Development of auditory cognition in 5- to 10-year-old children: Focus on speech-in-babble-noise perception

Jérémie Ginzburg, Lesly Fornoni, Pierre-Emmanuel Aguera, Caroline Pierre, Anne Caclin, Annie Moulin

Development of auditory cognition in 5- to 10-year-old children: focus on speech-in-babble-noise perception

J. Ginzburg[1], L. Fornoni[1], P.E. Aguera[1], C. Pierre[1], A. Caclin[1], A. Moulin[1]

[1] Université Claude Bernard Lyon 1, CNRS, INSERM, Centre de Recherche en Neurosciences de Lyon CRNL U1028 UMR5292, F-69500, Bron, France

Corresponding author: J. Ginzburg, jeremie.ginzburg@inserm.fr

Acknowledgments: We thank all the children who took part in the study, along with their parents, and their welcoming teachers. The data necessary to reproduce the analyses presented here are not publicly accessible due to privacy or ethical restrictions. The data are available on request from the corresponding author. The analytic code necessary to reproduce the analyses presented in this paper is publicly accessible at the following URL: https://github.com/jeremieginzburg/cd_analysis_sample_code. The materials necessary to attempt to replicate the findings presented here are not publicly accessible. The analyses presented here were not preregistered.

Funding: This work was conducted within the framework of the LabEx CeLyA (‘‘Centre Lyonnais d’Acoustique”, ANR-10-LABX-0060) and of the LabEx Cortex (‘‘Construction, Cognitive Function and Rehabilitation of the Cortex”, ANR-11-LABX-0042) of Université de Lyon, within the program ‘‘Investissements d’avenir” (ANR-16-IDEX-0005) operated by the French National Research Agency (ANR). This work was funded by a Pack Ambition Recherche (COGAUDYS project) from the Auvergne-Rhône-Alpes Region to AC and AM.

Abstract

Speech-in-noise perception is consistently reported to be impaired in learning disorders, which stresses the importance of documenting its developmental course in young children. In this cross-sectional study, ninety children (41 females, 5.5-11.6 years-old) and nineteen normal-hearing adults (15 females, 20-30 years-old), were tested with a newly-developed closed-set speech perception in babble-noise test, combining two levels of phonological difficulty and two noise levels. Results showed that speech-in-babble-noise perception takes a definite maturation step around 7 years-of-age (d = 1.17, grade effect) and is not mature at 10 years-of-age when compared to young adults (d = .94, group effect). Developmental trajectories of both accuracy and response times were evaluated, with influences of psycholinguistic factors, to foster the development of adequate screening tests.

The ability to make perceptual sense of a complex sound environment (Auditory scene analysis, Bregman, 1994) is a key component of auditory cognition and a major element for harmonious cognitive development in children. Hearing screening has become widely implemented in numerous countries, beginning at birth and continuing at regular intervals throughout children's primary school years. Its primary objective is to identify peripheral hearing loss, which is of paramount importance (Skarżyński & Piotrowska, 2012). However, these screenings typically do not routinely include the assessment of speech perception in noisy environments, despite an estimated 7% of children with normal audiograms experiencing specific difficulties in listening in noise (Bamiou et al., 2001; Wilson & Arnott, 2013). Speech-in-noise perception, along with other key auditory cognition processes (e.g., auditory short-term memory, STM), are consistently reported to be impaired in learning disorders (Bradlow et al., 2003; Majerus & Cowan, 2016; Plaza et al., 2002; Ziegler et al., 2005, 2009). This fact stresses the need for systematically screening such listening difficulties specific to noisy environments (Skarzynski et al., 2015), as they are often associated with learning disorders that will sometimes be diagnosed years later. Dyslexia often eludes diagnosis until formal reading instruction begins. For instance, Boets et al. (2011) found a link between speech-in-noise perception deficits in kindergarten and dyslexia diagnosis in 3rd grade, with frequency modulation sensitivity and speech-in-noise perception predicting reading ability growth. Guzek & Iwanicka-Pronicka (2022) emphasized the importance of screening for auditory processing disorders at age 6, with speech perception in babble noise being the most commonly impaired among 1012 children tested, with and without auditory processing disorder.

Children display more difficulties than adults in recognizing speech when faced with background sounds, and this trend is consistent across various stimuli and listening conditions (Kidd & Colburn, 2017). Studies on children's speech recognition in noise reveal that these differences between children and adults persist noticeably until adolescence, with the specific

age range varying depending on the nature of the noise. Various factors contribute collectively to the progressive improvement of speech perception in noise as children develop. These include the ability to leverage spatial separation between the target and masker (D. K. Brown et al., 2010; Ellen Peng & Litovsky, 2021; Litovsky, 2005), frequency differences between the target and masker (Flaherty et al., 2019; Leibold et al., 2018; Leibold & Buss, 2019), and the use of glimpses in the noise masker (Buss et al., 2017). Beyond these low-level factors, immature allocation of attention is suggested to play a pivotal role in children's pronounced difficulties with speech in noise (Leibold & Buss, 2019). At the cortical level, it has been proposed that children's challenges in taking advantage of acoustic cues to discriminate between target and masker, may correspond to the maturation timeline of the thalamus, thalamo-cortical radiations, and the primary auditory cortex, with structural maturation concluding around 10 years of age (Devous et al., 2006; Pujol et al., 2006; Werner et al., 2012). Since speech in noise maturation has been observed to continue until adolescence, it is conceivable that this late maturation aligns with the delayed development of the secondary auditory cortex and prefrontal regions, known to underlie attentional processes involved in speech-in-noise perception and maturing into adulthood (Kolk & Rakic, 2022).

Screening this process throughout development poses a significant challenge, given that both the nature of the speech perception task and the balance of energetic/informational masking within the noise masker exert substantial influence on the resulting scores (Buss et al., 2016, 2022; Corbin et al., 2016). Classically, energetic masking refers to noise maskers with the same spectro-temporal structure as the signal, hence reducing the availability of signal cues and affecting its neural representation, mostly at a peripheral level (e.g., white noise, speech-shaped noise). In contrast, informational masking refers typically to the situation of attending to one talker in the midst of another or several talkers (termed speech-on-speech masking, or babble noise or cocktail party noise), with masking occurring even when the target talker is

perfectly audible. Most studies in children have used energetic masking (mostly speech-shaped noise) and showed that differences between children and adults remain noticeable until around 9–10 years of age (for a review on development of speech recognition in steady-state noise, see Leibold & Buss, 2019). However, in realistic environments, environmental noises and speech sources from other children within the class add up to form a noise masker closer to a multi-talker babble noise than to speech-shaped broadband noise. In addition, listening to the teacher's voice in a classroom poses challenges due to voice reflections on various surfaces, creating multiple speech signals at the child's ear with different delays. While short reverberation times (<0.5 s) are not problematic, longer delays can lead to temporal smearing and partial masking, reducing word recognition (Wróblewski et al., 2012). Moreover, classroom reverberation times often exceed the recommended 0.6 s, sometimes reaching up to 1.5 s (Anderson, 2004; Crandell & Smaldino, 2000), making the noise close to speech-on-speech. Two-talker babble noise is known to be more challenging than speech-shaped noise in young adults as well as in children (Corbin et al., 2016; Lewis et al., 1988) and various studies indicate a more extended maturation of speech in multi-talker noise than speech in speech-shaped noise (Corbin et al., 2016; Elliott, 1979; Hall et al., 2002). Hall et al. (2002) used a forced-choice, picture-pointing task to measure the recognition of spondee words (words comprising two equally stressed syllables, e.g., "popcorn") within the presence of either speech-shaped noise or speech from two talkers. Participants were children aged 5 to 10 years and adults aged 19 to 48 years. On average, children needed an additional 3 dB to match the performance level of adults in the noise masker. In contrast, there was an 8-dB difference between adults and children in the two-talker masker to reach similar performance. Using a similar task to assess recognition of consonant-vowel stimuli differing only by consonants, Leibold & Buss (2013) showed that performance was adult-like at 11 years of age for the speech-shaped noise. However, in a two-talker masker, performance was lower than adults until

111 the oldest age group (13 years old). Corbin et al. (2016) measured children's speech reception
112 thresholds (SRT, thresholds for 50% correct recognition) of monosyllabic words by varying the
113 masker's intensity with an open-set paradigm. Comparably to the previously described studies,
114 children reached adult-like performance by 10 years of age when the masker was a speech-
115 shaped noise, while children's performance increased up until 13 years of age in a two-talker
116 masker. These studies suggest that with a speech masker comprising two talkers, regardless of
117 the stimuli used for recognition and of the task (open/closed set), children's performance reach
118 adult's performance later than for speech-shaped noise.

119 Babble noise is considered to be the most challenging noise for speech perception and
120 the closest to ecological conditions. Assessing speech perception in babble noise might be more
121 sensitive to listening difficulties associated with learning disorders than using speech shaped
122 noise. The number of talkers (Freyman et al., 2004; Rosen et al., 2013; Simpson & Cooke,
123 2005), the distance in fundamental frequency and the difference in language type (native versus
124 non native, Calandruccio et al., 2016; Calandruccio & Zhou, 2014) between the target talker
125 and the multi-talker competitive babble, have all shown major influence on the masking
126 efficiency in adults. As the number of competitive talkers increase from 1 to 6, consonant
127 identification scores of vowel-consonant-vowel syllables in young normally hearing adults
128 decrease for up to 6 talkers, and then remain fairly stable (or slightly increase) from 8 to 128
129 talkers (Simpson & Cooke, 2005). Similarly, Rosen et al. (2013) obtained a sharp decrease in
130 natural sentence identification scores from 1 to 2 talkers maskers, and a small increase of scores
131 with 4 to 16 multi-talker babble as in Simpson & Cooke (2005)'s study. This observed increase
132 in scores with an increasing number of talkers is unexpected as one would anticipate a decrease
133 in scores due to reduced gaps in noise amplitude, which would decrease glimpsing opportunities
134 that would help speech perception. These results thus indicate that while energetic masking
135 increases with more talkers, informational masking decreases because words spoken in babble

noise become progressively less comprehensible as the number of talkers increases. Despite this trend, it is important to note that multi-talker babble retains a significant element of informational masking. Even with a high number of talkers, such as 16 (Rosen et al., 2013) or 128 (Simpson & Cooke, 2005), the babble noise remains more challenging for speech perception than a speech-shaped noise masker. Research on the developmental trajectory of speech perception in two-talker babble noise is less common compared to studies on speech-shaped noise. To our knowledge, none have explored this aspect using babble noise involving more than two talkers, despite the indication that this balance between informational and energetic masking aligns more closely with children's real-world auditory environments.

