas concentrada. Mil gracias a Nadia, por tener tanto talento y ser tan crack, y enseñarme que hay luz fuera de la Academia. Marta Font, gracias por transmitirnos tranquilidad y perspectiva a todos, incluso cuando creemos que el mundo está ardiendo alrededor nuestro. Gracias, felicidades y mucha suerte a Stef, la nueva mamá y doctor-to-be del Brainlab: gracias por hacerme entender ein bisschen besser la cultura alemana. Gracias Giannina por siempre ser tan llana y amable, por tener siempre una sonrisa para todos, y por compartir mis penitas y agobios en el tramo final de la tesis. ¡Ánimo que tú lo tienes casi también! Gracias Marta por ser la siguiente andaluza en venir y hacerme compañía entre las líneas enemigas, gracias por coger el relevo del hospital y hacerlo tan bien. No te desanimes por las dificultades que vendrán, ¡tú puedes con ellas! Gracias Siham por demostrarme que se pueden aprender cosas de cero incluso sola, y dominarlas como has hecho tú; gracias por siempre reírte de las barbaridades que digo en catalán. Gracias Martina por siempre encargarte de organizar cosas cuando a los demás nos da pereza, gracias por siempre ser tan abierta compartiendo cosas de tu vida, y por dar la turra culturizando a la gente sobre la FFR. Thanks to Emma, for being the Greek version of myself, for sharing my sense of humor and always being so truly honest about everything. *Hombre, ehque* I will miss you. Grazie Enrico por siempre estar dispuesto a debatir sobre cualquier cosa y hacerme aprender sobre nuevos puntos de vista, y por enseñarme a hacer el mejor café en cafetera italiana. Gracias Yesenia por venir desde tan lejos dispuesta a aprender de todos y a dar lo mejor de ti: créeme, estás en el mejor sitio. Y, por último, gracias a los nuevos cracks del lab de arriba: Laura y Albeeeeeeeeert ♫. Gracias Laura por interesarte siempre por saber de la vida de todos, y por hacerme repasar hasta los conceptos más básicos de la FFR para poder explicártelos (ahora los tengo bien frescos para la defensa). Gracias Albert por ser simplemente Albert, gracias de verdad por llegar el último pero encajar y hacerte querer como si llevases años. Echaré mucho de menos tus abrazos de oso. Si me he dejado a alguien, os pido perdón. Brainlab, OS ADMIRO A TODOS

MUCHÍSIMO. Soy increíblemente afortunada por haberos conocido a todos, y siempre podréis contar conmigo. Me da una pena enorme irme, pero espero poder volver pronto.

También gracias al team del fondo del pasillo. Gracias María, por ser una amiga tan increíble, tan atenta y tan generosa. Gracias por acogerme en tu familia como si hubiera nacido en ella. Gracias por inspirarme y demostrarme que uno puede sobreponerse y salir más fuerte de todo; que aunque se vivan momentos muy duros uno puede con ellos. Eres una profesional como la copa de un pino, y tengo claro que lucharás para llegar a ser de las mejores sin importar en qué. Bendisiones por haberte conocido, no me faltes nunca. Gracias a Montse, por vivir sus dramas de final de tesis paralelos a los míos, por compartir agobios juntas y hacerme sentir tan comprendida. Gracias al Dr. Ottino por enseñarme lo peligroso que es Los Angeles y aun así aceptar quedar conmigo allí de noche.

Last but not least at the UB, thanks to all the people who were in the Brainlab for a short time, but who I wish they had stayed forever: Laura Winkler, Laura Jacxsens, Ken, Chris, Bautista… Thanks for all the great memories I have with each of you (almost all of them always related to beers, food and bravas from *La Enfadada*), I just wish these memories are not the last ones we share.

En cuanto a la gente del Hospital San Joan de Déu, quiero agradecer enormemente a la Dra. Gómez-Roig por abrirnos las puertas del hospital y brindarnos la oportunidad de trabajar allí. Gracias Lola, por confiar en nosotros y hacer del área de obstetricia y ginecología nuestra segunda casa. Gracias Érica, por coordinar, organizar y sacar adelante las mil tareas del proyecto, por convertirte en amiga y compartir tantos almuerzos y cotilleos. Gracias Carolina, por ayudarme en los malos ratos de revisión de papeleo, y por ser la nueva neuropsico con la que tener otros temas para hablar y despejarme de tanta FFR. Gracias Diana, por compartir proyectos con nosotros y trabajar tan duro codo con codo. Gracias Piedad, por compartir tardes y tardes y tardes con nosotros, y dejarnos desordenarte los despachos siempre con una sonrisa.

Finally, I also want to thank Bharath Chandrasekaran for giving me the opportunity to work with him, along with his amazing team, during my stay in Pittsburgh. Thank you for including me in all the meetings and making me feel like one of you from day one. Thanks also to Tobias Teichert, for taking the time to have meetings with me and inviting me from the very beginning to visit his lab and participate in everything. Special thanks also to Razan for always feeding me and for making possible my best moments in the US. Thanks to Aidah, Filip, Nikola, Sam, Benji and the others, for welcoming me so well into your lives and for making me enjoy a truly Thanksgiving as if I were surrounded by family.

Gracias a toda la gente que también vive lejos de mí, pero que se lo curran para estar siempre presentes y tratar de averiguar cuándo bajo a Sevilla o cuándo voy a visitarles a la punta del mundo en la que sea que viven en cada momento. Gracias Ismael, Bea, Nerea, Kev, Alberto, Julián, Dani, David, Álvaro y María… Gracias por llamarme siempre "la catalana", por interesaros por todo lo que hago, y por siempre tener tiempo para contarme las cosas que pasan a diario en vuestras vidas. Gracias por hacerme sentir como si nunca me hubiera ido. Gracias también a mi otra familia de Barcelona: andaluces, valencianos, canarios, alicantinos… Sois muchos como para poder mencionaros a todos, pero vosotros habéis hecho que esta etapa de vivir fuera de casa se sintiese llena de planes, de gente y de cosas nuevas que hacer. Gracias a vosotros no he perdido el acento ni una mijita.

Y para acabar, creo que habría escrito unos agradecimientos casi más largos que la propia tesis y aun así no bastarían para agradeceros a todos lo suficiente, ¡pero si sigo va a costar un riñón imprimirlo! Espero que hayáis disfrutado al leerlos tanto como yo he llorado al escribirlos. Os quiero mucho a todos, y gracias :)

INDEX

ABSTRACT

The process of language acquisition is a remarkable milestone in infant life observed across several cultures and populations, making oral language the most prevalent form of human communication. Regardless of cultural background or native language, infants show comparable developmental trajectories in acquiring language, suggesting a universal organic foundation for this ability. However, language acquisition is far from simple, as it involves the development of complex perceptual and cognitive abilities, linked to the maturation of the auditory system and the encoding of speech sounds. In this context, a well-developed ability to perceive and distinguish the phonemes that constitute the different languages serve as essential prerequisites for effective language acquisition and communication in infants. Therefore, assessing at birth the state of maturity of the abilities that will allow future language acquisition emerges as a primary aspect to study. The present thesis aimed at investigating the neural mechanisms underlying speech processing in neonates, as well as the potential effects of the prenatal acoustic environment on the development of these mechanisms.

Studying speech encoding abilities in newborns poses challenges due to the neonates' inability to provide overt responses. To overcome this, the frequency-following response (FFR) has emerged as a promising tool for exploring complex sound encoding. The FFR provides a non-invasive snapshot of sound processing in the brain and has proven potential to predict future language development. Additionally, the FFR appears to be well suited for the study of auditory plasticity, since it has been shown to be modulated by several factors both postnatally and prenatally, including context-dependent contingencies, different auditory experiences and short-term auditory training.

To investigate speech encoding, researchers have employed various speech stimuli in FFR studies (e.g., consonant-vowel syllable /da/, different single vowels with rising pitch, Mandarin syllables following the four lexical tones). These stimuli allowed to explore easily the neural responses to the voice pitch (mainly the fundamental frequency, F_0) in neonates, but do not allow to fully explore critical aspects of language encoding, such as how precisely the newborn brain encodes the temporal fine structure of auditory stimuli –particularly relevant to discriminate vowels. As a consequence, the extent to what neonates are able to encode the formant structure of speech sounds remains unclear.

Thus, the first objective of the present doctoral thesis was to design a new speech stimulus that simultaneously assessed fundamental frequency and temporal fine structure, to explore the functional maturity of the neural mechanisms involved in encoding voice pitch

contour and formant structure in newborns. We compared the FFRs recorded in a group of newborns and adults, and provided evidences on how newborns are able to encode voice pitch and formant structure at birth. We disclosed that newborns show a remarkable ability to encode the voice pitch of speech sounds with adult-like accuracy, and that they are able to encode the formant structure of the speech as well.

Given the exceptional plasticity of the neural mechanisms generating the FFR and its potential as a biomarker for early identification of speech encoding impairments, external factors that might influence speech encoding at birth were considered. Neonatal speech encoding is shaped by several factors that can affect the development of the auditory system, not only perinatally or long after birth, but also during the pregnancy period, including for example prematurity, prenatal clinical conditions, or prenatal environmental factors. Thus, a critical question arises regarding the potential influence of the fetal environment, particularly the acoustic environment, on the newborn's abilities to encode speech sounds. Consequently, the second and third studies of this thesis were focused on exploring the impact of prenatal music exposure and prenatal bilingualism environments, respectively, on the encoding of speech sounds at birth, by conducting FFR recordings in groups of newborns with different degrees of prenatal music and language exposure. The second study revealed that daily prenatal music exposure during the final trimester of pregnancy was linked to a stronger encoding of voice pitch at birth, suggesting that exposure to music during the prenatal period plays a facilitating role in tuning the infant's sensitivity to the fundamental frequency of human speech. The third study disclosed that newborns exposed to a monolingual prenatal environment during the final trimester of pregnancy exhibited a more accurate encoding of voice pitch and formant structure of speech sounds at birth. These findings were interpreted as the result of a less variable speech background in monolingually-exposed newborns, resulting in larger neural responses and more accurate encoding of voice pitch and formant structure at birth, while neonates bilingually-exposed would show an enhanced sensitivity to a broader range of frequencies without exhibiting an especially strong response at any of them.

Overall, the findings of the present thesis add significant value to the understanding of early language acquisition processes and highlight the importance of early acoustic experiences in shaping the development of auditory mechanisms. Further studies should continue to explore these acoustic environmental factors, as they hold large potential in designing early interventions for potential future speech and language disorders.

RESUM EN CATALÀ

L'adquisició de la parla és una de les fites més destacades de les primeres etapes de la vida infantil, que es presenta de forma pareguda en diverses cultures i poblacions, fent del llenguatge oral la forma més freqüent de comunicació humana. Independentment de l'origen cultural o de la llengua nativa, els nadons mostren trajectòries de desenvolupament comparables en l'adquisició del llenguatge, el que suggereix una base orgànica universal per a aquesta capacitat. No obstant això, l'adquisició del llenguatge dista de ser un procés simple, ja que implica el desenvolupament d'habilitats perceptives i cognitives complexes, lligades a la maduració del sistema auditiu i a la codificació dels sons de la parla. En aquest context, una capacitat ben desenvolupada per percebre i distingir els fonemes que constitueixen les diferents llengües serveix com a requisit previ essencial per a l'adquisició i la comunicació eficaç del llenguatge en els nadons. Per tant, la valoració en el naixement de l'estat de maduresa de les capacitats que permetran l'adquisició futura de la llengua sorgeix com un aspecte primordial a estudiar. La present tesi té com a objectiu investigar els mecanismes neuronals subjacents al processament de la parla en nounats, així com els efectes potencials de l'entorn acústic prenatal en el desenvolupament d'aquests mecanismes.

L'estudi de les habilitats de codificació de la parla en els nounats planteja reptes a causa de la incapacitat del nounat per proporcionar respostes obertes a requeriments dels qui fan la recerca. Per superar aquesta limitació, la resposta de seguiment de freqüència (RSF; FFR per les sigles en anglès) ha sorgit com una eina prometedora per explorar la codificació dels so complexos. La RSF proporciona una instantània no invasiva del processament del so al cervell i ha demostrat la seva capacitat per predir el desenvolupament futur del llenguatge. A més, la RSF sembla ser molt adequada per a l'estudi de la plasticitat auditiva, ja que s'ha trobat que està modulada per diversos factors tant en l'àmbit postnatal com prenatal, incloses les contingències dependents del context, diferents experiències auditives i un entrenament auditiu a curt termini.

Per investigar la codificació de la parla, els investigadors han emprat diversos estímuls de parla en estudis RSF (per exemple, síl·laba consonant-vocal /da/, diferents vocals simples amb to creixent, síl·labes del Mandarí seguint els quatre tons lèxics). Aquests estímuls han permès explorar fàcilment en nounats les respostes neuronals al to de la veu (principalment la freqüència fonamental, F_0), però no permeten explorar completament aspectes crítics de la codificació del llenguatge, com ara la precisió amb què el cervell nounat codifica l'estructura temporal fina d'estímuls auditius –especialment rellevants en la percepció del contingut vocàlic.

Com a conseqüència, encara no està clar fins a quin punt els nounats són capaços de codificar l'estructura de formants dels sons de la parla.

Així, el primer objectiu de la present tesi doctoral va ser dissenyar un nou estímul de la parla que avalués simultàniament la freqüència fonamental i l'estructura temporal fina, per explorar la maduresa funcional dels mecanismes neuronals implicats en la codificació del contorn de to de veu i l'estructura del formant en els nounats. Hem comparat els FFR enregistrats en un grup de nounats i adults i hem proporcionat evidències sobre com els nounats són capaços de codificar el to de la veu i l'estructura del formant al naixement. Vam trobar que els nounats mostren una capacitat notable per codificar el to de la veu dels sons de la parla amb una precisió semblant a l'adult, i que també són capaços de codificar l'estructura dels formants de la parla.

Un cop demostrat el punt de maduresa d'aquests mecanismes neuronals, i donada l'excepcional plasticitat del RSF i el seu potencial com a biomarcador per a la identificació precoç de les alteracions de la codificació de la parla, es van considerar factors externs que podrien influir en la codificació de la parla en néixer. La codificació de la parla neonatal està configurada per diversos factors que poden afectar el desenvolupament del sistema auditiu, no només durant el període perinatal o molt després del part, sinó també durant el període de l'embaràs, com ara la prematuritat, les condicions clíniques prenatals o els factors ambientals prenatals. Així, es planteja una qüestió interessant sobre la influència potencial d'aquest entorn fetal, especialment l'entorn acústic, en les capacitats del nounat per codificar els sons de la parla. En conseqüència, el segon i el tercer estudi es van centrar a explorar l'impacte de l'exposició prenatal a la música i els entorns de bilingüisme prenatal, respectivament, en la codificació dels sons de la parla en néixer, mitjançant la realització d'enregistraments RSF en grups de nounats amb diferents graus d'exposició prenatal a la música i al llenguatge. El segon estudi va mostrar que l'exposició diària de música durant el darrer trimestre de l'embaràs estava relacionada amb una codificació més forta del to de la veu al naixement, cosa que suggereix que l'exposició a la música durant el període prenatal té un paper facilitador en l'ajust de la sensibilitat del nadó a la freqüència fonamental de la parla humana, contribuint, per tant, a fomentar el processament i l'adquisició primerenca del llenguatge. D'altra banda, el tercer estudi va revelar que els nounats exposats a un entorn prenatal monolingüe durant l'últim trimestre de l'embaràs presentaven una codificació més precisa del to de la veu i de l'estructura de formants dels sons de la parla en néixer. Aquestes troballes es van interpretar com el resultat d'un fons de parla menys variable per a la codificació en nadons exposats de manera monolingüe, donant lloc a respostes neuronals més grans i a una codificació més precisa del to de la veu i de l'estructura del formant

en néixer, mentre que els nounats exposats a un entorn bilingüe mostrarien una sensibilitat millorada a una gamma més àmplia de freqüències sense mostrar una resposta especialment forta en cap d'elles.

En general, els resultats de la present tesi afegeixen un valor significatiu a la comprensió dels processos d'adquisició primerenca del llenguatge i posen de manifest la importància de les experiències acústiques primerenques en la configuració del desenvolupament dels mecanismes auditius i probablement, de l'adquisició futura del llenguatge. Els estudis posteriors haurien de continuar explorant aquestes dinàmiques, ja que tenen un gran potencial per dissenyar intervencions primerenques per als trastorns de la parla i el llenguatge.

FOREWORD

This work has been carried out at the *Brainlab-Cognitive Neuroscience Research Group* (Center of Excellence established by the Generalitat de Catalunya, 2021 SGR 00356), of the Department of Clinical Psychology and Psychobiology, Faculty of Psychology, University of Barcelona (UB, Barcelona, Catalonia, Spain), led by Dr. Carles Escera, and in the frame of the CERCA Center *Institut de Recerca Sant Joan de Déu* (IRSJD). Likewise, this work was carried in collaboration with the Department of Obstetrics and Gynecology, headed by Dr. María Dolores Gómez-Roig, of the Sant Joan de Déu (SJD) Barcelona Children's Hospital. The work was supported by the Spanish Ministry of Science and Innovation with the predoctoral fellowship FPI (PRE2018-084385) and the PGC2018-094765-B-I00 project (MCIN/AEI/ 10.13039/501100011033/FEDER "Una manera de hacer Europa"), the MDM-2017-0729-18-2 María de Maeztu Center of Excellence (MCIN/AEI/10.13039/501100011033), and the ICREA Acadèmia Distinguished Professorship awarded to Carles Escera.

LIST OF ORIGINAL PUBLICATIONS

STUDY I

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Scientific Reports*, 11, 6660.
DOI: https://doi.org/10.1038/s41598-021-85799-x

STUDY II

Arenillas-Alcón, S., Ribas-Prats, T., Puertollano, M., Mondéjar-Segovia, A., Gómez-Roig, M. D., Costa-Faidella, J., & Escera, C. (2023). Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: Evidence from neonatal frequency-following responses. *Developmental Science*, e13362.
DOI: https://doi.org/10.1111/desc.13362

STUDY III

Arenillas-Alcón, S., Gorina-Careta, N., Puertollano, M., Mondéjar-Segovia, A., Ijjou-Kadiri, S., Costa-Faidella, J., Gómez-Roig, M. D., & Escera, C. Effects of a prenatal bilingual environment on the neural encoding of speech sounds at birth: Evidence from frequency-following responses. *(Working paper)*

ABBREVIATIONS

ABR	Auditory Brainstem Response
AEP	Auditory Evoked Potential
EEG	Electroencephalography
ERP	Event-Related Potential
F_0	Fundamental Frequency
F_1	First Formant
F_2	Second Formant
FFR	Frequency-Following Response
FGR	Fetal Growth Restriction
fMRI	Functional Magnetic Resonance Imaging
Hz	Hertz
LGA	Large-for-Gestational Age
MEG	Magnetoencephalography
MMN	Mismatch Negativity
NICU	Neonatal Intensive Care Unit
NIRS	Near-Infrared Spectroscopy
PROMS	Profile of Music Perception Skills
RMS	Root-Mean-Square
SNR	Signal-to-Noise Ratio
TFS	Temporal Fine Structure
UNHS	Universal Newborn Hearing Screening

INTRODUCTION

Early speech encoding for language acquisition

Across different cultures and populations, oral language is acquired with remarkable ease, making it the most prevalent form of human communication. The process of language acquisition is so surprisingly quick and effortless that infants are able to produce full sentences in their mother tongue by the age of 3 years old (Kuhl, 2004). But, how can children go from babbling to producing complete sentences effortlessly, when adults have such a hard time learning new languages? Are there any specific neural mechanisms underlying language sound processing participating in the first weeks or months of life, or even before birth, that may contribute to the rapid emergence of language? Can we exert any influence on them, even months before birth?

Infants, regardless of their cultural background or mother tongue, exhibit comparable developmental trajectories in acquiring language, as supported by a growing body of research. These developmental processes involve a progression from the emergence of babbling (typically occurring between 5 and 10 months of age; Kuhl, 2004), to the production of their first words between 10 and 18 months (Feldman, 2019). Based on the consistent developmental pathways across various populations, languages, and cultures, research findings propose a universal organic foundation for language acquisition (Hoff, 2009; Kuhl, 2004; Sket et al., 2019). However, and despite the simplicity that may be assumed due to its universal basis, language acquisition is a complex and multifaceted process that involves the development of both perceptual and cognitive abilities. Language acquisition is also linked to the maturation of the auditory system and the encoding of speech sounds, making it a necessity for adequate speech perception to develop efficient auditory mechanisms to detect the subtle differences that characterize linguistic contrasts (Cabrera & Gervain, 2020; Werker & Tees, 2005).

Thus, speech encoding skills that are important for future language acquisition in infants should be carefully considered. Specifically, the ability to encode and distinguish speech phonemes is essential for communication and for an effective perception of language. Such importance is corroborated by studies showing that young infants, who are still in the process of acquiring language, are able to discriminate amongst almost all phonetic units present in different languages, whereas adults, who are already fluent in their mother tongue, cannot (Miyawaki et al., 1975). This ability depends on the sensitivity of the auditory system to the spectral and temporal characteristics of speech, which allow for the discrimination of different

speech sounds (Kuhl, 2010). Overall, a well-developed ability to encode and process speech sounds is a vital requirement for successful language acquisition in infants.

Speech discrimination relies on the sensitivity of the auditory system to encode and process the complex acoustic structure of speech sounds, i.e., the spectral and temporal characteristics of speech. Regarding vocalic sounds, the fundamental frequency (F_0; the lowest frequency of a periodic waveform that represents sound's periodicity) is related to the perception of voice pitch and facilitates speaker recognition (Benavides-Varela et al., 2012). Vowel formants (F_1, F_2, F_3…; certain harmonics with significant phonetic relevance that exhibit larger amplitude compared to the surrounding harmonics), in turn, serve as the primary cues to distinguish different vowels (Krizman & Kraus, 2019; Rosen, 1992). As far as it is known, infants' vowel perception becomes more language-specific at around 6 months of age and their perception of consonants in their native language increases at the age of around 11 months-old, being finally able to produce their first words at around 12 months of age (Kuhl, 2004). As evidence indicates, and despite its apparent complexity, the developmental window of language acquisition and speech processing has been well characterized during the first year of life (Kuhl, 2010; McCarthy et al., 2019). Nevertheless, the functional status of infants' speech encoding skills during the initial days after birth, when they are firstly exposed to the complexity and richness of the external acoustic environment, remains relatively understudied despite its relevance.

Assessment of early speech sounds encoding

Assessing speech encoding abilities in infants and newborns is a challenging task, as they cannot provide overt behavioral responses. However, several indirect techniques have been developed to explore the early stages of speech processing. The head-turn preference procedure, which relies on the tendency of infants to turn their heads toward a novel stimulus, has been used to evaluate infants' discrimination of speech sounds (Kuhl et al., 2003, 2006). Likewise, the high-amplitude sucking technique, also called non-nutritive sucking, based on the principle of habituation (i.e., decreased sucking behavior in response to a repeated auditory stimuli or a repeated pattern) and dishabituation (i.e., increased sucking behavior after the occurrence of a change in the pattern, or a new auditory stimulus, that captures the infant's attention), whereby newborns can modulate their sucking behavior in response to auditory stimuli, has been traditionally used for testing several aspects of speech processing in infants from birth to 4 months of age (Floccia et al., 1997). Another frequently used behavioral method when exploring speech processing is the habituation of visual fixation, in which the duration of

infants' visual fixation on each of the presented stimuli is measured, and a decrease in fixation-time over time (habituation) is interpreted as an evidence of stimulus recognition and discrimination (Polka et al., 1995).

In addition to behavioral procedures, non-invasive neurophysiological recordings are a valuable resource for researchers studying not only general cognitive processes, but specifically the early development of language. Thus, several non-invasive techniques have been used regarding the evaluation of language processing in infants, each of them with particular features that make them more or less suitable depending on the specific objectives of the study that is being carried out (Kuhl, 2010). Techniques such as near-infrared spectroscopy (NIRS) and functional magnetic resonance imaging (fMRI) provide good spatial resolution but limited temporal resolution, due to the slow nature of the changes that are registered (Aslin & Mehler, 2005; Gernsbacher & Kaschak, 2003). These limitations, together with their high cost of use, make these techniques more suitable for longer segments of speech such as sentences.

In contrast, the magnetoencephalography (MEG) technique provides excellent temporal resolution by measuring magnetic field changes associated with brain activity (Kuhl & Rivera-Gaxiola, 2008). However, its high cost, combined with the requirement to immobilize the newborn's head during the recording duration, has led researchers to a more widespread use of another, more economic, technique: the electroencephalography (EEG). EEG measures the activity of neural networks that fire in a synchronous and coordinated manner, by detecting the voltage changes that take place as a consequence of neural activity (Kuhl & Rivera-Gaxiola, 2008; Picton & Mazaheri, 2006). Specifically, to investigate speech and language encoding in populations of newborns, infants, and children, event-related potentials (ERP) have been increasingly used. ERPs reflect EEG activity that is time and phase-locked to the occurrence of a specific event, e.g., a sensory auditory stimulus, such a syllable or a word. ERPs are calculated by averaging multiple segments from the continuous EEG that are temporally aligned with a specific event, effectively reducing the impact of random noise or non-stimulus-related neural activity, and reflecting the systematic neural response to the experimental condition manipulated (Hoehl & Wahl, 2012). Although the spatial resolution of the activation source is limited, its low cost and precise time resolution render the ERP particularly well-suited for exploring the rapid and precisely-timed nature of human speech, specifically in participants unable to provide overt responses due to developmental or cognitive constraints, and so providing a more sensitive measure of the discriminative skills than behavior (Kuhl, 2010).

An extensive range of auditory evoked potentials (AEP) has been suggested to study speech encoding and detect language impairments. Several AEPs, including the N1/P2/N2 complex, the P300, and the mismatch negativity (MMN), have been widely used to study auditory processing (e.g., Alho et al., 1990; Näätänen et al., 2007; Näätänen & Picton, 1987). However, these AEPs are subject to several limitations, including their limited replicability, a notable variability across individuals in amplitude and latency, and the slow voltage fluctuations occurring hundreds of milliseconds after the evoking sound was presented, raising questions about their ability to accurately represent the characteristics of the auditory stimulus (Kraus & Nicol, 2014). To address these limitations, researchers have become interested in the auditory brainstem response (ABR) as an additional method for studying auditory processing. ABRs are AEPs whose activity originates in the brainstem in response to auditory stimuli and are characterized by a series of waveform peaks that correspond to different processing stages along the auditory pathway (Skoe et al., 2015). Interestingly, ABRs and otoacoustic emissions to simple auditory stimuli are fundamental in the universal newborn hearing screening (UNHS), currently employed worldwide to assess the integrity of the auditory pathway and hence providing the earliest possible diagnosis of infants suffering from hearing loss (Hoth et al., 2009; Moeller et al., 2006; Stuart et al., 1996). While ABRs have shown promise in investigating auditory processing, they are also limited in their ability to provide information about the encoding of complex sounds such as speech.

More recently, a variant of the ABR known as the frequency-following response (FFR) has been proposed as a promising tool for studying complex sound encoding, including those of speech (Skoe & Kraus, 2010). The FFR has shown potential in addressing the limitations of traditional AEPs, since it is less variable, more replicable and provides information about the neural encoding of complex acoustic features such as voice pitch and formant structure. These properties have made the FFR an increasingly popular tool for investigating auditory processing mechanisms in infants and newborns (Gorina-Careta et al., 2022; Lemos et al., 2021).

The Frequency-Following Response (FFR)

The Frequency-following response (FFR) is an AEP elicited by periodic sounds that reflects neural synchronization with the spectrotemporal components of the auditory eliciting signals along the ascending auditory pathway (Krizman & Kraus, 2019; Skoe & Kraus, 2010). The FFR can be recorded from the scalp by using noninvasive techniques, such as electroencephalography (EEG) or magnetoencephalography (MEG), and emerges 7 to 15 milliseconds after the beginning of the auditory stimulus, reflecting frequencies within the range of 100 to 1500 hertz (Hz)

(Krizman & Kraus, 2019; Moushegian et al., 1973; Skoe & Kraus, 2010). Only by replacing the conventional click or chirp sounds traditionally used in the UNHS with more complex periodic sounds, such as those encountered in speech or music, FFRs can be easily recorded using the same electrode montage already employed in ABR paradigms.

Thus, the FFR reflects the phase-locked neural activity to the periodicity of incoming sounds and faithfully reproduces the characteristics of the eliciting complex auditory stimuli, to an extent that the stimulus itself can be recognized when the neural signal is played through a loudspeaker (Picton, 2010) (see Figure 1). This particular feature highlights the potential of the FFR as a tool for studying the neural underpinnings of the mechanisms involved in sound processing, speech encoding and language development, and constitutes one of the principal reasons for its recent popularity.

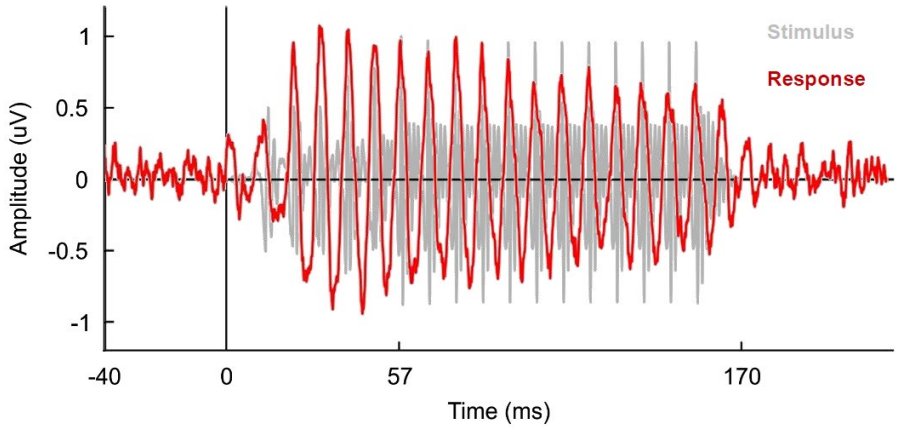

Figure 1. *An example of the frequency-following response (FFR).* Time-domain representation of the consonant-vowel /da/, with the stimulus waveform depicted in grey, overlapped with the neural signal represented in red (adapted from Ribas-Prats et al., 2019).

The FFR has long been considered a measure of neural sound encoding originated solely in the subcortical auditory system, but it is now widely accepted that the FFR reflects the synchronized neural activity of multiple generators throughout the entire auditory system, including subcortical and cortical levels (Coffey et al., 2016). The involvement and degree of contribution of the different parts of the ascending auditory pathway, including structures such as the auditory nerve, cochlear nucleus, inferior colliculus, thalamus, and auditory cortex, depend on the frequency of the incoming sound. Specifically, this frequency-dependent relation indicates that as information processing ascends through the auditory pathway, the range of frequencies reflected by each structure diminishes (Gorina-Careta et al., 2021). Additionally, the FFR can be obtained through passive and active listening paradigms and can be analyzed in both

temporal and spectral domains, providing an objective measure of the elemental acoustic features inherent to speech sounds, like time (latency of the start and the end of the stimulus), voice pitch (fundamental frequency), and timbre (harmonics and formants) (Kraus et al., 2017; Krizman & Kraus, 2019; Ribas-Prats et al., 2019; Skoe & Kraus, 2010). In particular, FFR analyses in the time domain offer information regarding the latency and amplitude of the neural signal, whilst analyses focused on the spectral domain will allow to study the magnitude of the frequency components of the encoded sound (Kraus et al., 2017). Consequently, the FFR has proven to have great potential for answering both basic and applied questions about mechanisms underlying sound encoding and language development.

Modulating factors of the FFR

The FFR appears to be well suited for the study of auditory plasticity, since it has been shown to be modulated by several factors, including context-dependent contingencies (Gorina-Careta et al., 2016), different auditory experiences along life (Musacchia et al., 2007; Wong et al., 2007), or even short-term auditory training (Carcagno & Plack, 2017). These modulations are thought to reflect the adaptive nature of the auditory system, which dynamically adjusts sound processing in constantly changing sensory environments.

In particular, FFR amplitude was found larger for sounds presented in a predictable sequence compared to those presented in an unpredictable succession, hence suggesting that the FFR is sensitive to context-dependent contingencies (Chandrasekaran et al., 2009). Moreover, life-long auditory experiences can also shape the FFR. Several studies have demonstrated that musicians show enhanced neural processing of sound in general, as evidenced by overall larger FFR amplitudes and shorter latencies compared to non-musicians, and in particular by functional improvements in pitch processing, superior detection of linguistic pitch manipulations and finer perception of prosody (Bidelman et al., 2011; Deguchi et al., 2012; Musacchia et al., 2007; Wong et al., 2007). Similarly, language experience has been shown to influence the FFR, with Mandarin individuals –raised in a tonal language environment– exhibiting more robust neural responses to speech sounds compared to American English –raised in a non-tonal environment– (Krishnan et al., 2005).

Nevertheless, long-life experiences are not exclusively responsible for modulating the FFR since short-term auditory training can also do so (Carcagno & Plack, 2017). Evidence for experience-dependent plasticity has been demonstrated by studies recording FFRs right before and after a training period in identification and discrimination tasks, with the “trained”

participants showing significant improvements as compared to the control group (Carcagno & Plack, 2011).

Overall, these studies demonstrated the plasticity of the auditory system and its dynamic nature as well as the sensitivity of the FFR to explore and evaluate it, and therefore suggested the potential use of the FFR as a powerful tool for investigate the current functional state of the auditory system and detect alterations in its functioning. This idea is supported by previous studies showing that disruptions in the FFR are related to deficits in phonological awareness (White-Schwoch, Woodruff Carr, et al., 2015), reading impairments (Lam et al., 2017), learning problems (King et al., 2002), dyslexia (Banai et al., 2009; Chandrasekaran et al., 2009; Hornickel & Kraus, 2013) or even autism spectrum disorder (Font-Alaminos et al., 2020; Otto-Meyer et al., 2018; Russo et al., 2008).

These findings led to the emergence of the idea that the FFR may provide a useful "biomarker" for identifying certain central auditory processing deficits by means of an objective electrophysiological technique (Kraus et al., 2017; Ribas-Prats et al., 2019; Richard et al., 2020), considering it is important not only to elaborate empirically based theories of auditory processing, but also to develop methods to identify children at risk. Moreover, and due to its sensitivity and potential as an early detection tool, the FFR could be a valuable instrument for identifying auditory processing deficits in vulnerable populations such as neonates, providing important information for research and clinical purposes. It should be emphasized that the FFR has shown to be adequate not only in evaluating the encoding of speech sounds at the present time, but also in predicting to a certain extent the future level of literacy one year ahead, even prior to infants being able to read (De Vos et al., 2020; White-Schwoch, Woodruff Carr, et al., 2015). This renders the FFR a potential tool for early diagnosis of learning disabilities that may appear during school-age years. Ultimately, early identification of speech encoding impairments could guide early interventions, leveraging the significant neural plasticity present during the first years of life to maximize their potential benefits.

The neonatal FFR

Despite the critical importance of the neonatal period for establishing the foundation for healthy development, little is known about the FFR characteristics during this first stage of life, not only for infants at risk of neurodevelopmental or language delay but also in healthy babies, and especially when compared with research conducted with older infants. Thus, while a considerable amount of research has investigated the FFR in school-age children, relatively few studies have focused on newborns and their unique auditory processing capabilities. This is

remarkable, given the enormous potential of this period for early interventions, and considering that the first study recording the neonatal FFR took place already in 1979 (Gardi et al., 1979). In this study, Gardi and colleagues recorded the FFRs elicited to 10 ms tone bursts in a group of 22 neonates aged 1-3 days old, proving the feasibility to effectively record the FFR during the first days of life and, additionally, its similarities in morphology with those FFRs obtained in adult participants.

However, it was not until the early 2010s when the research about neonatal FFR experienced a worthy expansion. At that time, Jeng and colleagues carried out a series of studies presenting several speech stimuli to a group of neonates, infants and adults, and recording their FFRs. Jeng's main findings not only corroborated the viability of recording the neonatal FFR, but also revealed that voice pitch encoding is innate and universal, as similar FFRs were recorded in American and Chinese newborns exposed to different Mandarin tones during their first three days of life (Jeng et al., 2011; Jeng, Lin, & Wang, 2016). However, an effect of linguistic experience was observed as well, with more robust neural responses obtained in Chinese adults compared to Chinese newborns (Jeng et al., 2011). This evidence represented a significant progress towards a better understanding of neural speech processing and served as a basis for further research trying to disentangle its developmental maturation process.

In spite of the relevance of studying the neonatal FFR, there are several constrains and particularities regarding FFR recording in newborns that probably restricted research in this field during these many years.

Challenges and practical considerations of the neonatal FFR

Although neonatal FFR recording procedures are not very different from those employed in any typical EEG recording in adults, these require additional characteristics and specifications that must be taken into account. It is crucial to address these practical challenges to ensure optimal conditions to record FFRs, hence ensuring reliable results and enhancing the clinical utility of this brain signal.

First, neonatal FFR requires the newborn to be in a state of complete rest or sleep during the entire recording process. This is because any muscle movements or activity can result in contamination of the recording, leading to inaccurate results. Therefore, it is essential to take measures to ensure that the newborn is comfortable and asleep during the FFR recording procedure, in line with standard protocols for FFR recordings that aim to minimize external artifacts interfering with the accurate capture of the brainstem's neural responses to auditory

stimuli. Additionally, and considering that neonatal FFR is recorded during the first days of life of the neonate, the recording session will probably be performed directly in the hospital room at the maternity ward where the baby is born, preferably at the newborn's bassinet, avoiding electrical interferences or skin-to-skin contact. These requirements may increase the recording time and present practical challenges in a hospital setting, considering the noise levels around and possible interruptions due to medical stuff and routine health controls.

Moreover, additional time will be necessary for cleaning and preparation. Considering the high electrical impedances of skin during the first days of life (Emery et al., 1991), as well as the residual substances from birth on the neonate's scalp, special attention must be paid on cleaning the placement area in which the scalp-electrodes will be located with a suitable abrasive gel. Likewise, and related to the residual substances, it would be necessary to perform the FFR recording after the UNHS has been passed, ensuring the integrity of the auditory pathway and the absence of amniotic fluid, mucus, or any other secretions from the pregnancy period that could be partially obstructing the ear canal and disrupting sound transmission.

Regarding technical considerations typical for low-voltage responses such as the FFR, the selected speech stimuli (e.g., phonemes, syllables) must be presented in a repetitive manner in order to obtain thousands of repetitions that will be averaged to obtain a clear and strong FFR. And, yet, in research with newborns time is of the essence (Jeng, Lin, Chou, et al., 2016). The need for repetitive presentation of the same speech stimulus (from 2000 to ideally 4000 repetitions, according to Gorina-Careta et al., 2022) that increases the total duration of the recording session, added to the need for the neonate to remain asleep during it, means that the amount of different stimuli that can be presented is limited. This poses an additional constraint to the aforementioned factors, as the properties of the stimuli to be used during the recording session and their duration must be selected with precision, based on the characteristics of speech encoding under study.

Thus, it is essential to address these practical challenges and to choose optimal stimuli for ensuring optimal conditions for FFR recordings, therefore obtaining reliable results and guaranteeing the successful implementation in hospital environments and the clinical utility of the FFR.