Defining the type of masker to use is only part of the story to assess speech-in-noise development in children. Indeed, most previous studies use open-set response paradigms, which can limit their application in children. For instance, in the first “babble noise in children study”, Elliott (1979) used sentences with different predictability of words, and the lower age limit considered for such test was 9 years of age. McCreery et al. (2020) showed a significant influence of vocabulary in sentence recognition in noise for children aged 5 to 6, compared to children older than 9 years of age. Open-set speech tests, that require verbal responses, need clear articulation by the child to allow precise scoring (Stiles et al. 2012), which can make them problematic for hearing-impaired young children, children with disorders of speech sound production (Cabbage & Hitchcock, 2022), or children whose verbal responses can be difficult to obtain (for instance in some children with non-speaking autism). By using picture-pointing response requiring to choose an answer amongst several foils, alternative forced choice tests (nAFC) alleviate the need for a verbal response and are very valuable in paediatric assessments (Leibold et al., 2024; Mendel, 2008; Vickers et al., 2018). Another advantage of alternative forced choice tests for children, in addition to not relying on a verbal response, is the lower sensitivity of this test format to top-down lexical influences, as shown in adults (Sommers et

al., 1997), and in young children (S. Jerger et al., 1983). Indeed, alternative forced-choice tests show better reproducibility than open response tests (Gelfand, 1998, 2003), are not sensitive to learning and are less sensitive to talker variability and lexical competition (Clopper et al., 2006; Sommers et al., 1997). By providing the child with a set of possible answers, closed-set tests restrict the set of potential answers, which can reduce the linguistic bias linked to language development (J. Jerger et al., 1968). Jerger et al's Pediatric speech intelligibility test (S. Jerger et al., 1981), which consists in a five-alternative forced choice test, with monosyllabic words presented as pictures, has been shown to be independent of receptive language abilities, in children aged 3 to 6 years. In adults, Clopper et al. (2006) showed that increasing the number of alternatives (from 6 to 12) increased the test difficulty, and the sensitivity of closed-set tests to talker variability and lexical competition.

In this study, our aim was to evaluate the developmental trajectory of speech perception in multiple-talker ($n > 2$) noise conditions, employing a task designed to address previously identified challenges. This approach enabled us to investigate the potential impact of crucial factors such as phonological proximity, fatigability, and various psycholinguistic factors on speech perception in noise. As this study focused on children, we adhered to the standard format of a four-alternative forced-choice test commonly employed for children (e.g., Calandruccio et al., 2014; Vickers et al., 2018) by using a French language adaptation for children of the English versions of 4AFC tests (Buss et al., 2016; Vickers et al., 2018). We chose a word recognition closed-set task over an open-response sentence-based test like QuickSIN^{TM} (Killion et al., 2004) to mitigate contextual influences. We used a 16-talker babble noise in order to bring us closer to real-world situations involving both energetic and informational masking, in contrast to the conventional speech-shaped noises used in traditional clinical audiometry. In order to evaluate the influence of the phonological proximity between target word and foils in nAFC tasks, we manipulated the task demands by modifying the phonological proximity between target word

and foils: the same target-word was presented with either phonologically close foils (imposing a more difficult phonological discrimination task), or phonologically distant foils (an easier discrimination task). Indeed, while closed-set paradigms reduce top-down lexical influences (Clopper et al., 2006; S. Jerger et al., 1983; Sommers et al., 1997), this type of task involves directly comparing an auditorily perceived word with the words represented by the pictures in the response set. The phonological proximity between the target words and foils has been shown to influence speech perception in noise for both adults and children in forced-choice paradigms (Clopper et al., 2006, Buss et al., 2016). The ability to discriminate efficiently between phonologically close and distant words is a phonological skill related to the broader concept of phonological awareness, defined as the ability to manipulate spoken language sound units (Anthony & Lonigan, 2004). Phonological awareness undergoes significant and early developmental changes with age (Snowling & Hulme, 1994; Stekić et al., 2023) and is related to literacy acquisition, often being impaired in dyslexia (e.g., Gijbels et al., 2024; Stahl & Murray, 1994; Stekić et al., 2023; Watson & Miller, 1993). Using an adaptive 4AFC in children from 5 to 13 years old, Buss et al. (2016) showed that phonological similarity had a significant effect for both speech-shaped noise and two-talker maskers. While phonological proximity affected children regardless of age for speech-shaped noise, younger children were more impacted by phonological proximity in the two-talker masker. In the present study, we used a 16-talker masker, a type of masker that has not been tested for such effects to our knowledge. Therefore, it is important to assess, in our dataset, whether phonological proximity between foils interacts with the developmental trajectory of speech-in-noise abilities, as it does with two-talker maskers, but not speech-shaped noise (Buss et al., 2016). Additionally, we employed a non-adaptive method with an equal number of trials for each phonological proximity condition, allowing for a thorough comparison of its effects.

To ensure strict comparability between the two phonological conditions, identical targets were randomly presented in both conditions, with fixed Signal-to-Noise Ratios (SNRs) in separate blocks. This approach was preferred over an adaptive method, involving trial-by-trial SNR modifications, which is considered more susceptible to fatigue and inattention (Wightman & Allen, 1992) and would not allow us a balanced design between the different levels of phonological proximity. Additionally, using this fixed SNR paradigm enabled the examination of potential training effects with task repetition and the assessment of psycholinguistic factors. Lastly, the use of a touch tablet allowed us to take advantage of the greater motivational aspect of such computerized tests in children, and gave us easy access to response time, in addition to accuracy data. This allowed us to investigate, in a systematic manner, developmental aspects of speech perception in babble noise, in 5 to 10 years old children, relative to a group of young normal-hearing adults.

Our expectations are as follows: Firstly, we anticipate a steep maturation trajectory in children's speech in babble noise abilities, with younger children demonstrating lower performance compared to older counterparts. Furthermore, we expect that children will not reach adult-level performance even at 10 years of age, given our use of babble noise instead of the typical speech-shaped noise. Secondly, we hypothesize a significant performance difference between phonological proximity conditions, highlighting the importance of this factor in alternative-forced choice tests. Lastly, as we used a forced-choice paradigm with carefully selected pictures and words specifically designed for young children, we do not anticipate observing any effects of learning, vocabulary or psycholinguistic factors throughout the task. Due to the clear and well-founded hypotheses, as well as the straightforward expectations regarding the direction of effects based on existing literature, this study qualifies as confirmatory in nature.

Methods

Participants

One hundred children (mean age = 7.6 years; min = 4.5 years, max = 10.6 years, 6 left-handed) attending a public primary school participated in this study in June 2019, . The children were tested in June 2019 during school hours. Participation in the study was proposed to a total of 114 children and they were included in the study only if both parents or legal guardians provided a written informed consent. In addition, prior to each test, verbal assent from each child was sought by the experimenters. The study was approved by the relevant services of the French public education services (Inspection de l'Education Nationale [IEN] and Direction Académique des Services de l'Education Nationale [DASEN] of the Isere department). Children from Kindergarten (KG) grade to 5th grade performed the experiment. Before testing, parents filled out a questionnaire about their child's level of education and their own level of education, the child's laterality (e.g., dominant hand), possible vision or auditory impairments, musical activities, bilingualism, learning disabilities, and 11 questions adapted from adults' musical listening questionnaires (Lévêque et al., 2018; Tillmann et al., 2014). Among the 100 children, parents' responses to questionnaires revealed that 5 children had a diagnosed learning disability and their data were excluded from the current analysis. Thirty children had seen at least once a speech-therapist, six had already worn ear tubes, eight had a few years of musical education, and three were bilingual. As we aimed here to explore the cognitive abilities of children among a representative set of the population, we present the results including data from all these children. Data from five children were excluded because of technical difficulties during recordings. Overall, data from 10 children were thus excluded from the analysis, the sample sizes per grade of the remaining 90 children are reported in Table 1. Data on participant ethnicity were not collected because the collection of such data is illegal in France and would require an exceptional waiver from government agencies that we did not seek. Proportion of parental level of education did not significantly differ by grade (see Figure 1 in Ginzburg et al.,

2022). Due to the limited number of children with musical education, a history of ear tubes, and bilingualism, these variables could not be included in the main analysis. However, for each of these factors, we separately conducted the main developmental trajectory analysis by excluding data from children with musical education, a history of ear tubes, or bilingualism. We found that none of these exclusions altered the main result, aside from a small decrease in statistical power.

Nineteen adults (four men), with a mean age of 24.11 years (SD = 3 years) were also included in the study. None of them presented any visual or auditory impairment, nor neurological or psychiatric troubles, and none of them had any musical background.

Table 1: Number of participants and mean age (SD in parentheses) for each grade and matching English and French labels for educational level.

English label	Kindergarten (KG)	1st Grade	2nd Grade	3rd Grade	4th Grade	5th Grade
French label	Grande Section de Maternelle (GS)	Cours Préparatoire (CP)	Cours Elémentaire 1 (CE1)	Cours Elémentaire 2 (CE2)	Cours Moyen 1 (CM1)	Cours Moyen 2 (CM2)
N (n females)	11 (5)	16 (9)	17 (8)	16 (6)	15 (5)	15 (8)
Mean age in years (SD)	5.01 (0.31)	6.03 (0.29)	6.93 (0.30)	8.01 (0.34)	8.92 (0.26)	9.94 (0.29)

Stimulus construction and task design

Each trial of the speech-in-babble task consisted in matching an aurally-presented word with its corresponding image among four images. The material consisted of 20 monosyllabic and 4 disyllabic words, easily representable by a drawing, selected as a function of their concreteness, their frequency of occurrence in the French language, their age of acquisition, and their phonological similarity. The 24 selected words were recorded by a French native

female talker, in a sound proof booth, using a Rode NT1 microphone. Sound files had an average duration of 540 ms (SD=171 ms) and their amplitudes were equated in root mean square (RMS) level. Each word had a corresponding hand-drawn image. From this list of 24 words, six sets of four phonologically close words were created, e.g., /ku/, /kul/, /ru/, /ruʒ/ (*cou, coule, roue, rouge* in French, corresponding to *neck, flow, wheel, red* in English). Within each phonologically close set, only consonants could change and up to three phonemes were modified. With the same 24 words, six sets of four phonologically distant words were created, e.g., /bɑ̃/, /flœʀ/, /mɛ̃/, /trwa/ (*banc, fleur, main, trois* in French, corresponding to *bench, flower, hand, three* in English), where several phonemes, including the vowels, were different. For the sake of clarity throughout the remainder of the article, we will refer to the phonologically distant condition as the "easy condition" (easy to discriminate) and to the phonologically close condition as the "difficult condition" (difficult to discriminate). The average of pairwise Levenshtein distance (i.e., the minimal possible number of insertions, deletions, and substitutions needed to transform one chain of phonemes into another) was 3.03 phonemes (SD = 0.4) for the easy condition and 1.39 phonemes (SD = 0.29) for the difficult condition. Using the French database 'Lexique' (New et al., 2004), we calculated the number of phonological neighbors (i.e., number of existing words by changing one phoneme without changing the others) for each of the 24 words used, and then extracted the number of unique phonological neighbors for each 4-word set. In the easy condition, the four words of each set were phonologically distant (each word having phonological neighbors different from the other three) thus the number of unique phonological neighbors within a set was relatively high (mean = 83, sd = 6.7). Conversely, in the difficult condition, the four words were quasi-phonological neighbors (each word in a set having several common phonological neighbors with the other three) thus having a relatively small number of unique phonological neighbors within a set (mean = 59.7, sd = 14.8). Lastly, we derived for each of the target words their frequency of

occurrence (per million occurrences) in the French language (mean = 85.9, SD=119) calculated from a movie subtitles corpus (New et al., 2007) (for details about the stimuli, see supplemental Tables S1 and S2).

Procedure

Children were tested by groups of five or six in the sport room of their school. Before testing, children sat in front of their tables, listening to the experimenter's instructions. The experimenter explained the speech-in-babble noise task with cardboard signs: children would see 4 images appear on the touchpad and at the same time, they would hear a word through their headphones. They had to find the spoken word in one of the four images and tap on it as quickly as possible. They had to ignore the people talking in the background (babble noise). The task was implemented on a touch-tablet and auditory stimuli were displayed binaurally through Sennheiser HD-250-Pro circum-aural headphones. Children underwent a training block comprising eight trials with a SNR of +3 dB. Then they underwent two 48-trial blocks with a SNR of +3 dB for the first block and a SNR of -3 dB for the second block. In the subsequent sections of the article, we will refer to the +3 dB SNR condition as the "favorable noise condition" and to the -3 dB SNR condition as the "unfavorable noise condition". The target-word was systematically presented at 66 dB SPL (A-weighted). During each trial, one of the four-image sets was presented, and 800 ms later, the target word was aurally presented. The subject had to tap with a finger on the matching image as quickly as possible but without any time limit. The next trial began immediately after the response. In both blocks, twelve four-image sets were used: six consisting of phonologically distant words (easy condition) and six consisting of phonologically close words (difficult condition, see above). Within a block, each set was presented four times, each time with a different target word (since the same words were used to create the easy and the difficult condition, each word was used twice as the target: once in each of the two phonological difficulty conditions). A block was thus divided in four series

of 12 trials with each of the four-image set presented once per series. The order of presentation of the 12 sets was randomized for each subject and each block, and the same randomized order was then used for the four series. This was done to avoid, as much as possible, the potential advantage linked to the limitation of the number of remaining response alternatives once items have been eliminated from the list of possibilities during the first series. The position of the images presented on the screen was randomized. Lastly, no feedback was given on a trial per trial basis. During the task, a babble noise made of 16 unintelligible French native male and female voices (Moulin et al., 2013) was presented in a continuous manner. The use of 16 talkers enabled us to achieve a balanced representation of both masculine and feminine voices, reducing the risk of any single voice standing out or being disproportionately closer or further from the signal voice. In contrast, with only a few talkers, the pitch (fundamental frequency) differences between the maskers and the signal can become especially relevant. The subjects could listen to several seconds of the babble noise before triggering the start of the block.

The duration of each block was around 2 minutes, so that the total hearing-in-noise test was performed in about 10 minutes briefing and training included. The unfavorable noise condition was always given last, as it was anticipated that the children listening behavior might be modified by the challenge represented by this difficult SNR (i.e., inducing discouragement before the next block).

Adults underwent the task individually in an experimental room in the lab with the same equipment as children. They performed the same eight-trial training block as the children and then eight 48-trial blocks with babble noise ranging from -12 dB SNR to +9 dB SNR by steps of 3 dB SNR in order to calculate their individual detection thresholds. As for children, each block was performed at a constant SNR with a constant-stimulus procedure. The order of the blocks was counterbalanced.

Data analyses

All data analysis were performed on R (R Core Team, 2019). For each participant, accuracy (percentage of correct responses) and response times (RT) from sound onset were extracted for each phonological difficulty and for each noise condition. Trials with RTs above 8 seconds and when the participant answered before the onset of the sound were excluded (i.e. below 800 ms), representing 1.4% of children's data and 0.1% of adult's data. For all analysis on RTs, we used response times on correct trials only and on trials with common words between the phonological easy and difficult condition (to avoid imbalance due to the fact that more trials were successful in easy than in difficult condition), which represented 75.4% of the data for children and 93.5% for adults.

Adult data: psychometric parameters

To evaluate if detection thresholds differed according to the phonological difficulty of the task, psychometric curves were fitted to responses as a function of the signal-to-noise ratio using a logistic function with a guess rate λ set to .25 (because of the 4AFC task's nature) for each difficulty condition and for each participant by using the *quickpsy* package in R (Linares & López-Moliner, 2016). It allowed the estimation of detection thresholds (i.e., SNR for which each participant obtains $\lambda + (1 - \lambda) * 0.5 = 62.5\%$ of correct responses thus half-way between chance level (25%) and perfect performance (100%)). Easy and difficult phonological conditions were compared for each subject with the *thresholdcomparisons* function from the *quickpsy* package in R which performs a bootstrap comparison between conditions to determine if the difference between them falls outside a 95% confidence interval. Comparison between conditions at the group level was tested by comparing the average detection thresholds across all subjects for both phonological difficulties with a pairwise t-test after checking for normality with a Shapiro-Wilk test.

Development of hearing in babble noise

Developmental aspects of children's performance were tested as a function of children's school grade, allowing for homogeneous groups composed of children being equally scholarly educated regarding reading and other abilities. In order to assess the effects of grade, phonological difficulty, and noise condition on children's performance, we used general linear models (GLM) for accuracy and general linear mixed models (GLMM) for response times (Bates et al., 2014). We thus analyzed the influence of three main effects on percent correct scores and RTs:

- Between-subject factor *Grade*: six levels (KG, 1st, 2nd, 3rd, 4th and 5th grade)
- Within-subject factor *Phonological Difficulty*: two levels (easy/difficult phonological condition)
- Within-subject factor *Noise condition*: two levels (unfavorable/favorable)

GLM(M) were fitted using the *lme4* package in R (Bates et al., 2014, p. 4). Statistical tests were performed using type II Wald chi-square analysis of variance using the *car* package in R (Fox & Weisberg, 2018), all effect-sizes were calculated for each fixed effect with the *esc* package in R (Lüdecke, 2019).

When a significant main effect or interaction was found, estimated marginal means post-hoc tests were performed using the *emmeans* package in R (Lenth, 2021) and corrected for multiple comparisons using the False Discovery Rate (FDR) method (Benjamini & Hochberg, 1995).

Two measures were considered to investigate children's performance:

- *Accuracy:* percent correct scores were fitted with a binomial distribution as GLMs can handle non-normally distributed data and, in particular, count data (Bates et al., 2014). Data were thus fitted with a binomial distribution with a logit link function

using the *glm* function from the core *stats* package in R. We did not use mixed models for the accuracy because one proportion of correct response was obtained for each condition and each subject, thus rendering the intra-subject variability null.

- *Response times*: As recommended by Lo and Andrews (2015), RTs were not log-transformed to fit a gaussian distribution which can produce additive effects. Since raw RTs distribution can be fitted either with an inverse gaussian or a gamma distribution, we compared four models, fitted with either a gamma or an inverse gaussian distribution and with either an identity link function or a log-link function. The model with the lowest AIC was then chosen. The random-effects structure was built according to Bates et al. (2015)'s specifications: we used the subject variable as an intercept and we started by including all within-subject fixed-effects as random slopes (i.e., all effects were allowed to vary across subjects) and dropped slopes factors one by one. After rejecting the non-converging models, we chose the model with the lowest AIC.

We also compared measures between the oldest children (5^{th} grade) and adults. We extracted adults' performance for the two noise conditions that children underwent (-3 and +3 dB SNR) and we performed similar GLM(M)s than the previously described ones (on accuracy and RTs), replacing the *Grade* factor by a *Group* factor with two levels: 5^{th} grade and adults. We kept the *phonological difficulty* and *noise condition* factors.

<u>Impact of parameters potentially influencing performance</u>

- Repetition and learning: Influence of trials repetition within a block

The effect of trials repetition within a block was assessed by analyzing children's accuracy (percentage of correct responses) as a function of the series of 4-image sets presented during a block (as each block consisted of four consecutive 12-trial series). Accuracies were calculated per subject, per block, and per series and were fitted using the *glm* function from the

core *stats* package in R with a binomial distribution and a logit link function with grade (KG to 5th grade), phonological difficulty (easy/difficult), noise condition (unfavorable/favorable) and series (1 to 4) as factors. In addition, the reproducibility within a block was assessed by a two-way random effect interclass correlation.

- Word/picture familiarity

To assess potential variations in word/picture familiarity across different grades and to rule out the influence of differences in picture familiarity on potential group variations in speech perception, we ranked the 24 target words based on average accuracy (percentage of correct responses) within each age group. The ranking ranged from the highest score (rank=1) to the lowest score (rank=24) for both easy and difficult phonological conditions, using the same method as Fallon et al. (2000). We then performed ranked correlations (Spearman method) between the ranks for the different age groups, and between the rank of each age group and the overall average ranking.