Typical speech stimulation in neonatal FFR studies

The study of neural encoding of speech in the auditory system requires the use of speech stimuli that can generate reliable and accurate neural responses in newborns. Hence, the choice

of the stimuli depends on the specific aims of the research study, and different options have been proposed and used across the literature. FFR research in neural speech encoding have widely employed Mandarin syllables following the four different lexical tones (Jeng, Lin, & Wang, 2016; Krishnan et al., 2004, 2005; Russo et al., 2008; Song et al., 2008; Xu, 1997), different single vowels with rising pitch (Jeng et al., 2010, 2011, 2013; Kujala et al., 2004) and the consonant-vowel syllable /da/ (Anderson et al., 2015; Cunningham et al., 2001; Kraus & Nicol, 2005; Musacchia et al., 2007, 2018; Novak et al., 1989; Pinto & Martinelli, 2018; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021; White-Schwoch, Davies, et al., 2015). Specifically, the syllable /da/ has been the most popular auditory stimulus for studying FFR in both infant and adult populations, due to its acoustic complexity and its presence in the phonetic repertoires of most languages (Lemos et al., 2021; Richard et al., 2020; Skoe & Kraus, 2010). According to Skoe and Kraus (2010), the interest in the /da/ syllable proceeds from its internal structure, characterized for an initial rapid stop burst (/d/) that creates a challenging encoding condition even for normal-hearing adults; followed by a short transition and a sustained portion (/a/) whose encoding characteristics along the auditory system have been well described.

However, although the acoustic complexity and universality of consonant-vowel /da/ have made it valuable for studying FFR in populations of infants and adults, these same properties may render it inadequate for extensively characterizing the neonatal FFR. The employment of /da/ or any of the previous stimuli allowed to explore easily in newborns the neural response to energy corresponding to the fundamental frequency (F_0) and some additional frequencies in lower range, i.e., the first harmonics. Nevertheless, they do not allow to investigate one of the critical aspects in language acquisition, which is how precisely the newborn's brain encodes the spectrotemporal fine structure (TFS) of the incoming auditory stimuli, which is particularly relevant in non-tonal languages like English or Spanish. This may be due to technical limitations of the stimulus, namely: (I) the short duration of the transition (47 ms) and the changes in the formant structure, leading to insufficient duration for compute spectral features with enough resolution (Anderson et al., 2015); (II) the high-frequency components present in the stimuli (e.g., /da/ F_2 = 1438–1214 Hz), whose decreased neural responses are challenging to identify (Picton, 2010); (III) the delay in the development of phase-locking to higher frequencies compared to lower ones (Anderson et al., 2015; Levi et al., 1995); and (IV) the type of analyses computed on the neural signal, which highlights representation of the F_0 while sacrificing the TFS representation (Aiken & Picton, 2008).

Consequently, it remains uncertain whether neonates are able of accurately detect and encode formant changes in complex auditory stimuli or whether the current stimulus

parameters are inadequate to properly assess this ability. Hence, the first aim of this PhD thesis was to design and develop a new speech stimulus that allows to simultaneously assess the neural encoding of both the fundamental frequency and the temporal fine structure of the speech sound, and thus explore the functional maturity of the mechanisms involved in encoding both voice pitch contour and formant structure in newborns.

Prenatal factors affecting neonatal speech encoding

In the attempt to use a FFR recording at birth with the ultimate goal of establishing its power as a biomarker and its capacity for early identification of speech encoding impairments that may benefit from early interventions, we began to consider the relevance of external factors that could be affect the FFR in newborns. Neonatal speech encoding is shaped by several factors that can affect the development of the auditory system, not only perinatally or long after birth, but also during the pregnancy period, including for example prematurity, prenatal clinical conditions, or prenatal environmental factors. Based on the exceptional plasticity of the FFR that have been previously discussed, a critical question arises regarding the potential influence of the fetal environment, particularly the acoustic environment, on the newborn's abilities to encode speech sounds. To what extent does the prenatal environment impact these abilities? Could we positively influence this in any way, so that we make the most of the pregnancy period?

Indeed, premature birth is known to be associated with a higher risk of auditory processing disorders and difficulties in language acquisition (Bosch, 2011). Additionally, in a recent study conducted by Madrid and colleagues (2021) with the syllable /da/ in preterm babies, it was reported a significant smaller amplitude of the FFR signal for the early preterm babies, and specifically, a significant enlargement of the spectral amplitude in F_0 and F_1 with increasing weeks of gestational age. This suggests that prematurity can negatively impact the development of the neural mechanisms involved in speech encoding.

In addition to prematurity, other prenatal conditions such as fetal growth restriction or macrosomia may also impact neonatal speech encoding. Fetal growth restriction (FGR) is an obstetric condition referring to the inability of the fetus to achieve its genetic growth potential (Lee et al., 2013), with a prevalence around 6-10% of all deliveries (Marsál, 2002). Apart from the high-risk of adverse neurobehavioral outcomes during life and the high mortality, a recent study by Ribas-Prats and colleagues (2021) also employing the consonant-vowel /da/, reported strong attenuations in the FFRs of neonates born after fetal growth restriction. Likewise, a further study of these authors, using the same procedure, focused on the high end of the birth weight continuum: the clinical population of newborns born large-for-gestational age (LGA).

Although the negative perinatal consequences of this condition are clear (e.g., obesity, cardiovascular diseases, type 2 diabetes, mortality), the risk of neurocognitive impairments is not yet well established. In their study, Ribas-Prats and colleagues (2023) reported significant smaller amplitudes at F_0, as well as a strong correlation between this F_0 amplitude and the neonatal body mass index. These findings reveal the relevance of prenatal factors on FFR and speech encoding and, eventually, support the potential clinical use of "disrupted" FFRs to identify neonates at risk and thus deriving those babies to exhaustive postnatal monitoring.

Prenatal clinical conditions have thus proven to have a considerable impact on speech processing at birth. Likewise, it has been suggested that prenatal environmental factors may also influence neonatal speech encoding. For instance, it has been known for a long time that fetal hearing experiences shape the preferences of newborns in such a way that neonates prefer their mother's voice (DeCasper & Fifer, 1980) and their native language (Moon et al., 1993), being even able to recognize stories only heard during pregnancy (DeCasper & Spence, 1986). Furthermore, not only behavioral studies have demonstrated the relevance of fetal auditory experiences, but also research employing a variety of neuroimaging techniques have brought to light their relevance with findings that support different hemispheric specialization (Vannasing et al., 2016), increase of volume bilaterally in the auditory cortices (Webb et al., 2015) and different brain activation in newborns when perceiving native or non-native languages (May et al., 2011). Moreover, findings from recent studies using ERP in neonates have indicated greater neural responses to lullabies (Partanen, Kujala, Tervaniemi, et al., 2013) and to changes in speech sounds (Partanen, Pakarinen, et al., 2013) heard during the prenatal period.

Overall, the expanding literature seems to indicate that sound exposure during critical periods of prenatal neural plasticity can significantly impact the future development of auditory processing and thus perception since the gestational period. After birth, sound encoding abilities will be enhanced by the wealth of auditory experiences, revealing that greater exposure to enriched linguistic contexts, as happens in bilingual contexts, environments with tonal languages or surrounded by music, yield more robust neural responses and may have a positive impact on the development of speech processing in infants (Bidelman et al., 2011; Jeng et al., 2011; Krishnan et al., 2005). But, can this positive influence of music extend to an earlier time and occur as early as in the prenatal period? Could this improvement in language development, reported as a consequence of postnatal exposure to music, be already enhanced and stimulated during pregnancy? Considering how pervasive is music exposure during infancy (Mendoza & Fausey, 2021), together with the relevance that prenatal music exposure has proven to have already on maternal, fetal and neonatal well-being (Brillo et al., 2021; Çatalgöl & Ceber Turfan,

2021; He et al., 2021; Poćwierz-Marciniak & Harciarek, 2021), it is tempting to consider that a musically enriched prenatal experience could also have enhancing effects on speech encoding abilities (Chorna et al., 2019; Gervain, 2018).

Therefore, it is suggested that exposure to musical stimulation while in utero may promote the development of the neural mechanisms involved in speech encoding, being associated with a better speech auditory processing at birth. However, and in spite of the relevance of this possible connection, previous research linking prenatal music exposure and language in neonates has predominantly relied on behavioral and physiological measures (e.g., registration of the number of movements, respiratory rate, heart rate accelerations or decelerations, feeding volume) or on the analysis of ERP components based on the brain's automatic detection of changes, instead of on neurophysiological responses that can precisely reveal the neural encoding of speech sounds at birth. Thus, the second aim of my thesis is to study the encoding of speech sounds at birth, by recording the FFR in two groups of newborns who were differently exposed to music during the pregnancy period.

In the same vein, could the effects of a multilingual environment during pregnancy have an impact on the neural mechanisms involved in language acquisition? Studies suggested that language heard in the prenatal period can impact speech perception in neonates, considering evidence revealing how newborns exposed to bilingual environments are able to distinguish between the two heard languages and to exhibit comparable preferences for both (Byers-Heinlein et al., 2010). Yet, similar to prenatal music exposure and its possible applications, it remains unexplored whether there are potential advantages in speech sound encoding among neonates exposed to bilingual environments, beyond their ability to just recognize and differentiate one language from another. Hence, and in the same line as above, the third aim of my thesis is to examine the neural encoding of speech at birth through the FFR in newborns who had been exposed to either monolingual or bilingual fetal environments during pregnancy.

OBJECTIVES

The present PhD thesis aimed at investigating the neural mechanisms underlying speech sound processing in neonates, with a particular focus on the role of voice pitch and vowel formant structure, as well as the potential effects of prenatal acoustic environment on the development of these mechanisms. To achieve this goal, three main studies have been designed:

Study I

In the first study we aim to characterize the functional maturity of the neural encoding mechanisms underlying voice pitch contour and formant structure processing in the neonate population, by using a non-invasive electrophysiological technique. To do so, we will create a new speech stimulus that allows the simultaneous recording of both components of speech, and establish a reference comparison with the adult population.

Study II

In the second study we are interested in the effects of a prenatal acoustic environment, such as the exposure to music content, on the encoding of speech sounds. Thus, we aim to investigate the encoding of the fundamental frequency of speech at birth, in a sample of newborns with different levels of music exposure during the last 3 months of the prenatal period. To accomplish this goal, we will record the FFRs elicited by two different speech stimuli: syllable /da/, the stimulus traditionally employed in most of the FFR studies; and /oa/, the two-vowel stimulus developed in Study I, and analyze both in a stimulus section with the same duration and fundamental frequency.

Study III

In the third study, we aim to examine the effects of prenatal language exposure on the neural encoding of speech stimuli at birth by recording the FFR in newborns who had been exposed to either monolingual or bilingual fetal environments during the pregnancy period. In order to explore the encoding accuracy of the fundamental frequency and the first formant depending on the language environment, we will employ the speech stimulus /oa/ designed and created in Study I to elicit the neonatal FFR.

STUDY I

Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses[1]

[1] For copyright issues (i.e., the intellectual property of the published paper belonging to the journal's Publisher and not to the authors), I here reproduce the first page of the published paper and the full unformatted manuscript as accepted (*pre-print*).

www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses

Sonia Arenillas-Alcón[1,2,3], Jordi Costa-Faidella[1,2,3]✉, Teresa Ribas-Prats[1,2,3], María Dolores Gómez-Roig[3,4] & Carles Escera[1,2,3]✉

Detailed neural encoding of voice pitch and formant structure plays a crucial role in speech perception, and is of key importance for an appropriate acquisition of the phonetic repertoire in infants since birth. However, the extent to what newborns are capable of extracting pitch and formant structure information from the temporal envelope and the temporal fine structure of speech sounds, respectively, remains unclear. Here, we recorded the frequency-following response (FFR) elicited by a novel two-vowel, rising-pitch-ending stimulus to simultaneously characterize voice pitch and formant structure encoding accuracy in a sample of neonates and adults. Data revealed that newborns tracked changes in voice pitch reliably and no differently than adults, but exhibited weaker signatures of formant structure encoding, particularly at higher formant frequency ranges. Thus, our results indicate a well-developed encoding of voice pitch at birth, while formant structure representation is maturing in a frequency-dependent manner. Furthermore, we demonstrate the feasibility to assess voice pitch and formant structure encoding within clinical evaluation times in a hospital setting, and suggest the possibility to use this novel stimulus as a tool for longitudinal developmental studies of the auditory system.

Spoken language is arguably the most prevalent form of human communication. Experimental evidence suggests a universal organic basis for language acquisition, based on the identical development of speech perception pathways observed across different populations, languages and cultures[1–3]. Speech perceptual skills have been well characterized along the lifespan, especially with regard to their maturation during the first year[4–6]. However, less is known about their functional state during the very first hours after birth, when humans newly encounter the rich and challenging complexity of the external acoustic environment. A highly efficient auditory system becomes hence a requisite for proper language acquisition, as the complex and dynamic acoustic signal of speech conveys only very slight spectral and temporal cues for speech sound discrimination[7].

Previous studies have shown that the auditory system of newborns and infants is able to handle several aspects related to pitch processing, such as tracking pitch contours[8–15], higher-order frequency direction relationships[16], processing a missing fundamental[17] or exhibiting relative pitch by discriminating transposed melodies[18,19]. Likewise, the newborn auditory system appears able to discriminate phonemes[20–22] even when only based upon vowel formant structure changes or duration[20,23,24]. And yet, the low level neural underpinning of these abilities remains to be established.

Using non-invasive electroencephalography recordings, auditory brainstem responses (ABR) evoked to acoustic transients, such as click stimuli, have successfully been used to assess the integrity of the auditory pathway[25–27]. However, periodic acoustic stimuli also elicit a particular brain response of subcortical and cortical origin, known as the frequency-following response (FFR)[28–30]. The FFR reflects with high fidelity the encoding of periodic temporal envelope modulations (FFR_{ENV}) and temporal fine structure harmonic constituents (FFR_{TFS}) of a stimulus (Aiken and Picton[31] following the terminology proposed by Krizman and Kraus[29]). In language studies, these

[1]Brainlab-Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, P. Vall d'Hebron 171, 08035 Barcelona, Catalonia, Spain. [2]Institute of Neurosciences, University of Barcelona, Barcelona, Catalonia, Spain. [3]Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat, Catalonia, Spain. [4]BCNatal-Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona, Barcelona, Catalonia, Spain. ✉email: jcostafaidella@ub.edu; cescera@ub.edu

Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses[2]

Sonia Arenillas-Alcón[1,2,3], Jordi Costa-Faidella[1,2,3,*], Teresa Ribas-Prats[1,2,3], María Dolores Gómez-Roig[3,4] & Carles Escera[1,2,3,*]

[1] Brainlab – Cognitive Neuroscience Research Group. Department of Clinical Psychology and Psychobiology, University of Barcelona (Catalonia, Spain)

[2] Institute of Neurosciences, University of Barcelona (Catalonia, Spain)

[3] Institut de Recerca Sant Joan de Déu, Esplugues de Llobregat (Catalonia, Spain)

[4] BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona (Catalonia, Spain)

*** Corresponding authors:**

Carles Escera, Jordi Costa-Faidella

Brainlab - Cognitive Neuroscience Research Group
Department of Clinical Psychology and Psychobiology
University of Barcelona
P. Vall d'Hebron 171, 08035 Barcelona
Catalonia-Spain

cescera@ub.edu, jcostafaidella@ub.edu

[2] Published in *Scientific Reports*

Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses

Abstract

Detailed neural encoding of voice pitch and formant structure plays a crucial role in speech perception, and is of key importance for an appropriate acquisition of the phonetic repertoire in infants since birth. However, the extent to what newborns are capable of extracting pitch and formant structure information from the temporal envelope and the temporal fine structure of speech sounds, respectively, remains unclear. Here, we recorded the frequency-following response (FFR) elicited by a novel two-vowel, rising-pitch-ending stimulus to simultaneously characterize voice pitch and formant structure encoding accuracy in a sample of neonates and adults. Data revealed that newborns tracked changes in voice pitch reliably and no differently than adults, but exhibited weaker signatures of formant structure encoding, particularly at higher formant frequency ranges. Thus, our results indicate a well-developed encoding of voice pitch at birth, while formant structure representation is maturing in a frequency-dependent manner. Furthermore, we demonstrate the feasibility to assess voice pitch and formant structure encoding within clinical evaluation times in a hospital setting, and suggest the possibility to use this novel stimulus as a tool for longitudinal developmental studies of the auditory system.

INTRODUCTION

Spoken language is arguably the most prevalent form of human communication. Experimental evidence suggests a universal organic basis for language acquisition, based on the identical development of speech perception pathways observed across different populations, languages and cultures (Hoff, 2009; Kuhl, 2004; Sket et al., 2019). Speech perceptual skills have been well characterized along the lifespan, especially with regard to their maturation during the first year (Kuhl, 2010; McCarthy et al., 2019; Zubiaurre-Elorza et al., 2018). However, less is known about their functional state during the very first hours after birth, when humans newly encounter the rich and challenging complexity of the external acoustic environment. A highly efficient auditory system becomes hence a requisite for proper language acquisition, as the complex and dynamic acoustic signal of speech conveys only very slight spectral and temporal cues for speech sound discrimination (Cabrera & Gervain, 2020).

Previous studies have shown that the auditory system of newborns and infants is able to handle several aspects related to pitch processing, such as tracking pitch contours (Jeng, Hu, et al., 2011; Jeng et al., 2013; Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Jeng et al., 2018; Pinto & Martinelli, 2018; Ribas-Prats et al., 2019; Richard et al., 2020), higher-order frequency direction relationships (Carral et al., 2005), processing a missing fundamental (He & Trainor, 2009) or exhibiting relative pitch by discriminating transposed melodies (Plantinga & Trainor, 2005; Tew et al., 2009). Likewise, the newborn auditory system appears able to discriminate phonemes (Cheour-Luhtanen et al., 1995; Kushnerenko et al., 2001; Thiede et al., 2019) even when only based upon vowel formant structure changes or duration (Cheour-Luhtanen et al., 1995; Leppänen et al., 1999; Partanen et al., 2013). And yet, the low level neural underpinning of these abilities remains to be established.

Using non-invasive electroencephalography recordings, auditory brainstem responses (ABR) evoked to acoustic transients, such as click stimuli, have successfully been used to assess the integrity of the auditory pathway (Hoth et al., 2009; Møller, 1999; Stuart et al., 1996). However, periodic acoustic stimuli also elicit a particular brain response of subcortical and cortical origin, known as the frequency-following response (FFR; Coffey et al., 2019; Krizman & Kraus, 2019; Skoe & Kraus, 2010). The FFR reflects with high fidelity the encoding of periodic temporal envelope modulations (FFR_{ENV}) and temporal fine structure harmonic constituents (FFR_{TFS}) of a stimulus (Aiken & Picton, 2008) following the terminology proposed by Krizman and Kraus (2019). In language studies, these two components of the FFR have been respectively regarded as indexes of two perceptual properties of speech sounds: voice pitch contour and formant structure (Aiken & Picton, 2008; Krizman & Kraus, 2019).

FFR recordings are increasingly considered a valuable tool to index the current functional state of the auditory system and to predict the future development of language (Schochat et al., 2017), since disruptions in the FFR elicited by speech sounds relate to deficits in phonological awareness, reading impairments and dyslexia (Chandrasekaran et al., 2009; King et al., 2002; Lam et al., 2017; Rosenthal, 2020). The potential of the FFR as a biomarker for auditory deficits and their relation to literacy skills has thus been proposed (Hornickel & Kraus, 2013; Kraus & White-Schwoch, 2016; Levi et al., 1995; Ribas-Prats et al., 2019; Richard et al., 2020; White-Schwoch & Kraus, 2013; White-Schwoch, Woodruff Carr, et al., 2015). However, most developmental studies on the FFR targeted babies of several months of age (e.g., Anderson et al., 2015; Jeng et al., 2010; Musacchia et al., 2018), toddlers, infants or years-old children (e.g., Cunningham et al., 2001; Hornickel et al., 2012; Hornickel & Kraus, 2013; Johnson et al., 2008; Musacchia et al., 2018; Russo et al., 2008; White-Schwoch, Davies, et al., 2015; White-Schwoch, Woodruff Carr, et al., 2015), with only a few published reports on newborns (Gardi et al., 1979; Jeng, Hu, et al., 2011; Jeng et al., 2013; Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Jeng et al., 2018; Ribas-Prats et al., 2019). Thus, knowledge about the expected speech perceptual skills in newborns, who are more vulnerable than older age groups to hearing damage (Richard et al., 2020; Sket et al., 2019), may aid the early detection of language impairments and guide appropriate interventions benefitting from the massive neural plasticity during the first years of life (Johnson et al., 2008; Krishnan et al., 2005; Moore & Guan, 2001; Moore & Linthicum, 2007; Song et al., 2008).

Moreover, while newborn studies have focused on the assessment and maturation of pitch processing through the analysis of the FFR_{ENV}, to date, none addressed formant structure encoding in a systematic manner (Jeng, Hu, et al., 2011; Ribas-Prats et al., 2019; Richard et al., 2020). To the best of our knowledge, only a recent study from our lab (Ribas-Prats et al., 2019), providing a normative newborn database of FFR_{ENV} properties, attempted, as a secondary aim, to reveal whether the neonate auditory system was able to discriminate sounds differing in their fine structure (/da/ vs. /ga/). While the results were negative, the apparent lack of formant structure encoding may be due to several reasons: a) the short duration of the consonant transition (47 ms) and its formant change, which limit the resolution of the computed spectral information (Anderson et al., 2015); b) the high frequency content of the stimuli (/da/ F_2 = 1438-1214 Hz, /ga/ F_2 = 1801-1214 Hz), which elicits diminished FFR amplitudes that are difficult to recognize (Picton, 2010); c) the fact that phase-locking to higher frequencies develops later than to lower ones (Anderson et al., 2015; Levi et al., 1995); and, ultimately, d) the nature of the analyzed signal (FFR_{ENV}), which emphasizes temporal envelope information representation at the expense of temporal fine structure (Aiken & Picton, 2008). Thus, it still remains unclear whether newborns

cannot yet precisely track formant changes in complex sounds or if stimulation parameters used so far were not suited to reveal this ability.

Furthermore, in newborn FFR research, time is of the essence. Recording time constraints determine what stimulus encoding abilities can be studied, and it is even a more challenging issue when newborn research is conducted in a hospital setting. In addition to the ease of waking up or disturbing sleep, the hospital environment requires frequent and continuous access to the baby and the mother for routine tests to discard serious health issues, interventions, in-depth health evaluations, and any other kind of neonatal care. Taking this into account, it would be unsuitable to carry out recording sessions lasting between 40 (Partanen et al., 2013) and 50 (Jeng, Lin, & Wang, 2016) min, typical of speech-sound discrimination studies. The most adequate session duration would be between 20 and 30 min (Anderson et al., 2015; White-Schwoch, Woodruff Carr, et al., 2015). Therefore, devising a new single stimulus that would allow a proper assessment of both the FFR_{ENV} and FFR_{TFS} simultaneously within recording times compatible with infant research is required. This would provide a snapshot of the functional state of speech sound processing mechanisms at birth and, ultimately, help better understand how the encoding of this complex auditory signal matures.

Thus, the aim of the present study was to characterize the functional maturity of voice pitch contour and formant structure encoding mechanisms in the newborn population with non-invasive electrophysiological recordings, using the adult population as a reference. To that end, we developed a novel speech stimulus that allows the simultaneous assessment of both components of the speech signal through analyzing the FFR_{ENV} and the FFR_{TFS}, and which is at the same time compatible with clinical evaluation time constrains in a hospital setting.

FFRs were recorded to a novel two-vowel (/oa/) speech stimulus with a rising pitch ending. In order to estimate voice pitch encoding from the temporal envelope of the recorded neural response, we computed the FFR_{ENV} and analyzed it (spectral measures at the fundamental frequency [F_0] peak; pitch tracking measures extracted from stimulus-to-response cross-correlations and signal autocorrelations) accounting for the different steady and rising pitch sections of the stimulus (both during the /a/ vowel). In order to estimate formant structure encoding from the temporal fine structure of the recorded response, we computed the FFR_{TFS} and analyzed it (spectral measures at the first formant [F_1] frequency peak) accounting for the different /o/ and /a/ vowel sections of the stimulus (both during the steady pitch section) (Anderson et al., 2015; Hornickel et al., 2012; Jeng, Hu, et al., 2011; Jeng et al., 2013; Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Jeng et al., 2018; Krizman & Kraus, 2019; Musacchia et al., 2007; Ribas-Prats et al., 2019).

Given that the human auditory system is able to encode changes in voice pitch with great precision starting from the first days of life (Jeng, Hu, et al., 2011), we would expect no significant differences between newborns and adults in spectral amplitudes at F_0 or in pitch tracking accuracy measures computed from the FFR_{ENV}. Because the capacity of neurons in the auditory system to phase-lock their activity to higher sound frequencies develops later than to lower ones and continues improving during the first year of life (Anderson et al., 2015; Levi et al., 1995), we would expect newborns to exhibit overall smaller FFR_{TFS} spectral amplitudes at F_1 peak frequencies than adults. However, we had no clear hypotheses as to whether the newborn FFR_{TFS} would reflect a discriminative encoding of vowel formant structure as it does in adults (Sanfins et al., 2020) and, if so, whether that discriminative encoding would depend on F_1 center frequency. The evidence suggesting that the newborn auditory system discriminates vowel changes (Cheour-Luhtanen et al., 1995; Kushnerenko et al., 2001; Leppänen et al., 1999; Partanen et al., 2013; Thiede et al., 2019), is based upon recordings of event-related potentials (ERPs) reflecting higher-order auditory system computations. But, the fundamentally different nature of the FFR as a phase-locked neural response reflecting the acoustic waveform with high precision, and the fact that this low-level encoding of acoustic features is immature at birth, at least for higher frequencies, precluded us at this point to have strong hypotheses on the newborn's auditory system's capabilities.

RESULTS

Temporal envelope-following responses (FFR_{ENV}) and temporal fine structure-following responses (FFR_{TFS}) elicited by a two-vowel syllable, rising-pitch ending, /oa/ stimulus (Fig. 1A) were collected from 34 newborns and 18 adult participants. In order to assess voice pitch contour and formant structure encoding in depth, neural responses were analyzed according to the sound features of the different stimulus sections. Below, we provide descriptive statistics and comparisons for a comprehensive number of parameters extracted from the FFR_{ENV} and FFR_{TFS} (see Methods for a detailed description). Statistically non-significant results can be found in Suppl. Table 1.

Neural transmission delay

Neural lag. Newborns showed a significantly longer neural lag (an estimation of FFR latency) compared to adults ($U_{(50)} = 59$, $p < 0.001$, Cohen's $d = 0.659$). Descriptive statistics can be found in Table 1.

Assessment of voice pitch encoding from FFR_{ENV}

In order to determine the strength of the representation of the F_0 and assess the accuracy in tracking F_0 changes, our /oa/ stimulus was devised to feature a steady pitch during its initial section (113 Hz; 0-160 ms) and a linearly increasing pitch during its final section (113-154 Hz; 160-250 ms) (Fig. 1A). To accentuate the FFR components corresponding to the encoding of the stimulus envelope (mainly the F_0) while suppressing those related to the fine structure, thus controlling for vowel changes that occur along the different sections of the stimulus, we computed the FFR_{ENV}. Grand-average FFR_{ENV} waveforms are shown in Fig. 1B for both groups separately (newborns and adults). All descriptive statistics for FFR_{ENV} derived parameters can be found in Table 1.

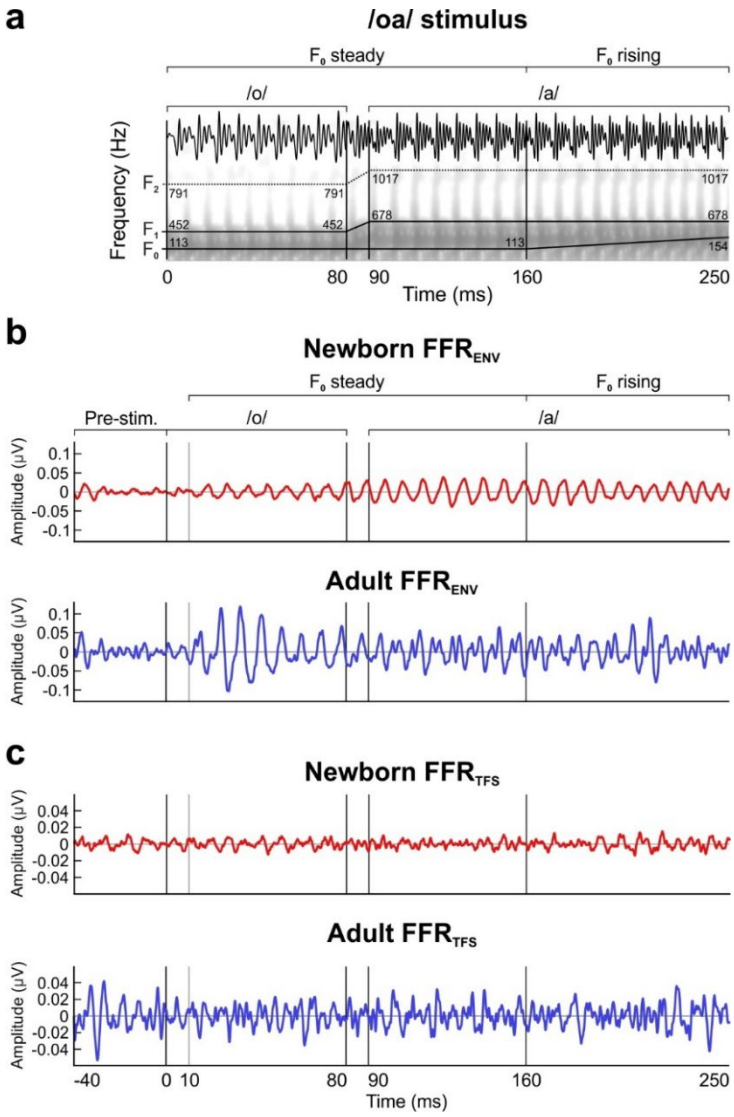

Fig. 1. Temporal representation of the stimulus (a); FFR_{ENV} (b) and FFR_{TFS} (c). **a**: Time waveform (top) and spectrogram of the /oa/ stimulus with schematic overlay of the formant structure trajectory (targeted F_0 and F_1 in solid lines; non-analyzed F_2 depicted in dotted line). **b**: Grand averaged time-domain waveform of the FFR_{ENV} from newborns (top red) and adults (bottom blue), obtained by averaging the neural responses to the two stimulus polarities. **c**: Grand averaged time-domain waveform of the FFR_{TFS} from newborns (top red) and adults (bottom blue), obtained by subtracting the neural responses to the two stimulus polarities.

Table 1. Descriptive statistics for FFR_{ENV} derived parameters: neural lag; F_0 spectral amplitude and SNR computed for the steady pitch section; stimulus-to-response cross-correlation, pitch error and pitch strength computed separately for each section of the stimulus (/a/ steady section; /a/ rising section).

Measure	Mean	SD	Median	Q_1	Q_3	IQR	Minimum	Maximum
Neural lag (ms; from 10-250ms)								
Newborns	9.33	1.74	9.26	8.70	10.01	1.31	3.53	12.60
Adults	6.26	1.21	5.85	5.55	6.47	0.92	5.10	9.75
F_0 Spectral Amplitude (μV; from 10-160ms)								
Newborns	0.01	0.01	0.01	<0.01	0.02	0.01	<0.01	0.03
Adults	0.02	0.01	0.02	0.01	0.02	0.01	0.01	0.04
F_0 SNR (from 10-160 ms)								
Newborns	4.36	4.72	5.86	0.82	7.83	7.01	-11.07	10.95
Adults	4.53	3.43	4.70	3.01	7.26	4.25	-3.52	9.40
/a/ steady section (90-160ms)								
Cross-correlation (Pearson's r)								
Newborns	0.18	0.06	0.18	0.12	0.23	0.11	0.06	0.30
Adults	0.20	0.05	0.20	0.17	0.23	0.07	0.07	0.27
Pitch error (Hz)								
Newborns	11.66	7.15	10.18	5.64	16.55	10.90	2.76	28.73
Adults	9.45	3.40	9.77	6.45	12.07	5.63	3.27	15.29
Pitch strength (r)								
Newborns	0.60	0.18	0.57	0.45	0.74	0.29	0.32	0.88
Adults	0.55	0.09	0.55	0.46	0.62	0.15	0.43	0.76
/a/ rising section (160-250ms)								
Cross-correlation (Pearson's r)								
Newborns	0.11	0.04	0.10	0.08	0.13	0.05	0.05	0.18
Adults	0.10	0.02	0.10	0.09	0.12	0.03	0.06	0.15
Pitch error (Hz)								
Newborns	11.72	7.08	10.44	5.50	16.25	10.75	2.60	28.15
Adults	9.50	3.34	9.87	6.55	12.00	5.45	3.13	15.03
Pitch strength (r)								
Newborns	0.60	0.18	0.56	0.45	0.74	0.29	0.32	0.88
Adults	0.55	0.09	0.55	0.46	0.61	0.15	0.44	0.75

SD = standard deviation. Q_1 = first quartile (25th percentile). Q_3 = third quartile (75th percentile). IQR = interquartile range.

The *spectral amplitude at F_0 peak* (113 Hz) during the steady pitch section of the stimulus (10-160 ms) was calculated as an indicator of the magnitude of neural phase-locking at that specific frequency (White-Schwoch, Davies, et al., 2015). Newborns exhibited significantly reduced spectral amplitudes at F_0 peak as compared to adults ($t_{(50)}$ = -3.079, p = 0.003, Cohen's d = -0.831). The corresponding amplitude spectra in the frequency domain computed along the steady pitch stimulus section is shown in Fig. 2A. Fig. 2B illustrates the distribution of F_0 spectral amplitude values obtained for each group.

The *signal-to-noise ratio* (SNR) at F_0 peak during the steady pitch section of the stimulus was taken as an estimation of the relative spectral magnitude of the response. No significant group differences were found. Fig. 2C illustrates the distribution of F_0 SNR values obtained per group.

The *stimulus-to-response cross-correlation* was taken as a measure of the accuracy with which the FFR_{ENV} reproduced the stimulus waveform, separately for the /a/ steady and /a/ rising pitch contour stimulus sections. Lower stimulus-to-response cross-correlation values were obtained during the rising pitch section (mean ± SD; /a/ rising = 0.11 ± 0.03) as compared to the steady pitch section (mean ± SD; /a/ steady = 0.18 ± 0.06) ($Z = -5.774$, $p < 0.001$, Cohen's d = 0.801). No significant group differences or group per stimulus section interaction were found.

We then computed the *pitch error* per pitch section separately, in order to determine pitch-tracking accuracy of the F_0 contour (Krizman & Kraus, 2019; Ribas-Prats et al., 2019). Neither significant group or stimulus section differences nor group per stimulus section interaction were found (see Fig. 3A for spectrogram and Fig. 3B for pitch track).

Pitch strength was taken as a measure of periodicity and the magnitude of neural phase-locking of the response (Jeng et al., 2013) and was also computed separately per stimulus pitch section. Neither significant group or stimulus section differences nor group per stimulus section interaction were found.

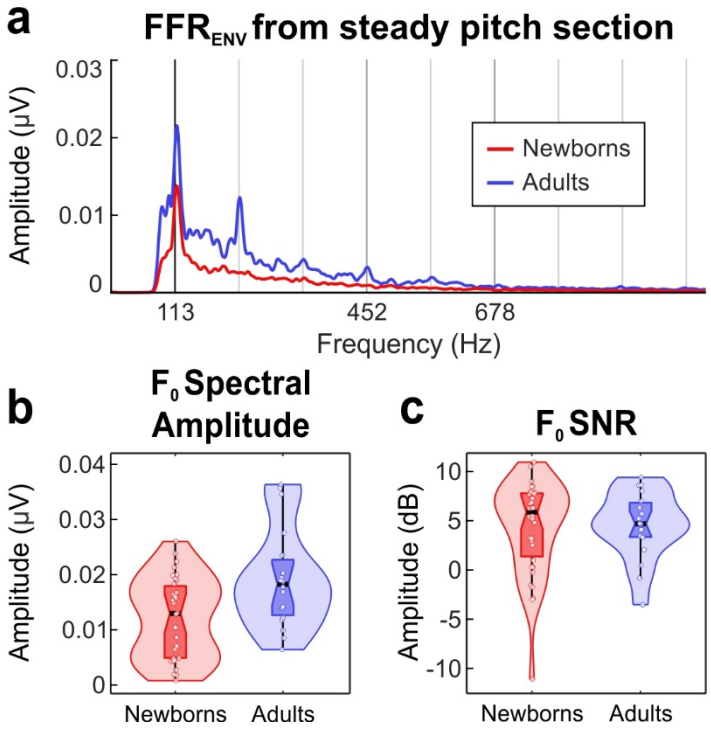

Fig. 2. Amplitude FFR_{ENV} spectra (a) and **data distributions** (violin plots) of F_0 spectral amplitude (b) and F_0 SNR (c) parameters extracted from the steady pitch section of the stimulus, by averaging the neural responses to the two stimulus polarities from both groups separately. Scatter plots show all tested participants in each group. In each plot, horizontal black line and vertical black line indicate the median and the interquartile range, respectively.

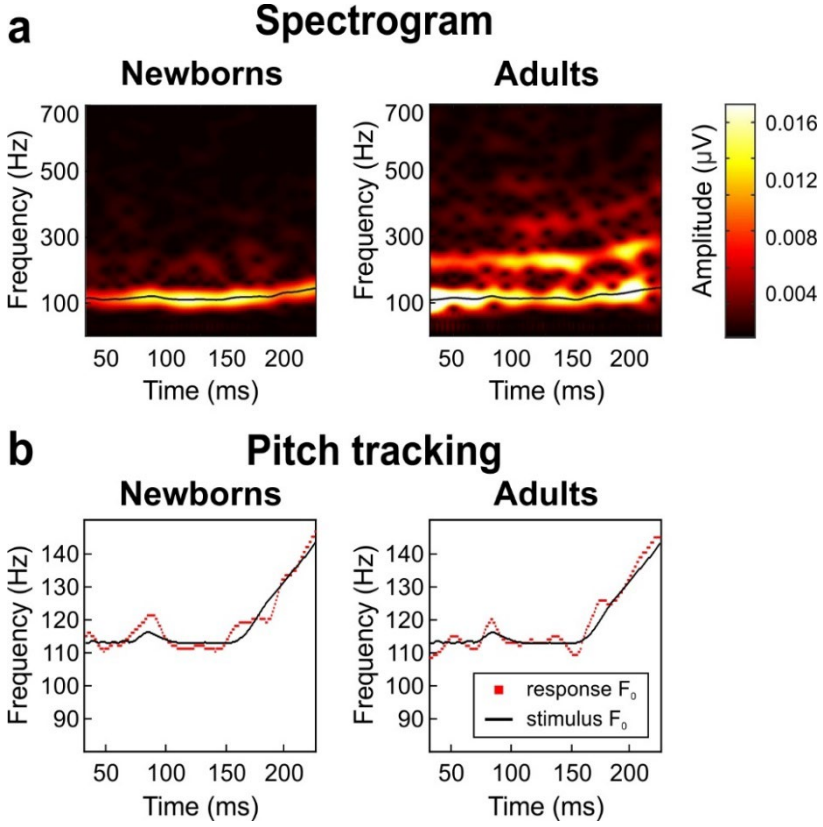

Fig. 3. Spectrogram (a) and **pitch tracking** (b) extracted from the newborns (left) and adults (right) FFR_{ENV}s grand averages elicited by /oa/ stimulus. **a**: The color scale from white to black represent the spectral amplitude in μV; dark colors represent smallest amplitude values, while light ones represent the highest. **b**: F_0s extracted from the stimulus is represented in solid black line, F_0s extracted from the FFR_{ENV}s elicited by the stimulus depicted in dotted red line.