- Psycholinguistic parameters

The effect of two psycholinguistic factors (word frequency of occurrence and number of phonological neighbors, see above) calculated from the Lexique database (New et al., 2004), was assessed on children's accuracy. For frequency of occurrence, mean accuracy was calculated for each target-word (24 in total), each phonological difficulty (easy/difficult) and each noise condition (unfavorable/favorable). A GLMM was fitted on proportion data (number of correct responses over number of trials) and fitted with a binomial distribution, a logit link function and target-word as random factor. In each GLMM, the effect of the frequency of occurrence factor was tested in addition with the grade factor (six levels: kindergarten to 5th grade), the phonological difficulty factor (easy/difficult) and the noise condition factor (unfavorable/favorable).

For phonological neighbors, mean accuracy was calculated for each four-image set (six sets of the easy difficulty, six sets for the difficult difficulty). One GLMM was fitted for each difficulty level, as by construction, the number of unique phonological neighbors for a set is smaller for easy sets than difficult sets. These two GLMMs were performed on accuracy and fitted with a binomial distribution, a logit link function, and four-image set as random factor. In each GLMM, the effect of the phonological neighbors factor was tested in addition with the grade factor (six levels: kindergarten to 5th grade) and the noise condition factor (unfavorable/favorable). Statistical tests were performed using type II Wald chi-square analysis of variance using the *car* package in R (Fox & Weisberg, 2018).

All effect sizes are reported as Cohen's d (Cohen, 1992).

Results

Adults data: psychometric parameters

Detection thresholds were computed for each participant and for each phonological condition. At the individual level, 74% of the adult participants displayed significantly different detection thresholds between the easy and the difficult conditions (fourteen out of the nineteen participants, Figure 1a). At the group level, detection thresholds averaged across participants were significantly different between the easy and the difficult condition (mean threshold for the easy condition = -10.6 dB; mean threshold for the difficult condition = -6.3 dB; $t(18) = 10.57$, $p < .001$, Figure 1b). Thus, in adults, to achieve similar performance in easy and difficult phonological conditions, an increase of ~4.3 dB of SNR is necessary in the difficult condition relative to the easy one. The average scores achieved by adults at the SNR levels used with children were as follows: in the favorable noise condition (+3 dB SNR), 97% (SD = 16%) for the easy condition and 95% (SD = 22%) for the difficult condition. In the unfavorable noise condition (-3 dB SNR), the scores were 96% (SD = 19%) for the easy condition and 86% (SD = 35%) for the difficult condition. As the adults' data were collected across multiple SNRs to

specifically obtain psychometric parameters using accuracy, no RT analysis was performed for all SNRs in the adults' data. However, adults' RTs were analyzed in comparison to children's RTs for the same fixed SNR levels (see the Comparison between children and adults' performance section).

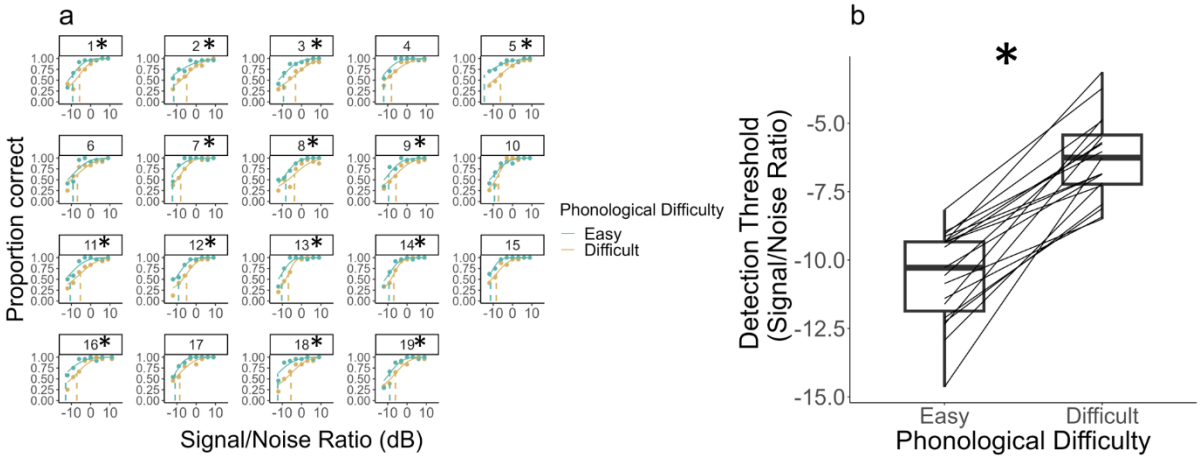

Fig.1: subject-level and group-level effect of the phonological difficulty on psychometric thresholds in adults. a: fitted psychometric functions on the proportion of correct response for each subject (numbered panels) and each phonological difficulty (blue lines: easy/orange lines: difficult) as a function of the SNR during the task. Dotted lines represent the projection of the estimated SNR threshold needed to achieve 62.5% correct response for each condition. Asterisks above the participant number indicate significantly different thresholds for the two conditions at the individual level. b: average detection thresholds across participants as a function of the phonological difficulty. Individual data are represented by black lines. Asterisk indicate a significant difference between the two conditions.

Development of hearing-in-babble noise

GLM(M) were performed on accuracy data and response times with Grade (kindergarten to 5th grade), Phonological Difficulty (easy/difficult), and Noise Condition (unfavorable/favorable) as factors. Mean accuracy and response times are shown in Figure 2a and 2b. Complete statistical results are shown in Table 2 and we present significant effects, interactions, and their follow-up post-hoc tests below.

Table 2: Type II Wald chi-square analysis of deviance results on the GLM with binomial modeling for the accuracy measure and on the GLMM with inverse gaussian modeling for the correct response time measure for children data. All main effects were tested for each measure: Grade (six

levels: KG to 5th grade), Phonological Difficulty (two levels: easy/difficult), Noise condition (two levels: unfavorable/favorable) and their interactions (see text for details). *Significant effects are depicted in bold font (p-values < .05). df: degrees of freedom. χ^2: chi-square test. p: p-value. Cohen's d: standardized mean difference.*

Measure	*Factor*	*Df*	*χ^2*	*p*	*Cohen's d*
Accuracy	**Grade**	**5**	**22.90**	**3.5e - 04**	**1.17**
	Phonological Difficulty	**1**	**61.90**	**3.6e - 15**	**2.97**
	Noise condition	**1**	**32.02**	**1.5e - 08**	**1.49**
	Grade:Difficulty	5	0.50	.99	0.15
	Grade:Noise	5	1.30	.94	0.24
	Difficulty:Noise	1	2.43	.12	0.33
	Grade:Difficulty:Noise	5	0.83	.98	0.20
Response times	**Grade**	**5**	**12.90**	**.024**	**0.49**
	Phonological Difficulty	**1**	**9.49**	**.0021**	**0.69**
	Noise Condition	1	0.0080	.93	0.02
	Grade:Difficulty	5	3.43	.63	0.10
	Grade:Noise	5	3.52	.62	0.10
	Difficulty:Noise	1	0.40	.53	0.13
	Grade:Difficulty:Noise	5	0.39	.99	0.0012

Accuracy. The following model was used for accuracy: *correct responses/number of trials ~ grade* noise * difficulty, family = binomial(link=logit).* The main effect of Grade was significant ($\chi^2(5) = 22.90$, $p < .001$, $d = 1.17$) revealing better performance for older children. Post-hoc tests revealed a significantly lower performance for KG children compared to 3rd, 4th, and 5th graders ($p < .0029$), significantly lower performance for 1st graders compared to 5th graders ($p < .030$), marginally significant lower performance for 2nd graders compared to 5th graders ($p = .060$). The main effect of Phonological Difficulty was significant ($\chi^2(1) = 61.90$, $p < .001$, $d = 2.97$) with lower performance in the difficult condition. The main effect of Noise

Condition was also significant ($\chi^2(1) = 32.02$, $p < .001$, $d = 1.49$) with lower performance for the unfavorable SNR (-3 dB). No interaction effect was significant ($p > .12$).

Response times. The model that best fitted our data was the one using an inverse gaussian distribution with a log-link function, with random effect structure including Difficulty and SNR as slopes (*response times ~ grade*difficulty*noise (noise + difficulty | subject), family=inverse.gaussian(link=log)).* The main effect of grade was significant ($\chi^2(5) = 12.90$, $p = .024$, $d = .49$). Post-hoc tests only revealed a marginally significantly faster response time for 5th graders compared to KG ($p = .050$) and for 5th graders compared to 2nd graders ($p = .050$). Interestingly, longer response times were obtained for the easy phonological condition compared to the difficult one (main effect of Phonological Difficulty ($\chi^2(1) = 9.49$, $p = .0020$, $d = .69$)). No effect of Noise Condition or interaction effects were found ($p > .53$).

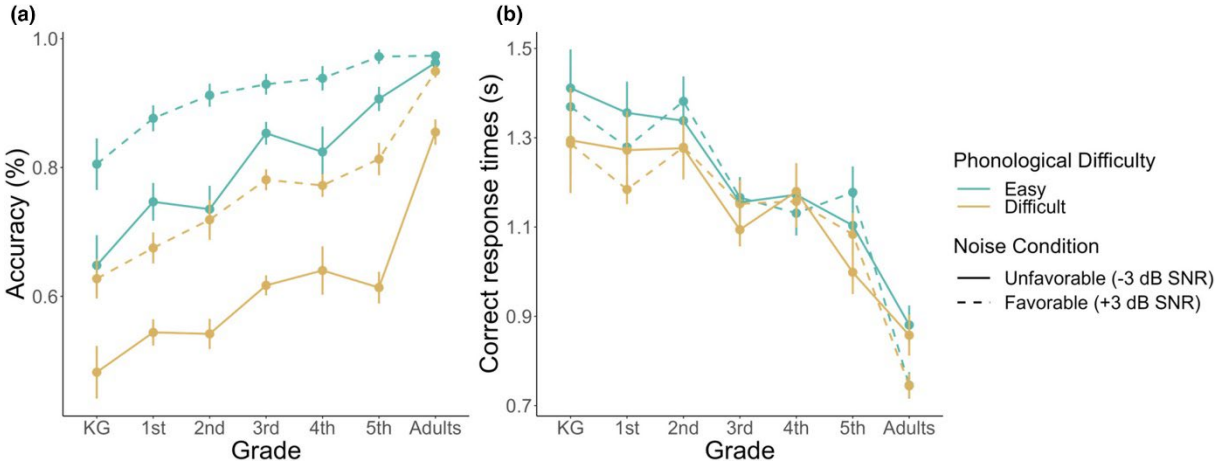

Fig.2: (a) mean performance (% of correct response) and (b) mean correct response times after onset of the sound (in seconds) as a function of Noise Condition and Phonological Difficulty for all grades and for adults. Response times are for words correctly recognized in both easy and difficult conditions (per participant). Error bars represent standard deviations around the mean. Blue color represents the easy condition; orange color represents the difficult condition. Solid lines represent the favorable noise condition, dashed lines the unfavorable noise condition. KG: Kindergarten. Adult's average response times for the easy and difficult condition in the favorable noise condition were similar, so that in the figure (b), the triangle for the easy condition hides the one for the difficult condition.