Assessment of formant structure encoding from FFR_{TFS}

In order to determine the ability of the participants to encode the formant structure of speech sounds, the /oa/ stimulus featured two sections with steady pitch but differing in their formant structure: the /o/ section (10-80 ms; F_1 = 452 Hz) and the /a/ steady pitch section (90-160 ms; F_1 = 678 Hz). In order to emphasize temporal fine structure components of the response while diminishing the contribution of responses to the temporal envelope, we computed the FFR_{TFS} (Aiken & Picton, 2008; Krizman & Kraus, 2019). Grand-average FFR_{TFS} waveforms are shown in Fig. 1C for both groups separately. The frequency spectrum of the /o/ section and the /a/ steady pitch section are shown in Fig. 4A for both groups. All descriptive statistics can be found in Table 2.

Spectral amplitudes and SNRs from the FFR_{TFS} were retrieved separately from neural responses during the /o/ section (10-80 ms) and the /a/ steady pitch section (90-160 ms), selecting the spectral peaks corresponding to stimulus F_1 frequencies (452 Hz [/o/] and 678 Hz [/a/]), as indicators of the magnitude (absolute and relative) of phase-locking at the selected frequencies.

Spectral amplitude at /o/ vowel F_1. Spectral amplitudes at the /o/ vowel F_1 (452 Hz) are illustrated in Fig. 4B (*left*). A main effect of group revealed significantly smaller spectral amplitudes at 452 Hz in newborns as compared to adults ($F_{(1,50)} = 85.778$, $p < 0.001$, $\eta p2 = 0.632$). A main effect of stimulus section showed a significantly larger spectral amplitude value at 452 Hz during the /o/ vs. /a/ steady pitch sections ($F_{(1,50)} = 25.529$, $p < 0.001$, $\eta p2 = 0.338$). The group per stimulus section interaction was significant as well ($F_{(1,50)} = 18.603$, $p < 0.001$, $\eta p2 = 0.271$). *Post-hoc* tests computed to determine the direction of the interaction revealed higher spectral amplitudes in adults at 452 Hz during the /o/ vs. /a/ sections ($t_{(17)} = 3.803$, $p = 0.001$, Cohen's $d = 0.896$), but no significant differences were found in newborns.

Table 2. Descriptive statistics for FFR_{TFS} derived parameters: spectral amplitude and SNR for each formant peak frequency (452 Hz; 678 Hz) computed separately during the two vowel sections (/o/; /a/).

Measure	Mean	SD	Median	Q_1	Q_3	IQR	Minimum	Maximum
Spectral Amplitude /o/ section at 452 Hz (µV)								
Newborns	0.0019	0.0010	0.0017	0.0010	0.0024	0.0014	0.0004	0.0053
Adults	0.0100	0.0052	0.0087	0.0065	0.0134	0.0069	0.0030	0.0211
Spectral Amplitude /a/ steady section at 452 Hz (µV)								
Newborns	0.0016	0.0011	0.0014	0.0006	0.0022	0.0016	0.0002	0.0046
Adults	0.0067	0.0036	0.0068	0.0034	0.0091	0.0057	0.0011	0.0145
Spectral Amplitude /o/ section at 678 Hz (µV)								
Newborns	0.0006	0.0004	0.0005	0.0004	0.0008	0.0004	0.0000	0.0017
Adults	0.0014	0.0008	0.0015	0.0008	0.0018	0.0010	0.0003	0.0038
Spectral Amplitude /a/ steady section at 678 Hz (µV)								
Newborns	0.0008	0.0005	0.0008	0.0004	0.0011	0.0007	0.0001	0.0023
Adults	0.0042	0.0022	0.0038	0.0024	0.0063	0.0039	0.0001	0.0087
SNR /o/ section at 452 Hz								
Newborns	1.33	3.22	2.11	-0.18	4.03	4.21	-7.10	5.87
Adults	5.20	1.24	5.37	4.58	6.38	1.80	2.18	6.63
SNR /a/ steady section at 452 Hz								
Newborns	-0.39	4.11	0.74	-3.37	2.75	6.13	-11.25	5.25
Adults	4.02	2.31	4.81	3.36	5.54	2.17	-2.21	6.39
SNR /o/ section at 678 Hz								
Newborns	-2.08	4.80	-1.26	-4.59	1.40	5.99	-19.01	3.37
Adults	-0.66	4.62	0.66	-2.28	1.93	4.21	-11.35	5.86
SNR /a/ steady section at 678 Hz								
Newborns	-0.43	4.75	0.90	-3.84	2.84	6.68	-15.68	5.38
Adults	5.12	1.14	5.23	4.21	6.05	1.84	3.14	6.95

SD = standard deviation. Q_1 = first quartile (25th percentile). Q_3 = third quartile (75th percentile). IQR = interquartile range.

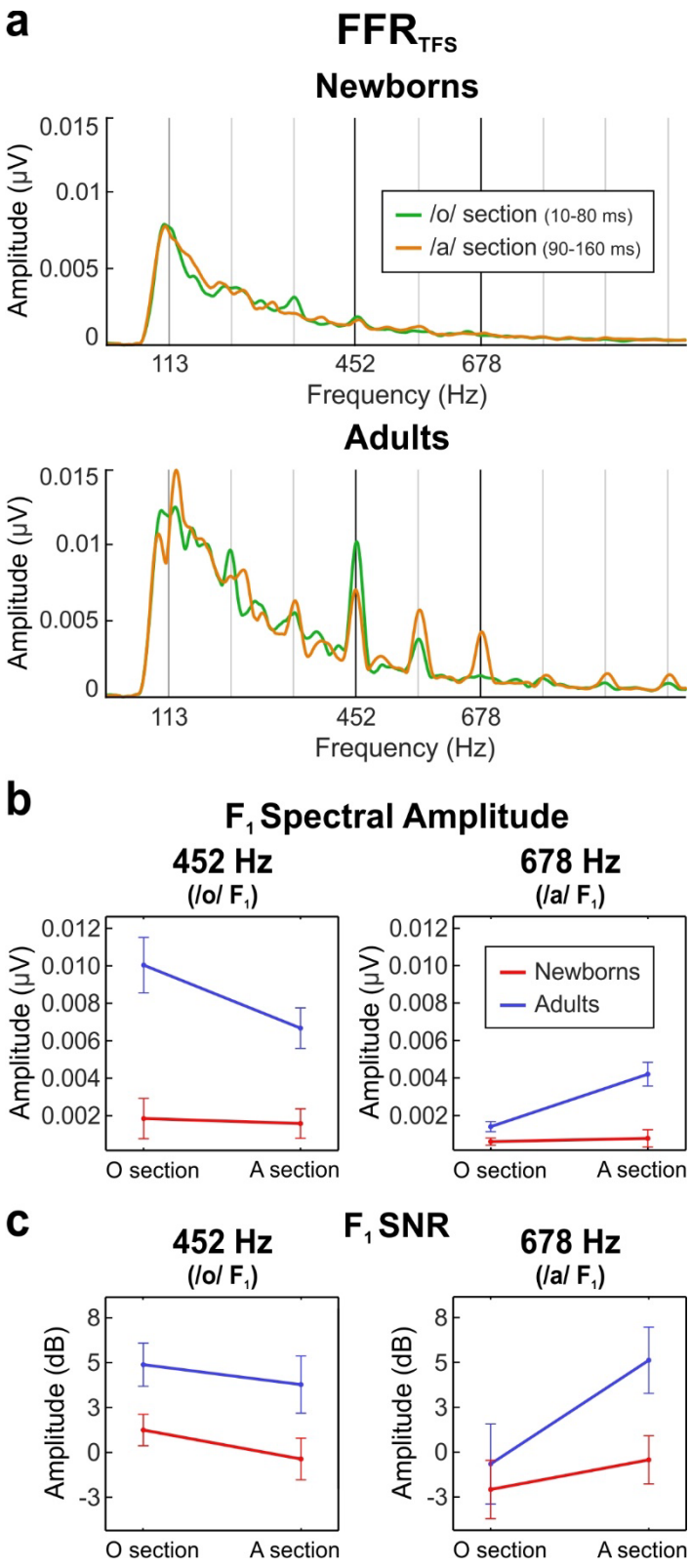

Fig. 4. Formant structure encoding in newborns and adults. **a**: Amplitude FFR_{TFS} spectra extracted from the /o/ vowel section (green) and the /a/ vowel section (orange) from the stimulus, plotted separately for newborns (top) and adults (bottom). **b**: Main effects graphic of F_1 spectral amplitude at 452 Hz (left) and 678 Hz (right) during the /o/ vowel section and the /a/ vowel section, plotted in red and blue lines for newborns and adults, respectively. **c**: Main effects of F_1 SNR at 452 Hz (left) and 678 Hz (right) during the /o/ vowel section and the /a/ vowel section, depicting neural responses from newborns (red) and adults (blue).

Spectral amplitude at /a/ vowel F_1. Spectral amplitudes at the /a/ vowel F_1 (678 Hz) are illustrated in Fig. 4B (*right*). A main effect of group revealed significantly smaller spectral amplitudes at 678 Hz in newborns as compared to adults ($F_{(1,50)} = 79.157$, $p < 0.001$, $\eta p2 = 0.613$). A main effect of stimulus section showed a significantly larger spectral amplitude value at 678 Hz during the /a/ steady pitch vs. /o/ sections ($F_{(1,50)} = 64.555$, $p < 0.001$, $\eta p2 = 0.564$). The group per stimulus section interaction was significant as well ($F_{(1,50)} = 50.252$, $p < 0.001$, $\eta p2 = 0.501$). *Post-hoc* tests computed to determine the direction of the interaction revealed higher spectral amplitudes in adults at 678 Hz during the /a/ steady pitch vs. /o/ sections ($t_{(17)} = -5.845$, $p < 0.001$, Cohen's $d = -1.378$), but no significant differences were found in newborns.

SNR at /o/ vowel F_1. SNR values at the /o/ vowel F_1 (452 Hz) are illustrated in Fig. 4C (*left*). A main effect of group revealed significantly smaller SNR values at 452 Hz in newborns as compared to adults ($F_{(1,50)} = 47.213$, $p < 0.001$, $\eta p2 = 0.486$). A main effect of stimulus section showed a significantly larger SNR value at 452 Hz during the /o/ vs. /a/ steady pitch sections ($F_{(1,50)} = 4.207$, $p = 0.046$, $\eta p2 = 0.078$). No significant group per stimulus section interaction was found.

SNR at /a/ vowel F_1. SNR values at the /a/ vowel F_1 (678 Hz) are illustrated in Fig. 4C (*right*). A main effect of group revealed significantly smaller SNR values at 678 Hz in newborns as compared to adults ($F_{(1,50)} = 17.136$, $p < 0.001$, $\eta p2 = 0.255$). A main effect of stimulus section showed a significantly larger SNR value at 678 Hz during the /a/ steady pitch vs. /o/ sections ($F_{(1,50)} = 15.414$, $p < 0.001$, $\eta p2 = 0.236$). The group per stimulus section interaction was significant as well ($F_{(1,50)} = 4.753$, $p = 0.034$, $\eta p2 = 0.087$). *Post-hoc* tests computed to determine the direction of the interaction revealed higher SNR values in adults at 678 Hz during the /a/ steady pitch vs. /o/ sections ($t_{(17)} = -5.656$, $p < 0.001$, Cohen's $d = -1.333$), but no significant differences were found in newborns.

It should be noted that some SNR values, especially those of newborns at 678 Hz peak, were very close to zero. In order to ascertain whether there was a measurable signal when expected (at 452 Hz during the /o/ section and at 678 Hz during the /a/ section), we submitted the SNR values, per group and per condition separately, to one-tailed, one sample t-tests against zero. Results demonstrated that newborns had a measurable signal for lower frequency formants, as shown by significant differences in SNR at 452 Hz during the /o/ section ($t_{(33)} = 2.407$, $p = 0.022$, Cohen's $d = 0.414$), but no clear response at 678 Hz during the /a/ section ($p = 0.602$). Intriguingly, the SNR value at 678 Hz during the /o/ section was negative and significantly different from zero (mean ± SD: -2.08 ± 4.80 dB; $t_{(33)} = -2.530$, $p = 0.016$, Cohen's $d = -0.433$). Adult participants exhibited a measurable signal in the two conditions: at the 452 Hz peak during the /o/ section ($t_{(17)}$

= 17.737, $p < 0.001$, Cohen's $d = 4.194$) and at the 678 Hz during the /a/ section ($t_{(17)} = 19.043$, $p < 0.001$, Cohen's $d = 4.491$).

DISCUSSION

We hereby provide an in-depth characterization of the neural encoding of speech sound features that newborns exhibit during their first hours of life, by comparing FFRs from healthy newborns and normal-hearing adult participants elicited by a novel, two-vowel /oa/ stimulus, with a rising pitch ending. Regarding the FFR parameters indexing voice pitch encoding, extracted from the FFR_{ENV}, our results support previous findings showing no significant differences in voice pitch encoding ability at birth as compared to adults, as can be appreciated from the SNR values at F_0 peak as well as in pitch tracking measures, such as stimulus-to-response cross-correlation, pitch error and pitch strength. Concerning the FFR parameters indexing formant structure encoding, extracted from the FFR_{TFS}, as expected, newborns exhibited overall diminished amplitudes than adults at both F_1 peaks of interest (452 and 678 Hz). On the other hand, obtained SNR values in newborns were higher at 452 Hz (/o/ F_1) during the /o/ section than during the /a/ section but not different at 678 Hz (/a/ F_1), revealing the functional state of formant structure encoding mechanisms, which appear to be partially developed but still to mature, especially at higher frequency ranges. Furthermore, our results prove the feasibility to record and assess simultaneously both voice pitch and formant structure encoding within a thirty-minute period, a time-span compatible with clinical settings that allows obtaining the FFR_{ENV} and the FFR_{TFS} in large samples of newborns.

Considerations on the mother's womb acting as an acoustic filter and speech perceptual skills at birth

Speech perception abilities are crucial for early phonetic discrimination (Kuhl, 2004, 2010; McCarthy et al., 2019; Werker & Curtin, 2005). Human hearing begins approximately at the 26th week of fetal life and most of the development takes place between the 26th and 28th week of gestation (Anbuhl et al., 2016; Granier-Deferre et al., 2011; May et al., 2011; Moore & Linthicum, 2007; Ruben, 1995), when hair cells and their connections to the cochlea are mature enough to tune in to specific frequencies. In this regard, previous research showed that fetuses can hear and remember language sounds and may learn about several sound properties while in the womb (May et al., 2011). Studies in newborns have shown a preference for their mother's voice (DeCasper & Fifer, 1980) and for their native language (Moon et al., 1993, 2013), as well as behavioral recognition of children's stories heard only during pregnancy (DeCasper & Spence, 1986). But, what speech sound features do babies rely upon to exhibit such identification skills? Considering that the mother's womb acts as a low-pass filter, the sounds available to a fetus during the gestation period are dominated by a low frequency content (<500 Hz; Gerhardt & Abrams,

2000; Jeng, 2017; McCarthy et al., 2019; Parga et al., 2018), while higher frequency ranges, which characterize most of the temporal fine structure of speech (Banai et al., 2009; Hornickel et al., 2012), would only be fully available at birth. Indeed, neonates may base their preferences on pitch contours and slow temporal dynamics, features available during pregnancy (Sambeth et al., 2008; Stefanics et al., 2009; D. Zhang et al., 2017). Furthermore, albeit previous studies have shown neural signatures of vowel change detection for vowel pairs differing only in second formant (F_2) frequencies in newborns (Cheour-Luhtanen et al., 1995) and 6 month-old babies (Cheour et al., 1998), recent electrophysiological (McCarthy et al., 2019) and behavioral (Curtin et al., 2009) evidence suggests that infant vowel discrimination relies more strongly on F_1 (usually below 800 Hz) than F_2 frequency differences. For instance, in a comprehensive study, McCarthy et al. (2019) analyzed neural responses to vowel changes using all pairs of a set of 7 English vowels, and showed that phonetic development from 4 to 11 months-old exhibits an increasing sensitivity to higher-frequency acoustic information (i.e., infants progressively rely less on F_1 changes and more on F_2 changes). Importantly, while youngest infants (4-5 months-old) neural responses appeared to reflect vowel acoustics (i.e., larger acoustic changes were reflected by larger neural response changes), those from older infants (10-11 months-old) seemed to represent putatively categorical changes (i.e., vowel space maps recreated from neural data showed large differences between vowel pairs with small acoustic differences). Intriguingly, a close inspection of their data (particularly at Fig. 4) strongly suggests that vowel pairs with lower F_1 frequency content (/ɔ/ vs. /u/; <500 Hz) are represented in youngest infants' vowel space farther apart from each other than vowel pairs with higher F_1 frequency content (/a/ vs. /ɛ/; >500 Hz), a pattern not apparent in older infants. However, the authors did not explicitly test this hypothesis. In fact, to the best of our knowledge, there is no behavioral or neurophysiological study in newborns or young infants explicitly testing vowel discrimination as a function of formant frequency. This may constitute an exciting avenue for future research linking auditory neural responses to auditory pathway and vowel discrimination development.

Regarding our data, in view of the above and taking into account that 1) the chosen first formants of our stimulus fall below (/o/ F_1) and above (/a/ F_1) the 500 Hz filter cut-off; 2) FFR spectral amplitudes increase with age (Sanfins et al., 2020); 3) FFR spectral amplitudes diminish along the frequency axis (Picton, 2010); and 4) FFRs are plastically modulated by experience (Jeng, Hu, et al., 2011; Jeng, 2017; Krishnan et al., 2005; Sanfins et al., 2020), it appears reasonable to expect certain degree of response in newborns at the lower frequency formant (452 Hz) and a fast decay of spectral power at the higher frequency formant (678 Hz). In any case, it seems plausible that certain speech sound processing skills were already mature at birth due to a greater exposure during pregnancy, while others would still be undeveloped.

Functional maturity state differences across speech perceptual skills at birth

A first indicator of auditory system's functional maturity is auditory transmission delay (Fria & Doyle, 1984; Fuess et al., 2002). Measuring wave V latencies and stimulus-to-response neural lags (which were consistent with activity generated in the brainstem; Fria & Doyle, 1984) we found, in agreement with previous literature, shortened delays in adult participants, which may be due to the increasing myelination and age-related changes in synaptic function (Anderson et al., 2015; Jeng et al., 2010; Jeng, Lin, Chou, et al., 2016; Moore & Guan, 2001).

However, even with a still maturing transmission speed, our results demonstrate that newborns accurately encode the F_0 of speech sounds as well as track changes in voice pitch during immediate postnatal hours, in line with previous studies (Jeng, Hu, et al., 2011; Jeng et al., 2013; Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Jeng et al., 2018; Ribas-Prats et al., 2019). Although spectral amplitudes at the F_0 peak were smaller in newborns as compared to adults, no significant differences were found with the adult sample when choosing relative amplitude measurements (i.e., SNR). Thus, the higher spectral amplitude values for adults could be due to the fact that, even during the pre-stimulus period, they also presented a higher spectral noise level (pre-stimulus root mean square: newborns = 0.03 ± 0.01 μV; adults = 0.05 ± 0.02 μV; $U_{(50)} = 571$, $p < 0.001$).

On the other hand, our results indicate a differential processing of formant structure in newborns in comparison to adults. Similar to the results on the FFR_{ENV}, neonates showed significantly smaller FFR_{TFS} absolute spectral amplitude values, but also smaller relative measures such as the SNR. However, our data demonstrate that newborns can encode the fine structure of speech sounds to a certain extent, with some limitation for higher frequency ranges, as evidenced by the fact that their SNR values were higher at 452 Hz (/o/ F_1) during the /o/ section than during the /a/ section, but at 678 Hz (/a/ F_1) they were not significantly different from zero. Although the SNR at 678 Hz during the /o/ section was negative in newborns, when analyzing the amplitude of the frequency spectrum (Fig. 4A) we observed that spectral amplitudes at 678 Hz during either of the two sections were very weak. Because of the reduced spectral amplitude and its large standard deviation, we considered this negative value as negligible, probably due to a noisy signal at higher frequencies rather than to active inhibition.

We considered the possibility that our results regarding formant structure encoding could be influenced by the internal structure of the stimulus, i.e., the /o/ section always preceded the /a/ section. As infants and neonates seem to preferentially use rhythmic cues to segment syllables and words from the acoustic stream (Fló et al., 2019; Otte et al., 2013), newborns may be more sensitive to sound onsets than codas. According to the temporal sampling framework hypotheses, put forward by Goswami (2011), rhythmic amplitude envelope modulations would entrain cortical oscillatory activity to exert a preferential processing of syllable onsets. However, there is

no obvious reason why such preferential onset processing should be apparent only at formant structure encoding and not at pitch encoding. Therefore, in order to shed some light on this possible confounding factor, we decided to statistically compare the SNR values at F_0 during the /o/ steady pitch section (10-80 ms) vs. the /a/ steady pitch section (90-160 ms), using a paired-samples t-test for each group of age. Our results showed that there were no significant differences in the SNR values at F_0 between stimulus steady pitch sections for either of the two groups (newborns: $t_{(33)} = -1.466$, $p = 0.152$, Cohen's $d = -0.251$; adults: $t_{(17)} = 0.797$, $p = 0.436$, Cohen's $d = 0.188$; for further statistical information, the reader is referred to Suppl. Table 1). Thus, no onset effect in pitch encoding was observable in any group. Moreover, given the rhythmic stimulation used in our study (SOA = 295 ms), half cycle of an entrained oscillation would last enough to cover, with the high excitability phase, both /o/ and /a/ steady pitch sections of our stimulus. Furthermore, the high frequency ranges we are dealing with in our FFR data (beyond 100 Hz) are more prone to elicit recordable subcortical activity than cortical (Bidelman, 2018; Chandrasekaran & Kraus, 2010; Gorina-Careta, Kurkela, Hämälainen, et al., 2021), and the modulation of phase-locking in subcortical neuronal ensembles by cortical oscillations has not been described, to the best of our knowledge, in the literature. Finally, in our study, the adult FFR_{TFS} SNR values at the formant peaks showed a double dissociation, being larger at the /o/ F_1 frequency during the /o/ section and at the /a/ F_1 frequency during the /a/ section, ruling out any onset effect. Therefore, given the pattern of results and the reviewed literature, an onset effect seems a negligible influencing factor in our results. In any case, further research studying the influence of vowel order should be carried out to help better clarify this possible confound (e.g., presenting an /ao/ syllable and comparing the pattern of results).

These results thus agree with the abovementioned notion that, due to the low-pass filter characteristics of the womb, fetuses are probably isolated from the mid and high frequency acoustic content of external sounds that characterizes most of the temporal fine structure of speech (Banai et al., 2009; Hornickel et al., 2012). Yet, while lacking the required prior experience for a mature perceptual system responding accurately to high frequencies, the ability to encode fine structure per se seems to be present at birth. Future testing with premature babies early exposed to natural sounds may shed more light on this issue.

Overall, our results are in line with the idea that humans, despite their limited experience to speech at birth, present mature functional mechanisms to detect changes in speech features at an unexpectedly early age (Eimas et al., 1971; Jeng, Lin, & Wang, 2016), and since alterations in the neural mechanisms underlying temporal envelope encoding are associated to several disabilities such as autism (Russo et al., 2008), dyslexia (Banai et al., 2009) or other learning problems (King et al., 2002), it is tempting to speculate that the encoding of temporal envelope information, such as its periodicity, may play a crucial role in the very first stages of language

acquisition (Jeng, Lin, & Wang, 2016). Temporal envelopes could provide a neural synchrony channel onto which separate neural representations of other speech features would anchor as parts of an ensemble that would, ultimately, give rise to a coherent unitary entity (Eggermont, 2001). Furthermore, there is increasing evidence that the FFR is a brain response that receives subcortical and cortical contributions in a frequency-specific manner, with frequencies below 150 Hz originating mainly from subcortical sources (Bidelman, 2018; Coffey et al., 2016, 2019; Gorina-Careta, Kurkela, Hämäläinen, et al., 2021; X. Zhang & Gong, 2019). Therefore, it is tempting to speculate that the effects observed here may reflect the increasing maturation of the subcortical auditory system from birth to adulthood.

The reported differences in formant structure encoding abilities found between newborns and adults open a window of opportunity to study the developmental progression of these skills. Considering that the gradual increase of phase-locking to high-frequencies is age-dependent (Anderson et al., 2015), understanding how inter-individual differences in development as revealed by FFR_{TFS} neural responses relate to the acquisition of formant encoding perceptual skills could be used to identify potential risks of future disabilities. Early impairment detection is thus critical to allow early interventions and to maximize the development of speech and listening competences, essential requirements for the acquisition of optimal literacy skills (Richard et al., 2020).

Considerations on speech stimuli commonly used for newborn FFR studies

In language FFR studies, the most commonly applied speech stimuli are mandarin syllables following the four different lexical tones (Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Krishnan et al., 2004, 2005; Russo et al., 2008; Song et al., 2008; Xu, 1997), and different single vowels with rising pitch (Jeng et al., 2010; Jeng, Hu, et al., 2011; Jeng et al., 2013; Kujala et al., 2004). The use of these stimuli focused the research field on assessing voice pitch encoding, putting the assessment of formant structure encoding aside. A notable exception is the widely used consonant-vowel syllable /da/ (Anderson et al., 2015; Cunningham et al., 2001; Kraus & Nicol, 2005; Musacchia et al., 2007, 2018; Novak et al., 1989; Pinto & Martinelli, 2018; Ribas-Prats et al., 2019; White-Schwoch, Davies, et al., 2015), which contains a fine structure change during the consonant-vowel transition. The relevance of using this stimulus relies on the fact that stop consonants are an important constraint in populations with literacy impairments (Kraus et al., 1996), and since stop bursts are rapid and low in amplitude in the /d/ consonant compared to vowels, even normal-hearing adults and children can find it difficult to discriminate it from other contrastive stop consonants (Skoe & Kraus, 2010). However, the short duration of the consonant transition and the high (and changing) frequency peak of the formants that compose it (e.g., the difference between /d/ and /g/ appears in the second formant: /da/ F_2 = 1438-1214 Hz, /ga/ F_2 =

1801-1214 Hz), render this type of stimuli suboptimal in the characterization of FFR responses, which exhibit a spectral power decay with increasing frequency (Picton, 2010), especially in populations with an immature encoding of the high frequency content of sounds, such as newborns (Anderson et al., 2015; Levi et al., 1995). Hence, while the phase locking to lower frequency sounds could in principle be safely assessed from the first hours of life (Anderson et al., 2015; Gardi et al., 1979) as we demonstrate here as well, the lack of prenatal experience to the high frequency content of sounds and the requirement of a later and greater maturation of the auditory system to encode them (Anderson et al., 2015; Gardi et al., 1979; Hornickel et al., 2012; Levi et al., 1995) pose some limitations in the design of stimuli suited to study formant structure encoding.

Therefore, we believe our newly designed /oa/ stimulus, with pitch variation and two vowel sections with different formant structure based on relatively lower frequency harmonic components and suitable durations for accurate spectral analyses, enables a proper assessment of speech sound temporal envelope (FFR_{ENV}) and temporal fine structure (FFR_{TFS}) encoding.

Conclusion

The present study provides the first evidence that neonates are able to encode not only the voice pitch of speech sounds and its changes with great accuracy, as has been demonstrated in previous research, but also the formant structure. Specifically, newborns show emerging formant structure encoding skills at lower frequency ranges but still immature encoding precision at higher frequency ranges. In addition, having already proved the feasibility of successfully recording temporal envelope and temporal fine structure in newborns, we here promote the use of this new stimulus as a powerful tool to perform a longitudinal assessment of speech encoding in babies from their very first hours of life throughout the first years of infant development.

METHODS

Participants

A sample of 34 healthy term newborns (17 females; mean gestational age = 40.19 ± 1.08 weeks; mean birth weight = 3379 ± 289 grams; aged 14-78 h after birth) was recruited from Sant Joan de Déu Hospital in Barcelona (Spain). Obstetric pathologies, high-risk gestations and risk factors related to hearing impairments (according to the criteria of the Joint Committee of Infant Hearing, 2019) were considered excluding factors. All newborns had Apgar scores higher than 8 at 1 and 5 minutes of life and had passed the standardized hearing screening test based on the automated auditory brainstem response system (ALGO 3i, Natus Medical Incorporated, San Carlos, CA). Six additional newborns were attempted to be recorded but finally not included in

the study because they woke up before concluding the recording session, and it was not possible to help them falling asleep again.

Additionally, 18 healthy young adult participants (14 females; mean age = 26.94 ± 3.78 years) with no self-reported history of neurological, psychiatric or hearing impairment, and with normal or corrected-to-normal visual acuity were included in the study for comparison. Taking into account previous research showing no differences between sexes for the encoding of frequencies until 720 Hz (Krizman et al., 2012, 2019), chances that data extracted from our selected range of analyzed frequencies (up to 678 Hz) were affected by sex condition were low. All participants underwent a screening pure tone audiometry to ensure a normal hearing level at 250, 500, 1000, 2000 and 4000 Hz. Excluding factors were mean hearing thresholds above 25 dB sound pressure level (SPL) or mean interaural hearing threshold differences larger than 20 dB SPL.

Both newborns and adults underwent a standard click-evoked auditory brainstem response test employing a standard *SmartEP* platform (Intelligent Hearing Systems, Miami, Fl, USA), with a 100 μs square-wave click stimulus delivered at 65 dB SPL for adults and 60 dB SPL for newborns. Following the precedent of Jeng, Chung et al. (2011), differences in stimulus intensities were chosen to compensate for the smaller ear canal volumes observed in young infants (Feigin et al., 1989; Keefe et al., 1993). All participants included in the sample had a reliably identifiable wave V. The mean latency of wave V was 8.70 (±0.42 SD) ms for newborns and 6.54 (±0.39 SD) ms for adults, and its mean amplitudes were 0.13 (±0.08 SD) μV for newborns and 0.29 (±0.12 SD) μV for adults (Suppl. Fig 1). All these values were comparable to those published previously (Ribas-Prats et al., 2019; Stuart et al., 1994).

The study was approved by the Ethical Committee of Clinical Research (CEIC) of the Sant Joan de Déu Foundation (Approval ID: PIC-53-17) and the Bioethics Committee of the University of Barcelona, and all adult participants and newborns' legal guardians gave informed consent in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). The data that support the findings of this study and the code used for data analysis are available upon reasonable request to the authors.

Stimulus

Inspired by the aforementioned previous stimuli limitations (e.g., short duration of consonant transitions and changing formants, high frequency content), a 250 ms two-vowel syllable stimulus with a rising pitch ending (/oa/) was created in *Praat* (Boersma & Weenink, 2020) (Fig. 1A). The /o/ vowel section (F_1 = 452 Hz; F_2 = 791 Hz) lasted from 0 to 80 ms, the /a/ vowel section (F_1 = 678Hz; F_2 = 1017 Hz) from 90 to 250 ms, and the /oa/ formant transition section from 80 to 90 ms. Stimulus pitch was kept steady at F_0 = 113 Hz from 0 to 160 ms and

increased linearly up to 154 Hz from 160 to 250 ms. We used 113 Hz F_0 instead of the common 100 Hz F_0 to avoid electric line noise harmonics by the European 50 Hz alternating current (Ribas-Prats et al., 2019). In order to maximize the detection of differences in vowel formant encoding in the FFR_{TFS}, formant peak frequencies coincided with harmonics of the fundamental.

Stimuli were delivered monaurally to the right ear with a stimulus-onset asynchrony (SOA) of 295 ms, in alternating polarities, at an intensity of 65 dB SPL for adults (Etymotic shielded earphones of 300Ω, ER, Elk Grove Village, IL, USA) and 60 dB SPL for newborns (same earphones connected to a Flexicoupler disposable adaptor, Natus Medical Incorporated, San Carlos, CA) using Intelligent Hearing Systems (Miami, Fl, USA). Differences in stimulus intensities were chosen for the same reason as in click stimulus.

Procedure

All newborns were recorded at the hospital room where they were resting with their mother. After the neonate passed the universal hearing screening test, the researcher started the recording session as soon as the newborn fell asleep, interrupting it to any sign of discomfort or sleep disruption and resuming it when the newborn was asleep again. The total mean duration of a test session was approximately 25 min (two click blocks x 2000 sweeps x 51.81 ms SOA, plus four /oa/ stimulus blocks x 1000 sweeps x 295 ms SOA, plus the duration of rejected sweeps), plus recording preparation time (around 5 min). Adult participants were tested in an acoustically shielded chamber in a laboratory facility located at the University of Barcelona, following the same procedure as in newborns with the exception of being awake with their eyes closed. Taking into account that the analyzed frequency content of neural responses recorded in the present study belongs to a higher frequency range than those characteristic of cortical sources (beyond 100 Hz; Chandrasekaran & Kraus, 2010), and that attentional modulations of the FFR seemingly affect only cortical sources (Bidelman, 2015; Coffey et al., 2016, 2019; Hartmann & Weisz, 2019), we can consider the contribution of alertness as a confounding factor in our results to be rather weak.

Data acquisition

FFRs were recorded from both newborns and adults with a *SmartEP* platform including the *cABR* and *Advanced Hearing Research* modules connected to a *Duet* amplifier (Intelligent Hearing Systems, Miami, Fl, USA), using three disposable snap Ag/AgCl electrodes placed in a vertical montage (ground electrode at the forehead; active at Fpz; online reference at the right mastoid, ipsilateral to the stimulated ear). All electrode impedances were kept <7 kΩ. The continuous signal was acquired at a sampling rate of 13333 Hz with an online bandpass filter from 30 to 1500 Hz and epoched from -40.95 (pre-stimulus period) to 249.975 ms relative to stimulus

onset. A total of 4000 artifact-free responses were obtained for each participant after automatic rejection of any sweep with voltage values exceeding ±30 μV.

FFR processing

Data was bandpass filtered offline from 80 to 1500 Hz. In order to assess voice pitch encoding, it was necessary to accentuate the FFR components corresponding to the encoding of the stimulus envelope, such as the fundamental frequency (F_0). Thus, neural responses were averaged by adding sweeps corresponding to the two stimulus polarities [(Rarefaction+Condensation)/2], yielding the envelope-following response (FFR_{ENV}). This procedure also aids in minimizing the contribution of putative cochlear microphonics. On the other hand, to properly evaluate formant structure representation, it was necessary to emphasize the FFR components highlighting the encoding of the stimulus temporal fine structure, such as vowel formants (F_1, F_2), and minimize the contribution of activity related to the envelope. To this aim, the responses to stimuli of alternating polarities were subtracted [(Rarefaction–Condensation)/2], yielding the temporal fine structure-following response (FFR_{TFS}; Aiken & Picton, 2008; Krizman & Kraus, 2019). In this study, only the FFR_{TFS} spectral peaks corresponding to F_1 frequencies were analyzed, since those from F_2 frequencies belonged to a very high frequency range that elicits weak neural responses difficult to record and, therefore, could not be reliably observed in all participants, especially in newborns. All data were analyzed using MATLAB R2019b (The Mathworks, 2019).

FFR parameters and statistical analysis

To give a comprehensive description of FFR properties both in newborns and adults, we computed several parameters, which we briefly detail below (see Ribas-Prats et al. (2019) for a full description of procedure, scripts and routines). All statistical analyses were performed on SPSS 25.0 (Corp., n.d.). Descriptive statistics are shown as mean, standard deviation (SD), median, first (Q_1) and third (Q_3) quartiles, interquartile range (IQR), and minimum and maximum values of the parameters for each group of age. The Kolmogorov-Smirnov test with the Lilliefors' significance correction was selected to check the normal distribution of the samples. Results were considered significant when $p < 0.05$. Contrast statistics, as well as p values and effect sizes obtained from statistically significant comparisons are reported in the Results sections. Statistically non-significant results and normality tests are reported in Suppl. Table 1.

Neural transmission delay

Neural lag was taken as an estimation of FFR latency due to the auditory system's neural transmission delay (Ribas-Prats et al., 2019), and was extracted from a cross-correlation of the entire stimulus with the neural response (10-250 ms), selecting the time lag that corresponds to

the maximum cross-correlation value. The obtained values were non-normally distributed, so a Mann-Whitney U test was used to assess for significant group differences (i.e., whether newborns showed a different transmission delay than adults).

Voice pitch encoding

To determine the abilities of newborns (by comparison with adults) to encode the voice pitch contour of the auditory stimulus presented, several parameters were extracted from the FFR_{ENV}:

Spectral amplitude at F_0 peak (113 Hz) was calculated as an indicator of the magnitude of neural phase-locking at that specific frequency (White-Schwoch, Davies, et al., 2015) only during the steady pitch section of the stimulus (10-160 ms), due to the continuous variation in pitch frequency throughout the rising section (160-250 ms). Since the obtained values were normally distributed, we employed a two-samples T-test to assess for significant group differences (i.e., whether newborns showed different spectral amplitudes of the signal at F_0 peak than adults).

Signal-to-noise ratio (SNR) at F_0 peak was taken as an estimation of the relative spectral magnitude of the response, taking into account not only the amplitude value of the signal at the frequency peak of interest (113 Hz) but also around that peak. Therefore, we divided the mean amplitude within a ±5 Hz frequency window centered at the peak of the frequency of interest (F_0) by the mean amplitude within two 28 Hz wide frequency windows (flanks) centered at ±19 Hz from the frequency of interest (e.g., for F_0 = 113 Hz, the mean amplitude from 108 to 118 Hz divided by the average of the mean amplitude from 80 to 108 Hz and the mean amplitude from 118 to 146 Hz). In order to ascertain group differences in the magnitude of the F_0 encoding and discern whether newborns had different responses to voice pitch than adults, we used Mann-Whitney U tests because the obtained values were non-normally distributed.

Stimulus-to-response cross-correlation. In order to assess the accuracy with which the FFR_{ENV} reproduces the stimulus waveform, we calculated the normalized cross-correlation between each individual's neural response and the stimulus, separately for the /a/ steady (90-160 ms) and /a/ rising pitch contour stimulus sections (160-250 ms) (Krizman & Kraus, 2019). The maximum value reached within a time lag of 3 to 10 ms (corresponding to the neural lag) was selected (Pearson's r; values from -1 to 1). The obtained values were non-normally distributed. Therefore, to test for putative between-subjects differences (i.e., whether newborns showed a different overall stimulus-response correlation than adults), a Mann-Whitney U test was used, with Age (newborns; adults) as grouping variable and Stimulus Section (/a/ steady; /a/ rising) as contrast variable. To test for putative within-subjects differences (i.e., whether stimulus-response correlations were different depending on stimulus pitch contour), a Wilcoxon test for two related

samples comparing the correlation values obtained for each stimulus section (/a/ steady; /a/ rising) was used. Finally, to test for a putative interaction between factors (i.e., whether newborns showed a different correlation value depending on stimulus pitch section than adults), a Mann-Whitney U test was used taking Age (newborns; adults) as grouping variable and the difference between the two conditions of the Stimulus Section (/a/ steady – /a/ rising) as contrast variable.