Comparison between children and adults' performance

Comparison of children's performance and adult's psychometric parameters

In order to obtain a representative adult psychometric curve, we fitted a psychometric curve to accuracy averaged across adult data separately for the easy and the difficult phonological conditions (respectively left and right panels in Figure 3). We then projected children's performance on the overall adult psychometric curve as a function of their grade and of the SNR used (-3 dB and +3 dB, respectively top and bottom panels in Figure 3). We thus obtained the value of the SNR at which an adult would have to be tested to obtain the same performance as children. This representation allows to highlight the relative deficit of speech in noise of children in comparison to adults, using the same task and stimuli. Furthermore, it allows a better comparison across studies that use different languages and parameters, where the young normally-hearing adult group acts as a reference. Indeed, the « SNR gap » (i.e., the difference between the SNR used to test children and the SNR corresponding to the same performance in adult data) goes up to 10 dB for the youngest children and the minimal one is 1 dB for the oldest children in the easy condition and for a testing SNR of -3 dB.

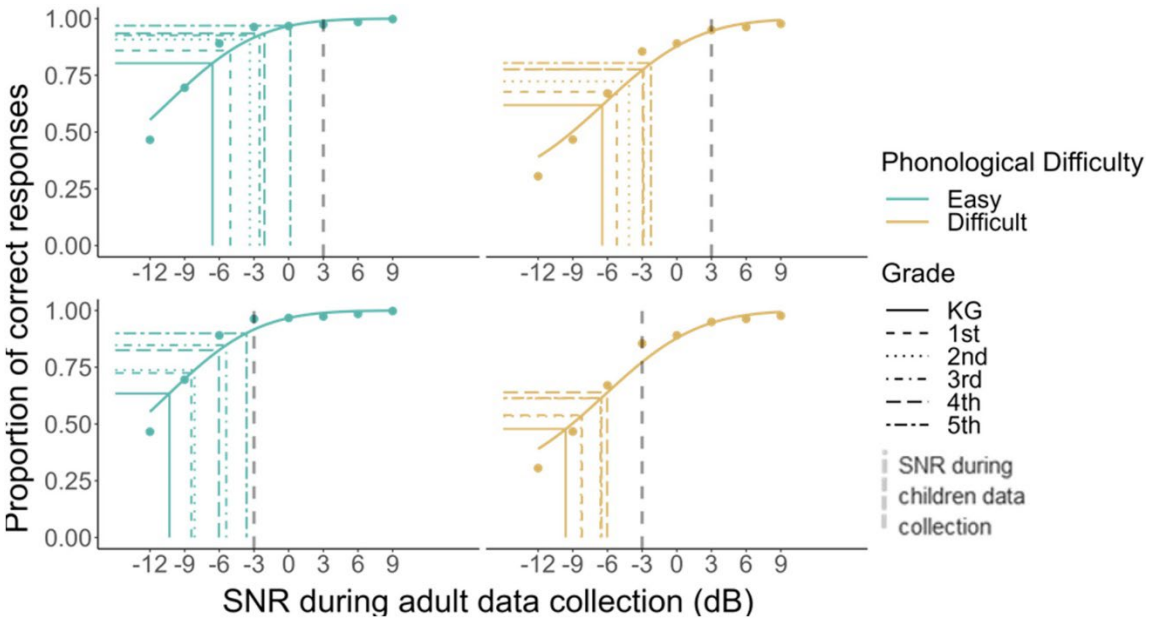

Fig.3: projection of children's performance by grade on the adult's psychometric function (derived from the average data of the 19 adult participants). Left panels and blue lines for the easy condition; right panels and orange lines for the difficult condition. Top panels: the projected performance of children when they underwent the -3 dB SNR condition. Bottom panels: the projected performance of children when they underwent the +3 dB SNR condition.

544 *Filled circles represent the average performance for adults in the corresponding SNR*
545 *condition.*

546 *Comparison between performance of oldest children and performance of adults*

547 To compare performance of the oldest children (5th grade) with performance of adults,
548 we performed GLM(M)s on the same measures as for children (accuracy and response times)
549 with Group (adults (n = 19) and children (5th graders, n = 15)), Phonological Difficulty
550 (easy/difficult), and Noise Condition (unfavorable/favorable) as factors. Mean accuracy and
551 response times are shown in Figure 2a and 2b. Complete statistical results are shown in Table
552 3 and we present significant effects, interactions, and their follow-up post-hoc tests below.

553

***Table 3:** Type II Wald chi-square analysis of deviance results on the GLM with binomial modeling for the accuracy measure and on the GLMM with inverse gaussian modeling for the correct response times measure, to compare data in the oldest children with adult data. All main effects were tested for each measure: Group (two levels: 5th grade vs. adults), Phonological Difficulty (two levels: easy/difficult), Noise Condition (two levels: unfavorable/favorable) and their interactions (see text for details).* *Significant effects are depicted in bold font (p-values < .05). df: degrees of freedom. X^2: chi-square test. P: p-value. Cohen's d: standardized mean difference.*

Measure	*Factor*	*Df*	*χ^2*	*p*	*Cohen's d*
Accuracy	**Group**	**1**	**6.16**	**.013**	**0.94**
	Phonological Difficulty	**1**	**9.55**	**.0020**	**1.25**
	Noise Condition	**1**	**4.04**	**.044**	**0.74**
	Group:Difficulty	**1**	**4.08**	**.043**	**0.74**
	Group:Noise	1	1.19	.28	0.38
	Difficulty: Noise	1	1.93	.17	0.49
	Group:Difficulty: Noise	1	0.32	.57	0.20
Response times	**Group**	**1**	**26.92**	**2.1e– 07**	**3.9**
	Phonological Difficulty	1	3.20	.074	0.53
	Noise Condition	1	2.24	.13	0.65

Group:Difficulty	1	2.31	.13	0.54
Group:Noise	**1**	**12.63**	**3.8 -04**	**1.54**
Difficulty:**Noise**	1	1.13	.29	0.37
Group:Difficulty:**Noise**	1	0.0020	.97	0.015

554

555 *Accuracy.* The following model was used for accuracy: *correct responses/number of*
556 *trials ~ group*difficulty*noise, family = binomial(link=logit)).* All main effects were
557 significant: Group ($\chi^2(1) = 6.16$, $p = .013$, $d = 0.94$ with adults having better performance than
558 5th graders, Phonological Difficulty ($\chi^2(1) = 9.55$, $p = .0020$; $d = 1.25$) with better performance
559 in the easy condition, and Noise Condition ($\chi^2(1) = 4.04$, $p = .044$, $d = .74$) with better
560 performance for the favorable noise condition. In addition, a significant Group-by-Difficulty
561 interaction was found ($\chi^2(1) = 4.08$, $p = .043$, $d = 0.74$) with significantly better performance
562 in the easy condition compared to the difficult one for 5th graders ($p < .001$) but not for adults
563 ($p = .29$). Note that the absence of difficulty effect in the adult group was expected as in the
564 present analysis, we are comparing children and adults' data recorded at exactly the same
565 SNR levels, and therefore, there is a ceiling effect for the adult's data.

566 *Response times*. The model that best fitted our data was the one using an inverse
567 gaussian distribution with a log-link function, and a random effect structure including Difficulty
568 and SNR as slopes (*response times ~ group*difficulty*noise + (noise + difficulty | subject),*
569 *family = inverse.gaussian(link=log)).* There was a marginally significant effect of Phonological
570 Difficulty ($\chi^2(1) = 3.20$, $p = .074$, $d = 0.65$) for which longer response times were found for the
571 easy condition compared to the difficult condition. Critically, the main effect of Group was
572 significant ($\chi^2(1) = 26.92$, $p < .001$, $d = 3.90$) as well as the Group-by-Noise interaction ($\chi^2(1)$
573 $= 12.63$, $p < .001$, $d = 1.54$), that revealed longer response times for the unfavorable noise
574 condition (-3 dB) for children ($p = .018$) but not for adults ($p = .66$).

575 ***Impact of learning, familiarity, and psycholinguistic parameters***

576 *Learning in the course of a block: Influence of the repetition of the trials*

577 The model used in this analysis of accuracy is the following: *correct responses/number*
578 *of trials ~ grade*difficulty*noise*series, family = binomial(link=logit)).* As expected from the
579 developmental trajectory analysis above, effects of Grade ($\chi^2(5) = 22.90$, $p < .001$, $d = 1.17$),
580 Phonological Difficulty ($\chi^2(5) = 62.19$, $p < .001$, $d = 2.99$) and Noise Condition ($\chi^2(5) = 32.18$,
581 $p < .001$, $d = 1.49$) were found. No effect of series was observed ($\chi^2(3) = 0.32$, $p = .96$, $d = 0.12$)
582 indicating an absence of significant increase or decrease of performance within a block. No
583 interaction effect was found (all $p > .45$). Reproducibility was good with an interclass
584 correlation coefficient of .60 ($F = 2.57$, $p < .001$) for the unfavorable noise condition and of .68
585 ($F = 3.14$, $p < .001$) for the favorable noise condition.

586 *Word picture familiarity*

587 Ranking of performance per word were correlated between groups, and all pairwise
588 correlations were significant ($p < .001$, r ranging from .69 to .87). Correlations between the
589 ranking of each group with the overall ranking were also significant ($p < .001$, r ranging from
590 .83 (KG) to .92 (2^{nd} grade)).

591 *Influence of psycholinguistic parameters*

592 *Frequency of occurrence*

593 The effect of the Frequency of occurrence of target words, Grade (kindergarten to 5^{th}
594 grade), Phonological Difficulty (easy/difficult), and Noise Condition (unfavorable/favorable)
595 factors was tested on accuracy. Except for the expected Grade ($\chi^2(5) = 24.30$, $p < .001$, $d =$
596 0.86), Phonological Difficulty ($\chi^2(1) = 66.74$, $p < .001$, $d = 3.39$), and Noise Condition ($\chi^2(1)$
597 $= 34.61$, $p < .001$, $d = 1.58$) effects (already found in the previous analysis), no other significant

598 effect or interaction was found (all $p > .11$), with, in particular, no significant effect of the
599 frequency of occurrence.

600 *Phonological neighbors*

601 The effect of the Number of unique phonological neighbors within each 4-images set,
602 Grade (kindergarten to 5th grade), and Noise Condition (unfavorable/favorable) factors were
603 tested on accuracy. One model was fitted for each Phonological Difficulty (easy/difficult). The
604 expected effect of Grade was found for the easy ($\chi^2(5) = 11.16$, $p = .048$, $d = 0.43$) and the
605 difficult conditions ($\chi^2(5) = 12.49$, $p = .029$, $d = 0.47$), as well as the Noise Condition effect for
606 the easy ($\chi^2(1) = 9.55$, $p = .0020$, $d = 0.69$) and the difficult ($\chi^2(1) = 26$, $p < .001$, $d = 1.27$)
607 conditions. All other effects or interactions were non-significant ($p > .11$).

608 **Discussion**

609 The present study depicts the developmental aspects of speech perception in multi-talker
610 babble noise, in French language, in primary school children, using a newly developed French
611 closed-set test with a 4AFC paradigm. Additionally, the exploration of children's response
612 times allowed to unveil an effect of competition between phonologically similar words within
613 the lexicon. Finally, we explored the impact on performance of several factors that have
614 implications for the design of central hearing screening test: reproducibility, word familiarity,
615 and psycholinguistic factors.

616 *Developmental trajectory of speech-in-babble-noise perception*

617 The main results show definite steps into the development, with significant differences
618 in accuracy between 5-year-olds (KG) and 7-, 9-, and 10-year-olds (3rd, 4th and 5th grade),
619 between 6-year-olds (1st grade) and 9-10-year-olds (4th and 5th grade), and between 7- and 10-
620 year-olds (2nd and 5th grade), revealing a major maturation step around 7 years of age (2nd
621 grade). The accuracy obtained by the oldest children (10-year-olds, 5th grade) were equivalent

622 to a difference of more than 3 dB SNR on average with normally hearing young adults, showing
623 that speech perception in one of the most common, but challenging situation (i.e., speech-in-
624 babble noise), is far from mature during nearly the entirety of primary and secondary school
625 attendance years. Those results are in agreement with Wilson et al. (2010)'s large normative
626 study, using an open set task in babble noise (Word in Noise test). Indeed, the authors showed
627 that the largest improvement in recognition performance occurred between 6- and 7-year-olds,
628 then a slight improvement between 7 and 9-year-olds, and performance remained fairly constant
629 between 9- and 12-year-olds, with a significant difference between 12-year-olds and young
630 adults.

631 It could be argued that phonological awareness maturation could account for the lower
632 scores of the youngest children, especially for Kindergarten children. For instance, using an
633 oddity non-sense syllable recognition task (presented with 3 auditory non-sense syllables - two
634 identical and one odd), Hnath-Chisolm et al. (1998) obtained significantly lower accuracy for
635 5- to 7-year-olds compared to 7- to-9-, or 9-to-11-year-olds. Those differences could possibly
636 be attributed to cognitive and phonological development for children under 7 years (Hnath-
637 Chisolm et al., 1998). In the present study, even though we obtained the expected greater
638 accuracy for phonologically distant words than for close words, we did not observe any
639 interaction between grade and difficulty in the children's data analysis. This fairly parallel
640 evolution of accuracy in easy versus difficult phonological conditions with age (see also Figure
641 4) argues against an important influence of phonological awareness maturation in the evolution
642 of accuracy as a function of age, at least between 5 and 10 years of age. However, we observed
643 a weakly significant interaction between difficulty and group when comparing children's data
644 with adults' data (with a difference between difficult and easy conditions in children only).
645 Therefore, we cannot dismiss the potential impact of later maturation of phonological
646 awareness on the current test. However, the lack of a difficulty effect in adults could

alternatively be attributed to a ceiling effect, particularly influenced by the favorable SNR condition (+3 dB SNR). Additionally, there might be a potential ceiling effect on response times for older children, as indicated by the statistical weakness of the interaction.

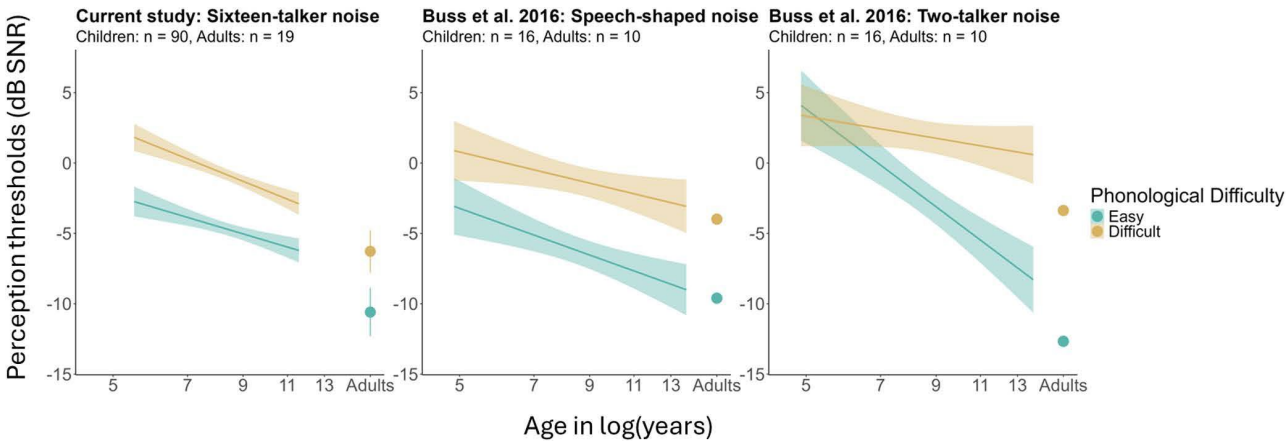

Fig.4: comparison between the current study and Buss et al. (2016). For the current study a SNR threshold was derived for each child and each condition from the 2 points obtained along the psychometric curve and adding 2 other points: one at 100% correct score for a 15 dB SNR and one at 25% chance level for a -15 dB SNR. We then fitted a linear regression as a function of the decimal logarithm of age in years, as in Buss et al. (2022). For the Buss et al. (2016)'s figures, data were recovered from their Figure 2 using the Engauge Digitizer software (https://markummitchell.github.io/engauge-digitizer/). For the current study, the effect of age was significant ($p < .001$) and slopes were -0.53 and -0.75 for easy and difficult conditions respectively. For Buss et al.'s speech-shaped noise, the effect of age was significant ($p < .014$) and slopes were -0.64 and -0.47 for the easy and difficulty conditions respectively. For the two-talker noise, the effect of age was significant only for the easy condition ($p < .001$) but not for the difficult condition ($p = .14$). Slopes were -1.43 and -0.03. The objective of this comparison with the existing literature is to evaluate the contribution of the informational/energetic masking component within our 16-talker babble noise. This is in contrast to the study we are comparing with (Buss et al., 2016), which employed a two-talker masker with a pronounced informational masking component and speech-shaped noise with a strong energetic masking component.

The difference between phonologically similar and dissimilar words obtained here (as measured with the procedure described in Figure 4, e.g., a threshold difference of 3.8 dB SNR in children and 4.3 dB in adults) appears to be lower than the one obtained by Buss et al. (2016) in a 4AFC task, in a speech shaped noise (about 5 dB in children and 7 dB in adults). Differences in language (French versus English), in the degree of similarity between the target words and the foils, and in the type of noise used, could account for that. In addition, Buss et al. (2016) used the 2 conditions in 2 different blocks, determining the SNR threshold using an adaptative

675 method, whereas we used a fixed SNR method, where easy target words and difficult trials were
676 randomly intermixed. Both in the present study and in Buss et al.'s data (Figure 4), performance
677 improved with age for both easy and difficult phonological conditions. It is interesting to note
678 though, that our results match better with Buss et al. (2016)'s results in speech shaped noise
679 than with their results in 2-talker masker. Indeed, in the 2-talker masker, they obtained a
680 significantly sharper improvement of accuracy with age in the easy condition than in the
681 difficult condition, with a difference of almost 10 dB SNR between both conditions in adults.
682 This similarity between Buss et al. (2016)'s results in speech shaped noise and ours can be
683 easily attributed to the 16-talker babble noise we used, that involves both energetic and
684 informational masking, and is closer, in performance, to speech shaped noise than the strong
685 informational masking of a 2 talker noise (Rosen et al., 2013). The 16-talker babble noise thus
686 offers a good mix between energetic and informational masking. The definite existence of an
687 informational masking component, involving central auditory processing, in this 16-talker
688 babble noise, has been shown in a case of venous cerebral infarct in an adult patient (Bourgeois-
689 Vionnet et al., 2020). Indeed, this patient showed a specific deficit in the 16-talker babble noise
690 speech perception in the ear contralateral to the infarct, without any deficit in a speech shaped
691 noise of the same spectrum, using the same test as in the present study.

692 The informational masking component could explain the absence of mature
693 performance at 10 years of age, as already pointed out by Elliott (1979), who observed
694 significant differences between 13-year-olds and 17-year-olds in speech-in-babble-noise
695 perception. Indeed, speech-in-babble-noise perception depends on a wide range of cognitive
696 abilities, whose development extend well beyond 11 years of age, such as sustained attention
697 (Betts et al., 2006; Hoyer et al., 2021; Thillay et al., 2015), distractibility (Hoyer et al., 2021;
698 Wetzel et al., 2009), impulsivity and motor control (Booth et al., 2003; Wright et al., 2003),
699 verbal short-term memory (Alloway et al., 2006; Gathercole et al., 2004), and the ability to use

pitch cues to focus on the target, including pitch short-term memory (Ginzburg et al., 2022). This explains the great difference in the amount of noise that children and young adults can sustain to achieve the same performance, with more than 3 dB of difference in SNR between 10-year-olds and young adults, and more than 9 dB between 5-year-olds and adults in the difficult condition (see Figure 3). The latter result is very close to the 10 dB threshold difference reported between 5-year-olds and young adults, in a 2-talker masker, in Buss et al. (2019). This emphasizes the relevance of using a babble-noise in a speech perception screening test as speech in babble noise is often associated with several deficits, such as attention or short-term memory impairments, and would be more sensitive to all those deficits than perception in speech-shaped noise.