We also computed the normalized autocorrelation of the neural response, as well as that of the stimulus, in 40 ms sliding bins, to extract pitch error and pitch strength values.

Pitch error per stimulus section was used to determine pitch-tracking accuracy of the F_0 contour (Krizman & Kraus, 2019; Ribas-Prats et al., 2019) (corresponding to the autocorrelation peak lag per bin) by averaging the absolute Euclidian distance between the stimulus F_0 contour and the response F_0 per pitch section separately (steady [10-160 ms]; rising [160-250 ms]; starting from the onset of the section + the individual neural lag; values in Hz). Since obtained values were non-normally distributed, to determine between-subject effects, within-subjects effects and interaction, we followed the same procedure as with the stimulus-to-response cross-correlation explained above.

Pitch strength per stimulus section was taken as a measure of periodicity and the magnitude of neural phase-locking of the response (Jeng et al., 2013), , and calculated by averaging the obtained peak autocorrelation value of the response across bins, per pitch section separately (steady; rising; starting from the onset of the section + the individual neural lag; values from -1 to 1). Values were non-normally distributed, thus an identical method with the same factors as employed above in cross-correlation and pitch error parameters was used to determine between-subject effects, within-subjects effects and interaction.

Formant structure encoding

Regarding the encoding of the perceptual quality of formant structure, several parameters were retrieved from the FFR_{TFS}.

Spectral amplitudes at spectral peaks corresponding to stimulus F_1 frequencies (452 Hz [/o/] and 678 Hz [/a/]) were retrieved separately from neural responses to the /o/ section (10-80 ms) and the /a/ steady section (90-160 ms). All values were normally distributed, so an ANOVA test was conducted. Regarding the spectral amplitude at 452 Hz, a) the Group variable (newborns; adults) was chosen as between-subjects factor, to examine whether newborns showed different amplitude values at 452 Hz than adults; b) Stimulus Section (/o/ section; /a/ section) as within-subjects factor, in order to test whether spectral amplitudes at 452 Hz were different depending on stimulus vowel section; c) Interaction between factors was analyzed to ascertain whether newborns showed a different amplitude value at 452 Hz depending on stimulus vowel section

than adults. Pursuing an identical purpose, we conducted again the same test to examine differences at 678 Hz. The transition from /o/ vowel to /a/ vowel was not analyzed due to its short duration (10 ms).

Following the same procedure as with the spectral amplitude, *SNRs* at spectral peaks corresponding to stimulus F_1 frequencies (452 Hz [/o/] and 678 Hz [/a/]) were also retrieved separately from responses to the /o/ and the /a/ steady section, using an identical method to calculate it as described above for the FFR_{ENV}. All values were normally distributed, so ANOVA tests on 452 Hz and 678 Hz were conducted with the same factors and objectives as described above for F_1 spectral amplitudes analyses.

All analyses were additionally computed by excluding participants with extreme values (more than three interquartile ranges; N=9; 4 newborns + 5 adults). As the statistical results obtained did not alter the main findings of the study, we decided to keep all participants within the reported analyses to better represent the inherent variability of our samples (results excluding extreme values are reported in Suppl. Tables 2-7).

ACKNOWLEDGEMENTS

The authors would like to thank to all babies and their mothers who selflessly participated in this study. This work was supported by the Spanish Science and University Ministry PGC2018-094765-B-I00 project and the MDM-2017-0729-18-2 Maria de Maeztu Center of Excellence (MINECO-FEDER), the 2017SGR-974 Excellence Research Group of the Generalitat de Catalunya, and the ICREA Acadèmia Distinguished Professorship awarded to Carles Escera.

AUTHOR CONTRIBUTIONS

S.A.A.: Conceptualization, Methodology, Investigation, Formal analysis, Writing-Original draft, Writing-Review & Editing, Visualization. **J.C.F.:** Conceptualization, Methodology, Writing-Original draft, Writing-Review & Editing, Supervision. **T.R.P.:** Conceptualization, Methodology, Software. **M.D.G.R.:** Funding acquisition, Resources. **C.E.:** Conceptualization, Methodology, Funding acquisition, Writing-Review & Editing, Supervision.

ADDITIONAL INFORMATION

COMPETING INTERESTS: The authors declare no competing interests.

REFERENCES

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1–2), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Anbuhl, K. L., Uhler, K. M., Werner, L. A., & Tollin, D. J. (2016). Early Development of the Human Auditory System. In R. A. Polin, S. H. Abman, D. Rowitch, & W. E. Benitz (Eds.), *Fetal and Neonatal Physiology* (5th ed., pp. 1396–1410). Elsevier. https://doi.org/10.1016/B978-0-323-35214-7.00138-4

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2015). Development of subcortical speech representation in human infants. *The Journal of the Acoustical Society of America*, *137*(6), 3346–3355. https://doi.org/10.1121/1.4921032

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and Subcortical Auditory Function. *Cerebral Cortex*, *19*, 2699–2707. https://doi.org/10.1093/cercor/bhp024

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research*, *323*, 68–80. https://doi.org/10.1016/j.heares.2015.01.011

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage*, *175*(March), 56–69. https://doi.org/10.1016/j.neuroimage.2018.03.060

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer* (Version 6.1.09; p. Version 6.1.09). http://www.praat.org/

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances*, *6*(30), eaba7830. https://doi.org/10.1126/sciadv.aba7830

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory "primitive intelligence" already present at birth. *European Journal of Neuroscience*, *21*(11), 3201–3204. https://doi.org/10.1111/j.1460-9568.2005.04144.x

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, *64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009.10.006

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, *47*, 236–246.

Cheour-Luhtanen, M., Alho, K., Kujala, T., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Näätänen, R. (1995). Mismatch negativity indicates vowel discrimination in newborns. *Hearing Journal*, *82*, 53–58.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1*(5), 351–353. https://doi.org/10.1038/1561

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms11070

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-13003-w

Corp., I. (n.d.). *SPSS 25.0*.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology*, *112*(5), 758–767. https://doi.org/10.1016/S1388-2457(01)00465-5

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–731. https://doi.org/10.1111/j.1467-7687.2009.00814.x

DeCasper, A. J., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voice. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Eggermont, J. J. (2001). Between sound and perception: reviewing the search for a neural code. *Hearing Research*, *157*(1–2), 1–42. https://doi.org/10.1016/S0378-5955(01)00259-3

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech Perception in Infants. *Science*, *171*, 303–306.

Feigin, J. A., Kopun, J. G., Stelmachowicz, P. G., & Gorga, M. P. (1989). Probe-tube microphone measures of ear-canal sound pressure levels in infants and children. *Ear and Hearing*, *10*(4), 254–258. https://doi.org/10.1097/00003446-198908000-00008

Fló, A., Brusini, P., Macagno, F., Nespor, M., Mehler, J., & Ferry, A. L. (2019). Newborns are sensitive to multiple cues for word segmentation in continuous speech. *Developmental Science*, *22*, 1–16. https://doi.org/10.1111/desc.12802

Fria, T. J., & Doyle, W. J. (1984). Maturation of the auditory brain stem response (ABR): Additional perspectives. *Ear and Hearing*, *5*(6), 361–365. https://doi.org/10.1097/00003446-198411000-00008

Fuess, V. L., Bento, R. F., & da Silveira, J. A. (2002). Delay in maturation of the auditory pathway and its relationship to language acquisition disorders. *Ear, Nose, & Throat Journal*, *81*(10), 706–712.

Gardi, J., Salamy, A., & Mendelson, T. (1979). Scalp-Recorded Frequency-Following Responses in Neonates. *International Journal of Audiology*, *18*(6), 494–506.

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal Exposures to Sound and Vibroacoustic Stimulation. *Journal of Perinatology*, *20*, 20–29.

Gorina-Careta, N., Kurkela, J. L. O., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (MEG) study. *NeuroImage*, *231*, 117866. https://doi.org/10.1016/j.neuroimage.2021.117866

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*(1), 3–10. https://doi.org/10.1016/j.tics.2010.10.001

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Hartmann, T., & Weisz, N. (2019). Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. *NeuroImage*, *203*, 116185.

https://doi.org/10.1016/j.neuroimage.2019.116185

He, C., & Trainor, L. J. (2009). Finding the pitch of the missing fundamental in infants. *The Journal of Neuroscience*, *29*(24), 7718–7722. https://doi.org/10.1523/JNEUROSCI.0157-09.2009

Hoff, E. (2009). Language Development at an Early Age: Learning Mechanisms and Outcomes from Birth to Five Years. *Encyclopedia on Early Childhood Development*, 7–10. http://www.child-encyclopedia.com/sites/default/files/dossiers-complets/en/language-development-and-literacy.pdf

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., & Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *NeuroReport*, *23*(1), 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *The Journal of Neuroscience*, *33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

Hoth, S., Neumann, K., Weißschuh, H., Bräunert, J., Böttcher, P., Hornberger, C., Maul, H., Beedgen, B., Buschmann, K., Sohn, C., Hoffmann, G., & Plinkert, P. (2009). Universal newborn hearing screening. Methodical aspects. *HNO*, *57*(1), 29–36. https://doi.org/10.1007/s00106-008-1872-1

Jeng, F. C. (2017). Infant and Childhood Development: Intersections Between Development and Language Experience. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 17–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_2

Jeng, F. C., Chung, H. K., Lin, C. Der, Dickman, B., & Hu, J. (2011). Exponential modeling of human frequency-following responses to voice pitch. *International Journal of Audiology*, *50*(9), 582–593. https://doi.org/10.3109/14992027.2011.582164

Jeng, F. C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C. Der. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in american and chinese neonates and adults. *Ear and Hearing*, *32*(6), 699–707. https://doi.org/10.1097/AUD.0b013e31821cc0df

Jeng, F. C., Lin, C.-D., & Wang, T.-C. (2016). Subcortical neural representation to Mandarin pitch contours in American and Chinese newborns. *The Journal of the Acoustical Society of America*, *139*(6), 190–195. https://doi.org/10.1121/1.4953998

Jeng, F. C., Lin, C. Der, Chou, M. S., Hollister, G. R., Sabol, J. T., Mayhugh, G. N., Wang, T. C., & Wang, C. Y. (2016). Development of subcortical pitch representation in three-month-old Chinese infants. *Perceptual and Motor Skills*, *122*(1), 123–135. https://doi.org/10.1177/0031512516631054

Jeng, F. C., Nance, B., Montgomery-Reagan, K., & Lin, C.-D. (2018). Exponential Modeling of Frequency-Following Responses in American Neonates and Adults. *Journal of the American Academy of Audiology*, *29*, 125–134. https://doi.org/10.3766/jaaa.16135

Jeng, F. C., Peris, K. S., Hu, J., & Lin, C.-D. (2013). Evaluation of an Automated Procedure for Detecting Frequency-Following Responses in American and Chinese Neonates. *Perceptual and Motor Skills*, *116*(2), 456–465. https://doi.org/10.2466/24.10.PMS.116.2.456-465

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.-D., & Chung, H.-K. (2010). Early Maturation of Frequency-Following Responses to Voice Pitch in Infants with Normal

Hearing. *Perceptual and Motor Skills*, *111*(3), 765–784. https://doi.org/10.2466/10.22.24.PMS.111.6.765-784

Johnson, K. L., Nicol, T., Zecker, S. G., & Kraus, N. (2008). Developmental Plasticity in the Human Auditory Brainstem. *The Journal of Neuroscience*, *28*(15), 4000–4007. https://doi.org/10.1523/JNEUROSCI.0012-08.2008

Joint Committee on Infant Hearing. (2019). Year 2019 Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention Programs. *Journal of Early Hearing Detection and Intervention*, *4*(2), 1–44. https://doi.org/10.1542/peds.106.4.798

Keefe, D. H., Bulen, J. C., Arehart, K. H., & Burns, E. M. (1993). Ear-canal impedance and reflection coefficient in human infants and adults. *Journal of the Acoustical Society of America*, *94*(5), 2617–2638. https://doi.org/10.1121/1.407347

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, *319*, 111–115. https://doi.org/10.1016/S0304-3940(01)02556-3

Kraus, N., McGee, T. J., Carrell, T. D., Zecker, S. G., Nicol, T. G., & Koch, D. B. (1996). Auditory Neurophysiologic Responses and Discrimination Deficits in Children with Learning Problems. *Science*, *273*(5277), 971–973.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical "what" and "where" pathways in the auditory system. *Trends in Neurosciences*, *28*(4), 176–181. https://doi.org/10.1016/j.tins.2005.02.003

Kraus, N., & White-Schwoch, T. (2016). Newborn hearing screening 2.0. *Hearing Journal*, *69*(11), 44–46.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, *189*(1–2), 1–12. https://doi.org/10.1016/S0378-5955(03)00402-7

Krizman, J., Bonacina, S., & Kraus, N. (2019). Sex differences in subcortical auditory processing emerge across development. *Hearing Research*, *380*, 166–174. https://doi.org/10.1016/j.heares.2019.07.002

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research*, *382*, 166–174. https://doi.org/10.1016/j.heares.2019.107779

Krizman, J., Skoe, E., & Kraus, N. (2012). Sex differences in auditory subcortical function. *Clinical Neurophysiology*, *123*(3), 590–597. https://doi.org/10.1016/j.clinph.2011.07.037

Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.cortex.2009.08.003.Predictive

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport*, *15*(13), 2089–2092. https://doi.org/10.1097/00001756-200409150-00018

Kushnerenko, E., Cheour, M., Čeponiené, R., Fellman, V., Renlund, M., Soininen, K., Alku, P., Koskinen, M., Sainio, K., & Näätänen, R. (2001). Central auditory processing of durational changes in complex speech patterns by newborns: An event-related brain potential study. *Developmental Neuropsychology*, *19*(1), 83–97. https://doi.org/10.1207/S15326942DN1901_6

Lam, S. S.-Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia*, *103*, 162–167. https://doi.org/10.1016/j.physbeh.2017.03.040

Leppänen, P. H. T., Pihko, E., Eklund, K. M., & Lyytinen, H. (1999). Cortical responses of infants with and without a genetic risk for dyslexia: II. Group effects. *NeuroReport*, *10*, 969–973. https://doi.org/10.1097/00001756-199904060-00002

Levi, E. C., Folsom, R. C., & Dobie, R. A. (1995). Coherence analysis of envelope-following responses (EFRs) and frequency-following responses (FFRs) in infants and adults. *Hearing Research*, *89*, 21–27.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*(222), 1–9. https://doi.org/10.3389/fpsyg.2011.00222

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*, *9*, 1–7. https://doi.org/10.1038/s41598-019-55085-y

Møller, A. R. (1999). Neural mechanisms of BAEP. *Electroencephalography and Clinical Neurophysiology. Supplement*, *49*, 27–35.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paedriatrica*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098.Language

Moore, J. K., & Guan, Y.-L. (2001). Cytoarchitectural and axonal maturation in human auditory cortex. *JARO - Journal of the Association for Research in Otolaryngology*, *2*(4), 297–311. https://doi.org/10.1007/s101620010052

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, *46*(9), 460–478. https://doi.org/10.1080/14992020701383019

Musacchia, G., Ortiz-Mantilla, S., Roesler, C. P., Rajendran, S., Morgan-Byrne, J., & Benasich, A. A. (2018). Effects of noise and age on the infant brainstem response to speech. *Clinical Neurophysiology*, *129*(12), 2623–2634. https://doi.org/10.1016/j.clinph.2018.08.005

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan, H. G. (1989). Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephalography and Clinical Neurophysiology*, *73*(4), 295–305. https://doi.org/10.1016/0013-4694(89)90108-9

Otte, R. A., Winkler, I., Braeken, M. A. K. A., Stekelenburg, J. J., Van der Stelt, O., & Van den Bergh, B. R. H. (2013). Detecting violations of temporal regularities in waking and sleeping two-month-old infants. *Biological Psychology*, *92*, 315–322. https://doi.org/10.1016/j.biopsycho.2012.09.009

Parga, J. J., Daland, R., Kesavan, K., Macey, P. M., Zeltzer, L., & Harper, R. M. (2018). A description of externally recorded womb sounds in human subjects during gestation. *PLoS ONE*, *13*(5), 1–14. https://doi.org/10.1371/journal.pone.0197045

Partanen, E., Pakarinen, S., Kujala, T., & Huotilainen, M. (2013). Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clinical Neurophysiology*, *124*(8), 1578–1585. https://doi.org/10.1016/j.clinph.2013.02.014

Picton, T. W. (2010). *Human Auditory Evoked Potentials*. Plural Publishing Inc.

Pinto, E. S. M., & Martinelli, M. C. (2018). Brainstem auditory evoked potentials with speech stimulus in neonates. *Brazilian Journal of Otorhinolaryngology*, *86*(2), 191–200. https://doi.org/10.1016/j.bjorl.2018.11.006

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: infants use a relative pitch code. *Cognition*, *98*, 1–11. https://doi.org/10.1016/j.cognition.2004.09.008

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28–39. https://doi.org/10.1016/j.heares.2018.11.001

Richard, C., Neel, M., Jeanvoine, A., Connell, S., Gehred, A., & Maitre, N. L. (2020). Characteristics of the Frequency-Following Response to Speech in Neonates and Potential Applicability in Clinical Practice: A Systematic Review. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1618–1635.

Rosenthal, M. A. (2020). A systematic review of the voice-tagging hypothesis of speech-in-noise perception. *Neuropsychologia*, *136*(November 2019), 107256. https://doi.org/10.1016/j.neuropsychologia.2019.107256

Ruben, R. J. (1995). The ontogeny of human hearing. *International Journal of Pediatric Otorhinolaryngology*, *32*, 199–204. https://doi.org/10.1016/0165-5876(94)01159-U

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, *119*(8), 1720–1731. https://doi.org/10.1016/j.clinph.2008.01.108

Sambeth, A., Ruohio, K., Alku, P., Fellman, V., & Huotilainen, M. (2008). Sleeping newborns extract prosody from continuous speech. *Clinical Neurophysiology*, *119*(2), 332–341. https://doi.org/10.1016/j.clinph.2007.09.144

Sanfins, M. D., Garcia, M. V., Biaggio, E. P. V., & Skarzynski, P. H. (2020). The Frequency Following Response: Evaluations in Different Age Groups. In S. Hatzopoulos, A. Ciorba, & P. H. Skarzynski (Eds.), *The Human Auditory System - Basic Features and Updates on Audiological Diagnosis and Therapy* (pp. 335–203). IntechOpen.

Schochat, E., Rocha-Muniz, C. N., & Filippini, R. (2017). Understanding Auditory Processing Disorder Through the FFR. In N. Kraus, S. Anderson, T. White-Schwoch, R. Fay, & A. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 225–250). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_5

Sket, G. M., Overfeld, J., Styner, M., Gilmore, J. H., Entringer, S., Wadhwa, P. D., Rasmussen, J. M., & Buss, C. (2019). Neonatal White Matter Maturation Is Associated With Infant Language Development. *Frontiers in Human Neuroscience*, *13*(434). https://doi.org/10.3389/fnhum.2019.00434

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear and Hearing*, *31*(3), 302–324. papers2://publication/uuid/07E666C5-A423-43C0-85EC-BA2067A7F59F

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short- term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131.Plasticity

Stefanics, G., Háden, G. P., Sziller, I., Balázs, L., Beke, A., & Winkler, I. (2009). Newborn infants process pitch intervals. *Clinical Neurophysiology*, *120*(2), 304–308. https://doi.org/10.1016/j.clinph.2008.11.020

Stuart, A., Yang, E. Y., & Botea, M. (1996). Neonatal auditory brainstem responses recorded from four electrode montages. *Journal of Communication Disorders*, *29*(2), 125–139. https://doi.org/10.1016/0021-9924(95)00018-6

Stuart, A., Yang, E. Y., & Green, W. B. (1994). Neonatal Auditory Brainstem Response Thresholds To Air- and Bone-Conducted Clicks : 0 to 96 Hours Postpartum. *Journal of the American Academy of Audiology*, *5*, 163–172.

Tew, S., Fujioka, T., He, C., & Trainor, L. (2009). Neural representation of transposed melody in infants at 6 months of age. *Annals of the New York Academy of Sciences*, *1169*, 287–290. https://doi.org/10.1111/j.1749-6632.2009.04845.x

The Mathworks, I. (2019). *MATLAB R2019b* (p. Natick, Massachusetts).

Thiede, A., Virtala, P., Ala-Kurikka, I., Partanen, E., Huotilainen, M., Mikkola, K., Leppänen, P. H. T., & Kujala, T. (2019). An extensive pattern of atypical neural speech-sound discrimination in newborns at risk of dyslexia. *Clinical Neurophysiology*, *130*, 634–646. https://doi.org/10.1016/j.clinph.2019.01.019

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4

White-Schwoch, Tracis, & Kraus, N. (2013). Physiologic discrimination of stop consonants relates to phonological skills in pre-readers: A biomarker for subsequent reading ability? *Frontiers in Human Neuroscience*, *7*, 1–9. https://doi.org/10.3389/fnhum.2013.00899

White-Schwoch, Travis, Davies, E. C., Thompson, E. C., Woodruff Carr, K., Nicol, T., Bradlow, A. R., & Kraus, N. (2015). Auditory-neurophysiological responses to speech during early childhood: Effects of background noise. *Hearing Research*, *328*, 34–47.

White-Schwoch, Travis, Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, *13*(7). https://doi.org/10.1371/journal.pbio.1002196

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83. https://doi.org/10.1006/jpho.1996.0034

Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters*, *658*, 62–66. https://doi.org/10.1016/j.neulet.2017.08.047

Zhang, X., & Gong, Q. (2019). Frequency-following responses to complex tones at different frequencies reflect different source configurations. *Frontiers in Neuroscience*, *13*, 1–18. https://doi.org/10.3389/fnins.2019.00130

Zubiaurre-Elorza, L., Linke, A. C., Herzmann, C., Wild, C. J., Duffy, H., Lee, D. S. C., Han, V. K., & Cusack, R. (2018). Auditory structural connectivity in preterm and healthy term infants during the first postnatal year. *Developmental Psychobiology*, *60*(3), 256—264. https://doi.org/10.1002/dev.21610

STUDY II

Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: evidence from neonatal frequency-following responses[3]

[3] For copyright issues (i.e., the intellectual property of the published paper belonging to the journal's Publisher and not to the authors), I here reproduce the first page of the published paper and the full unformatted manuscript as accepted (*pre-print*).

Check for updates

Received: 30 April 2022 | Revised: 14 December 2022 | Accepted: 15 December 2022

DOI: 10.1111/desc.13362

RESEARCH ARTICLE

Developmental Science WILEY

Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: Evidence from neonatal frequency-following responses

Sonia Arenillas-Alcón[1,2,3] | Teresa Ribas-Prats[1,2,3] | Marta Puertollano[1,2,3] | Alejandro Mondéjar-Segovia[1,2] | María Dolores Gómez-Roig[3,4] | Jordi Costa-Faidella[1,2,3] | Carles Escera[1,2,3]

[1]Brainlab – Cognitive Neuroscience Research Group, Department of Clinical Psychology and Psychobiology, University of Barcelona, Catalonia, Spain

[2]Institute of Neurosciences, University of Barcelona, Catalonia, Spain

[3]Institut de Recerca Sant Joan de Déu, Catalonia, Spain

[4]BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona, Catalonia, Spain

Correspondence
Carles Escera and Jordi Costa-Faidella, Department of Clinical Psychology and Psychobiology, University of Barcelona. Passeig Vall d'Hebron 171, 08035 Barcelona, Catalonia, Spain.
Email: cescera@ub.edu and jcostafaidella@ub.edu

Funding information
Spanish Ministry of Science and Innovation, Grant/Award Numbers: PGC2018-094765-B-I00, MCIN/AEI/10.13039/501100011033/FEDER "UnamaneradehacerEuropa"; María de Maeztu Center of Excellence, Grant/Award Number: MDM-2017-0729-18-2(MCIN/AEI/10.13039/501100011033); ICREA Acadèmia Distinguished Professorship awarded to Carles Escera

Abstract

Fetal hearing experiences shape the linguistic and musical preferences of neonates. From the very first moment after birth, newborns prefer their native language, recognize their mother's voice, and show a greater responsiveness to lullabies presented during pregnancy. Yet, the neural underpinnings of this experience inducing plasticity have remained elusive. Here we recorded the frequency-following response (FFR), an auditory evoked potential elicited to periodic complex sounds, to show that prenatal music exposure is associated to enhanced neural encoding of speech stimuli periodicity, which relates to the perceptual experience of pitch. FFRs were recorded in a sample of 60 healthy neonates born at term and aged 12–72 hours. The sample was divided into two groups according to their prenatal musical exposure (29 daily musically exposed; 31 not-daily musically exposed). Prenatal exposure was assessed retrospectively by a questionnaire in which mothers reported how often they sang or listened to music through loudspeakers during the last trimester of pregnancy. The FFR was recorded to either a /da/ or an /oa/ speech-syllable stimulus. Analyses were centered on stimuli sections of identical duration (113 ms) and fundamental frequency (F_0 = 113 Hz). Neural encoding of stimuli periodicity was quantified as the FFR spectral amplitude at the stimulus F_0. Data revealed that newborns exposed daily to music exhibit larger spectral amplitudes at F_0 as compared to not-daily musically-exposed newborns, regardless of the eliciting stimulus. Our results suggest that prenatal music exposure facilitates the tuning to human speech fundamental frequency, which may support early language processing and acquisition.

KEYWORDS
infants, language, newborns, plasticity, prenatal, speech auditory brainstem response

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Developmental Science* published by John Wiley & Sons Ltd.

14677687, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/desc.13362 by Readcube (Labtiva Inc.), Wiley Online Library on [09/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: evidence from neonatal frequency-following responses[4]

Sonia Arenillas-Alcón[1,2,3], Teresa Ribas-Prats[1,2,3], Marta Puertollano[1,2,3], Alejandro Mondéjar-Segovia[1,2], María Dolores Gómez-Roig[3,4], Jordi Costa-Faidella[1,2,3,*] & Carles Escera[1,2,3,*]

[1] Brainlab – Cognitive Neuroscience Research Group. Department of Clinical Psychology and Psychobiology, University of Barcelona (Catalonia, Spain)

[2] Institute of Neurosciences, University of Barcelona (Catalonia, Spain)

[3] Institut de Recerca Sant Joan de Déu, Santa Rosa 39-57, 08950 Esplugues de Llobregat (Catalonia, Spain)

[4] BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona (Catalonia, Spain)

*** Corresponding authors:**

Carles Escera, PhD. Department of Clinical Psychology and Psychobiology, University of Barcelona. Passeig Vall d'Hebron 171, 08035 Barcelona (Catalonia, Spain). Tel: +34 933 125 048. Fax: +34 934 021 584. E-mail: cescera@ub.edu

Jordi Costa-Faidella, PhD. Department of Clinical Psychology and Psychobiology, University of Barcelona. Passeig Vall d'Hebron 171, 08035 Barcelona (Catalonia, Spain). Tel: +34 933 125 058. Fax: +34 934 021 584. E-mail: jcostafaidella@ub.edu

Conflict of Interest Disclosure: The authors have no conflict of interest to declare.

Data Availability Statement: The data that support the findings of this study and materials employed are publicly available on OSF (https://osf.io/6avxq/).

Funding/Support Statement: This work was supported by the Spanish Ministry of Science and Innovation PGC2018-094765-B-I00 project (MCIN/AEI/10.13039/501100011033/ FEDER "Una manera de hacer Europa"), the MDM-2017-0729-18-2 María de Maeztu Center of Excellence (MCIN/AEI/10.13039/501100011033), and the ICREA Acadèmia Distinguished Professorship awarded to Carles Escera.

Acknowledgements: The authors would like to thank to all mothers and newborns whose generous participation contributed to make this study possible.

[4] Published in *Developmental Science*

Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: evidence from neonatal frequency-following responses

Research Highlights

- Frequency-following responses to speech were collected from a sample of neonates prenatally exposed to music daily and compared to neonates not-daily exposed to music.
- Neonates who experienced daily prenatal music exposure exhibit enhanced frequency-following responses to the periodicity of speech sounds.
- Prenatal music exposure is associated to a fine-tuned encoding of human speech fundamental frequency, which may facilitate early language processing and acquisition.

Abstract

Fetal hearing experiences shape the linguistic and musical preferences of neonates. From the very first moment after birth, newborns prefer their native language, recognize their mother's voice and show a greater responsiveness to lullabies presented during pregnancy. Yet, the neural underpinnings of this experience inducing plasticity have remained elusive. Here we recorded the frequency-following response (FFR), an auditory evoked potential elicited to periodic complex sounds, to show that prenatal music exposure is associated to enhanced neural encoding of speech stimuli periodicity, which relates to the perceptual experience of pitch. FFRs were recorded in a sample of 60 healthy neonates born at term and aged 12-72 hours. The sample was divided in two groups according to their prenatal musical exposure (29 daily musically exposed; 31 not-daily musically-exposed). Prenatal exposure was assessed retrospectively by a questionnaire in which mothers reported how often they sung or listened to music through loudspeakers during the last trimester of pregnancy. The FFR was recorded to either a /da/ or an /oa/ speech-syllable stimulus. Analyses were centered on stimuli sections of identical duration (113 ms) and fundamental frequency (F_0 = 113 Hz). Neural encoding of stimuli periodicity was quantified as the FFR spectral amplitude at the stimulus F_0. Data revealed that newborns exposed daily to music exhibit larger spectral amplitudes at F_0 as compared to not-daily musically-exposed newborns, regardless of the eliciting stimulus. Our results suggest that prenatal music exposure facilitates the tuning to human speech fundamental frequency, which may support early language processing and acquisition.

Keywords (6): infants, language, newborns, plasticity, prenatal, speech auditory brainstem response

INTRODUCTION

Fetal hearing experiences shape the linguistic and musical preferences of newborns (Chorna et al., 2019; Gervain, 2018; May et al., 2011; Partanen et al., 2013). Behavioral studies have shown that newborns prefer their mother's voice (DeCasper & Fifer, 1980) and their native language (Moon et al., 1993), and even recognize stories only heard during pregnancy (DeCasper & Spence, 1986), proving that babies respond differently to native and non-native sounds just a few hours after birth (Moon et al., 2012). Likewise, recent studies using a range of neuroimaging techniques such as cranial ultrasonography, functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS) demonstrated the influence of hearing experiences on the neonate's brain through several findings, such as distinct hemisphere specialization (Vannasing et al., 2016), differential brain activation in newborns for native and non-native languages (May et al., 2011), and bilateral volume increase of auditory cortices (Webb et al., 2015). Furthermore, evidence from neurophysiological studies in newborns using event-related potentials (ERP) revealed greater neural activation to lullabies (Partanen, Kujala, Tervaniemi, et al., 2013) and to changes in speech sounds (Partanen et al., 2013) heard during pregnancy. Research thus suggests a paramount influence of exposure to sound during prenatal neural plasticity windows (Gilmore et al., 2018), molding auditory processing and perception since gestation.

This growing body of evidence points to a very early form of auditory learning that takes place in utero, shaping the infant's future neurodevelopment and processing of language. This is hardly surprising, considering that most of hearing development occurs between the 26th and the 28th weeks of pregnancy (Anbuhl et al., 2016; Granier-Deferre et al., 2011; May et al., 2011; Moore & Linthicum, 2007; Ruben, 1995), and that by the third trimester of gestation, the sense of hearing is already functional in some aspects similarly to that of adults (Ullal-Gupta et al., 2013).

Thus, previous research has revealed that the newborn's brain, albeit its limited language and auditory experience, is already able to encode and perceive different components of speech, such as pitch, in an adult-like manner (Arenillas-Alcón et al., 2021a; Cabrera & Gervain, 2020; Jeng et al., 2011). Pitch is defined as the perceptual attribute of the periodicity rate of a sound waveform that allows sounds to be ordered in a musical scale (Plack et al., 2014). Thus, pitch relates to the perception of sound periodicity, and hence mainly depends on the lowest frequency of a periodic waveform, i.e., the so-called fundamental frequency (F_0; Krizman & Kraus, 2019; Plack et al., 2014). The accurate encoding and tracking of F_0 plays an essential role in the future acquisition of language and sound processing, including the perception of melodies, harmony in music or prosody in speech, as well as language comprehension in noisy environments, perception

of the emotional content of a conversation, phoneme acquisition in tonal languages, recognition of speakers or speech segmentation, among others (Arenillas-Alcón et al., 2021a; Benavides-Varela et al., 2012; Cabrera & Gervain, 2020; Gervain, 2018; Musacchia et al., 2007; Partanen et al., 2013; Plack et al., 2014; Ribas-Prats et al., 2021). Because the mother's womb acts as a low-pass filter, only allowing the transmission of sound frequencies below 500 Hz (Gerhardt & Abrams, 2000; Jeng, 2017; McCarthy et al., 2019; Parga et al., 2018), sounds available to the fetus are mainly dominated by these frequency ranges. The consequences of this prenatal exposure to low-frequencies, which are typical from human speech (from 100-255 Hz; Traunmüller & Eriksson, 1995), support the idea of an increased sensitivity of the auditory system regarding low frequency ranges and may offer an explanation to the remarkable adult-like status observed already at birth in the encoding of F_0 (Arenillas-Alcón et al., 2021a).

After birth, F_0 processing abilities are enhanced by a wealth of auditory experiences, entailing auditory training periods and language or music exposure (Carcagno & Plack, 2011). For instance, a greater exposure to enriched linguistic contexts, as occurs in bilingual environments or with tonal languages, which employ pitch to convey word meaning, has been found to yield a more robust neural encoding of F_0 (Bidelman et al., 2011; Jeng et al., 2011; Krishnan et al., 2005). Likewise, musicians from different ages show enhanced neural encoding of F_0 (Musacchia et al., 2007; Wong et al., 2007), exhibit superior detection of linguistic pitch manipulations (Bidelman et al., 2011; Deguchi et al., 2012; Magne et al., 2006; Schön et al., 2004), as well as finer perception of prosody (Thompson et al., 2003). Importantly, musical exposure is pervasive in infancy (Mendoza & Fausey, 2021). Regardless of socioeconomic status, ethnicity and technical developments such as recorded music availability, most parents direct singing to their offspring on a daily basis (Yan et al., 2021), with arousing, pleasing and soothing effects (Cirelli & Trehub, 2020). Lullabies, characterized by an overall lower pitch, reduced accentuation and slower tempos than e.g., play songs (Tsang & Conrad, 2010), exert an especially relaxing effect on the infant, even if unknown and stemming from different cultures (Bainbridge et al., 2021). Interestingly, infant-directed speech (i.e., "parentese") is characterized, among other features, by an overall higher pitch and pitch variability, as well by purer and less harsh vocal timbres (Hilton et al., 2022), which emphasize the periodicity of the sound waveform over noise. These characteristics have led some authors to conclude that "the constellation of acoustic features that characterize infant-directedness in speech, across cultures, are rather musical" (Hilton et al., 2022). Hence, while lullabies use lower pitches to exert relaxing effects, infant-directed speech, which supports language acquisition (Ma et al., 2011; Trainor & Desjardins, 2002) and coordinates communicative interactions with infants (Mehr et al., 2021), resembles other song types in its use of a higher pitch and pitch variation

Thus, according to previous findings showing an enhanced F_0 encoding due to a higher language and music exposure, and considering the pervasive presence of music and infant-directed speech in early life periods and their characteristic pitch modulation patterns, it seems likely to consider that a musically enriched prenatal experience could also have enhancing effects on F_0 encoding skills (Chorna et al., 2019; Gervain, 2018).

Studies investigating F_0 encoding have gained interest in using the frequency-following response (FFR) as a precise neural activity correlate of early processing stages in the auditory pathway. The FFR is an auditory evoked potential originating from combined cortical and subcortical sources that mimics with high fidelity the acoustic features of the eliciting auditory stimulus (Coffey et al., 2019; Gorina-Careta et al., 2019, 2021; Skoe & Kraus, 2010), providing a non-invasive lens into sound processing in the brain. The growing attention it has obtained stems from its potential to predict the future development of language (Schochat et al., 2017), considering that abnormal FFR patterns in children have been related to reading impairments, learning problems, deficits in phonological awareness, dyslexia and even to clinical conditions such as autism (Banai et al., 2009; Basu et al., 2010; Chandrasekaran et al., 2009; Font-Alaminos et al., 2020; Hornickel et al., 2012; King et al., 2002; Lam et al., 2017; Otto-Meyer et al., 2018; Rosenthal, 2020). Indeed, the FFR is sensitive not only to several clinical conditions, but also to many different auditory contexts, such as training or musical experience (Carcagno & Plack, 2017; Gorina-Careta et al., 2019; Kraus & Chandrasekaran, 2010; Russo et al., 2005; Song et al., 2008). As a result of the abovementioned evidence, together with the feasibility to record the FFR in newborns (Arenillas-Alcón et al., 2021a; Gardi et al., 1979; Gorina-Careta et al., 2022; Jeng et al., 2011, 2016; Ribas-Prats et al., 2019; Ribas-Prats et al., 2021; Richard et al., 2020), the idea that the FFR could become a potential biomarker for identifying auditory and speech processing impairments has recently emerged (Arenillas-Alcón et al., 2021a, 2021b; Coffey et al., 2016; Font-Alaminos et al., 2020; Ribas-Prats et al., 2019; Richard et al., 2020; Schochat et al., 2017).

Considering the enhancement of pitch processing by musical training and exposure and its potential to foster language development (Bidelman et al., 2011; Musacchia et al., 2007; Wong et al., 2007), together with the importance of prenatal musical exposure on fetal and neonatal well-being (Brillo et al., 2021; Çatalgöl & Ceber Turfan, 2021; He et al., 2021; Poćwierz-Marciniak & Harciarek, 2021), it appears reasonable to expect that prenatal music exposure could be related to language and speech encoding abilities at birth. However, previous findings relating prenatal music exposure and language in newborns have been mostly based on behavioral measures such as the register of the number of movements, heart rate accelerations or decelerations, respiratory rate or feeding volume; or through the analysis of ERP components related to the brain's automatic detection of changes, rather than neurophysiological responses that accurately reflect the neural encoding of speech sounds.

The present study was hence set to investigate the encoding of the fundamental frequency of speech stimuli at birth through recording the FFR in newborns with different degrees of exposure to music during the prenatal period. We hypothesized better F_0 encoding in the group of neonates with daily exposure to music during pregnancy, i.e., a significant increase in the magnitude of the neural signal at the stimulus F_0. Should this hypothesis be confirmed, our findings would support early neural plasticity in audition and, critically, would point out to the relevance of prenatal music exposure to facilitate the tuning of the fetus' auditory system to human speech F_0, which is crucial for a successful future language acquisition.