Response times reveals competition processes within the lexicon

Response times reflected the expected slower speed of processing from the youngest to the eldest children (with about 500 ms of difference in response times observed here), due to processing speed maturation (Kail, 1991), in addition with slower response times for the oldest children as compared to adults. In a similar 4AFC task, Rigler et al. (2015) obtained a 400 ms difference between 16-year-olds and 9-year-olds having to click on the picture of a target word amongst 4 pictures, revealing that processing speed, as in the present study, undergoes maturation well beyond 10 years of age. Unexpectedly, we observed longer response times for the easy condition compared to the difficult condition, whereas for accuracy, the expected higher accuracy for the easy condition over the difficult one was observed in both children and adults.

One mechanism that could explain these counterintuitive results is the organisation of mental lexicon with distant words engendering a greater lexical competition within the mental lexicon than phonologically close words. Using eye tracking with the visual world paradigm (introduced by Huettig et al., 2011), Rigler et al. (2015) showed more activation for competition

words for 9-year-olds than for 16-year-olds. Indeed, according to most models for spoken word recognition (e.g., Frauenfelder & Floccia, 1999, for a review), words are organised within the mental lexicon as clusters of phonologically similar words (e.g., the neighbourhood activation model, Luce & Pisoni, 1998) and words sharing the same onset sounds (e.g. the Cohort model Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989). As a word begins to be heard, several potential candidates are activated within the mental lexicon and the process of selecting the correct target involves competition within the several activated candidates (e.g., the trace model, Dahan et al., 2001; McClelland & Elman, 1986). Several studies, using open response tests, have shown that words with numerous phonological neighbours are more difficult to recognize and take a longer time to process, especially when those phonological neighbours are of higher occurrence frequency than the target word, than words with a low number of phonological neighbours (Dirks et al., 2001; Kirk et al., 1995; Krull et al., 2010). We would thus expect a greater response time when more numerous candidates are activated. In the present task, candidates are not only activated by the word heard, but, as well, by the three other words displayed in the pictures (i.e., the target and the three foils). Here, in the difficult condition, the words are phonologically close, so the phonological patterns that are activated by the target words and the foils are similar, and are less numerous than the phonological patterns activated in the easy condition. Indeed, because the words, in the easy condition, are phonologically dissimilar, they would activate four different clusters of phonological patterns, demanding a greater number of comparisons between the target word and all the phonological similar words to the foils. In other words, a difficult set (e.g., /ku/, /kul/, /ru/, /ruʒ/) would activate fewer candidates in the participant's lexicon because all of these words share common phonological neighbors. Conversely, an easy set (e.g., /bɑ̃/, /flœʀ/, /mɛ̃/, /trwa/) would activate a larger number of candidates due to the dissimilarity between each word in the set. Thus, the

749 greater number of activated candidates could account for the longer response times in the 'easy'
750 condition, even if this condition shows a higher accuracy.

751 *Central hearing screening test design: impact of reproducibility, word familiarity, and*
752 *psycholinguistic factors*

753 Another major result obtained here is the absence of short-term practice effects. Indeed,
754 by using a fixed SNR ratio method, rather than an adaptative one, we were able to assess the
755 potential changes from the beginning of a block (12 first sets) and the end (12 last sets). Indeed,
756 although we avoided the major source of practice effect by not using trial per trial visual
757 feedback (Munro & Lutman, 2005), there are still a wide range of reasons why substantial
758 modifications could occur. Firstly, the same sets of 4 pictures are used throughout, so more
759 familiarity with the specific phonetic contrasts to listen to would be expected as the sets repeat
760 themselves, especially in the difficult condition. Secondly, as the same feminine voice for the
761 target words is used throughout, better identification and separation of the talker's voice from
762 the babble background could be expected with practice. Those factors could account for a
763 significant improvement of the accuracy with time. Conversely, a deterioration of accuracy
764 could account for lack of maintenance of the necessary listening effort, involving a strong
765 focused attention component, during the entire 2-minute-long blocks, for such a challenging
766 task. We did not observe any significant difference in accuracy between series, whether the
767 whole population was considered, or only the youngest ones (i.e., first quartile, < 7.3 years old).
768 This shows the absence of learning during the course of one block with the task used here, and,
769 as well, the absence of loss of motivation during the course of a block, which is particularly
770 important when addressing young children. Indeed, to maintain motivation, Fallon et al., (2000)
771 used a "game-like" automated visual feedback for correct and incorrect responses in a 4AFC
772 task in babble noise. The authors obtained a significant improvement between the first 20 trials
773 (82%) and the last 20 trials (86%) of a 40-word block at constant SNR in their children and

adult groups. In order to be able to use the test later in an adaptative manner, a stability of scores, at least within one block, is recommended.

Performance is likely to be influenced by familiarity with the target words depending on children's age (G. D. Brown & Watson, 1987), as contextual influences on those words develop. To assess this possibility, Wilson et al. (2010) checked the performance of their youngest groups in quiet, and obtained significant lower performance in the 6-year-olds versus the older children. They concluded that part of the 6-year-olds increased threshold was probably due to less familiarity for the target words. Although the pictures we used were designed so that 5-year-olds could recognize them easily, a growing familiarity as age increases is difficult to rule out. To address this question, we ranked each target word according to its score and correlated the ranking in each age group with the overall ranking, as in Fallon et al. (2000), who used a 4AFC task as well. We obtained quite similar results (r ranging from .83 to .92) as in their study ($r > .90$), showing that the relative difficulty of each target word didn't depend on age, and that the improvement of accuracy observed with age was not likely due to an increase in familiarity of the target words with age.

Although the experimental design was not aimed at testing specifically the influence of psycholinguistic factors on the test accuracy, it is important to ascertain whether they have substantial influence on the accuracy at a test aimed at screening for speech in babble noise perception deficits. Indeed, the influence of such psycholinguistic parameters is rarely considered when developing auditory perception tests. Yet, those parameters can introduce unwanted variability in audiological speech perception accuracy (Moulin et al., 2017), especially in interaction with cognitive status and lexical knowledge (Braza et al., 2022; Moulin & Richard, 2015). Elliot et al. (1983) specifically designed a 4AFC task to test the potential influence of occurrence frequency. By comparing scores for the same target words presented with foils of high and low occurrence frequencies, the authors observed a word frequency effect

799 that was significantly greater in adults than in 3- to 7-year-olds. Furthermore, this frequency
800 effect, in children, increased significantly with receptive vocabulary knowledge, which argues
801 for the need to control for that effect when developing audiological tests of speech perception
802 (Elliott et al., 1983). In the present test, no effect of target-word occurrence frequency was
803 found, nor any effect of phonological neighbourhood density. To ensure that scores reflect more
804 auditory perception (Kosky & Boothroyd, 2003; Mendel, 2008) than mental lexicon
805 development, it is important, for audiological assessment, to ensure that there is no significant
806 influence of lexical factors (such as occurrence frequencies and/or phonological density) of the
807 stimuli chosen on the scores.

808 The limitations of this study lie in the relatively moderate sample size in each grade, the
809 limited age range (focusing on primary school years) and the absence of other thorough clinical
810 investigations (e.g., pure-tone audiometry, including high frequencies audiometry, lexical
811 knowledge and other cognitive investigations). However, these investigations would go beyond
812 the aims of the present study. Indeed, to establish normative data with the aim of a screening
813 test, we would need at least 40 children per age group. Several refinements in the words used
814 could be made as well, in order to further decrease the duration of the test, and being able to
815 use it in children younger than 5 years of age.

816 Be that as it may, this study represents a stepping stone towards a screening test. Indeed,
817 this study showed the developmental course of speech perception in a babble noise that
818 combines both energetic and informational masking, in primary school children, and confirms
819 that processes involved in speech perception in babble noise (in particular informational
820 masking) show a definite maturation step around 7 years of age and that they are far from mature
821 at 10 years of age. We also unveiled the effect of lexicon competition processes on speed
822 processing and checked for crucial factors that can influence such central hearing screening
823 tests, such as word occurrence frequency and phonological neighborhood density. This resulted

in a preliminary version of a screening test, that is quick (around 2 minutes for one block), reproducible, engaging for children, devoid of short-term learning effect and that has therefore good potential for screening speech perception in babble noise deficits.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children : Are They Separable? *Child Development*, *77*(6), 1698-1716. https://doi.org/10.1111/j.1467-8624.2006.00968.x

Anderson, K. (2004). The Problem of Classroom Acoustics : The Typical Classroom Soundscape Is a Barrier to Learning. *Seminars in Hearing*, *25*(02), 117-129. https://doi.org/10.1055/s-2004-828663

Anthony, J. L., & Lonigan, C. J. (2004). The Nature of Phonological Awareness : Converging Evidence From Four Studies of Preschool and Early Grade School Children. *Journal of Educational Psychology*, *96*(1), 43-55. https://doi.org/10.1037/0022-0663.96.1.43

Bamiou, D., Musiek, F., & Luxon, L. (2001). Aetiology and clinical presentations of auditory processing disorders—A review. *Archives of Disease in Childhood*, *85*(5), 361-365. https://doi.org/10.1136/adc.85.5.361

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. https://doi.org/10.48550/ARXIV.1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate : A practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

Betts, J., Mckay, J., Maruff, P., & Anderson, V. (2006). The development of sustained attention in children : The effect of age and task load. *Child Neuropsychology*, *12*(3), 205-221.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., & Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, *32*(2), 560-570. https://doi.org/10.1016/j.ridd.2010.12.020

Booth, J. R., Burman, D. D., Meyer, J. R., Lei, Z., Trommer, B. L., Davenport, N. D., Li, W., Parrish, T. B., Gitelman, D. R., & Mesulam, M. M. (2003). Neural development of selective attention and response inhibition. *Neuroimage*, *20*(2), 737-751.