METHODS

Participants

A sample of 29 newborns daily-exposed to music during pregnancy (DE; 11 females; mean gestational age = 39.85 ± 0.79 weeks; mean birth weight = 3329 ± 256 g) and 31 not-daily exposed (NDE; 16 females; mean gestational age = 39.92 ± 1.06 weeks; mean birth weight = 3367 ± 305 g) was recruited from *SJD Barcelona Children's Hospital* in Barcelona (Spain), based on a successful completion of a musical exposure questionnaire filled out by the babies' mothers (see below). No significant differences are reported across groups in gestational age ($t_{(58)}$ = -0.307, p = 0.760) or birth weight ($t_{(58)}$ = -0.525, p = 0.602). All newborns passed positively the universal hearing screening as part of the hospital routine, based on the detection of the auditory brainstem responses (ALGO 3i, Natus Medical Incorporated, San Carlos, SA), and obtained Apgar scores higher than 8 at 1 and 5 min of life. High-risk gestations as well as newborns with obstetric pathologies or other risk factors related to hearing impairment according to the Joint Committee of Infant Hearing (2019), were excluded from the recruitment.

Additionally, to double-check the integrity of the auditory pathway as well as the neural transmission time –as performed in previous studies from our laboratory–, both groups of newborns received a standard click-evoked auditory brainstem response (ABR) test. The click stimulus had a duration of 100 μs and was presented at a rate of 19.30 Hz, at an intensity of 60 dB sound pressure level (SPL) until a total of 4000 artifact-free sweeps, divided in two runs of 2000, were collected. Identification of a reliable wave V peak was a requirement for all newborns to proceed to the experiment. The study was approved by the Ethical Committee of Clinical Research (CEIC) of the Sant Joan de Déu Foundation (Approval ID: PIC-53-17), and required the mothers to fill out a sociodemographic questionnaire and to sign an informed consent prior to the participation, in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Because FFRs are modulated by socioeconomic status (Krizman et al., 2021; Skoe et al., 2013), we checked for group differences in the mothers' educational level and employment status, considering these variables as proxies available in the collected sociodemographic questionnaire. Summary statistics and analysis, as well as a graphical representation of the distribution, can be found in Supp. Fig. 1 and Supp. Table 1, respectively. In short, we did not find any differences across groups (Educational level of the mother (DE vs. NDE): $\chi^2_{(2)} = 0.229$, $p = 0.892$; Employment status of the mother (DE vs. NDE): $\chi^2_{(2)} = 2.092$, $p = 0.351$), suggesting that the socioeconomic status is not a confounding factor in our data analyses.

Musical exposure

The musical exposure that newborns underwent during pregnancy was assessed by a short retrospective questionnaire delivered to the babies' mothers. Based on previous research in newborns that conceptualizes musical exposure in terms of hours of exposure per week (Coffey et al., 2017; Musacchia et al., 2007; Strait et al., 2012; Zuk et al., 2013), mothers were asked the frequency with which they used to sing or listen to music through loudspeakers during the last three months before delivery (an English version of the musical questionnaire employed can be found in the Appendix). They were instructed to spurn situations in which music exposure was not intentional, such as ambient music in shops, elevators or restaurants. Instead, they were told to consider as an "exposed day" those days with periods in which their exposure to music through loudspeakers and/or singing was a minimum of around 30 minutes, regardless whether they were solely listening to it and/or singing, or while carrying out other activities (e.g., cleaning, cooking, exercising). Newborns were then classified into two groups attending to their musical exposure. The DE group included 29 neonates whose mothers were listening to music through loudspeakers and singing on a daily basis. The NDE group included 31 neonates whose mothers did not sing or listen to music through loudspeakers on a daily basis. Summary statistics of the mother's reports can be found in Table 1.

We acknowledge that our study would have benefited from allowing answers in a continuous manner (i.e., number of days per month), rather than in incremental steps (Daily, Weekly, Once every two weeks, Monthly and Never), enabling a correlation approach in data analysis. With the collected data, though, such an analysis would be subject to statistical issues given the distribution of counts per answer. Counts per answer to days in a month with music listening or singing exhibited a left skewed distribution, revealing that about half the sample (29/60) listened to music through loudspeakers and/or sang daily, while the rest (31/60) did so on a lesser degree. Interestingly, the number of participants who listened to music weekly (N=24) was similar to that of participants who listened to music daily (N=29). Taking into account that group size was well balanced between DE (N=29) and NDE (N=31), and that NDE contained a

77.4% of participants (24/31) with weekly musical exposure, any differences between groups we may found would be even more conclusive.

Table 1. Summary statistics of mother's reports: frequency of responses for listening to music and singing as reported by the newborns' mothers through the musical questionnaire.

	DE		NDE	
	Frequency	% of the sample	Frequency	% of the sample
Listen to music				
Daily	29	100	-	-
Weekly	-	-	24	77.4
Once every two weeks	-	-	3	9.7
Monthly	-	-	0	0
NA/Never	-	-	4	12.9
Sing				
Daily	29	100	-	-
Weekly	-	-	12	38.7
Once every two weeks	-	-	4	12.9
Monthly	-	-	2	6.5
NA/Never	-	-	13	41.9
Total	29	100	31	100

Stimuli

Neonatal FFRs were recorded to two different speech stimuli: a consonant-vowel syllable /da/ (Ribas-Prats et al., 2019; Ribas-Prats et al., 2021) and a two-vowel syllable /oa/ (Arenillas-Alcón et al., 2021a) in a group design (i.e., a single newborn was stimulated either with the /da/ or the /oa/ stimulus). The stimulus /da/ was chosen since it is the most commonly employed in FFR research with newborns (Lemos et al., 2021; Ribas-Prats et al., 2019; Ribas-Prats et al., 2021; Richard et al., 2020). This stimulus, created by Klatt-based synthesizer (Klatt, 1980) and modified by *Praat* (Boersma & Weenink, 2020), has a duration of 170 ms, divided in 10 ms of onset period, 47 ms of consonant transition and 113 ms for the /a/ vowel section (consonant transition: 10-57 ms, F_0 = 113 Hz, F_1 = 553-688 Hz; /a/ vowel section: 57-170 ms, F_0 = 113 Hz, F_1 = 688 Hz; Fig. 1A), and was presented at a rate of 3.7 Hz. The F_0 was kept steady at 113 Hz during the whole duration of the stimulus. Of the total number of newborns whose FFR were recorded with the /da/ stimulus, 8 of them were considered to belong to the DE group, and 11 belonged to the NDE group.

In turn, the stimulus /oa/ (Arenillas-Alcón et al., 2021b), was created in *Praat* (Boersma & Weenink, 2020) with a total duration of 250 ms divided in two vowel sections as well as two different F_0 sections (/o/ section: 10-80 ms, F_0 = 113 Hz, F_1 = 452 Hz; /a/ steady section = 90-160 ms, F_0 = 113 Hz, F_1 = 678 Hz; /a/ rising section = 160-250 ms, F_0 = 113-154 ms, F_1 = 678 Hz; Fig. 1B), and was presented at a rate of 3.39 Hz. Attending to its pitch, F_0 was kept steady at 113

Hz during the first part of the stimulus (0-160 ms) and increased linearly until 154 Hz (160-250 ms). Of the total number of newborns whose FFR were recorded with the /oa/ stimulus, 21 of them were considered to belong to the DE group, and 20 belonged to the NDE group.

For the present study, only a section with equal duration (113 ms) and steady fundamental frequency (113 Hz) was chosen for analysis in both stimuli (/da/ from 57-170 ms, F_0 = 113 Hz; /oa/ from 47-160 ms, F_0 = 113 Hz). Both speech sounds were delivered monaurally to the right ear, in alternating polarities, at 60 dB SPL of intensity with an earphone connected to a Flexicoupler disposable adaptor (Natus Medical Incorporated, San Carlos, CA).

Procedure and data acquisition

After passing the universal neonatal hearing screening, newborns were tested in their crib at the hospital room while they were sleeping. Four disposable Ag/AgCl electrodes were placed in a vertical montage (active at Fpz, ground at forehead, one reference at each mastoid; Supp. Fig. 2A) with impedances below 7 kΩ, though in the current study only data referenced to the electrode located at the right mastoid (ipsilateral to the auditory stimulation) was considered for analyses. Click and speech stimuli were presented by using a *SmartEP* platform connected to a *Duet* amplifier, that includes the *cABR* and the *Advanced Hearing Research* modules (Intelligent Hearing Systems, Miami, Fl, USA).

Following procedures from previous newborn studies carried out in our laboratory, four blocks of 1000 artifact-free responses to the /da/ or /oa/ stimulus, respectively were recorded after the ABR blocks described above. Any electrical activity exceeding ±30 μV was automatically rejected until a total of 4000 presentations to each stimulus were collected. The total mean duration of the recording session was approximately 25 min (sessions with /da/: two click blocks x 2000 repetitions x 51.81 ms of stimulus-onset asynchrony (SOA) + four /da/ blocks x 1000 repetitions x 270 ms SOA + duration of rejected repetitions; sessions with /oa/: two click blocks x 2000 repetitions x 51.81 ms SOA + four /oa/ blocks x 1000 repetitions x 295 ms SOA + duration of rejected repetitions). The continuous electroencephalography signal was acquired at a sampling rate of 13333 Hz, bandpass filtered online from 30-1500 Hz and epoched and averaged online from -40.95 ms (pre-stimulus period, for both stimuli) to 229.32 ms (/da/ stimulus), or to 249.975 ms (/oa/ stimulus).

Data processing and analysis

Data were bandpass filtered offline from 80 to 1500 Hz. To emphasize the encoding of the stimulus (F_0) and to minimize the contribution of cochlear microphonics, neural responses elicited to the two opposite stimulus polarities were averaged [(Condensation + Rarefaction)/2] (Aiken & Picton, 2008; Krizman & Kraus, 2019), to obtain the envelope-following FFR. All parameters

were computed with custom scripts from Matlab R2019b (The Mathworks, 2019) developed in our laboratory and used in similar analyses performed in previous studies.

Parameters extracted from ABR

Wave V. The latency of wave V peak in the ABR was determined by automatically identifying the major positive peak from 8-9.90 ms and its corresponding amplitude, and was taken as an estimation of the brainstem conduction time from the stimulus reception in the cochlea to the inferior colliculus in the midbrain (Ribas-Prats et al., 2019; Stuart et al., 1994). Additionally, the automatic detection was visually reviewed to identify peaks slightly outside the established time range.

Parameters extracted from FFR

Neural lag. Neural lag was considered as an inference of the neural transmission delay of the auditory system, since this value provides evidence of the amount of time passed from the reception of the stimulus at the cochlea until the start of the neural phase-locking (Arenillas-Alcón et al., 2021a; Jeng et al., 2010; Liu et al., 2015; Ribas-Prats et al., 2019; Ribas-Prats et al., 2021). It was calculated by computing a cross-correlation between the auditory stimulus and the neural response within a 3-13 ms time window, obtaining a correlation value at each time lag. The neural lag was obtained by selecting the time lag corresponding to the maximum cross-correlation value.

Pre-stimulus root mean square (RMS) amplitude. The RMS of the pre-stimulus period was taken as an indicator of the overall magnitude of neural activity along time, and used to discard electrophysiological differences in the pre-stimulus region (Liu et al., 2015; Ribas-Prats et al., 2019; Ribas-Prats et al., 2021; White-Schwoch et al., 2015). It was calculated by squaring each point of the pre-stimulus region of the neural response (from -40 to 0 ms), computing the mean of the obtained values and calculating the square root of the obtained average.

Spectral amplitude at F_0. Spectral amplitude at F_0 (113 Hz) was considered as a measure of the magnitude of the neural phase-locking at the specific chosen frequency, obtaining an indicator of the response strength (Arenillas-Alcón et al., 2021a; Ribas-Prats et al., 2019; Ribas-Prats et al., 2021; White-Schwoch et al., 2015). It was computed using a fast Fourier transform (FFT; Cooley & Tukey, 1965) to obtain the neural response frequency structure, and calculating the mean amplitude within a ±5 Hz window centered at the stimulus F_0 peak.

Statistical analysis

Statistical analyses were performed on SPSS 25.0 (Corp., n.d.). Descriptive statistics include the mean, standard deviation (SD), median, first (Q_1) and third (Q_3) quartiles, interquartile range (IQR), and minimum and maximum values of the computed parameter for each group of

newborns (DE; NDE). The Shapiro-Wilk test was selected to check the normal distribution of the samples. Depending on the normality of the samples, two-tailed independent samples *t*-tests or *Mann-Whitney U* tests were computed to check for significant differences in comparisons between groups, reporting Cohen's *d* for effect size. Results were considered significant when $p < 0.05$.

A univariate general linear ANOVA was conducted to examine and control for different effects depending on the type of stimulus used. To assess effects of Group (DE and NDE), this variable was introduced in the ANOVA as fixed factor; for the effect of Stimulus (/da/ and /oa/), the variable was taken as random factor. Partial eta squared effect (η^2_p) sizes are reported.

RESULTS

The ABR and the FFRs elicited by the /da/ (Fig. 1A) and the /oa/ (Fig. 1B) stimuli were successfully collected from a total of 60 newborns, which were divided into two groups according to their musical exposure during the last trimester of pregnancy (Daily musical exposure: DE; Not-Daily musical exposure: NDE). Neonatal FFR parameters (neural lag, pre-stimulus RMS, spectral amplitude at stimulus F_0) were analyzed in a section of both stimuli identical in duration (113 ms) and fundamental frequency (113 Hz). Table 2 reports the descriptive statistics for the FFR parameters analyzed separately in both groups; statistics for each stimulus separately can be found in Supp. Table 2 (/da/) and Supp. Table 3 (/oa/).

ABR. From the ABRs, the wave V could be identified in all newborns recruited. For the daily musically exposed (DE) group, the mean latency of wave V was 8.582 ± 0.336 ms, and its mean amplitude was 0.132 ± 0.062 μV. The not-daily musically exposed (NDE) group exhibited a wave V latency of 8.482 ± 0.368 ms, and a mean amplitude of 0.096 + 0.091 μV. In Supp. Fig. 2B the grand-average of the ABR waveform is shown; violin plots of the group distribution for wave V latency and amplitude are depicted in Supp. Fig. 2C and 2D, respectively. No significant group differences were found for wave V latency ($U = 347.000$, $p = 0.129$, Cohen's $d = 0.400$) or amplitude ($U = 329.000$, $p = 0.075$, Cohen's $d = 0.473$).

Neural lag. DE and NDE groups did not exhibit significant differences in neural transmission delay values ($t_{(58)} = 0.459$, $p = 0.648$, Cohen's $d = 0.119$).

Pre-stimulus root mean square (RMS) amplitude. No different background neural activity was found across groups ($U = 356.000$, $p = 0.167$, Cohen's $d = 0.363$).

Spectral amplitude at F_0. Spectral representation of the neonatal FFR extracted from the equivalent analyzed sections of each stimulus is shown in Fig. 2A. Violin plots and graphic bars illustrating the distribution of spectral amplitudes values at the fundamental frequency (F_0) for

each group and stimulus are provided in Fig. 2B and 2C. Statistical analyses revealed significant differences between groups, indicating that newborns exposed daily to music during the last trimester of pregnancy exhibited a larger spectral amplitude at the stimulus F_0 as compared to the less exposed group ($U = 275.000$, $p = 0.010$, Cohen's $d = 0.707$). Furthermore, in order to control for the influence of the stimulus type (/da/ or /oa/) on the stimulus F_0 spectral amplitude values, a univariate ANOVA was computed. A main effect of Group revealed significantly greater spectral amplitudes at stimulus F_0 for the DE newborns ($F = 6.750$, $p = 0.012$, $\eta^2_p = 0.108$). A significant main effect of Stimulus was also observed, caused by a larger spectral amplitude to the /da/ stimulus as compared to the /oa/ ($F = 14.152$, $p < 0.001$, $\eta^2_p = 0.202$). However, no interaction between the two factors was found ($F = 0.132$, p = 0.718, $\eta^2_p = 0.002$). Consequently, our data reveals that newborns exposed to music on a daily basis during the last trimester of pregnancy show a greater magnitude of neural phase-locking to speech stimuli F_0 than not-daily exposed neonates. Importantly, this effect was observed regardless of the type of speech stimulus, /da/ or /oa/, used to elicit the FFR, and despite the neural response was in itself remarkably different between stimuli (Fig. 2C).

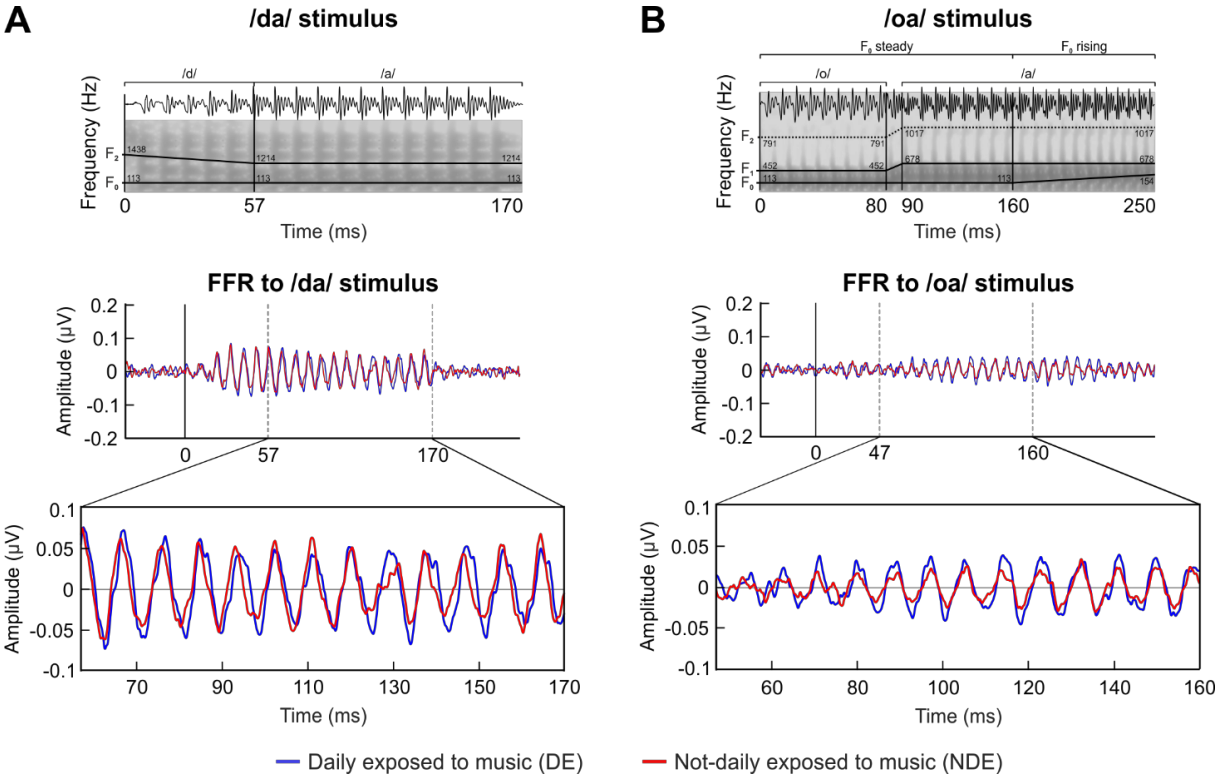

Fig. 1. Temporal representations of the FFR elicited in the DE (blue) and the NDE (red) groups separately. **A (top)**: Acoustic waveform and spectrogram of the /da/ stimulus with a schematic overlay of the formant structure trajectory; **A (center)**: Grand averaged time-domain waveform of the FFR elicited by the /da/ stimulus. **A (bottom)**: Zoom of the equivalent analyzed section (57-170 ms) of the waveform neural response. **B (top)**: Acoustic waveform and spectrogram of the /oa/ stimulus with schematic overlay of the formant structure trajectory; **B (center)**: Grand averaged time-domain waveform of the FFR elicited by the /oa/ stimulus. **B (bottom)**: Zoom of the equivalent analyzed section (47-160 ms) of the waveform neural response.

Table 2. Descriptive statistics for DE (N=29) **and NDE** (N=31) **groups in FFR parameters**: neural lag, Root-Mean-Square from pre-stimulus section, spectral amplitude at F_0 peak, computed for the steady pitch section of each stimulus (/da/; /oa/).

Measure	Mean	SD	Median	Q_1	Q_3	IQR	Minimum	Maximum
Neural lag (ms)								
DE	7.585	1.396	7.425	6.675	8.475	1.800	4.650	10.575
NDE	7.389	1.873	7.425	5.850	8.100	2.250	4.650	12.975
Pre-stimulus RMS (μV)								
DE	0.033	0.018	0.031	0.020	0.038	0.018	0.015	0.088
NDE	0.026	0.010	0.022	0.020	0.032	0.012	0.011	0.047
F_0 Spectral Amplitude (nV)								
DE	20.068	9.464	18.622	12.696	25.702	13.005	5.480	49.073
NDE	14.420	8.840	11.600	7.697	20.131	12.435	2.838	35.041

SD = standard deviation. Q_1 = first quartile (25th percentile). Q_3 = third quartile (75th percentile). IQR = interquartile range.
Note: Descriptive statistics in FFR parameters for each stimulus separately can be found in Supp. Table 2 (/da/) and Supp. Table 3 (/oa/).

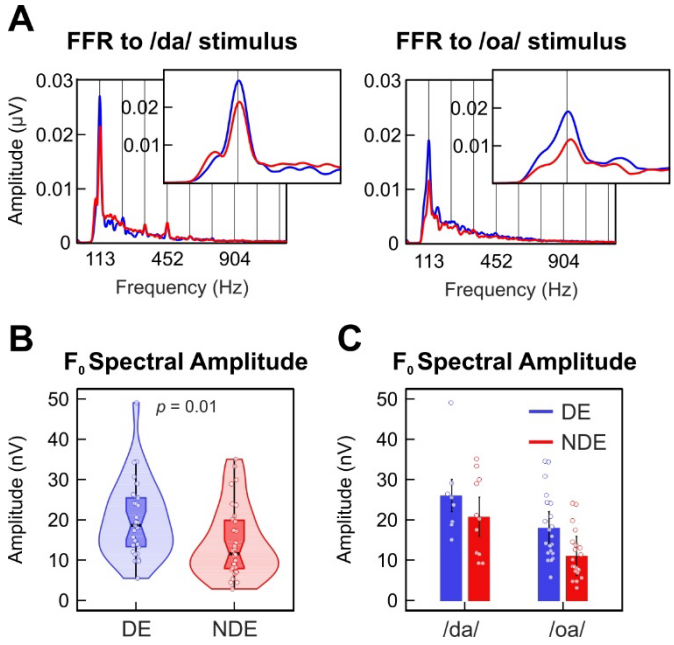

Fig. 2. Spectral representation of the FFR and data distribution of the F_0 spectral amplitude for the DE (blue) and the NDE (red) groups. **A (left)**: FFR frequency spectra extracted from the equivalent /da/ analyzed section of the neural response. **A (right)**: FFR frequency spectra extracted from the equivalent /oa/ analyzed section of the neural response. **B**: Violin plots of F_0 spectral amplitude. Horizontal black lines and vertical black lines indicate the median and 1.5 times the interquartile range (IQR), respectively. Dark colored boxes illustrate the IQR, while violin plots outlines show kernel probability density, i.e., the proportion of data located there. Scatter plots show all tested participants in each group. **C**: Vertical color bars represent the mean F_0 spectral amplitude separately for each group (DE; NDE) exposed to the /da/ (left) or the /oa/ (right) stimuli. Vertical black lines illustrate the mean standard error. Scatter plots show all tested participants in each group.

An additional ANOVA excluding one outlier newborn in the DE group (with a F_0 spectral amplitude elicited to the /da/ stimulus equal to 2.18 times the interquartile range) yielded similar results, except for a Group effect p value marginally above the significance level (Group effect: $F = 3.919$, $p = 0.053$, $\eta^2_p = 0.067$; Stimulus effect: $F = 10.270$, $p = 0.002$, $\eta^2_p = 0.157$; Interaction:

F = 1.294, p = 0.260, η^2_p = 0.023). Considering that the data was not normally distributed, we also conducted a non-parametric analysis in order to test for Group and Stimulus main effects, yielding significant differences in both factors (Group main effect: U = 275.000, p = 0.016, Cohen's d = 0.587; Stimulus main effect: U = 200.000, p = 0.005, Cohen's d = 0.835). Detailed descriptive statistics are reported in Supp. Table 4.

DISCUSSION

In the current study, we disclose the effects of prenatal music exposure on speech stimuli F_0 encoding at birth, through the analysis of the neonatal frequency-following response (FFR) elicited by two different periodic speech stimuli (/da/ and /oa/) in stimulus sections of identical pitch. Our results indicate that daily exposure to music, during the last three months of pregnancy, is related to a stronger encoding of speech stimuli F_0 content at birth. This was evidenced as greater neural response amplitudes at F_0 in the daily exposed (DE) group, in absence of distinct background neural activity and regardless of the specific stimulus used to elicit the FFR. Moreover, no significant differences between groups were found in auditory neural transmission time measures, such as wave V latency or neural lag. Thus, the results of the present study suggest that the ability to perceive and process pitch at birth, that mainly depends on the neural encoding of the F_0 of the incoming sounds (Krizman & Kraus, 2019; Plack et al., 2014) is, to a certain extent, modulated by the auditory experiences while in the womb.

Several factors support the value of the present results. First, the sample of newborns was homogeneous in terms of gestational age and birth weight, factors known to affect the FFR (Ribas-Prats et al., 2021). Second, we found no differences across groups segregated by musical exposure in background neural activity, as measured by pre-stimulus RMS amplitudes. This highlights the observed group differences in FFR spectral amplitudes elicited by speech stimuli. Third, auditory neural transmission time measures did not differ across groups, as evidenced by ABR wave V latencies and the FFR neural lag. This ensures that the processing of the acoustic input along the ascending auditory pathway is homogeneous across groups, being all newborns tested normal in terms of typical hearing screening measures implemented in hospital routines. Furthermore, this indicates that prenatal music exposure does not have an impact on neural transmission times immediately after birth, in contrast with the faster processing speed of auditory and speech stimuli reported in musically trained adults (Schochat et al., 2017). Finally, and most importantly, our study design featured two speech stimuli of different characteristics except for their periodicity, which were delivered to two different sets of newborns. Albeit FFR spectral amplitudes were different across stimuli, with the /da/ stimulus yielding overall higher spectral amplitudes at F_0 than the /oa/ stimulus, in line with previous studies (see spectral amplitudes for /da/ in Ribas-Prats et al., 2019; and for /oa/ in Arenillas-Alcón et al., 2021a), we found no

interaction between stimulus type and musical exposure. Therefore, regardless of the syllable that a newborn heard, fetuses that were daily exposed to music while in the womb showed a stronger neural encoding of speech F_0 at birth.

Accurate F_0 tracking is fundamental to pitch processing, as pitch directly relates to the periodicity of a sound waveform (Plack et al., 2014). Sound pitch patterns organized in time are the basis of musical melody, and simultaneously sounding pitches produce musical harmony. In human oral communication, pitch is a crucial cue to recognize different speakers, segment speech in linguistically relevant units and segregate speech from a noisy background, but it is an essential feature in tonal language semantics and in linguistic prosody, which conveys intonational meaning as well as emotional content (Arenillas-Alcón et al., 2021a; Benavides-Varela et al., 2012; Cabrera & Gervain, 2020; Gervain, 2018; Musacchia et al., 2007; Partanen et al., 2013; Plack et al., 2014; Ribas-Prats et al., 2021). Moreover, pitch cues underlie the discrimination between infant-directed and adult-directed speech (Fernald & Kuhl, 1987). In fact, the acoustic characteristics of infant-directed speech, particularly regarding pitch (i.e., higher and more variable) and purer vocal timbres (Hilton et al., 2022), resemble those of music. Because infant-directed speech aids language acquisition (Ma et al., 2011; Trainor & Desjardins, 2002), we hypothesize that exposure to music might as well provide similar benefits. Unfortunately, our questionnaire did not collect any data on the amount of speech exposure while in the womb, nor whether speech was directed to the fetus. Nevertheless, pitch appears to be a very important feature of speech to babies, as its variations directly call for their attention, providing a substantial wealth of information and aiding language development in multiple ways.

It is worth noting that linguistic prosody and non-vocal musical patterns are readily available to fetuses due to the low-pass filter characteristics of the mother's womb (Gerhardt & Abrams, 2000; Jeng, 2017; McCarthy et al., 2019; Parga et al., 2018). These characteristics may explain why healthy newborns exhibit an adult-like neural encoding of speech F_0, while the encoding of other speech features, based on sound frequencies higher than the womb's low-pass cutoff, are still undeveloped (Arenillas-Alcón et al., 2021a). It is also the most plausible explanation for the wealth of studies showing the influence of prenatal acoustic experiences in shaping the sound preferences of newborns (Chorna et al., 2019; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Gervain, 2018; May et al., 2011; Moon et al., 1993, 2012; Partanen et al., 2013). Interestingly, if plasticity in the fetus' auditory system is mainly driven by F_0 variation information, our findings stress the importance of a shared pitch processing mechanism across speech and non-speech auditory domains that can be modulated before birth and assessed after birth through non-invasive FFR recordings.

Timing and rhythmic structure are other important characteristics of both music and speech which are also available to the fetus. However, we overlooked them in the present study as we focused only on F_0 neural representation. Both refer to the temporal organization of acoustic events (i.e., energy changes in the acoustic signal), and are crucial to find structure and meaning in speech and music (Iversen et al., 2008). Musical rhythm perception is associated with phonological awareness (Flaugnacco et al., 2014), with the production of complex syntax and reorganization of grammatical information (Gordon et al., 2015), and with developmental conditions such as dyslexia and developmental language disorders (for a review from an auditory neuroscience perspective, see Goswami, 2022). Moreover, training in temporal processing and rhythmic skills with musical material appears to exert a beneficial effect in children with developmental dyslexia (Flaugnacco et al., 2015). Therefore, future studies on prenatal musical exposure should measure as well the neural representation of musical and speech rhythmic patterns in newborns and, in longitudinal designs, relate them to language development.

The present study suggests potential implications for infants at risk for language development conditions. Children who experience language disorders and clinical conditions with affected linguistic functions exhibit a weaker neural encoding of F_0 (Banai et al., 2009; Basu et al., 2010; Chandrasekaran et al., 2009; Font-Alaminos et al., 2020; Hornickel et al., 2012; King et al., 2002; Lam et al., 2017; Otto-Meyer et al., 2018; Rosenthal, 2020). Moreover, musicians from different ages show, as compared to non-musicians, improved non-vocal musical processing and, crucially, speech processing (Bidelman et al., 2011; Deguchi et al., 2012; Magne et al., 2006; Musacchia et al., 2007; Schön et al., 2004; Thompson et al., 2003; Wong et al., 2007). Thus, prenatal interventions based on musical stimulation could prove beneficial to ameliorate future language conditions. However, much basic research is needed before designing any evidence-based intervention. Our results link prenatal musical exposure to F_0 neural encoding, but do not provide any causal explanation nor detail on the most relevant constituents of musical exposure. For instance, the type of music that the mother and the fetus are exposed to, and its intensity, are both important factors in determining the impact of the exposure on speech encoding abilities and their general well-being (Gerhardt & Abrams, 2000; Wright et al., 2022). Musical features such as tempo, meter, melodic frequency range, musical notes, syllabic contour and presence or absence of singing differ across music genres (Teie, 2016), and are based on acoustic features (i.e., pitch, intensity, timing…) that are readily available to the fetus despite the low-pass filter characteristics of the womb. The relative impact of these variables on fetal hearing development is currently unknown. Unfortunately, although we collected information about musical genres that the mothers who participated in this study mostly heard or sung, the small size of the sample precludes us to consider this intriguing variable. Thus, longitudinal studies controlling for the amount, intensity and genre/characteristics of musical exposure such as timing, assessing F_0

neural encoding at birth and at several developmental stages, and relating these variables to measures of language acquisition and brain development (as the critical developmental windows for neuroplasticity already start in utero; Gilmore et al., 2018) should provide the needed evidence to adequately inform music-based interventions. Also, future longitudinal studies should also take into account the prenatal exposure to infant-directed speech, as its acoustic characteristics, especially regarding pitch, resemble those of music (Hilton et al., 2022).

Moreover, music-based interventions have extensively been used in numerous conditions of neurodevelopmental risk at neonatal intensive care units (NICU), especially in preterm newborns (Chorna et al., 2018; Lordier et al., 2019; Olischar et al., 2011; Palazzi et al., 2021; Partanen et al., 2022; Standley, 2012), proving their effectiveness on fetal and neonatal well-being and, in addition and reciprocally, on the mother's comfort (Brillo et al., 2021; Çatalgöl & Ceber Turfan, 2021; García González et al., 2017; He et al., 2021; Poćwierz-Marciniak & Harciarek, 2021). In fetal growth restriction (FGR), an obstetric condition that affects 6-10% of all deliveries (Marsál, 2002), neural F_0 encoding was found attenuated in neonates right after birth (Ribas-Prats et al., 2021). Since babies at risk of FGR are routinely identified around the third trimester of pregnancy (Melamed et al., 2021) and, as observed in the present study, musical exposure during the last trimester of pregnancy can enhance F_0 processing, future studies could aim to disentangle the effectiveness of prenatal musical interventions in this obstetric condition, in line with recent results showing the impact of environmental manipulations on stress reduction and prevention in FGR (Crovetto et al., 2022).

We here interpreted our findings as resulting from neural plasticity mechanisms occurring before birth due to a prenatal exposure to musical stimulation. However, there are alternative viewpoints to consider. First, and in line with fetal music-based interventions, a reduction of maternal and fetus stress due to a higher exposure to music could induce neuroplastic changes in the auditory system. Also, a recent FFR study demonstrated an increased neural representation of sound F_0 in noisy environments in individuals with better musical abilities, despite no musical training (Mankel & Bidelman, 2018). Considering the heritability of some musical abilities, with some genes linked to music perception, singing and music memory (Tan et al., 2014), our results might as well be explained in part by the possibility that mothers more prone to listen to music and sing, give birth to babies with better sound encoding abilities.

Finally, despite being confident about our results due to the abovementioned reasons, we are fully aware of an existing, important limitation of our study: musical exposure was assessed by a short (approx. 5 min answer time), retrospective questionnaire provided at the time of delivery, with a spoken description of the content of the questionnaire. This poses, at least, two factors not adequately controlled. First, the actual frequency in which mothers listened to music

or sung, as we rely on their reports referring to the last trimester of pregnancy. The present study only documented number of days per month of exposure in closed categories (Daily; Weekly; Twice a month; Monthly; Never) rather than in a continuous fashion. Furthermore, although a minimum period of exposure to music had to occur to be considered as valid, the questionnaire did not address the exact amount of music exposure within a day (e.g., 30 min versus 3 hours per day). Second, the intensity of the music reaching the womb. Future studies should address these limitations. For instance, by collecting large amounts of data from a maternal diary of musical exposure during the last trimester of pregnancy, which could also inspect the music genre variable, and include an additional musical abilities test (such as PROMS; Law & Zentner, 2012) to evaluate the putative link between F_0 encoding abilities in newborns and parental musical abilities. Additionally, a controlled experimental design in which the experimenter could define musical exposure (location and intensity of a loudspeaker, distance to the loudspeaker, amount of hours of exposure, etc.), could be implemented. Moreover, since our study used a cross-sectional design, we can only establish a link between the measured prenatal music exposure and F_0 neural encoding. Longitudinal and intervention work are of the essence to determine whether prenatal exposure to music, and no other possible contributing factors, is associated with neuroplasticity in utero and through early development, and its impact on future language development.

CONCLUSION

In conclusion, acknowledging the aforementioned limitations, our findings support the idea that daily musical exposure during the last trimester of pregnancy is associated with enhanced encoding of low-frequency sound components, such as those typical of the fundamental frequency of human speech, that relate to pitch perception. However, future studies should address open questions, such as what acoustic and musical parameters provide greater benefits or what is the actual nature of the neuroplastic changes that the fetus undergoes with musical exposure, before music-based prenatal interventions are implemented to alleviate putative language disorders.

AUTHOR CONTRIBUTIONS

S.A.A.: Conceptualization, Methodology, Investigation, Formal analysis, Writing-Original draft, Writing-Review & Editing, Visualization. **T.R.P.:** Methodology, Investigation. **M.P.:** Methodology, Investigation. **A.M.S.:** Methodology, Investigation. **M.D.G.R.:** Funding acquisition, Resources. **J.C.F.:** Conceptualization, Methodology, Writing-Original draft, Writing-Review & Editing, Supervision. **C.E.:** Conceptualization, Methodology, Writing-Review & Editing, Funding acquisition, Resources, Supervision.

REFERENCES

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*(1–2), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Anbuhl, K. L., Uhler, K. M., Werner, L. A., & Tollin, D. J. (2016). Early Development of the Human Auditory System. In R. A. Polin, S. H. Abman, D. Rowitch, & W. E. Benitz (Eds.), *Fetal and Neonatal Physiology* (5th ed., pp. 1396–1410). Elsevier. https://doi.org/10.1016/B978-0-323-35214-7.00138-4

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021a). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Scientific Reports, 11*, 6660. https://doi.org/10.1038/s41598-021-85799-x

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021b). Neural Encoding of Vocalic Sounds in Newborns. *Hearing Journal, 74*(7), 10–11.

Bainbridge, C. M., Bertolo, M., Youngers, J., Atwood, S., Yurdum, L., Simson, J., Lopez, K., Xing, F., Martin, A., & Mehr, S. A. (2021). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behaviour, 5*(2), 256–264. https://doi.org/10.1038/s41562-020-00963-z

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and Subcortical Auditory Function. *Cerebral Cortex, 19*, 2699–2707. https://doi.org/10.1093/cercor/bhp024

Basu, M., Krishnan, A., & Weber-Fox, C. (2010). Brainstem correlates of temporal auditory processing in children with specific language impairment. *Developmental Science, 13*(1), 77–91. https://doi.org/10.1111/j.1467-7687.2009.00849.x

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the United States of America, 109*(44), 17908–17913. https://doi.org/10.1073/pnas.1205413109

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434.

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer* (Version 6.1.09; p. Version 6.1.09). http://www.praat.org/

Brillo, E., Tosto, V., Ceccagnoli, A., Nikolova, N., Pinzaglia, V., Bordoni, F., Spano, F., Bini, V., Giardina, I., & Di Renzo, G. C. (2021). The effect of prenatal exposure to music on fetal movements and fetal heart rate: a pilot study. *Journal of Maternal-Fetal and Neonatal Medicine, 34*(14), 2274–2282. https://doi.org/10.1080/14767058.2019.1663817

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances, 6*(30), eaba7830. https://doi.org/10.1126/sciadv.aba7830

Carcagno, S., & Plack, C. J. (2011). Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. *Journal of the Association for Research in Otolaryngology, 12*, 89–100. https://doi.org/10.1007/s10162-010-0236-1

Carcagno, S., & Plack, C. J. (2017). Short-Term Learning and Memory: Training and Perceptual Learning. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (Vol. 61, pp. 75–100). Springer Nature.

Çatalgöl, Ş., & Ceber Turfan, E. (2021). The effects of music therapy applied to pregnant women on maternal, fetal, and neonatal results: A randomized controlled study. *Health Care for Women International*, 1–17. https://doi.org/10.1080/07399332.2021.1944150

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron, 64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009.10.006

Chorna, O, Filippa, M., Sa De Almeida, J., Lordier, L., Monaci, M. G., Hüppi, P., Grandjean, D., & Guzzetta, A. (2019). Neuroprocessing Mechanisms of Music during Fetal and Neonatal Development: A Role in Neuroplasticity and Neurodevelopment. *Neural Plasticity, 2019*, 1–9. https://doi.org/10.1155/2019/3972918

Chorna, Olena, Hamm, E., Shrivastava, H., & Maitre, N. L. (2018). Feasibility of ERP biomarker use to study effects of mother's voice exposure on speech sound differentiation of preterm infants. *Developmental Neuropsychology, 43*(2), 123–134. https://doi.org/10.1080/87565641.2018.1433671.Feasibility

Cirelli, L. K., & Trehub, S. E. (2020). Familiar Songs Reduce Infant Distress. *Developmental Psychology, 56*(5), 861–868. https://doi.org/10.1037/dev0000917

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications, 7*, 1–11. https://doi.org/10.1038/ncomms11070

Coffey, E. B. J., Musacchia, G., & Zatorre, R. J. (2017). Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. *The Journal of Neuroscience, 37*(4), 830–838. https://doi.org/10.1523/JNEUROSCI.1265-16.2016

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications, 10*(1), 1–10. https://doi.org/10.1038/s41467-019-13003-w

Cooley, J. W., & Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation, 19*(90), 297.

Corp., I. (n.d.). *SPSS 25.0*.

Crovetto, F., Crispi, F., & Gratacós, E. (2022). Mediterranean Diet or Mindfulness-Based Stress Reduction and Prevention of Small-for-Gestational-Age BirthWeights in Newborns--Reply. *JAMA*, 327(13), 1293–1294. https://doi.org/10.1159/000502652

DeCasper, A. J., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voice. *Science, 208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development, 9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-

musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research, 1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034

Fernald, A., & Kuhl, P. (1987). Acoustic Determinants of Infant Preference for Motherese Speech. *Infant Behavior and Development, 10*, 279–293.

Flaugnacco, E., Lopez, L., Terribili, C., Montico, M., Zoia, S., & Schön, D. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE, 10*(9), e0138715. https://doi.org/10.1371/journal.pone.0138715

Flaugnacco, E., Lopez, L., Terribili, C., Zoia, S., Buda, S., Tilli, S., Monasta, L., Montico, M., Sila, A., Ronfani, L., & Schön, D. (2014). Rhythm perception and production predict reading abilities in developmental dyslexia. *Frontiers in Human Neuroscience, 8*, 392. https://doi.org/10.3389/fnhum.2014.00392

Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., & Escera, C. (2020). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biological Psychology, 149*, 107807. https://doi.org/10.1016/j.biopsycho.2019.107807

García González, J., Ventura Miranda, M. I., Manchón García, F., Pallarés Ruiz, T. I., Marin Gascón, M. L., Requena Mullor, M., Alarcón Rodriguez, R., & Parron Carreño, T. (2017). Effects of prenatal music stimulation on fetal cardiac state, newborn anthropometric measurements and vital signs of pregnant women: A randomized controlled trial. *Complementary Therapies in Clinical Practice, 27*, 61–67. https://doi.org/10.1016/j.ctcp.2017.03.004

Gardi, J., Salamy, A., & Mendelson, T. (1979). Scalp-Recorded Frequency-Following Responses in Neonates. *International Journal of Audiology, 18*(6), 494–506.

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal Exposures to Sound and Vibroacoustic Stimulation. *Journal of Perinatology, 20*, 20–29.

Gervain, J. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences, 21*, 62–67. https://doi.org/10.1016/j.cobeha.2018.02.004

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience, 19*(3), 123–137. https://doi.org/10.1038/nrn.2018.1

Gordon, R. L., Jacobs, M. S., Schuele, C. M., & McAuley, J. D. (2015). Perspectives on the rhythm-grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences, 1337*, 16–25. https://doi.org/10.1111/nyas.12683

Gorina-Careta, N., Kurkela, J. L. O., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (MEG) study. *NeuroImage, 231*, 117866. https://doi.org/10.1016/j.neuroimage.2021.117866

Gorina-Careta, N., Ribas-Prats, T., Arenillas-Alcón, S., Puertollano, M., Gómez-Roig, M. D., & Escera, C. (2022). Neonatal Frequency-Following Responses: A Methodological Framework for Clinical Applications. *Seminars in Hearing, 43*(3), 162–176. https://doi.org/10.1055/s-0042-1756162

Gorina-Careta, N., Ribas-Prats, T., Costa-Faidella, J., & Escera, C. (2019). Auditory Frequency-Following Responses. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational*

Neuroscience (pp. 1–13). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-7320-6_100689-1

Goswami, U. (2022). Language acquisition and speech rhythm patterns: an auditory neuroscience perspective. *Royal Society Open Science, 9*, 211855. https://doi.org/10.1098/rsos.211855

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science, 14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

He, H., Huang, J., Zhao, X., & Li, Z. (2021). The effect of prenatal music therapy on fetal and neonatal status: A systematic review and meta-analysis. *Complementary Therapies in Medicine, 60*, 102756. https://doi.org/10.1016/j.ctim.2021.102756

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., Simson, J., Knox, D., Glowacki, L., Alemu, E., Galbarczyk, A., Jasienska, G., Ross, C. T., Neff, M. B., Martin, A., Cirelli, L. K., Trehub, S. E., Song, J., Kim, M., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01410-x

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., & Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *NeuroReport, 23*(1), 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd

Iversen, J. R., Patel, A. D., & Ohgushi, K. (2008). Perception of rhythmic grouping depends on auditory experience. *The Journal of the Acoustical Society of America, 124*, 2263. https://doi.org/10.1121/1.2973189

Jeng, F. C. (2017). Infant and Childhood Development: Intersections Between Development and Language Experience. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 17–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_2

Jeng, F. C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C. Der. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in american and chinese neonates and adults. *Ear and Hearing, 32*(6), 699–707. https://doi.org/10.1097/AUD.0b013e31821cc0df

Jeng, F. C., Lin, C.-D., & Wang, T.-C. (2016). Subcortical neural representation to Mandarin pitch contours in American and Chinese newborns. *The Journal of the Acoustical Society of America, 139*(6), 190–195. https://doi.org/10.1121/1.4953998

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.-D., & Chung, H.-K. (2010). Early Maturation of Frequency-Following Responses to Voice Pitch in Infants with Normal Hearing. *Perceptual and Motor Skills, 111*(3), 765–784. https://doi.org/10.2466/10.22.24.PMS.111.6.765-784

Joint Committee on Infant Hearing. (2019). Year 2019 Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention Programs. *Journal of Early Hearing Detection and Intervention, 4*(2), 1–44. https://doi.org/10.1542/peds.106.4.798

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters, 319*, 111–115. https://doi.org/10.1016/S0304-3940(01)02556-3

Klatt, D. (1980). Software for a cascade/parallel formant synthesizer. *The Journal of the Acoustical Society of America, 67*, 971–995.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599–605. https://doi.org/10.1038/nrn2882

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Krizman, J., Bonacina, S., Otto-Meyer, R., & Kraus, N. (2021). Non-stimulus-evoked activity as a measure of neural noise in the frequency-following response. *Journal of Neuroscience Methods, 362*, 109290. https://doi.org/10.1016/j.jneumeth.2021.109290

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research, 382*, 166–174. https://doi.org/10.1016/j.heares.2019.107779

Lam, S. S.-Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia, 103*, 162–167. https://doi.org/10.1016/j.physbeh.2017.03.040

Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE, 7*(12), e52508. https://doi.org/10.1371/journal.pone.0052508

Lemos, F. A., Da Silva Nunes, A. D., De Souza Evangelista, C. K., Escera, C., Meira Taveira, K. V., & Andreoli Balen, S. (2021). Frequency-Following Response in Newborns and Infants: A Systematic Review of Acquisition Parameters. *Journal of Speech, Language, and Hearing Research*, 1–18.

Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Frontiers in Human Neuroscience, 8*, 1–19. https://doi.org/10.3389/fnhum.2014.01029

Lordier, L., Loukas, S., Grouiller, F., Vollenweider, A., Vasung, L., Meskaldij, D.-E., Lejeune, F., Pascale Pittet, M., Borradori-Tolsa, C., Lazeyras, F., Grandjean, D., Van De Ville, D., & Hüppi, P. S. (2019). Music processing in preterm and full-term newborns: A psychophysiological interaction (PPI) approach in neonatal fMRI. *NeuroImage, 185*, 857–864. https://doi.org/10.1016/j.neuroimage.2018.03.078

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word Learning in Infant- and Adult-Directed Speech. *Language Learning and Development, 7*(3), 185–201.

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience, 18*(2), 199–211.

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *PNAS, 115*(51), 13129–13134. https://doi.org/10.1073/pnas.1811793115

Marsál, K. (2002). Intrauterine growth restriction. *Current Opinion in Obstetrics & Gynecology, 14*(2), 127–135.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology, 2*(222), 1–9. https://doi.org/10.3389/fpsyg.2011.00222

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports, 9*, 1–7. https://doi.org/10.1038/s41598-019-55085-y

Mehr, S. A., Krasnow, M. M., Bryant, G. A., & Hagen, E. H. (2021). Origins of music in credible signaling. *The Behavioral and Brain Sciences, 44*, e60. https://doi.org/10.1017/S0140525X20000345.Origins

Melamed, N., Baschat, A., Yinon, Y., Athanasiadis, A., Mecacci, F., Figueras, F., Berghella, V., Nazareth, A., Tahlak, M., McIntyre, H. D., Da Silva Costa, F., Kihara, A. B., Hadar, E., McAuliffe, F., Hanson, M., Ma, R. C., Gooden, R., Sheiner, E., Kapur, A., … Hod, M. (2021). FIGO (International Federation of Gynecology and Obstetrics) initiative on fetal growth: Best practice advice for screening, diagnosis, and management of fetal growth restriction. *International Journal of Gynecology and Obstetrics, 152*(Suppl 1), 3–57. https://doi.org/10.1002/ijgo.13522

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science, 24*, e13122. https://doi.org/10.1111/desc.13122

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2012). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paedriatrica, 102*, 156–160. https://doi.org/10.1111/apa.12098

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology, 46*(9), 460–478. https://doi.org/10.1080/14992020701383019

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America, 104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Olischar, M., Shoemark, H., Holton, T., Weninger, M., & Hunt, R. W. (2011). The influence of music on aEEG activity in neurologically healthy newborns ≥32 weeks' gestational age. *Acta Paediatrica, 100*(5), 670–675. https://doi.org/10.1111/j.1651-2227.2011.02171.x

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research, 236*, 733–743. https://doi.org/10.1007/s00221-017-5164-4

Palazzi, A., Filippa, M., Meschini, R., & Piccinini, C. A. (2021). Music therapy enhances preterm infant's signs of engagement and sustains maternal singing in the NICU. *Infant Behavior and Development, 64*, 101596. https://doi.org/10.1016/j.infbeh.2021.101596

Parga, J. J., Daland, R., Kesavan, K., Macey, P. M., Zeltzer, L., & Harper, R. M. (2018). A description of externally recorded womb sounds in human subjects during gestation. *PLoS ONE, 13*(5), 1–14. https://doi.org/10.1371/journal.pone.0197045

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *PNAS, 110*(37), 15145–15150. https://doi.org/10.1073/pnas.1302159110

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. *PLoS ONE, 8*(10), e78946. https://doi.org/10.1371/journal.pone.0078946

Partanen, E., Mårtensson, G., Hugoson, P., Huotilainen, M., Fellman, V., & Ådén, U. (2022). Auditory Processing of the Brain Is Enhanced by Parental Singing for Preterm Infants. *Frontiers in Neural Circuits, 16*, 772008. https://doi.org/10.3389/fnins.2022.772008

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research, 307*, 53–64. https://doi.org/10.1016/j.heares.2013.07.020

Poćwierz-Marciniak, I., & Harciarek, M. (2021). The Effect of Musical Stimulation and Mother's Voice on the Early Development of Musical Abilities: A Neuropsychological Perspective. *International Journal of Environmental Research and Public Health, 18*, 8467. https://doi.org/10.3390/ijerph18168467

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research, 371*, 28–39. https://doi.org/10.1016/j.heares.2018.11.001

Ribas-Prats, T., Arenillas-Alcón, S., Lip-Sosa, D. L., Costa-Faidella, J., Mazarico, E., Gómez-Roig, M. D., & Escera, C. (2021). Deficient neural encoding of speech sounds in term neonates born after fetal growth restriction. *Developmental Science*, 1–15. https://doi.org/10.1111/desc.13189

Richard, C., Neel, M., Jeanvoine, A., Connell, S., Gehred, A., & Maitre, N. L. (2020). Characteristics of the Frequency-Following Response to Speech in Neonates and Potential Applicability in Clinical Practice: A Systematic Review. *Journal of Speech, Language, and Hearing Research, 63*(5), 1618–1635.

Rosenthal, M. A. (2020). A systematic review of the voice-tagging hypothesis of speech-in-noise perception. *Neuropsychologia, 136*(November 2019), 107256. https://doi.org/10.1016/j.neuropsychologia.2019.107256

Ruben, R. J. (1995). The ontogeny of human hearing. International Journal of Pediatric *Otorhinolaryngology, 32*, 199–204. https://doi.org/10.1016/0165-5876(94)01159-U

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research, 156*, 95–103. https://doi.org/10.1016/j.bbr.2004.05.012

Schochat, E., Rocha-Muniz, C. N., & Filippini, R. (2017). Understanding Auditory Processing Disorder Through the FFR. In N. Kraus, S. Anderson, T. White-Schwoch, R. Fay, & A. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 225–250). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_5

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41*, 341–349. https://doi.org/10.1111/1469-8986.00172.x

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear and Hearing, 31*(3), 302–324. papers2://publication/uuid/07E666C5-A423-43C0-85EC-BA2067A7F59F

Skoe, E., Krizman, J., & Kraus, N. (2013). The impoverished brain: Disparities in maternal education affect the neural response to sound. *Journal of Neuroscience, 33*(44), 17221–17231. https://doi.org/10.1523/JNEUROSCI.2102-13.2013

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short- term linguistic training. *Journal of Cognitive Neuroscience, 20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131.Plasticity

Standley, J. (2012). Music therapy research in the NICU: An updated meta-analysis. *Neonatal Network, 31*(5), 311–316. https://doi.org/10.1891/0730-0832.31.5.311

Strait, D. L., Parbery-Clark, A., Hittner, E., & Kraus, N. (2012). Musical training during early childhood enhances the neural encoding of speech in noise. *Brain and Language, 123*(3), 191–201. https://doi.org/10.1016/j.bandl.2012.09.001

Stuart, A., Yang, E. Y., & Green, W. B. (1994). Neonatal Auditory Brainstem Response Thresholds To Air- and Bone-Conducted Clicks: 0 to 96 Hours Postpartum. *Journal of the American Academy of Audiology, 5*, 163–172.

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology, 5*, 658. https://doi.org/10.3389/fpsyg.2014.00658

Teie, D. (2016). A Comparative Analysis of the Universal Elements of Music and the Fetal Environment. *Frontiers in Psychology, 7*, 1158. https://doi.org/10.3389/fpsyg.2016.01158

The Mathworks, I. (2019). *MATLAB R2019b* (p. Natick, Massachusetts).

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Annals of the New York Academy of Sciences, 999*, 503–2.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Pyschonomic Bulletin & Review, 9*(2), 335–340.

Traunmüller, H., & Eriksson, A. (1995). *The frequency range of the voice fundamental in the speech of male and female adults. 2.*

Tsang, C. D., & Conrad, N. J. (2010). Does the message matter? The effect of song type on infants' pitch preferences for lullabies and playsongs. *Infant Behavior and Development, 33*, 96–100. https://doi.org/10.1016/j.infbeh.2009.11.006

Ullal-Gupta, S., Vanden Bosch der Nederlanden, C. M., Tichko, P., Lahav, A., & Hannon, E. E. (2013). Linking prenatal experience to the emerging musical mind. *Frontiers in Systems Neuroscience, 7*(48), 1–7. https://doi.org/10.3389/fnsys.2013.00048

Vannasing, P., Florea, O., González-Frankenberger, B., Tremblay, J., Paquette, N., Safi, D., Wallois, F., Lepore, F., Béland, R., Lassonde, M., & Gallagher, A. (2016). Distinct hemispheric specializations for native and non-native languages in one-day-old newborns identified by fNIRS. *Neuropsychologia, 84*, 63–69. https://doi.org/10.1016/j.neuropsychologia.2016.01.038

Webb, A. R., Heller, H. T., Benson, C. B., & Lahav, A. (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proceedings of the National Academy of Sciences of the United States of America, 112*(10), 3152–3157. https://doi.org/10.1073/pnas.1414924112

White-Schwoch, T., Davies, E. C., Thompson, E. C., Woodruff Carr, K., Nicol, T., Bradlow, A. R., & Kraus, N. (2015). Auditory-neurophysiological responses to speech during early childhood: Effects of background noise. *Hearing Research, 328*, 34–47.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420–422. https://doi.org/10.1038/nn1872.Musical

Wright, S. E., Bégel, V., & Palmer, C. (2022). Physiological influences of music in perception and action. In J. T. Enns (Ed.), *Elements in Perception*. Cambridge University Press. https://doi.org/10.1017/9781009043359

Yan, R., Jessani, G., Spelke, E. S., de Villiers, P., de Villiers, J., & Mehr, S. A. (2021). Across demographics and recent history, most parents sing to their infants and toddlers daily. *Philosophical Transactions of the Royal Society B: Biological Sciences, 376*(1840), 20210089. https://doi.org/10.1098/rstb.2021.0089

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D. E., Tallal, P., & Gaab, N. (2013). Enhanced Syllable Discrimination Thresholds in Musicians. *PLoS ONE, 8*(12), 1–8. https://doi.org/10.1371/journal.pone.0080546

STUDY III

Effects of a prenatal bilingual environment on the neural encoding of speech sounds: Evidence from neonatal frequency-following responses

Effects of a prenatal bilingual environment on the neural encoding of speech sounds at birth: Evidence from frequency-following responses[5]

Sonia Arenillas-Alcón[1,2,3], Natàlia Gorina-Careta[1,2,3], Marta Puertollano[1,2,3], Alejandro Mondéjar-Segovia[1,2], Siham Ijjou-Kadiri[1,2], Jordi Costa-Faidella[1,2,3,*], María Dolores Gómez-Roig[3,4] & Carles Escera[1,2,3,*]

[1] Brainlab – Cognitive Neuroscience Research Group. Departament de Psicologia Clinica i Psicobiologia, Universitat de Barcelona (Barcelona, Spain)

[2] Institut de Neurociènces, Universitat de Barcelona (Barcelona, Spain)

[3] Institut de Recerca Sant Joan de Déu, Santa Rosa 39-57, 08950 Esplugues de Llobregat (Catalonia, Spain)

[4] BCNatal – Barcelona Center for Maternal Fetal and Neonatal Medicine (Hospital Sant Joan de Déu and Hospital Clínic), University of Barcelona (Catalonia, Spain)

*** Corresponding authors:**

Carles Escera, PhD. Department of Clinical Psychology and Psychobiology, University of Barcelona. Passeig Vall d'Hebron 171, 08035 Barcelona (Catalonia, Spain). Tel: +34 933 125 048. Fax: +34 934 021 584. E-mail: cescera@ub.edu

Jordi Costa-Faidella, PhD. Department of Clinical Psychology and Psychobiology, University of Barcelona. Passeig Vall d'Hebron 171, 08035 Barcelona (Catalonia, Spain). Tel: +34 933 125 058. Fax: +34 934 021 584. E-mail: jcostafaidella@ub.edu

Conflict of Interest Disclosure: The authors declare no conflict of interest.

Data Availability Statement: The data that support the findings of this study are available from the corresponding authors Carles Escera (cescera@ub.edu) and Jordi Costa-Faidella (jcostafaidella@ub.edu) upon reasonable request.

Funding/Support Statement: This work was supported by the Spanish Ministry of Science and Innovation PGC2018-094765-B-I00 project (MCIN/AEI/10.13039/501100011033/ FEDER "Una manera de hacer Europa"), the MDM-2017-0729-18-2 María de Maeztu Center of Excellence (MCIN/AEI/10.13039/501100011033), and the ICREA Acadèmia Distinguished Professorship awarded to Carles Escera.

Acknowledgements: The authors would like to thank to all families who selflessly participated in this study and made it possible.

[5] Working paper

Effects of a prenatal bilingual environment on the neural encoding of speech sounds at birth: Evidence from frequency-following responses

Abstract

Beginning in the earliest stages of infancy, language exposure affects how the auditory system processes sounds. Even though the newborns' auditory systems are not biased toward their parents' native tongues, exposure to specific language environments during the first year of life obviously changes how babies perceive speech. However, the early phases of language development have already started even before birth, with certain plastic changes occurring even during the prenatal stage to the infants' musical and language prenatal hearing experiences. As bilingualism, relative to a monolingual environment, has been shown to enhance brain evoked responses to speech in children and adults, the present study sought to investigate whether a bilingual acoustic environment during the third trimester of pregnancy could modulate the neonate's ability to encode speech sounds. For testing that, the frequency-following response (FFR), an auditory evoked potential sensitive to the complex spectrotemporal dynamics of speech sounds, was recorded in a sample of 131 healthy term neonates during their first days of life. Newborns were divided into two groups according to their prenatal language exposure as reported by their mothers through a questionnaire (53 exposed to a monolingual fetal acoustic environment; 78 bilingual-exposed). The FFR was recorded to an /oa/ stimulus and quantified as the spectral amplitude and signal-to-noise ratio (SNR) at the stimulus fundamental frequency (F_0). Analyses disclosed that monolingual-exposed newborns exhibited larger SNR at F_0 as compared to the bilingual group, as well as significantly greater spectral amplitudes to the vowels formant structure. Our results suggest that prenatal language experience has a modulatory effect on the neural responses to speech sounds at birth, and specifically, that a monolingual acoustic environment generates a less variable background for the neonates to encode and process speech sounds. Our results support the current hypothesis that bilingual infants commence the process of language acquisition by separating languages from birth by demonstrating that, whilst the exposure of the fetus to a single language would produce more robust and larger amplitude responses at certain specific frequencies, bilingually exposed newborn's auditory system shows a major sensitivity to a wider range of frequencies without generating a particularly strong response at any of them.

Keywords: speech brainstem responses, bilingualism, newborns, early language acquisition

INTRODUCTION

The process of language acquisition has long been a point of uncertainty in research exploring the roots of human language. Particularly, regarding the open debate of nature versus nurture framework. The nature perspective postulates that language, like other aspects of human knowledge, is primarily determined by genetic factors (Chomsky, 1959; Fodor, 1985; Pinker, 1984), while the nurture viewpoint emphasizes the role of experience in language learning (Piaget, 1977; Tomasello, 2000). Thus, researchers have conducted extensive investigations to understand the initial state and process of language acquisition, providing insights into how environmental and genetic factors interact to fashion language and cognitive functions, and the mechanisms underlying brain plasticity (Barkat et al., 2011; Weaver et al., 2004; Werker & Hensch, 2015; Werker & Tees, 2005). It is now widely accepted that both genetic and experiential factors contribute to language development (Gervain & Mehler, 2010; Werker & Curtin, 2005), and researchers are interested in understanding how these factors interact during human development.

Infants at birth already exhibit advanced speech perception and language learning abilities. Newborns manifest a preference for speech over non-speech sounds (Vouloumanos & Werker, 2007), can discriminate between different languages based on their speech rhythms (Ramus et al., 2000), detect word boundaries (Christophe et al., 2001), discriminate words with different patterns of stress (Sansavini et al., 1997), or even distinguish consonant sounds (Cabrera & Gervain, 2020) and encode voice pitch (Arenillas-Alcón et al., 2021) in an adult-like manner. These findings suggest the impact that prenatal experiences may have on speech perception from an early age.

Interestingly, the prenatal period is not devoid of language experience. Hearing becomes operational and experiences most of its development around the 26th to 28th week of gestation (Anbuhl et al., 2016; Granier-Deferre et al., 2011; May et al., 2011; Moore & Linthicum, 2007; Ruben, 1995), allowing for exposure to the maternal speech signal in utero. Although the exact characteristics of prenatal language experience are not fully known, intrauterine recordings from animal models and simulations suggest that the maternal tissues act as a low-pass filter (Gerhardt & Abrams, 2000), attenuating around 30 dB for frequencies over 600-1000 Hz. Evidence indicates that this prenatal experience, despite attenuation and variations due to the womb, shapes speech perception and linguistic preferences of newborns, as shown by studies revealing that neonates are able to recognize a story heard frequently while in the womb (DeCasper & Spence, 1986), prefer the voice of their mother (DeCasper & Fifer, 1980) and prefer their native language (Moon et al., 1993). Additionally, recent findings suggest that prenatal learning extends beyond these common preferences, disclosing that infants acquire specific knowledge regarding the prosodic patterns inherent in their native language (Gervain, 2018).

Consequently, it has been proposed that language heard in utero has an impact on speech perception in newborns. Such impact is evidenced by the ability of neonates exposed to bilingual environments to differentiate between the two languages and exhibit equal preferences for each (Byers-Heinlein et al., 2010). In the present study we wonder whether, beyond their ability to recognize and differentiate one language from another, neonates exposed to more than one language during pregnancy show a superior and more accurate processing of speech sounds at birth compared to newborns who have only been exposed to one language.

In this regard, the frequency-following response (FFR) can provide insights into the underlying neural mechanisms associated with prenatal language experience, shedding light on how early linguistic exposure shapes the speech encoding abilities of newborns. The FFR is an auditory evoked potential elicited by periodic complex sounds that reflects neural synchronization with the auditory eliciting signals along the ascending auditory pathway (Krizman & Kraus, 2019; Skoe & Kraus, 2010), providing an accurate snapshot of the neural encoding of speech sounds. The interest in the neonatal FFR arises from its potential to serve as a predictive measure for language development in the future (Schochat et al., 2017), since alterations in FFR patterns in children have been associated with difficulties in reading and learning, dyslexia, impairments in phonological awareness or even autism (Banai et al., 2009; Basu et al., 2010; Chandrasekaran et al., 2009; Font-Alaminos et al., 2020; Hornickel et al., 2012; King et al., 2002; Lam et al., 2017; Otto-Meyer et al., 2018; Rosenthal, 2020). Interestingly, the FFR seems also to be able to reflect the impact of a wide range of auditory contexts, including training interventions, musical experience or bilingualism (Carcagno & Plack, 2017; Gorina-Careta et al., 2019; Kraus & Chandrasekaran, 2010; Krizman, Marian, et al., 2012; Krizman et al., 2015; Russo et al., 2005; Skoe et al., 2017; Song et al., 2008), or even the effects at birth of prenatal fetal auditory experiences such as music exposure (Arenillas-Alcón et al., 2023).

Therefore, the FFR to speech stimuli can be recorded in newborns to investigate whether a prenatal bilingual exposure leads to a more accurate processing of speech sounds at birth. Although the precise mechanisms underlying the neural specialization for speech processing and the influence of specific linguistic environment remain open questions, it is clear that experience plays a significant role in shaping the early development of language processing, even during the prenatal period. For this reason, the main objective of this study was to examine the neural encoding of speech stimuli at birth by recording the FFR in newborns who had been exposed to either monolingual or bilingual fetal environments during the prenatal period. This study would contribute to our knowledge of how prenatal language exposure may impact the processing of speech sounds at birth, helping to better understand the early stages of language acquisition and provide insights into populations at risk of language impairments (e.g., preterm babies, infants

born after fetal growth restriction), thereby paving the way for further research in this specific domain.

METHODS

Participants

A sample of 131 newborns was recruited from *SJD Barcelona Children's Hospital* in Barcelona (Spain), and divided into two groups according to the response of their mothers to the question "Did you communicate using more than one language during the last 3 months of pregnancy?" in a language usage questionnaire. Based on the collected responses, a total of 53 newborns were assigned to the group exposed to a monolingual fetal acoustic environment (MON; 27 females; mean gestational age = 39.93 ± 1.03 weeks; mean birth weight = 3321 ± 272 g) and 78 to the bilingual-exposed group (BIL; 34 females; mean gestational age = 39.71 ± 0.99 weeks; mean birth weight = 3327 ± 323 g). No significant differences are reported across groups in gestational age (U = 1868.500, p = 0.351) or birth weight (t = -0.104, p = 0.917). All neonates obtained Apgar scores higher than 8 at 1 and 5 min of life and passed adequately the universal newborn hearing screening (UNHS) prior to the recruitment. According to the recommendations of the Joint Committee of Infant Hearing (2019), newborns born from high-risk gestations, after obstetric pathologies or any other kind of risk factors related to hearing impairment were excluded from the recruitment.

Additionally, as performed in previous research from our laboratory (Arenillas-Alcón et al., 2021; Arenillas-Alcón et al., 2023; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021), both groups of newborns received a standard click-evoked auditory brainstem response (ABR) test to ensure the integrity of the auditory pathway. A click-stimulus, with a duration of 100 μs, was employed during the test, presented at a rate of 19.30 Hz with an intensity of 60 dB sound pressure level (SPL) until a total of 4000 artifact-free repetitions were collected. A prerequisite for participation in the experiment for all newborns was the successful identification of the wave V peak. This study counted with the approval of the Ethical Committee of Clinical Research (CEIC) of the Sant Joan de Déu Foundation (Approval ID: PIC-53-17), and required the mothers to fill out a sociodemographic questionnaire and to sign an informed consent prior to the participation, in line with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimulus

Neonatal FFRs were elicited by a two-vowel stimulus with a rising pitch ending (/oa/; Arenillas-Alcón et al., 2021). The /oa/ stimulus was created in *Praat* (Boersma & Weenink, 2020) and had a total length of 250 ms divided into three different sections, depending on its

fundamental frequency (F_0) as well as on its formant content (/o/ section: 10-80 ms, F_0 = 113 Hz, F_1 = 452 Hz, F_2 = 791 Hz; /a/ steady section = 90-160 ms, F_0 = 113 Hz, F_1 = 678 Hz, F_2 = 1017 Hz; /a/ rising section = 160-250 ms, F_0 = 113-154 Hz, F_1 = 678 Hz, F_2 = 1017 Hz; Fig. 1A). Considering the stimulus formant content, we focused our analyses exclusively on the spectral peaks that corresponded to F_1 frequencies. This decision was made because the spectral peaks associated with F_2 frequencies belong to a significant high frequency range, resulting in elicited neural responses relatively weak and challenging to be accurately observed in newborns. The /oa/ stimulus was presented at a rate of 3.39 Hz in alternating polarities, and delivered monaurally to the right ear at 60 dB SPL of intensity with an earphone connected to a Flexicoupler disposable adaptor (Natus Medical Incorporated, San Carlos, CA).

Procedure and data acquisition

After the successful completion of the UNHS, neonates were tested at the hospital room while they were sleeping in their bassinet. Four disposable Ag/AgCl electrodes were placed in a vertical montage configuration (active at Fpz, ground at forehead, one reference at each mastoid and online referenced to the right mastoid; as shown in Fig. 1B) ensuring impedances below 7 kΩ. For the purpose of analyses in the current study, only data referenced to the electrode at the right mastoid (ipsilateral to the auditory stimulation) was considered. The presentation of click and speech stimuli was done by using a *SmartEP* platform connected to a *Duet* amplifier, which incorporated the *cABR* and the *Advanced Hearing Research* modules (Intelligent Hearing Systems, Miami, Fl, USA).

Following established protocols from previous newborn studies conducted in our laboratory, the experimental procedure involved the recording of two blocks of click stimuli, followed by four blocks of 1000 artifact-free responses to the /oa/ stimulus. Any electrical activity surpassing ±30 μV threshold was automatically rejected until a total of 4000 presentations was collected. The total mean duration of the recording session was approximately 25 min [2 click blocks x 2000 repetitions x 51.81 ms SOA + 4 /oa/ blocks x 1000 repetitions x 295 ms of stimulus-onset asynchrony (SOA)] including the duration of rejected sweeps. The continuous electroencephalography signal was acquired at a sampling rate of 13333 Hz with an online bandpass filter with cutoff frequencies from 30 to 1500Hz, and online epoched and averaged from -40.95 ms (pre-stimulus period) to 249.975 ms.

Data processing and analysis

An offline bandpass filter with cutoff frequencies from 80 to 1500 Hz was applied. In order to highlight the encoding of the stimulus fundamental frequency (F_0) and to reduce the contribution of cochlear microphonics, neural responses were averaged by combining the two opposite stimulus polarities [(Condensation + Rarefaction)/2], obtaining the envelope-following

response (FFR_{ENV}). To accurately assess the representation of formant structure, it was essential to emphasize the FFR components associated with the encoding of the temporal fine structure, such as the stimulus first formant (F_1), while reducing the impact of envelope-related activity. To achieve this, the neural responses with alternating polarities were subtracted [(Condensation - Rarefaction)/2], yielding the temporal fine structure-following response (FFR_{TFS}; Aiken & Picton, 2008; Krizman & Kraus, 2019). Detailed information regarding the analyzed parameters from the neonatal FFR can be found below. All parameters were computed using custom scripts in Matlab R2019b (The Mathworks, 2019), developed in our laboratory and previously employed in similar analyses in former studies.

Neural lag. Neural lag served as an indicator of the neural transmission delay within the auditory system, and was assessed to estimate the time passed from cochlear stimulus reception to the onset of neural phase-locking (Arenillas-Alcón et al., 2021; Arenillas-Alcón et al., 2023; Jeng et al., 2010; Liu et al., 2015; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021). To calculate the neural lag, a cross-correlation analysis was computed between the auditory stimulus and the neural response, within a time window of 3-13 ms, generating a correlation value for each time lag. The neural lag was determined by identifying the time lag corresponding to the highest cross-correlation value.

Pre-stimulus root mean square (RMS) amplitude. The RMS of the pre-stimulus period was employed as a measure of the general magnitude of neural activity over time, and to dismiss electrophysiological disparities in the pre-stimulus region (Arenillas-Alcón et al., 2023; Liu et al., 2015; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021; White-Schwoch et al., 2015). This measure was computed by squaring each data point within the pre-stimulus region of the neural response (from -40 to 0 ms), calculating the mean of the squared values and subsequently obtaining the square root of the resulting average.

Voice pitch encoding from FFR_{ENV}

Spectral amplitude at F_0. Spectral amplitude at F_0 (113 Hz) was used as a quantitative measure of the neural phase-locking strength at the specific frequency of interest (Arenillas-Alcón et al., 2021; Arenillas-Alcón et al., 2023; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021; White-Schwoch et al., 2015). It was computed by applying a fast Fourier transform (FFT; Cooley & Tukey, 1965) to obtain the frequency spectrum of the neural response, and then calculating the average amplitude within a ±5 Hz window centered around the peak of the stimulus F_0.

Signal-to-noise ratio at F_0. The estimation of the relative spectral magnitude of the response was determined by calculating the signal-to-noise ratio (SNR) at the peak frequency of interest. This measure attended not only to the amplitude value at the specific frequency peak (113 Hz) but also to the surrounded frequencies. To obtain the SNR, the mean amplitude within

a ±5 Hz frequency window centered at the peak of the frequency of interest was divided by the mean amplitude within two additional 28 Hz wide frequency windows (flanks), which were centered at ±19 Hz from the frequency of interest. For instance, in the case of the F_0 = 113 Hz, the SNR was calculated by dividing the mean amplitude from 108 Hz to 118 Hz by the average of the mean amplitude from 80 Hz to 108 Hz and the mean amplitude from 118 Hz to 146 Hz.

Formant structure encoding from FFR_{TFS}

Spectral amplitudes at F_1 peaks. To assess spectral amplitudes at the specific spectral peaks regarding the stimulus F_1 frequencies (452 Hz [/o/] and 678 Hz [/a/]), the neural responses corresponding to the /o/ section (10–80 ms time window) and the /a/ steady section (90-160 ms time window) were individually analyzed and the respective amplitudes were extracted.

SNRs at F_1 peaks. The method for calculating the SNR at F_1 peaks was identical to that explained above for the F_0. Thus, SNRs at spectral peaks that correspond to the stimulus F1 frequencies (452 Hz [/o/] and 678 Hz [/a/]) were likewise obtained separately from the neural responses to the /o/ and the /a/ steady sections.

Statistical analysis

Statistical analyses were conducted using SPSS 25.0 (Corp., n.d.). Descriptive statistics were calculated, including the mean, standard deviation (SD), median, first quartile (Q_1), third quartile (Q_3), interquartile range (IQR), and minimum and maximum values, for each computed parameter within the two groups of newborns (MON; BIL). The Kolmogorov-Smirnov test was used to assess the normal distribution of the samples. Depending on the normality of the data, two-tailed independent samples *t*-tests or *Mann-Whitney U* tests were conducted to evaluate significant differences between groups, with Cohen's *d* being reported as the effect size. Results were considered statistically significant when $p < 0.05$.

RESULTS

Frequency following responses (FFR) elicited by a two-vowel speech stimulus /oa/ (Fig. 1A) were collected from a total sample of 131 newborns divided into two groups according to their prenatal fetal exposure to a monolingual (MON) or bilingual (BIL) environment. To comprehensively evaluate the neonates' ability to encode the pitch and vowel formant structure of speech sounds, the neural responses to the fundamental frequency (F_0) and the vowels' first formant (F_1) were analyzed taking into account the distinct sound characteristics of the different stimulus sections. All detailed descriptive statistics from the parameters analyzed can be found in Table 1.

Table 1. Descriptive statistics for MON (n=53) **and BIL** (n=78) **groups in FFR parameters**: neural lag, Root-Mean-Square from pre-stimulus section, spectral amplitude at F_0 and F_1 peaks, Signal-to-noise ratio at F_0 and F_1 peaks.

Measure	Mean	SD	Median	Q_1	Q_3	IQR	Minimum	Maximum
Neural lag (ms)								
Monolingual	7.913	1.233	7.950	7.238	8.738	1.500	4.200	11.025
Bilingual	8.016	1.394	7.800	6.975	8.850	1.875	5.325	12.975
Pre-stimulus RMS (nV)								
Monolingual	30.186	15.723	26.945	18.821	35.969	17.148	13.533	88.444
Bilingual	29.851	11.921	29.009	21.013	35.727	14.714	13.556	71.361
F_0 Spectral Amplitude (nV)								
Monolingual	10.015	5.184	9.042	6.495	12.270	5.775	1.503	26.110
Bilingual	9.186	4.924	7.709	6.063	11.236	5.173	2.568	23.960
SNR F_0								
Monolingual	4.607	3.512	5.245	2.589	7.377	4.788	-4.351	10.770
Bilingual	3.200	3.741	3.967	1.294	5.504	4.210	-6.739	10.224
F_1 Spectral Amplitude /o/ section at 452 Hz (nV)								
Monolingual	3.537	5.375	2.564	1.093	3.701	2.608	0.118	35.135
Bilingual	1.975	1.519	1.533	0.990	2.491	1.501	0.369	9.263
F_1 Spectral Amplitude /a/ steady section at 452 Hz (nV)								
Monolingual	2.118	1.373	1.863	1.354	2.619	1.265	0.279	7.869
Bilingual	1.866	1.314	1.731	0.946	2.434	1.488	0.242	9.405
F_1 Spectral Amplitude /o/ section at 678 Hz (nV)								
Monolingual	0.819	0.963	0.599	0.388	0.937	0.549	0.080	6.424
Bilingual	0.645	0.542	0.557	0.372	0.743	0.371	0.120	4.481
F_1 Spectral Amplitude /a/ steady section at 678 Hz (nV)								
Monolingual	2.519	5.948	1.063	0.555	1.918	1.363	0.167	38.348
Bilingual	0.929	0.726	0.759	0.603	1.353	0.750	0.061	5.068
SNR /o/ section at 452 Hz								
Monolingual	0.516	5.435	1.928	-1.678	4.406	6.084	-18.893	6.388
Bilingual	0.599	3.643	1.262	-1.514	3.772	5.286	-9.345	6.043
SNR /a/ steady section at 452 Hz								
Monolingual	0.913	3.448	1.231	-0.718	3.214	3.932	-8.834	6.078
Bilingual	0.161	4.116	1.360	-2.556	3.284	5.840	-13.965	6.196
SNR /o/ section at 678 Hz								
Monolingual	-1.112	4.395	-0.174	-4.008	2.002	6.010	-14.753	6.236
Bilingual	-1.008	3.538	-0.748	-2.852	1.766	4.618	-10.866	4.321
SNR /a/ steady section at 678 Hz								
Monolingual	1.431	4.305	2.744	-0.975	4.597	5.572	-14.170	6.660
Bilingual	0.142	4.792	1.802	-2.724	3.661	6.385	-16.269	5.982

SD = standard deviation. Q_1 = first quartile (25th percentile). Q_3 = third quartile (75th percentile). IQR = interquartile range.

Neural transmission delay. No significant differences were found across groups in neural lag parameter ($t_{(129)}$ = -0.435, p = 0.664, Cohen's d = 0.078).

Pre-stimulus root mean square (RMS) amplitude. There were no statistically significant differences observed between the groups in relation to the background neural activity prior to the auditory stimulation ($U_{(129)}$ = 2184.000, p = 0.583, Cohen's d = 0.024).

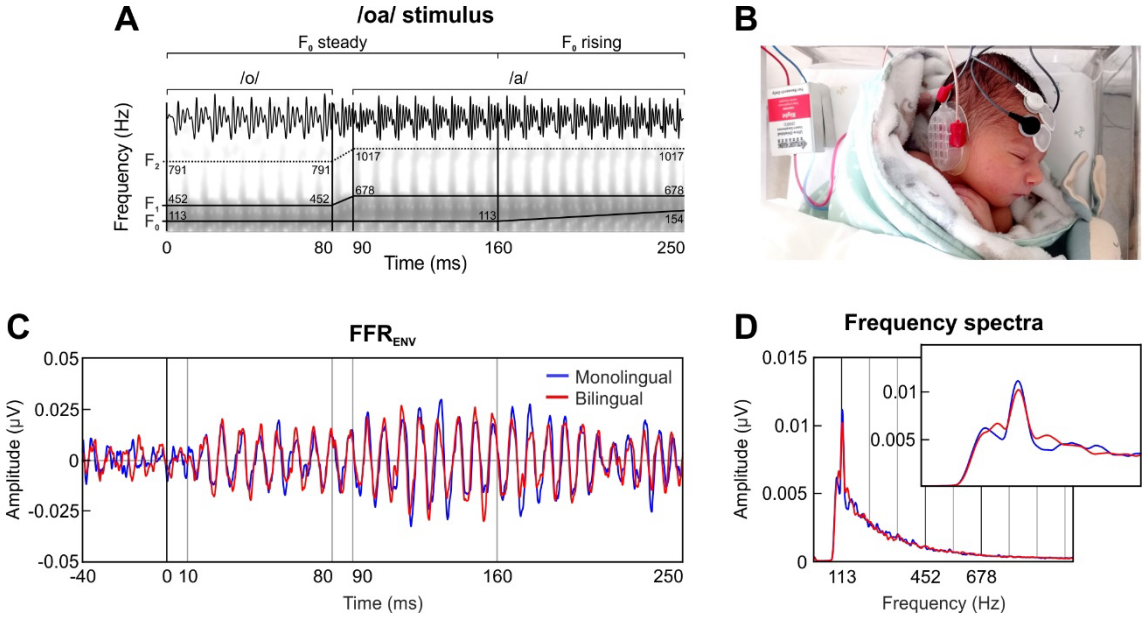

Fig. 1. A: Temporal and spectral representation of the two-vowel auditory stimulus /oa/, with a schematic representation of its fundamental frequency (F_0) and formant structure (F_1, F_2). **B:** Recording setup of the four disposable electrodes placed in a vertical montage (active located at Fpz, ground at forehead, references at each mastoid). Reproduced with the written consent of the neonate's parents. **C:** Grand-averaged waveform of the FFR_{ENV} in the time domain, retrieved separately for the group exposed to a monolingual fetal acoustic environment (blue) and the bilingual-exposed group (red). **D:** Frequency spectra of the FFR_{ENV} extracted from the steady pitch section of the stimulus (10-160 ms). The inset zooms in a narrower frequency band to illustrate the effect around the F_0 peak.

Voice pitch encoding from FFR_{ENV}

The grand-averaged FFR_{ENV} waveform for each group is illustrated in Fig. 1C. To assess the robustness of the voice pitch representation, the /oa/ stimulus incorporated an initial section (10-160 ms) with a steady fundamental frequency (F_0) of 113 Hz.

Spectral amplitude at F_0. The grand-averaged spectral representation of the neonatal FFR extracted from each group is depicted in Fig. 1D. Considering the encoding of the F_0 in the steady pitch section of the stimulus, no differences were found across groups ($U_{(129)}$ = 1809.000, p = 0.226, Cohen's d = 0.164).

Signal-to-noise ratio at F_0. Statistical analyses revealed significant differences between groups, indicating that newborns exposed to a monolingual prenatal fetal environment exhibited significantly larger SNR values as compared to the bilingual exposed neonates ($t_{(129)} = 2.166$, $p = 0.032$, Cohen's $d = 0.388$).

Formant structure encoding from FFR_{TFS}

The grand-averaged FFR_{TFS} waveform for each group is shown in Fig. 2A. To evaluate the newborns' ability to encode the formant structure of speech sounds, the /oa/ stimulus included two sections with the same voice pitch but different fine-structure. Specifically, the /o/ section (10–80 ms), which was characterized by a F_1 of 452 Hz, and the /a/ steady section (90–160 ms), which was characterized by a F_1 frequency of 678 Hz.

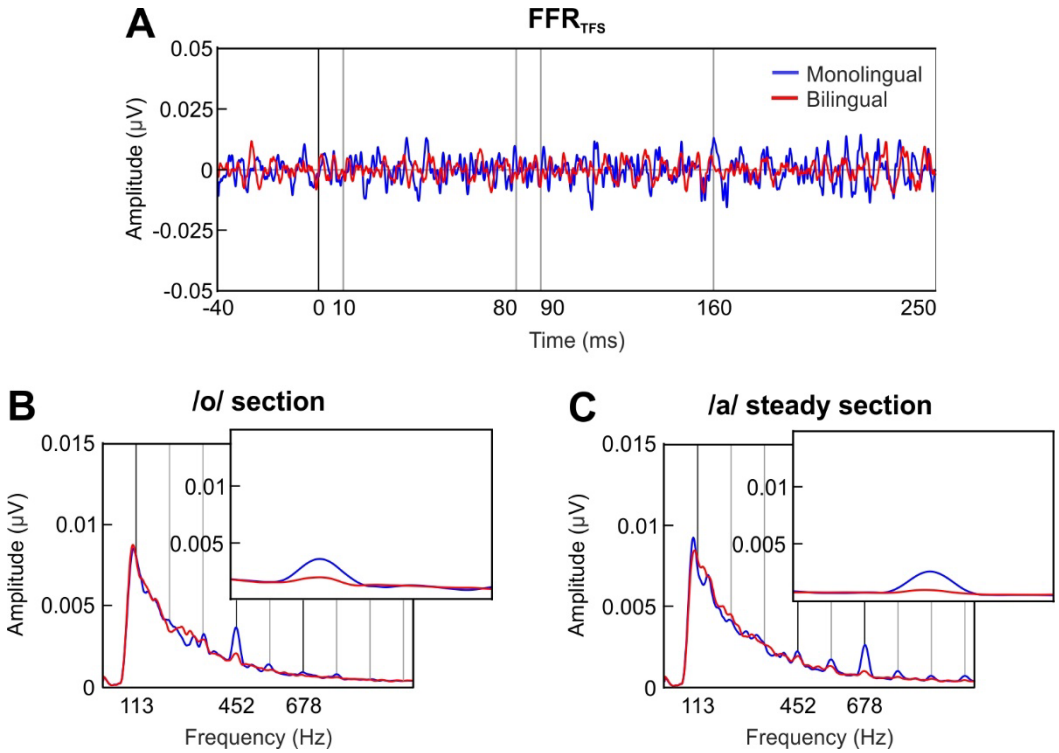

Fig. 2. Formant structure encoding. A: Grand-averaged waveform of the FFR_{TFS} in the time domain, retrieved separately for the group exposed to a monolingual fetal acoustic environment (blue) and the bilingual-exposed group (red). **B:** Frequency spectra of the FFR_{TFS} extracted from the /o/ section of the stimulus (10-80 ms). The inset zooms in a narrower frequency band to illustrate the effect around the /o/ F_1 peak (452 Hz) during the /o/ section. **C:** Frequency spectra of the FFR_{TFS} extracted from the /a/ steady section of the stimulus (90-160 ms). The inset zooms in a narrower frequency band to illustrate the effect around the /a/ F_1 peak (678 Hz) during the /a/ steady section.

Spectral amplitudes at F_1 peaks. The grand-averages of the FFR_{TFS} spectral amplitudes during the /o/ section (10-80 ms) are illustrated in Fig 2B for each group separately, while the spectral representations during the /a/ steady section (90-160 ms) are depicted in Fig. 2C. During the /o/ section, monolingual exposed neonates showed significantly larger spectral amplitudes at

its corresponding frequency (/o/ F_1 = 452 Hz; $U_{(129)}$ = 1560.000, p = 0.017, Cohen's d = 0.396) as compared to the bilingual newborns. However, no significant differences were found across groups for the /a/ F_1 (678 Hz) during the same /o/ section ($U_{(129)}$ = 1867.000, p = 0.348, Cohen's d = 0.223). The same tendency was observed when considering the encoding of the /a/ steady section: monolingual exposed newborns exhibited higher spectral amplitudes at its corresponding frequency (/a/ F_1 = 678 Hz; $U_{(129)}$ = 1586.000, p = 0.024, Cohen's d = 0.375) as compared to the bilinguals; likewise, no significant differences were found across groups when encoding the /o/ F_1 (452 Hz) during the /a/ section ($U_{(129)}$ = 1839.000, p = 0.285, Cohen's d = 0.187).

SNRs at F_1 peaks. No significant differences between newborns exposed to monolingual and bilingual environments were found in SNR values for any of F_1 peaks (452 Hz; 678 Hz) during any of the two vowel sections (/o/ vowel; /a/ steady section). Thus, statistical analyses revealed no differences during the /o/ section (10-80 ms) for neither the /o/ F_1 (452 Hz; $U_{(129)}$ = 1885.000, p = 0.393, Cohen's d = 0.018) nor the /a/ F_1 (678 Hz; $U_{(129)}$ = 2088.000, p = 0.922, Cohen's d = 0.026). As well, no differences were found during the /a/ steady section (90-160 ms) neither for the /o/ F_1 ($U_{(129)}$ = 1919.000, p = 0.0.488, Cohen's d = 0.198) nor the /a/ F_1 ($U_{(129)}$ = 1708.000, p = 0.092, Cohen's d = 0.283).

DISCUSSION

The present study investigated the impact of prenatal exposure to a bilingual fetal environment on the neural encoding of a speech syllable in neonates, as indexed by the frequency-following response (FFR). A total sample of 131 healthy-term newborns was divided into two groups according to their monolingual or bilingual prenatal exposure, as reported by their mothers through a questionnaire. The FFRs elicited to the two-vowel speech stimulus /oa/ (Arenillas-Alcón et al., 2021) were recorded in order to evaluate the neural responses to the stimulus' fundamental frequency (F_0 = 113 Hz; related to voice pitch encoding) and the first formant of each vowel (F_1 = 452 Hz for /o/ vowel; F_1 = 678 Hz for /a/ steady vowel; related to formant structure encoding).

While our results did not reveal significant group differences in neural lag and pre-stimulus root mean square (RMS) amplitude, implying comparable neural transmission delays and absence of a distinct background neural activity, our data did reveal significant differences in the encoding of voice pitch (F_0) and formant structure (F_1) of speech sounds. Notably, and contrary to our hypothesis, the monolingual-exposed group exhibited a greater response in both cases. These findings contribute novel insights into the effects of prenatal language exposure on the neural encoding of speech sounds at birth.

Specifically, considering the encoding of voice pitch, we found significant differences in the signal-to-noise ratio (SNR) at F_0 between the groups, with neonates exposed to a monolingual environment showing higher SNR values, but no differences in the spectral amplitudes at F_0. In other words, although there was a clearer and less noisy encoding of voice pitch in the monolingual exposed newborns, both groups exhibited similar overall F_0 amplitudes in the neural response. This supports the idea that neural mechanisms underlying voice pitch encoding are already mature at birth, pointing out the crucial role of pitch for the perception of both speech and non-speech sounds, for conveying prosodic information, facilitating speaker recognition and speech segmentation, accelerating phoneme acquisition in tonal languages, helping with language comprehension in noisy environments or even contributing with the emotional state perception in a conversation (Arenillas-Alcón et al., 2021; Benavides-Varela et al., 2012; Cabrera & Gervain, 2020; Gervain, 2018; Musacchia et al., 2007; Partanen, Kujala, Näätänen, et al., 2013; Plack et al., 2014; Ribas-Prats et al., 2021; Rosen, 1992; Skuk & Schweinberger, 2014). The more specific encoding of pitch in the monolingual-exposed group may be due to a more intensive exposure to an exclusive and single language during the fetal period, as it is known that the ability to discriminate different F_0 or different F_0 contours drastically improves with practice and exposure (Carcagno & Plack, 2017). In contrast, newborns exposed to a bilingual environment may experience more variable acoustic conditions during the pregnancy period, since studies have shown that bilingual individuals, particularly females, exhibit different pitch frequency ranges depending on the language they speak (Ordin & Mennen, 2017). In fact, not only different languages have distinct overall height pitch levels [e.g., Polish was found to have a higher pitch compared to American English (Majewski et al., 1972); Mandarin, a higher pitch than English (Keating & Kuo, 2012); Japanese, a higher pitch than Dutch (Van Bezooijen, 1995); or Slavic languages, a higher pitch than Germanic ones (Andreeva et al., 2014)], but even speakers from two phonologically similar dialects exhibit differences in their height pitch levels (e.g., two different dialects of Mandarin; Deutsch et al., 2009), and, importantly, the same bilingual individual, especially females, exhibit different pitch height depending on the language spoken (Ordin & Mennen, 2017). These findings highlight the greater variability of acoustic speech inputs to which the fetus of bilingual mothers would be exposed.

In line with this, the spectral amplitude at F_1 peaks revealed interesting patterns, with monolingual-exposed neonates exhibiting higher spectral amplitudes for the corresponding frequencies during both the /o/ and /a/ steady sections. This indicates a more robust encoding of the formant structure (FFR_{TFS}) of neonates exposed to a prenatal monolingual environment compared to the bilingual-exposed ones. These findings may seem contrary to previous research supporting enhanced auditory processing in bilingual individuals (Krizman, Marian, et al., 2012; Krizman et al., 2015). However, it is important to consider that previous studies typically involve

bilingual infants who were several months old, whereas our study focuses on newborn infants. The age difference in participants may account for the divergence in results. Furthermore, our findings can potentially be explained by the nature of bilingual environments, which are often characterized as more demanding, dynamic and phonologically rich, thus requiring strengthened attention to all linguistic stimuli (Krizman, Marian, et al., 2012). With continued exposure to these complex linguistic contexts, the auditory system gradually becomes finely tuned to process sound more efficiently (Krizman, Marian, et al., 2012), explaining the enhanced flexibility and encoding abilities reported in bilingual individuals. Therefore, the disparities in our findings may be attributed to the specific age group studied and the unique demands of monolingual versus bilingual environments: whilst the exposure of the fetus to a single language would produce more robust and larger amplitude responses, the auditory system of a bilingually exposed newborn would show a major sensitivity to a wider range of frequencies without yet generating a particularly strong response at any of them. A monolingual environment would offer a more consistent and less complex phonetic condition than a bilingual environment, possibly leading to a more effective and accurate encoding of the specific vowel sound characteristics. Further investigation into the developmental trajectories of auditory processing in different populations of newborns, with different prenatal auditory experiences, may shed more light on this issue.

Nonetheless, overall these findings emphasize the potential importance of prenatal linguistic exposure in shaping the neural mechanisms underlying language acquisition and highlight the sensitivity of the FFR in capturing these subtle changes. The results add to a growing body of research that suggests a role for prenatal fetal experiences in molding language development (Arenillas-Alcón et al., 2023; Gervain, 2015, 2018; Moon et al., 2012; Partanen, Kujala, Tervaniemi, et al., 2013).

The potential implications of our study might extend to the design of intervention strategies in cases of pregnancies at risk for developmental language-related disorders (e.g., in babies born after fetal growth restriction (FGR), who appear to exhibit attenuated F_0 encoding at birth; Ribas-Prats et al., 2021), in which prenatal language exposure could be a risk or a protective factor. By better understanding the prenatal factors that influence neural speech encoding, interventions could potentially be designed to optimize the linguistic environment even before birth.

Even so, while our study has yielded valuable insights into the effects of prenatal language exposure on speech encoding in newborns, it is important to acknowledge some limitations that need further consideration. By understanding these limitations, we can better contextualize the findings and identify avenues for future research. Thus, while we sought to reduce potential confounding factors, it remains challenging to isolate the specific effects of monolingual versus

bilingual prenatal fetal exposure, given the multiple genetic and environmental factors at play. Despite having a large sample size, the diversity inherent to the newborns can lead to significant variability in FFR responses (Richard et al., 2020), making it a challenge to isolate the effects of bilingual exposure. Furthermore, the in-utero acoustic environment is complex and influenced by many factors other than maternal speech, such as internal sounds from the mother's body and external noises (Gerhardt & Abrams, 2000).

Particularly, and having already discarded significant differences between groups in gestational age (p = 0.351) and birth weight (p = 0.917), we accounted for prenatal musical exposure, considering previous research demonstrating its significant impact on speech encoding at birth (Arenillas-Alcón et al., 2023; Partanen, Kujala, Tervaniemi, et al., 2013; Partanen et al., 2022). Taking into account parameters reflecting F_0 (113 Hz) and F_1 (452 Hz; 678 Hz) encoding and the different levels of exposure to music during the pregnancy period, the chi-squared test (χ^2) conducted revealed no significant differences across groups in those parameters depending on the musical exposure (χ^2 = 2.019, p = 0.364). In addition to prenatal music exposure, and although sex differences are not clear in relation to speech encoding accuracy (females seem to have a greater representation of the higher frequencies up to 1130 Hz than males, but no differences until 720 Hz; Krizman et al., 2019; Krizman, Skoe, et al., 2012), we also discarded significant differences across groups in the sex of the newborns (χ^2 = 0.686, p = 0.408). Finally, our analyses did not reveal differences in the maternal educational level (χ^2 = 2.368, p = 0.668), an element closely tied to the linguistic environment a fetus is exposed to, which in turn may affect their language acquisition and development (Hoff, 2003; Rowe, 2008). Socioeconomic level could potentially constrain the complexity and richness of language exposure, and its influence on speech encoding in newborns cannot be disregarded.

Yet, despite the lack of differences across groups in these factors, the multifactorial nature of human development makes it difficult to completely eliminate their potential confounding effects. Therefore, our findings should be interpreted within this context and acknowledging these limitations. Longitudinal studies following the same individuals from birth into childhood and adulthood are needed to better understand the long-term impacts of prenatal bilingual exposure on speech processing and language development. Besides, our study provides a valuable basis, although further research is necessary to improve our understanding of the specific effects of fetal language exposure, and to disentangle the complexity of genetic, environmental, and experiential factors.

CONCLUSION

Overall, the present study contributes significant insights into the impact of prenatal bilingual exposure on the neural encoding of speech sounds at birth, thereby increasing our

knowledge of the early stages of language development. While no significant group differences were observed in parameters indexing neural transmission delay nor magnitude of neural activity prior to the auditory stimulation, we did find notable differences in the encoding of voice pitch and formant structure, with the monolingual-exposed group exhibiting greater responses in both cases. These findings suggest that intensive exposure to a single language during the fetal period may result in more precise and accurate encoding of speech characteristics. Our study highlights the remarkable plasticity and learning potential of the human brain even before birth, emphasizing the complex interaction between genetic and environmental factors in shaping our cognitive abilities and linguistic development. Further research exploring this field may provide specific strategies and interventions suitable for those at risk of language-related developmental disorders, in order to optimize the linguistic environment even before birth.

AUTHOR CONTRIBUTIONS

S.A.A.: Conceptualization, Methodology, Investigation, Formal analysis, Writing-Original draft, Writing-Review & Editing, Visualization. **N.G.C:** Conceptualization, Methodology, Investigation, Formal analysis. **M.P.:** Methodology, Investigation. **A.M.S.:** Methodology, Investigation. **S.I.K.:** Methodology, Investigation. **J.C.F.:** Conceptualization, Methodology, Writing-Review & Editing, Supervision. **M.D.G.R.:** Funding acquisition, Resources. **C.E.:** Conceptualization, Methodology, Writing-Review & Editing, Funding acquisition, Resources, Supervision.

REFERENCES

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1–2), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Anbuhl, K. L., Uhler, K. M., Werner, L. A., & Tollin, D. J. (2016). Early Development of the Human Auditory System. In R. A. Polin, S. H. Abman, D. Rowitch, & W. E. Benitz (Eds.), *Fetal and Neonatal Physiology* (5th ed., pp. 1396–1410). Elsevier. https://doi.org/10.1016/B978-0-323-35214-7.00138-4

Andreeva, B., Demenko, G., Möbius, B., Zimmerer, F., Jügler, J., & Jastrzebska, M. (2014). Differences of pitch profiles in Germanic and Slavic languages. In *Proceedings of Interspeech*. ISCA.

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Scientific Reports*, *11*, 6660. https://doi.org/10.1038/s41598-021-85799-x

Arenillas-Alcón, S., Ribas-Prats, T., Puertollano, M., Mondéjar-Segovia, A., Gómez-Roig, M. D., Costa-Faidella, J., & Escera, C. (2023). Prenatal daily musical exposure is associated with enhanced neural representation of speech fundamental frequency: Evidence from neonatal frequency-following responses. *Developmental Science*, e13362.

https://doi.org/10.1111/desc.13362

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and Subcortical Auditory Function. *Cerebral Cortex*, *19*, 2699–2707. https://doi.org/10.1093/cercor/bhp024

Barkat, T., Polley, D., & Hensch, T. (2011). A critical period for auditory thalamocortical connectivity. *Nature Neuroscience*, *14*, 1189–1194.

Basu, M., Krishnan, A., & Weber-Fox, C. (2010). Brainstem correlates of temporal auditory processing in children with specific language impairment. *Developmental Science*, *13*(1), 77–91. https://doi.org/10.1111/j.1467-7687.2009.00849.x

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17908–17913. https://doi.org/10.1073/pnas.1205413109

Boersma, P., & Weenink, D. (2020). *Praat: doing phonetics by computer* (Version 6.1.09; p. Version 6.1.09). http://www.praat.org/

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science*, *21*(3), 343–348. https://doi.org/10.1177/0956797609360758

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances*, *6*(30), eaba7830. https://doi.org/10.1126/sciadv.aba7830

Carcagno, S., & Plack, C. J. (2017). Short-Term Learning and Memory: Training and Perceptual Learning. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (Vol. 61, pp. 75–100). Springer Nature.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, *64*(3), 311–319. https://doi.org/10.1016/j.neuron.2009.10.006

Chomsky, N. (1959). A review of B.F. Skinner's verbal behavior. *Language*, *35*, 26–58.

Christophe, A., Mehler, J., & Sebastian-Galles, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*, 358–394.

Cooley, J. W., & Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, *19*(90), 297.

Corp., I. (n.d.). *SPSS 25.0*.

DeCasper, A. J., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voice. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Deutsch, D., Le, J., Shen, J., & Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *The Journal of the Acoustical Society of America*, *125*, EL208–EL213.

Fodor, J. A. (1985). Precis of the modularity of mind. *Behavioral and Brain Sciences*, *8*, 1–42.

Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., & Escera, C. (2020). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biological Psychology*, *149*, 107807. https://doi.org/10.1016/j.biopsycho.2019.107807

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal Exposures to Sound and Vibroacoustic Stimulation. *Journal of Perinatology*, *20*, 20–29.

Gervain, J. (2015). Plasticity in early language acquisition: the effects of prenatal and early childhood experience. *Current Opinion in Neurobiology*, *35*, 13–20. https://doi.org/10.1016/j.conb.2015.05.004

Gervain, J. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences*, *21*, 62–67. https://doi.org/10.1016/j.cobeha.2018.02.004

Gervain, J., & Mehler, J. (2010). Speech Perception and Language Acquisition in the First Year of Life. *Annual Review of Psychology*, *61*, 191–218. https://doi.org/10.1146/annurev.psych.093008.100408

Gorina-Careta, N., Ribas-Prats, T., Costa-Faidella, J., & Escera, C. (2019). Auditory Frequency-Following Responses. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational Neuroscience* (pp. 1–13). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-7320-6_100689-1

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., & Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *NeuroReport*, *23*(1), 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.-D., & Chung, H.-K. (2010). Early Maturation of Frequency-Following Responses to Voice Pitch in Infants with Normal Hearing. *Perceptual and Motor Skills*, *111*(3), 765–784. https://doi.org/10.2466/10.22.24.PMS.111.6.765-784

Joint Committee on Infant Hearing. (2019). Year 2019 Position Statement: Principles and Guidelines for Early Hearing Detection and Intervention Programs. *Journal of Early Hearing Detection and Intervention*, *4*(2), 1–44. https://doi.org/10.1542/peds.106.4.798

Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, *132*, 1050–1060.

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, *319*, 111–115. https://doi.org/10.1016/S0304-3940(01)02556-3

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*, 599–605. https://doi.org/10.1038/nrn2882

Krizman, J., Bonacina, S., & Kraus, N. (2019). Sex differences in subcortical auditory processing

emerge across development. *Hearing Research*, *380*, 166–174. https://doi.org/10.1016/j.heares.2019.07.002

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research*, *382*, 166–174. https://doi.org/10.1016/j.heares.2019.107779

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *PNAS*, *109*(20), 7877–7881. https://doi.org/10.1073/pnas.1201575109

Krizman, J., Skoe, E., & Kraus, N. (2012). Sex differences in auditory subcortical function. *Clinical Neurophysiology*, *123*(3), 590–597. https://doi.org/10.1016/j.clinph.2011.07.037

Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015). Neural processing of speech in children is influenced by bilingual experience. *Neuroscience Letters*, *0*, 48–53. https://doi.org/10.1016/j.neulet.2014.11.011.Neural

Lam, S. S.-Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia*, *103*, 162–167. https://doi.org/10.1016/j.physbeh.2017.03.040

Liu, F., Maggu, A. R., Lau, J. C. Y., & Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Frontiers in Human Neuroscience*, *8*, 1–19. https://doi.org/10.3389/fnhum.2014.01029

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica*, *25*, 119–125.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech? *Frontiers in Psychology*, *2*(222), 1–9. https://doi.org/10.3389/fpsyg.2011.00222

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.

Moon, C., Lagercrantz, H., & Kuhl, P. K. (2012). Language experienced in utero affects vowel perception after birth: a two-country study. *Acta Paedriatrica*, *102*, 156–160. https://doi.org/10.1111/apa.12098

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, *46*(9), 460–478. https://doi.org/10.1080/14992020701383019

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Ordin, M., & Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech, Language, and Hearing Research*, *60*, 1493–1506. https://doi.org/10.1044/2016_JSLHR-S-16-0315

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*, 733–743. https://doi.org/10.1007/s00221-017-5164-4

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013).

Learning-induced neural plasticity of speech processing before birth. *PNAS*, *110*(37), 15145–15150. https://doi.org/10.1073/pnas.1302159110

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. *PLoS ONE*, *8*(10), e78946. https://doi.org/10.1371/journal.pone.0078946

Partanen, E., Mårtensson, G., Hugoson, P., Huotilainen, M., Fellman, V., & Ådén, U. (2022). Auditory Processing of the Brain Is Enhanced by Parental Singing for Preterm Infants. *Frontiers in Neural Circuits*, *16*, 772008. https://doi.org/10.3389/fnins.2022.772008

Piaget, J. (1977). The language and thought of the child. In *The Essential Piaget* (pp. 66–83). Basic Books Inc.

Pinker, S. (1984). *Language Learnability and Language Development*. Harvard University Press.

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, *307*, 53–64. https://doi.org/10.1016/j.heares.2013.07.020

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, *288*, 349–351.

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28–39. https://doi.org/10.1016/j.heares.2018.11.001

Ribas-Prats, T., Arenillas-Alcón, S., Pérez-Cruz, M., Costa-Faidella, J., Gómez-Roig, M. D., & Escera, C. (2023). Speech-encoding deficits in neonates born large-for-gestational age as revealed with the frequency-following response. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000001330

Ribas-Prats, T., Arenillas-Alcón, S., Lip-Sosa, D. L., Costa-Faidella, J., Mazarico, E., Gómez-Roig, M. D., & Escera, C. (2021). Deficient neural encoding of speech sounds in term neonates born after fetal growth restriction. *Developmental Science*, 1–15. https://doi.org/10.1111/desc.13189

Richard, C., Neel, M., Jeanvoine, A., Connell, S., Gehred, A., & Maitre, N. L. (2020). Characteristics of the Frequency-Following Response to Speech in Neonates and Potential Applicability in Clinical Practice: A Systematic Review. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1618–1635.

Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions: Biological Sciences*, *336*, 367–373. https://doi.org/10.1098/rstb.1992.0070

Rosenthal, M. A. (2020). A systematic review of the voice-tagging hypothesis of speech-in-noise perception. *Neuropsychologia*, *136*(November 2019), 107256. https://doi.org/10.1016/j.neuropsychologia.2019.107256

Rowe, M. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

Ruben, R. J. (1995). The ontogeny of human hearing. *International Journal of Pediatric Otorhinolaryngology*, *32*, 199–204. https://doi.org/10.1016/0165-5876(94)01159-U

Russo, N. M., Nicol, T. G., Zecker, S. G., Hayes, E. A., & Kraus, N. (2005). Auditory training improves neural timing in the human brainstem. *Behavioural Brain Research*, *156*, 95–103.

https://doi.org/10.1016/j.bbr.2004.05.012

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology*, *33*, 3–11.

Schochat, E., Rocha-Muniz, C. N., & Filippini, R. (2017). Understanding Auditory Processing Disorder Through the FFR. In N. Kraus, S. Anderson, T. White-Schwoch, R. Fay, & A. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 225–250). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_5

Skoe, E., Burakiewicz, E., Figueiredo, M., & Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience*, *349*, 278–290. https://doi.org/10.1016/j.neuroscience.2017.02.049

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear and Hearing*, *31*(3), 302–324. papers2://publication/uuid/07E666C5-A423-43C0-85EC-BA2067A7F59F

Skuk, V., & Schweinberger, S. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech Language and Hearing Research*, *57*(1), 285–296.

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short- term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131.Plasticity

The Mathworks, I. (2019). *MATLAB R2019b* (p. Natick, Massachusetts).

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, *74*, 209–253.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*, 253–265.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Developmental Science*, *10*, 159–164.

Weaver, I., Cervoni, N., Champagne, F., D'Alessio, A., Sharma, S., Seckl, J., Dymov, S., Szyf, M., & Meaney, M. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*, 847–854.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4

Werker, J. F., & Hensch, T. (2015). Critical periods in speech perception: new directions. *Annual Review of Psychology*, *66*, 173–196.

Werker, J. F., & Tees, R. C. (2005). Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain. *Developmental Psychobiology*, *46*(3), 233–251. https://doi.org/10.1002/dev.20060

White-Schwoch, T., Davies, E. C., Thompson, E. C., Woodruff Carr, K., Nicol, T., Bradlow, A. R., & Kraus, N. (2015). Auditory-neurophysiological responses to speech during early childhood: Effects of background noise. *Hearing Research*, *328*, 34–47.

SUMMARY OF RESULTS

In Study I, we recorded the temporal envelope-following responses (FFR_{ENV}) and the temporal fine structure-following responses (FFR_{TFS}) elicited by a two-vowel rising-pitch ending syllable /oa/, of our own creation (see Figure 1a of the corresponding publication), in a sample of 34 healthy-term newborns and 18 adult participants. To comprehensively evaluate the encoding of voice pitch contour and formant structure, we conducted analyses of neural responses based on the acoustic characteristics of the different stimulus sections, extracting several parameters in the temporal and spectral domains and providing descriptive statistics derived from the FFR_{ENV} and the FFR_{TFS} (see Table 1 and Table 2 of the corresponding publication). Regarding the FFR parameters indexing voice pitch encoding, we did not find significant differences in voice pitch encoding ability at birth as compared to adults, as can be appreciated from the signal-to-noise ratio (SNR) values at the F_0 spectral peak as well as in pitch tracking measures, such as stimulus-to-response cross-correlation, pitch strength and pitch error. Concerning the FFR parameters indexing formant structure encoding, newborns exhibited overall diminished spectral amplitudes than adults at both F_1 peaks of interest (452 Hz, corresponding to the /o/ vowel section F_1; and 678 Hz, corresponding to the /a/ vowel section F_1). On the other hand, obtained SNR values in newborns were higher at 452 Hz during the /o/ section than during the /a/ section, but not different between /o/ and /a/ sections at 678 Hz. Ultimately, the data revealed no different F_0 tracking accuracy between newborns and adults, but disclosed a weaker encoding of formant structure in newborns, especially at higher frequency ranges.

In Studies II and III we explored the influence of prenatal acoustic environments –music exposure and bilingual language exposure– on the neural encoding of speech sounds at birth.

In Study II we recorded the FFR by employing two different periodic speech stimuli (widely used /da/ stimulus; new /oa/ stimulus designed and created in Study I) and analyzed them within an equivalent section of equal duration (113 ms) and fundamental frequency (F_0 = 113 Hz). A total sample of 60 healthy-term newborns were divided into two groups according to their prenatal music exposure during the last 3 months of pregnancy (daily-exposed to music; not-daily exposed), based on the responses of the babies' mothers to a musical exposure questionnaire. Analyses conducted on the neonatal FFRs revealed that the newborns who were daily-exposed to music exhibited higher spectral amplitudes of neural responses at F_0. The same effect was found in both stimuli used to elicit the FFR (/da/ and /oa/), in absence of different background neural activity (i.e., lack of significant differences in the pre-stimulus root-mean-

square (RMS) amplitude). Furthermore, no statistically significant differences were observed between groups in measures that address auditory neural transmission times, including wave V latency or neural lag.

In Study III, we collected a sample of 131 healthy-term newborns and divided them into two groups according to their prenatal fetal exposure to a monolingual or bilingual environment during the last 3 months of pregnancy, as reported by their mothers. By analyzing their neonatal FFRs elicited to the two-vowel speech stimulus (/oa/), our data did not reveal significant group differences in neural lag and RMS amplitude. However, significant differences were found in the encoding of voice pitch (F_0) and formant structure (F_1) of speech sounds, with the monolingual-exposed group exhibiting greater responses in both cases. Specifically, considering the encoding of voice pitch, we found significant differences in the SNR at F_0 between the groups, with neonates exposed to a monolingual environment showing higher SNR values, but no differences in the spectral amplitudes at F_0. Concerning the formant structure encoding, monolingual-exposed neonates exhibited higher spectral amplitudes for the corresponding frequencies during both the /o/ and the /a/ steady sections.

DISCUSSION

The present doctoral thesis was set to explore the neural mechanisms involved in the processing of speech sounds in newborns. Specifically, it was focused on the role of voice pitch and vowel formant structure encoding, as well as the potential impact of different prenatal acoustic environments, such as music or language exposure, on the development of the underlying encoding mechanisms. Accurate processing of voice pitch and formant structure is crucial for speech perception and language development, since it would serve as the primary foundation that enables subsequent early phonetic discrimination (Kuhl, 2004, 2010; Kujala et al., 2023; McCarthy et al., 2019; Werker & Curtin, 2005). Despite this, no research to date had explored these aspects in a systematic manner in healthy-term newborns, nor considered the prenatal acoustic exposure context. Therefore, the first objective of this series of studies was to shed light on the functional maturity of these speech neural encoding mechanisms, by using the electrophysiological response known as frequency-following response (FFR). By employing this approach, we aimed to identify any potential variations in neural encoding of complex sounds at birth compared to adulthood.

Moreover, previous studies have indicated that human hearing starts around the 26th week of pregnancy (Anbuhl et al., 2016; Granier-Deferre et al., 2011; Moore & Linthicum, 2007; Ruben, 1995). Fetuses are thus capable of perceiving and remembering external auditory speech stimuli while in the womb (May et al., 2011; Partanen, Kujala, Näätänen, et al., 2013; Partanen, Kujala, Tervaniemi, et al., 2013). This suggests the critical role that the prenatal and neonatal period may have in laying the basis for an optimal development of the auditory system. Thus, and based on the plasticity of the FFR and speech encoding abilities, such as those due to short-term auditory training, musical and language experience, context-dependent contingencies, prenatal clinical conditions including FGR and LGA, prematurity, acoustic environmental factors and experiences, etc. (Bosch, 2011; Carcagno & Plack, 2017; Chandrasekaran et al., 2009; DeCasper & Fifer, 1980; DeCasper & Spence, 1986; Gorina-Careta et al., 2016; Krishnan et al., 2005; Madrid et al., 2021; Moon et al., 1993; Musacchia et al., 2007; Ribas-Prats et al., 2023; Ribas-Prats et al., 2021; Wong et al., 2007), we considered whether the fetal environment, particularly the acoustic environment reaching the fetus, could affect the neonatal abilities to encode speech sounds. Therefore, the second and third aims of this thesis were to study the impact of prenatal music exposure and prenatal bilingualism environments, respectively, on the encoding of speech sounds at birth. By examining the impact of different prenatal acoustic environments on the FFR elicited to speech sounds, we aimed to gain a deeper understanding of the early stages of speech processing and language acquisition in healthy-term newborns.

Moreover, the findings from this PhD thesis may inspire further investigations into the effects of prenatal acoustic environments on the development of speech encoding mechanisms and the long-term neurodevelopmental outcomes, in order to assess the predictive value of the FFR in identifying potential risks for speech encoding disorders. Understanding the relationship between prenatal acoustic exposure and speech sound processing in newborns could have important implications for early intervention strategies aimed at optimizing language acquisition and communication skills in infancy and beyond.

Therefore, the first step was to characterize the functional maturity of speech mechanisms at birth. This objective was accomplished in Study I by recording the FFR elicited to the new two-vowel /oa/ stimulus. The results of this first study showed, regarding the encoding of the voice pitch, that newborns exhibit an accurate encoding of the fundamental frequency (F_0) of speech sounds and demonstrate the ability to track voice pitch modulations shortly after birth, which is consistent with previous research (Jeng et al., 2011, 2013, 2018; Jeng, Lin, & Wang, 2016; Jeng, Lin, Chou, et al., 2016; Ribas-Prats et al., 2019, 2023; Ribas-Prats et al., 2021). While the spectral amplitudes at F_0 peak were smaller in newborns compared to adults, when focusing on relative amplitude measurements (such as signal-to-noise ratio; SNR), we did not find differences between the newborn and adult groups.

By contrast, our findings suggest distinct processing of formant structure in newborns when compared to adults. Parallel to the F_0 values regarding the encoding of voice pitch, neonates exhibited as well reduced absolute spectral amplitudes at F_1 compared to adults, but also reduced amplitudes in relative amplitude measures (i.e., SNR). This was in line with previous literature showing a systematic increase in amplitude at the harmonic content with age (Anderson et al., 2015). Nevertheless, our data revealed that newborns possess the ability to encode the temporal fine structure of speech sounds to a certain degree, albeit with limitations observed in higher frequency ranges. Since the greatest differences appeared in the higher frequency formant (/a/ F_1 = 678 Hz), we interpreted it as the result of the low-pass filter characteristics of the womb (Gerhardt & Abrams, 2000; Jeng, 2017; McCarthy et al., 2019; Parga et al., 2018). Consequently, fetuses would be likely isolated from the mid and high-frequency content of external sounds –which constitutes most of the temporal fine structure of speech sounds (Banai et al., 2009; Hornickel et al., 2012)–, and hence may lack the prior experience necessary for a fully developed perceptual system to accurately encode higher frequencies. To further investigate and shed light on this matter, we proposed future research to involve testing premature newborns who would have been exposed to natural sounds early in their development, providing additional insights into the role of prenatal acoustic exposure on the

encoding of speech sounds and the extent to which the maturation of these mechanisms is influenced by the nature and richness of the acoustic environment to which the neonate is exposed. Additionally, it would be of interest to focus on longitudinal studies to track the progression of these auditory mechanisms over time and elucidate how this development correlates with subsequent language acquisition milestones.

Ultimately, results obtained in Study I provide a foundation for future research by establishing the initial evidence of the newborns' ability to encode the different speech sound components, just like by proving the feasibility of accurately recording in a single session both the temporal envelope and temporal fine structure in newborns through the new /oa/ stimulus created. Based on the positive experience of our first study and motivated by our interest in the prenatal acoustic environment and its potential impact on the maturation of the neural encoding mechanisms, Studies II and III moved a step forward to explore this matter on healthy-term newborn population, specifically regarding music and bilingual language exposure during pregnancy.

In particular, results from Study II provided evidence indicating that daily prenatal music exposure during the final trimester of pregnancy was related to an enhanced F_0 encoding at birth, therefore demonstrating that the prenatal acoustic environment plays a role in shaping the newborn's capacity for voice pitch processing –that mainly depends on the neural encoding of the F_0– during the early stages of life. In this case, only F_0 encoding was considered for analyses, since our research design in this second study incorporated the use of two different speech stimuli (/da/ from Ribas-Prats et al., 2019; /oa/ from Study I), each of them with unique characteristics –except for their F_0–, which were presented to different newborns of the total sample of 60 neonates. It is important to note that each newborn received only one of the two stimuli, ensuring that they were not exposed to both. And yet, despite differences in FFR spectral amplitudes between stimuli, with the /da/ stimulus exhibiting higher overall spectral amplitudes compared to the /oa/ stimulus, we did not find significant interaction between stimulus type and music exposure. In other words, regardless of the specific auditory stimulus used to elicit the FFR, and their relative amplitudes, fetuses exposed to daily music during the prenatal period exhibited stronger F_0 encoding at birth.

In Study II, the use of two different speech stimuli was motivated by practical considerations, as the newborn sample was merged from different cohorts of participants from other ongoing studies in our lab. Firstly, and considering results from Study I which did not reveal evidence at birth of complete maturity of the fine structure encoding mechanisms, it seemed

more appropriate to focus our analyses on the encoding of voice pitch, which served also as the shared element between the two stimuli selected. Furthermore, voice pitch represents a crucial component of speech that is perceived and encoded similarly to adults at birth (Arenillas-Alcón et al., 2021; Cabrera & Gervain, 2020; Gorina-Careta et al., 2022; Jeng et al., 2011), with previous research indicating that this particular mechanism is susceptible to enhancements in different aspects within enriched auditory contexts (Bidelman et al., 2011; Carcagno & Plack, 2011; Deguchi et al., 2012; Jeng et al., 2011; Krishnan et al., 2005; Magne et al., 2006; Musacchia et al., 2007; Schön et al., 2004; Thompson et al., 2003; Wong et al., 2007). Consequently, the encoding of F_0 appeared to be the most suitable mechanism to investigate in this initial exploratory study on the impact of prenatal music exposure. Moreover, the inclusion of two distinct stimuli and the convergence of similar results regarding the impact of musical exposure lend additional support to the conclusions highlighting the relationship between prenatal music exposure and enhancements in F_0 encoding, reinforcing the notion that the prenatal acoustic environment might have a remarkable influence on the development of the newborn's ability to encode voice pitch during the first days of life.

In the third study of this thesis we found that prenatal language experiences also have a modulatory effect on the neural responses to speech sounds at birth. In this case, we recorded the FFRs elicited to the /oa/ stimulus in a sample of 131 newborns divided into two groups according to their prenatal fetal exposure to a bilingual or monolingual environment. Our results indicated that newborns exposed to a monolingual environment during the last months of the fetal period exhibited greater responses both at F_0 and F_1, i.e., a better encoding of both voice pitch and formant structure of the speech stimulus. We hypothesized it may be attributed to an extensive exposure to a single language during the pregnancy period, since it is well established that the ability to discriminate different F_0 or F_0 contours improves radically with practice and exposure (Carcagno & Plack, 2017). On the contrary, bilingual-exposed newborns may encounter greater variability of acoustic conditions during the prenatal period. Notably, and regarding pitch encoding, previous research has demonstrated that bilingual individuals –in particular females–, produce distinct pitch frequency ranges depending on the language they speak (Ordin & Mennen, 2017). This variability in pitch ranges extends not only across different languages (Andreeva et al., 2014; Majewski et al., 1972; Van Bezooijen, 1995) but also among speakers of phonologically similar dialects (Deutsch et al., 2009). Similarly, we can expect a comparable inference with regards to the formant structure encoding. Bilingual environments, known for their demanding, dynamic and phonologically rich nature, demand heightened attention to every aspect of the linguistic stimuli (Krizman et al., 2012). As the auditory system

adapts to such complex linguistic contexts through continued exposure, it gradually becomes finely attuned to process sounds more effectively, thereby accounting for the enhanced flexibility and encoding abilities observed in bilingual individuals (Krizman et al., 2012). Consequently, we presumed that while fetal exposure to a single language may result in more robust and greater neural responses, a newborn exposed to a bilingual context would exhibit superior sensitivity across a broader range of frequencies without necessarily generating a robust response to any specific frequency. In contrast, a monolingual environment, characterized by more consistent and less complex acoustic conditions, may foster a more precise encoding of specific vowel sound characteristics at birth. Further research of the developmental trajectories of auditory processing in diverse newborn populations exposed to different acoustic environments during pregnancy, may offer a greater comprehension of this matter.

The discrepancy in the apparent immature abilities to encode the formant structure (mainly F_1) reported in the first study of this thesis, as compared to the recognizable neural responses to F_1 observed in Study III (see Figure 2), is an intriguing finding. In Study I, based on the available data and the hypothesis we considered at the time, we initially interpreted our findings as a result of the inadequate prior experience necessary to fully develop mechanisms able to accurately encode higher frequencies, likely due to the acoustic filter of the mother's womb (Gerhardt & Abrams, 2000; Jeng, 2017; McCarthy et al., 2019; Parga et al., 2018). However, in light of the new findings from Study III, in which appears to be a clear encoding of the formant structure in the group of monolingual-exposed newborns, we determined to review the sample of neonates from Study I attending to their prenatal language exposure in those cases where the information was available. Considering the 34 newborns recruited, for 5 of them there was no information available regarding their prenatal fetal exposure to language, since their mothers did not complete the questionnaire at the time of delivery. Remarkably, we found out that from the rest of the sample, all were bilingually-exposed (24 newborns in total; 12 females) and only 5 neonates were exposed to a monolingual environment during the pregnancy time. This is reasonable rather than surprising, taking into account the recruitment of the sample was done in a hospital in Barcelona (Catalonia), where the predominant majority of the population is bilingual in Catalan and Spanish. Therefore, given that that most of the newborn sample included in Study I were bilingually-exposed, and that we have observed in Study III how bilingual-exposed neonates did not exhibit as precise F_1 encoding as monolingual-exposed infants, this could potentially offer an explanation for the apparent still-immature encoding mechanisms at high-frequency ranges reported in the first study. Moreover, this provides

further support for the impact of prenatal acoustic environments on the speech encoding abilities at birth.

It is interesting to perceive this re-interpretation of the previous findings not as a mistake in former results nor a reason to distrust them, but rather as an inherent part of learning and knowledge development. An illustrative example of this process is the history of the sources of the FFR itself. Traditionally, it was believed to originate from exclusively subcortical sources, whereas it is only recently that new evidences suggested that, in humans, the FFR implies multiple cortical and subcortical sources (Coffey et al., 2019; Gorina-Careta et al., 2021). This highlights the role of scientific research as an ever-evolving instrument, where conclusions and interpretations are continuously refined, contrasted, and improved in light of new findings and data, such as the exploration of the impact of prenatal musical and language exposure on speech encoding.

Together, the findings of the present thesis hold significant implications for our understanding of early language development, suggesting that the foundations for language acquisition may be laid even before birth. Thus, the results suggest potential implications for infants at risk for language development conditions, such as children who experience language disorders and clinical conditions with affected linguistic functions related to a weaker neural encoding of speech sounds (Banai et al., 2009; Basu et al., 2010; Chandrasekaran et al., 2009; Font-Alaminos et al., 2020; Hornickel et al., 2012; King et al., 2002; Lam et al., 2017; Otto-Meyer et al., 2018; Rosenthal, 2020). In particular, the implications of our studies on prenatal exposure might extend to the design of intervention strategies in cases of pregnancies at risk for developmental and language-related disorders (e.g., in babies born after FGR, who appear to exhibit attenuated F_0 encoding at birth; Ribas-Prats et al., 2021), in which prenatal music exposure may act as an enhancing and desirable factor, and prenatal language exposure may have modulatory effects on the speech encoding abilities depending on the different languages the fetus was exposed to.

However, and despite the promising findings, we acknowledge several limitations apart from those already mentioned above. Firstly, it would be important to consider alternative viewpoints and other potentially influential variables in the interpretation of our findings. In particular, and related to music exposure, the reduction of maternal and fetus stress reported as a consequence of a higher exposure to music could induce neuroplastic changes in the auditory system (García González et al., 2017). Likewise, and considering the heritability of some musical abilities, with some genes linked to music perception, singing and music memory (Tan

et al., 2014), our results might as well be explained in part by the possibility that mothers more prone to listen to music and sing, give birth to babies with better sound encoding abilities. In the same way, upon discarding significant differences between groups in gestational age and birth weight, known factors that appear to affect the FFR (Ribas-Prats et al., 2023; Ribas-Prats et al., 2021), we focused in the maternal educational and socioeconomic level. This aspect is closely linked to the linguistic environment a fetus is exposed to, and therefore may affect their language acquisition and development (Hoff, 2003; Rowe, 2008): specifically, socioeconomic level could potentially constrain the complexity and richness of language exposure. Yet, despite the lack of differences across groups in these factors, the multifactorial nature of human development makes it difficult to completely eliminate their potential confounding effects. Therefore, our findings should be interpreted within this context and acknowledging these limitations. Additionally, it is important to emphasize that our participant samples were all collected solely from a single hospital center in Barcelona (*Barcelona SJD Children's Hospital*, Catalonia, Spain), which limits the generalization and interpretations of our findings: it may not apply universally due to cultural, geographic or individual differences. Thus, the sample size for each study, while sufficient for initial investigation, needs to be expanded in future research to better represent the variability inherent to human development and prenatal experiences, ensuring the generalizability of the results. Ultimately, although our contribution is limited, we hope our work will motivate further studies trying to categorize and address exhaustively the previously discussed limitations.

Our results link prenatal musical exposure to greater F_0 neural encoding, and monolingual prenatal environments with a general enhanced encoding, but do not provide any causal explanation nor detail on the most relevant constituents of prenatal music or language exposure. While prenatal acoustic experiences appear to influence the neonatal encoding of speech sounds, language acquisition is also shaped by a multitude of postnatal experiences and interactions, which continue to influence the brain's plasticity throughout childhood. Ultimately, further research with larger and more variate samples, and exhaustive control over confounding variables would shed light on how these early beginnings interact with later experiences to shape the trajectory of language development. Such findings could inform strategies for promoting optimal language development, such as encouraging music exposure and introducing diverse language environments during pregnancy according to the specific situations.

CONCLUSIONS

The present doctoral thesis aimed to explore the neural mechanisms involved in the encoding of speech sounds at birth, as well as the potential impact of the prenatal acoustic environment on the maturation of these underlying mechanisms. Based on the three studies conducted, the main conclusions can be listed as follows:

I. Newborns show a remarkable ability to encode the pitch of speech sounds, as well as to track variations in pitch with great accuracy, exhibiting similar FFR_{ENV} parameter values at F_0 to the adult population. This supports the idea of the essential role of voice pitch encoding for the first stages of language acquisition.

II. Newborns are able to encode the formant structure of speech sounds, as revealed by different FFR_{TFS} parameter values obtained to the F_1 of different vowels. This supposes the first evidence of the functional maturity of formant structure encoding mechanisms at birth.

III. Newborns exposed daily to music during the final trimester of pregnancy exhibited a stronger encoding of voice pitch than newborns who underwent a lesser exposure, as revealed by larger FFR_{ENV} parameter values at F_0. This suggests that exposure to music during the prenatal period plays a facilitating role in tuning the infant's sensitivity to the F_0 of human speech.

IV. Newborns exposed to a monolingual prenatal environment during the final trimester of pregnancy exhibited a more accurate encoding of voice pitch and formant structure at birth than newborns exposed to a bilingual environment. The findings suggest that the higher acoustic variability encountered in a bilingual environment promotes an enhanced sensitivity to a broader range of frequencies, yet without exhibiting an especially strong response at any of them.

Overall, this PhD thesis adds significant value to the understanding of early language acquisition processes and the impacts of prenatal acoustic environments on them, suggesting that the foundations for language acquisition may be laid even before birth. The use of the FFR has proven to be a powerful tool in examining the neural encoding of complex sound stimuli and opens new avenues for future research and potential clinical applications. The findings highlight the importance of early acoustic experiences, both prenatally and postnatally, in shaping the development of auditory mechanisms and subsequently, language acquisition. Future studies

should continue to explore these dynamics, as they hold large potential in designing early interventions for speech and language disorders.

REFERENCES

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1–2), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology*, *77*(2), 151–155. https://doi.org/10.1016/0168-5597(90)90031-8

Anbuhl, K. L., Uhler, K. M., Werner, L. A., & Tollin, D. J. (2016). Early Development of the Human Auditory System. In R. A. Polin, S. H. Abman, D. Rowitch, & W. E. Benitz (Eds.), *Fetal and Neonatal Physiology* (5th ed., pp. 1396–1410). Elsevier. https://doi.org/10.1016/B978-0-323-35214-7.00138-4

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2015). Development of subcortical speech representation in human infants. *The Journal of the Acoustical Society of America*, *137*(6), 3346–3355. https://doi.org/10.1121/1.4921032

Andreeva, B., Demenko, G., Möbius, B., Zimmerer, F., Jügler, J., & Jastrzebska, M. (2014). Differences of pitch profiles in Germanic and Slavic languages. In *Proceedings of Interspeech*. ISCA.

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021). Neural Encoding of Vocalic Sounds in Newborns. *Hearing Journal*, *74*(7), 10–11.

Aslin, R. N., & Mehler, J. (2005). Near-infrared spectroscopy for functional studies of brain activity in human infants: promise, prospects, and challenges. *Journal of Biomedical Optics*, *10*(1), 11009.

Banai, K., Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Reading and Subcortical Auditory Function. *Cerebral Cortex*, *19*, 2699–2707. https://doi.org/10.1093/cercor/bhp024

Basu, M., Krishnan, A., & Weber-Fox, C. (2010). Brainstem correlates of temporal auditory processing in children with specific language impairment. *Developmental Science*, *13*(1), 77–91. https://doi.org/10.1111/j.1467-7687.2009.00849.x

Benavides-Varela, S., Hochmann, J. R., Macagno, F., Nespor, M., & Mehler, J. (2012). Newborn's

brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17908–17913. https://doi.org/10.1073/pnas.1205413109

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain Effects of Music and Language Experience on the Representation of Pitch in the Human Auditory Brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425–434.

Bosch, L. (2011). Precursors to language in preterm infants. Speech perception abilities in the first year of life. In *Progress in Brain Research* (1st ed., Vol. 189). Elsevier B.V. https://doi.org/10.1016/B978-0-444-53884-0.00028-2

Brillo, E., Tosto, V., Ceccagnoli, A., Nikolova, N., Pinzaglia, V., Bordoni, F., Spano, F., Bini, V., Giardina, I., & Di Renzo, G. C. (2021). The effect of prenatal exposure to music on fetal movements and fetal heart rate: a pilot study. *Journal of Maternal-Fetal and Neonatal Medicine*, *34*(14), 2274–2282. https://doi.org/10.1080/14767058.2019.1663817

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The Roots of Bilingualism in Newborns. *Psychological Science*, *21*(3), 343–348. https://doi.org/10.1177/0956797609360758

Cabrera, L., & Gervain, J. (2020). Speech perception at birth: The brain encodes fast and slow temporal information. *Science Advances*, *6*(30), eaba7830. https://doi.org/10.1126/sciadv.aba7830

Carcagno, S., & Plack, C. J. (2011). Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. *Journal of the Association for Research in Otolaryngology*, *12*, 89–100. https://doi.org/10.1007/s10162-010-0236-1

Carcagno, S., & Plack, C. J. (2017). Short-Term Learning and Memory: Training and Perceptual Learning. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (Vol. 61, pp. 75–100). Springer Nature.

Çatalgöl, Ş., & Ceber Turfan, E. (2021). The effects of music therapy applied to pregnant women on maternal, fetal, and neonatal results: A randomized controlled study. *Health Care for Women International*, 1–17. https://doi.org/10.1080/07399332.2021.1944150

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, *64*(3), 311–319.

https://doi.org/10.1016/j.neuron.2009.10.006

Chorna, O., Filippa, M., Sa De Almeida, J., Lordier, L., Monaci, M. G., Hüppi, P., Grandjean, D., & Guzzetta, A. (2019). Neuroprocessing Mechanisms of Music during Fetal and Neonatal Development: A Role in Neuroplasticity and Neurodevelopment. *Neural Plasticity*, *2019*, 1–9. https://doi.org/10.1155/2019/3972918

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*, 1–11. https://doi.org/10.1038/ncomms11070

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-13003-w

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: Deficits and strategies for improvement. *Clinical Neurophysiology*, *112*(5), 758–767. https://doi.org/10.1016/S1388-2457(01)00465-5

De Vos, A., Vanvooren, S., Ghesquière, P., & Wouters, J. (2020). Subcortical auditory neural synchronization is deficient in pre-reading children who develop dyslexia. *Developmental Science*, *23*(6), e12945.

DeCasper, A. J., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voice. *Science*, *208*(4448), 1174–1176. https://doi.org/10.1126/science.7375928

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior & Development*, *9*(2), 133–150. https://doi.org/10.1016/0163-6383(86)90025-1

Deguchi, C., Boureux, M., Sarlo, M., Besson, M., Grassi, M., Schön, D., & Colombo, L. (2012). Sentence pitch change detection in the native and unfamiliar language in musicians and non-musicians: Behavioral, electrophysiological and psychoacoustic study. *Brain Research*, *1455*, 75–89. https://doi.org/10.1016/j.brainres.2012.03.034

Deutsch, D., Le, J., Shen, J., & Henthorn, T. (2009). The pitch levels of female speech in two Chinese villages. *The Journal of the Acoustical Society of America*, *125*, EL208–EL213.

Emery, M. M., Hebert, A. A., Aguirre Vila-Coro, A., & Prager, T. C. (1991). The relationship

between skin maturation and electrical skin impedance. *Journal of Dermatological Science*, *2*(5), 336–340.

Feldman, H. M. (2019). How young children learn language and speech. *Pediatrics in Review*, *40*(8), 398–411. https://doi.org/10.1542/pir.2017-0325

Floccia, C., Christophe, A., & Bertoncini, J. (1997). High-amplitude sucking and newborns: the quest for underlying mechanisms. *Journal of Experimental Child Psychology*, *64*, 175–198.

Font-Alaminos, M., Cornella, M., Costa-Faidella, J., Hervás, A., Leung, S., Rueda, I., & Escera, C. (2020). Increased subcortical neural responses to repeating auditory stimulation in children with autism spectrum disorder. *Biological Psychology*, *149*, 107807. https://doi.org/10.1016/j.biopsycho.2019.107807

García González, J., Ventura Miranda, M. I., Manchón García, F., Pallarés Ruiz, T. I., Marin Gascón, M. L., Requena Mullor, M., Alarcón Rodriguez, R., & Parron Carreño, T. (2017). Effects of prenatal music stimulation on fetal cardiac state, newborn anthropometric measurements and vital signs of pregnant women: A randomized controlled trial. *Complementary Therapies in Clinical Practice*, *27*, 61–67. https://doi.org/10.1016/j.ctcp.2017.03.004

Gardi, J., Salamy, A., & Mendelson, T. (1979). Scalp-Recorded Frequency-Following Responses in Neonates. *International Journal of Audiology*, *18*(6), 494–506.

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal Exposures to Sound and Vibroacoustic Stimulation. *Journal of Perinatology*, *20*, 20–29.

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology*, *54*, 91–114.

Gervain, J. (2018). The role of prenatal experience in language development. *Current Opinion in Behavioral Sciences*, *21*, 62–67. https://doi.org/10.1016/j.cobeha.2018.02.004

Gorina-Careta, N., Kurkela, J. L. O., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: a magnetoencephalographic (MEG) study. *NeuroImage*, *231*, 117866. https://doi.org/10.1016/j.neuroimage.2021.117866

Gorina-Careta, N., Ribas-Prats, T., Arenillas-Alcón, S., Puertollano, M., Gómez-Roig, M. D., & Escera, C. (2022). Neonatal Frequency-Following Responses: A Methodological Framework for Clinical Applications. *Seminars in Hearing*, *43*(3), 162–176. https://doi.org/10.1055/s-0042-1756162

Gorina-Careta, N., Zarnowiec, K., Costa-Faidella, J., & Escera, C. (2016). Timing predictability enhances regularity encoding in the human subcortical auditory pathway. *Scientific Reports*, *6*, 1–9. https://doi.org/10.1038/srep37405

Granier-Deferre, C., Ribeiro, A., Jacquet, A. Y., & Bassereau, S. (2011). Near-term fetuses process temporal features of speech. *Developmental Science*, *14*(2), 336–352. https://doi.org/10.1111/j.1467-7687.2010.00978.x

He, H., Huang, J., Zhao, X., & Li, Z. (2021). The effect of prenatal music therapy on fetal and neonatal status: A systematic review and meta-analysis. *Complementary Therapies in Medicine*, *60*, 102756. https://doi.org/10.1016/j.ctim.2021.102756

Hoehl, S., & Wahl, S. (2012). Recording Infant ERP Data for Cognitive Research. *Developmentaal Neuropsychology*, *37*(3), 187–209. https://doi.org/10.1080/87565641.2011.627958

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, E. (2009). Language Development at an Early Age: Learning Mechanisms and Outcomes from Birth to Five Years. *Encyclopedia on Early Childhood Development*, 7–10. http://www.child-encyclopedia.com/sites/default/files/dossiers-complets/en/language-development-and-literacy.pdf

Hornickel, J., Anderson, S., Skoe, E., Yi, H.G., & Kraus, N. (2012). Subcortical representation of speech fine structure relates to reading ability. *NeuroReport*, *23*(1), 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *The Journal of Neuroscience*, *33*(8), 3500–3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013

Hoth, S., Neumann, K., Weißschuh, H., Bräunert, J., Böttcher, P., Hornberger, C., Maul, H., Beedgen, B., Buschmann, K., Sohn, C., Hoffmann, G., & Plinkert, P. (2009). Universal newborn hearing screening. Methodical aspects. *HNO*, *57*(1), 29–36. https://doi.org/10.1007/s00106-008-1872-1

Jeng, F. C. (2017). Infant and Childhood Development: Intersections Between Development and Language Experience. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The Frequency-Following Response: A Window into Human Communication* (pp. 17–

43). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6_2

Jeng, F. C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C.D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in american and chinese neonates and adults. *Ear and Hearing*, *32*(6), 699–707. https://doi.org/10.1097/AUD.0b013e31821cc0df

Jeng, F. C., Lin, C.D., & Wang, T.C. (2016). Subcortical neural representation to Mandarin pitch contours in American and Chinese newborns. *The Journal of the Acoustical Society of America*, *139*(6), 190–195. https://doi.org/10.1121/1.4953998

Jeng, F. C., Lin, C.D., Chou, M. S., Hollister, G. R., Sabol, J. T., Mayhugh, G. N., Wang, T. C., & Wang, C. Y. (2016). Development of subcortical pitch representation in three-month-old Chinese infants. *Perceptual and Motor Skills*, *122*(1), 123–135. https://doi.org/10.1177/0031512516631054

Jeng, F. C., Nance, B., Montgomery-Reagan, K., & Lin, C.D. (2018). Exponential Modeling of Frequency-Following Responses in American Neonates and Adults. *Journal of the American Academy of Audiology*, *29*, 125–134. https://doi.org/10.3766/jaaa.16135

Jeng, F. C., Peris, K. S., Hu, J., & Lin, C.D. (2013). Evaluation of an Automated Procedure for Detecting Frequency-Following Responses in American and Chinese Neonates. *Perceptual and Motor Skills*, *116*(2), 456–465. https://doi.org/10.2466/24.10.PMS.116.2.456-465

Jeng, F. C., Schnabel, E. A., Dickman, B. M., Hu, J., Li, X., Lin, C.D., & Chung, H.K. (2010). Early Maturation of Frequency-Following Responses to Voice Pitch in Infants with Normal Hearing. *Perceptual and Motor Skills*, *111*(3), 765–784. https://doi.org/10.2466/10.22.24.PMS.111.6.765-784

King, C., Warrier, C. M., Hayes, E., & Kraus, N. (2002). Deficits in auditory brainstem pathway encoding of speech sounds in children with learning problems. *Neuroscience Letters*, *319*, 111–115. https://doi.org/10.1016/S0304-3940(01)02556-3

Kraus, N., Anderson, S., White-Schwoch, T., Fay, R. R., & Popper, A. N. (Eds.). (2017). *The Frequency-Following Response: A Window into Human Communication* (Vol. 61). Springer Handbook of Auditory Research.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical “what” and “where” pathways in the auditory system. *Trends in Neurosciences*, *28*(4), 176–181. https://doi.org/10.1016/j.tins.2005.02.003

Kraus, N., & Nicol, T. (2014). The cognitive auditory system: the role of learning in shaping the biology of the auditory system. In A. Popper & R. Fay (Eds.), *Perspectives on Auditory Research* (pp. 299–319). Springer Handbook of Auditory Research.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing Research*, *189*(1–2), 1–12. https://doi.org/10.1016/S0378-5955(03)00402-7

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing Research*, *382*, 166–174. https://doi.org/10.1016/j.heares.2019.107779

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *PNAS*, *109*(20), 7877–7881. https://doi.org/10.1073/pnas.1201575109

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuhl, P. K. (2010). Brain Mechanisms in Early Language Acquisition. *Neuron*, *67*(5), 713–727. https://doi.org/10.1016/j.cortex.2009.08.003.Predictive

Kuhl, P. K., & Rivera-Gaxiola, M. (2008). Neural Substrates of Language Acquisition. *Annual Review of Neuroscience*, *31*, 511–534. https://doi.org/10.1146/annurev.neuro.30.051606.094321

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21.

Kuhl, P. K., Tsao, F.M., & Liu, H.M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *PNAS*, *100*(15), 9096–9101.

Kujala, A., Huotilainen, M., Hotakainen, M., Lennes, M., Parkkonen, L., Fellman, V., & Näätänen, R. (2004). Speech-sound discrimination in neonates as measured with MEG. *NeuroReport*, *15*(13), 2089–2092. https://doi.org/10.1097/00001756-200409150-00018

Kujala, T., Partanen, E., Virtala, P., & Winkler, I. (2023). Prerequisites of language acquisition in the newborn brain. *Trends in Neurosciences*, 1–12. https://doi.org/10.1016/j.tins.2023.05.011

Lam, S. S.Y., White-Schwoch, T., Zecker, S. G., Hornickel, J., & Kraus, N. (2017). Neural stability: A reflection of automaticity in reading. *Neuropsychologia*, *103*, 162–167. https://doi.org/10.1016/j.physbeh.2017.03.040

Lee, A. C., Katz, J., Blencowe, H., Cousens, S., Kozuki, N., Vogel, J. P., Adair, L., Baqui, A. H., Bhutta, Z. A., Caulfield, L. E., Christian, P., Clarke, S. E., Ezzati, M., Fawzi, W., Gonzalez, R., Huybregts, L., Kariuki, S., Kolsteren, P., Lusingu, J., … Black, R. E. (2013). National and regional estimates of term and preterm babies born small for gestational age in 138 low-income and middle-income countries in 2010. *The Lancet Global Health*, *1*(1), E26–E36.

Lemos, F. A., Da Silva Nunes, A. D., De Souza Evangelista, C. K., Escera, C., Meira Taveira, K. V., & Andreoli Balen, S. (2021). Frequency-Following Response in Newborns and Infants: A Systematic Review of Acquisition Parameters. *Journal of Speech, Language, and Hearing Research*, 1–18.

Levi, E. C., Folsom, R. C., & Dobie, R. A. (1995). Coherence analysis of envelope-following responses (EFRs) and frequency-following responses (FFRs) in infants and adults. *Hearing Research*, *89*, 21–27.

Madrid, A. M., Walker, K. A., Smith, S. B., Hood, L. J., & Prieve, B. A. (2021). Relationships between click auditory brainstem response and speech frequency following response with development in infants born preterm. *Hearing Research*, *407*, 108277. https://doi.org/10.1016/j.heares.2021.108277

Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children : behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*(2), 199–211.

Majewski, W., Hollien, H., & Zalewski, J. (1972). Speaking fundamental frequency of Polish adult males. *Phonetica*, *25*, 119–125.

Marsál, K. (2002). Intrauterine growth restriction. *Current Opinion in Obstetrics & Gynecology*, *14*(2), 127–135.

May, L., Byers-Heinlein, K., Gervain, J., & Werker, J. F. (2011). Language and the newborn brain: does prenatal language experience shape the neonate neural response to speech?

Frontiers in Psychology, *2*(222), 1–9. https://doi.org/10.3389/fpsyg.2011.00222

McCarthy, K. M., Skoruppa, K., & Iverson, P. (2019). Development of neural perceptual vowel spaces during the first year of life. *Scientific Reports*, *9*, 1–7. https://doi.org/10.1038/s41598-019-55085-y

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*, e13122. https://doi.org/10.1111/desc.13122

Miyawaki, K., Jenkins, J. J., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: the discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.

Moeller, M. P., White, K. R., & Shisler, L. (2006). Primary care Physicians' knowledge, attitudes, and practices related to Newborn Hearing Screening. *Pediatrics*, *118*, 1357e1370.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*(4), 495–500.

Moore, J. K., & Linthicum, F. H. (2007). The human auditory system: A timeline of development. *International Journal of Audiology*, *46*(9), 460–478. https://doi.org/10.1080/14992020701383019

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Laboratory note. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, *35*(6), 665–667. https://doi.org/10.1016/0013-4694(73)90223-x

Musacchia, G., Ortiz-Mantilla, S., Roesler, C. P., Rajendran, S., Morgan-Byrne, J., & Benasich, A. A. (2018). Effects of noise and age on the infant brainstem response to speech. *Clinical Neurophysiology*, *129*(12), 2623–2634. https://doi.org/10.1016/j.clinph.2018.08.005

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to

sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375–425. https://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan, H. G. (1989). Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephalography and Clinical Neurophysiology*, *73*(4), 295–305. https://doi.org/10.1016/0013-4694(89)90108-9

Ordin, M., & Mennen, I. (2017). Cross-linguistic differences in bilinguals' fundamental frequency ranges. *Journal of Speech, Language, and Hearing Research*, *60*, 1493–1506. https://doi.org/10.1044/2016_JSLHR-S-16-0315

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*, 733–743. https://doi.org/10.1007/s00221-017-5164-4

Parga, J. J., Daland, R., Kesavan, K., Macey, P. M., Zeltzer, L., & Harper, R. M. (2018). A description of externally recorded womb sounds in human subjects during gestation. *PLoS ONE*, *13*(5), 1–14. https://doi.org/10.1371/journal.pone.0197045

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *PNAS*, *110*(37), 15145–15150. https://doi.org/10.1073/pnas.1302159110

Partanen, E., Kujala, T., Tervaniemi, M., & Huotilainen, M. (2013). Prenatal music exposure induces long-term neural effects. *PLoS ONE*, *8*(10), e78946. https://doi.org/10.1371/journal.pone.0078946

Partanen, E., Pakarinen, S., Kujala, T., & Huotilainen, M. (2013). Infants' brain responses for speech sound changes in fast multifeature MMN paradigm. *Clinical Neurophysiology*, *124*(8), 1578–1585. https://doi.org/10.1016/j.clinph.2013.02.014

Picton, T. W. (2010). *Human Auditory Evoked Potentials*. Plural Publishing Inc.

Picton, T. W., & Mazaheri, A. (2006). Electroencephalography (EEG). In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*.

Pinto, E. S. M., & Martinelli, M. C. (2018). Brainstem auditory evoked potentials with speech stimulus in neonates. *Brazilian Journal of Otorhinolaryngology*, *86*(2), 191–200. https://doi.org/10.1016/j.bjorl.2018.11.006

Poćwierz-Marciniak, I., & Harciarek, M. (2021). The Effect of Musical Stimulation and Mother's Voice on the Early Development of Musical Abilities: A Neuropsychological Perspective. *International Journal of Environmental Research and Public Health*, *18*, 8467. https://doi.org/10.3390/ijerph18168467

Polka, L., Jusczyk, P. W., & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In W. Strange (Ed.), *Speech Perception and Language Experience: Issues in Cross-language Research* (pp. 49–89). York Press.

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28–39. https://doi.org/10.1016/j.heares.2018.11.001

Ribas-Prats, T., Arenillas-Alcón, S., Pérez-Cruz, M., Costa-Faidella, J., Gómez-Roig, M. D., & Escera, C. (2023). Speech-encoding deficits in neonates born large-for-gestational age as revealed with the frequency-following response. *Ear and Hearing*. https://doi.org/10.1097/AUD.0000000000001330

Ribas-Prats, T., Arenillas-Alcón, S., Lip-Sosa, D. L., Costa-Faidella, J., Mazarico, E., Gómez-Roig, M. D., & Escera, C. (2021). Deficient neural encoding of speech sounds in term neonates born after fetal growth restriction. *Developmental Science*, 1–15. https://doi.org/10.1111/desc.13189

Richard, C., Neel, M., Jeanvoine, A., Connell, S., Gehred, A., & Maitre, N. L. (2020). Characteristics of the Frequency-Following Response to Speech in Neonates and Potential Applicability in Clinical Practice: A Systematic Review. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1618–1635.

Rosen, S. (1992). Temporal Information in Speech: Acoustic, Auditory and Linguistic Aspects. *Philosophical Transactions: Biological Sciences*, *336*, 367–373. https://doi.org/10.1098/rstb.1992.0070

Rosenthal, M. A. (2020). A systematic review of the voice-tagging hypothesis of speech-in-noise perception. *Neuropsychologia*, *136*(November 2019), 107256. https://doi.org/10.1016/j.neuropsychologia.2019.107256

Rowe, M. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

Ruben, R. J. (1995). The ontogeny of human hearing. *International Journal of Pediatric Otorhinolaryngology*, *32*, 199–204. https://doi.org/10.1016/0165-5876(94)01159-U

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with Autism Spectrum Disorders. *Clinical Neurophysiology*, *119*(8), 1720–1731. https://doi.org/10.1016/j.clinph.2008.01.108

Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349. https://doi.org/10.1111/1469-8986.00172.x

Sket, G. M., Overfeld, J., Styner, M., Gilmore, J. H., Entringer, S., Wadhwa, P. D., Rasmussen, J. M., & Buss, C. (2019). Neonatal White Matter Maturation Is Associated With Infant Language Development. *Frontiers in Human Neuroscience*, *13*(434). https://doi.org/10.3389/fnhum.2019.00434

Skoe, E., & Kraus, N. (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear and Hearing*, *31*(3), 302–324. papers2://publication/uuid/07E666C5-A423-43C0-85EC-BA2067A7F59F

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and Plasticity of Auditory Brainstem Function Across the Lifespan. *Cerebral Cortex*, *25*, 1415–1426. https://doi.org/10.1093/cercor/bht311

Song, J. H., Skoe, E., Wong, P. C. M., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short- term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892–1902. https://doi.org/10.1162/jocn.2008.20131.Plasticity

Stuart, A., Yang, E. Y., & Botea, M. (1996). Neonatal auditory brainstem responses recorded from four electrode montages. *Journal of Communication Disorders*, *29*(2), 125–139. https://doi.org/10.1016/0021-9924(95)00018-6

Tan, Y. T., McPherson, G. E., Peretz, I., Berkovic, S. F., & Wilson, S. J. (2014). The genetic basis of music ability. *Frontiers in Psychology*, *5*, 658. https://doi.org/10.3389/fpsyg.2014.00658

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Annals of the New York Academy of Sciences*, *999*, 503–2.

Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*, 253–265.

Vannasing, P., Florea, O., González-Frankenberger, B., Tremblay, J., Paquette, N., Safi, D., Wallois, F., Lepore, F., Béland, R., Lassonde, M., & Gallagher, A. (2016). Distinct hemispheric specializations for native and non-native languages in one-day-old newborns identified by fNIRS. *Neuropsychologia*, *84*, 63–69. https://doi.org/10.1016/j.neuropsychologia.2016.01.038

Webb, A. R., Heller, H. T., Benson, C. B., & Lahav, A. (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3152–3157. https://doi.org/10.1073/pnas.1414924112

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. https://doi.org/10.1207/s15473341lld0102_4

Werker, J. F., & Tees, R. C. (2005). Speech Perception as a Window for Understanding Plasticity and Commitment in Language Systems of the Brain. *Developmental Psychobiology*, *46*(3), 233–251. https://doi.org/10.1002/dev.20060

White-Schwoch, T., Davies, E. C., Thompson, E. C., Woodruff Carr, K., Nicol, T., Bradlow, A. R., & Kraus, N. (2015). Auditory-neurophysiological responses to speech during early childhood: Effects of background noise. *Hearing Research*, *328*, 34–47.

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, *13*(7). https://doi.org/10.1371/journal.pbio.1002196

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422. https://doi.org/10.1038/nn1872.Musical

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*(1), 61–83. https://doi.org/10.1006/jpho.1996.0034