Bourgeois-Vionnet, J., Moulin, A., Hermier, M., Pralus, A., & Nighoghossian, N. (2020). A case of verbal and emotional prosody processing dissociation after a right temporal venous infarct. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *41*(6), 1615-1618. https://doi.org/10.1007/s10072-019-04175-w

Bradlow, A. R., Kraus, N., & Hayes, E. (2003). *Speaking Clearly for Children With Learning DisabilitiesSentence Perception in Noise*. 18.

Braza, M. D., Porter, H. L., Buss, E., Calandruccio, L., McCreery, R. W., & Leibold, L. J. (2022). Effects of word familiarity and receptive vocabulary size on speech-in-noise recognition among young adults with normal hearing. *PLOS ONE*, *17*(3), e0264581. https://doi.org/10.1371/journal.pone.0264581

Bregman, A. S. (1994). *Auditory scene analysis : The perceptual organization of sound*. MIT press.

Brown, D. K., Cameron, S., Martin, J. S., Watson, C., & Dillon, H. (2010). The North American Listening in Spatialized Noise—Sentences Test (NA LiSN-S) : Normative Data and Test-Retest Reliability Studies for Adolescents and Young Adults. *Journal of the American Academy of Audiology*, *21*(10), 629-641. https://doi.org/10.3766/jaaa.21.10.3

Brown, G. D., & Watson, F. L. (1987). First in, first out : Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & cognition*, *15*, 208-216.

875 Buss, E., Felder, J., Miller, M. K., Leibold, L. J., & Calandruccio, L. (2022). Can Closed-Set Word
876 Recognition Differentially Assess Vowel and Consonant Perception for School-Age Children
877 With and Without Hearing Loss? *Journal of Speech, Language, and Hearing Research*, *65*(10),
878 3934-3950. https://doi.org/10.1044/2022_JSLHR-20-00749
879 Buss, E., Hodge, S. E., Calandruccio, L., Leibold, L. J., & Grose, J. H. (2019). Masked Sentence
880 Recognition in Children, Young Adults, and Older Adults : Age-Dependent Effects of Semantic
881 Context and Masker Type. *Ear and Hearing*, *40*(5), 1117-1126.
882 https://doi.org/10.1097/AUD.0000000000000692
883 Buss, E., Leibold, L. J., & Hall, J. W. (2016). Effect of response context and masker type on word
884 recognition in school-age children and adults. *The Journal of the Acoustical Society of*
885 *America*, *140*(2), 968-977. https://doi.org/10.1121/1.4960587
886 Buss, E., Leibold, L. J., Porter, H. L., & Grose, J. H. (2017). Speech recognition in one- and two-talker
887 maskers in school-age children and adults : Development of perceptual masking and
888 glimpsing. *The Journal of the Acoustical Society of America*, *141*(4), 2650-2660.
889 https://doi.org/10.1121/1.4979936
890 Cabbage, K. L., & Hitchcock, E. R. (2022). Clinical Considerations for Speech Perception in School-Age
891 Children With Speech Sound Disorders : A Review of the Current Literature. *Language,*
892 *Speech & Hearing Services in Schools*, *53*(3), 768-785. https://doi.org/10.1044/2022_LSHSS-
893 21-00120
894 Calandruccio, L., Gomez, B., Buss, E., & Leibold, L. J. (2014). Development and Preliminary Evaluation
895 of a Pediatric Spanish–English Speech Perception Task. *American Journal of Audiology*, *23*(2),
896 158-172. https://doi.org/10.1044/2014_AJA-13-0055
897 Calandruccio, L., Leibold, L. J., & Buss, E. (2016). Linguistic Masking Release in School-Age Children
898 and Adults. *American Journal of Audiology*, *25*(1), 34-40. https://doi.org/10.1044/2015_AJA-
899 15-0053

Calandruccio, L., & Zhou, H. (2014). Increase in Speech Recognition due to Linguistic Mismatch Between Target and Masker Speech : Monolingual and Simultaneous Bilingual Performance. *Journal of Speech Language and Hearing Research*, 1089. https://doi.org/10.1044/2013_JSLHR-H-12-0378

Clopper, C. G., Pisoni, D. B., & Tierney, A. T. (2006). Effects of Open-Set and Closed-Set Task Demands on Spoken Word Recognition. *Journal of the American Academy of Audiology*, *17*(05), 331-349. https://doi.org/10.3766/jaaa.17.5.4

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Corbin, N. E., Bonino, A. Y., Buss, E., & Leibold, L. J. (2016). Development of Open-Set Word Recognition in Children : Speech-Shaped Noise and Two-Talker Speech Maskers. *Ear & Hearing*, *37*(1), 55-63. https://doi.org/10.1097/AUD.0000000000000201

Crandell, C. C., & Smaldino, J. J. (2000). Classroom Acoustics for Children With Normal Hearing and With Hearing Impairment. *Language, Speech, and Hearing Services in Schools*, *31*(4), 362-370. https://doi.org/10.1044/0161-1461.3104.362

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access : Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507-534. https://doi.org/10.1080/01690960143000074

Devous, M. D., Altuna, D., Furl, N., Cooper, W., Gabbert, G., Ngai, W. T., Chiu, S., Scott, J. M., Harris, T. S., Payne, J. K., & Tobey, E. A. (2006). Maturation of Speech and Language Functional Neuroanatomy in Pediatric Normal Controls. *Journal of Speech, Language, and Hearing Research*, *49*(4), 856-866. https://doi.org/10.1044/1092-4388(2006/061)

Dirks, D. D., Takayanagi, S., Moshfegh, A., Noffsinger, P. D., & Fausti, S. A. (2001). Examination of the neighborhood activation theory in normal and hearing-impaired listeners. *Ear and Hearing*, *22*(1), 1-13.

925 Ellen Peng, Z., & Litovsky, R. Y. (2021). The Role of Interaural Differences, Head Shadow, and Binaural
926 Redundancy in Binaural Intelligibility Benefits Among School-Aged Children. *Trends in*
927 *Hearing*, *25*, 233121652110453. https://doi.org/10.1177/23312165211045313
928 Elliott, L. L. (1979). Performance of children aged 9 to 17 years on a test of speech intelligibility in
929 noise using sentence material with controlled word predictability. *The Journal of the*
930 *Acoustical Society of America*, *66*(3), 651-653. https://doi.org/10.1121/1.383691
931 Elliott, L. L., Clifton, L. A., & Servi, D. G. (1983). Word frequency effects for a closed-set word
932 identification task. *Audiology*, *22*(3), 229-240. https://doi.org/10.3109/00206098309072787
933 Fallon, M., Trehub, S. E., & Schneider, B. A. (2000). Children's perception of speech in multitalker
934 babble. *The Journal of the Acoustical Society of America*, *108*(6), 3023-3029.
935 https://doi.org/10.1121/1.1323233
936 Flaherty, M. M., Buss, E., & Leibold, L. J. (2019). Developmental Effects in Children's Ability to Benefit
937 From F0 Differences Between Target and Masker Speech. *Ear & Hearing*, *40*(4), 927-937.
938 https://doi.org/10.1097/AUD.0000000000000673
939 Fox, J., & Weisberg, S. (2018). *An R companion to applied regression*. Sage publications.
940 Frauenfelder, U. H., & Floccia, C. (1999). The Recognition of Spoken Words. In A. D. Friederici,
941 *Language Comprehension : A Biological Perspective* (p. 1-40). Springer Berlin Heidelberg.
942 https://doi.org/10.1007/978-3-642-59967-5_1
943 Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and
944 auditory priming on informational masking in speech recognition. *The Journal of the*
945 *Acoustical Society of America*, *115*(5), 2246-2256. https://doi.org/10.1121/1.1689343
946 Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working
947 Memory From 4 to 15 Years of Age. *Developmental Psychology*, *40*(2), 177-190.
948 https://doi.org/10.1037/0012-1649.40.2.177
949 Gelfand, S. A. (1998). Optimizing the Reliability of Speech Recognition Scores. *Journal of Speech,*
950 *Language, and Hearing Research*, *41*(5), 1088-1102. https://doi.org/10.1044/jslhr.4105.1088

Gelfand, S. A. (2003). Tri-Word Presentations With Phonemic Scoring for Practical High-Reliability Speech Recognition Assessment. *Journal of Speech, Language, and Hearing Research*, *46*(2), 405-412. https://doi.org/10.1044/1092-4388(2003/033)

Gijbels, L., Burkhardt, A., Ma, W. A., & Yeatman, J. D. (2024). Rapid online assessment of reading and phonological awareness (ROAR-PA). *Scientific Reports*, *14*(1), 10249. https://doi.org/10.1038/s41598-024-60834-9

Ginzburg, J., Moulin, A., Fornoni, L., Talamini, F., Tillmann, B., & Caclin, A. (2022). Development of auditory cognition in 5- to 10-year-old children : Focus on musical and verbal short-term memory. *Developmental Science*, *25*(3). https://doi.org/10.1111/desc.13188

Guzek, A., & Iwanicka-Pronicka, K. (2022). Analysis of the auditory processing skills in 1,012 children aged 6–9 confirms the adequacy of APD testing in 6-year-olds. *PLOS ONE*, *17*(8), e0272723. https://doi.org/10.1371/journal.pone.0272723

Hall, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children: *Ear and Hearing*, *23*(2), 159-165. https://doi.org/10.1097/00003446-200204000-00008

Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech perception capacity. *Journal of Speech, Language, and Hearing Research: JSLHR*, *41*(1), 94-106. https://doi.org/10.1044/jslhr.4101.94

Hoyer, R. S., Elshafei, H., Hemmerlin, J., Bouet, R., & Bidet-Caulet, A. (2021). Why Are Children So Distractible? Development of Attention and Motor Control From Childhood to Adulthood. *Child Development*, *92*(4). https://doi.org/10.1111/cdev.13561

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing : A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171. https://doi.org/10.1016/j.actpsy.2010.11.003

Jerger, J., Speaks, C., & Trammell, J. L. (1968). A New Approach to Speech Audiometry. *Journal of Speech and Hearing Disorders*, *33*(4), 318-328. https://doi.org/10.1044/jshd.3304.318

977 Jerger, S., Jerger, J., & Abrams, S. (1983). Speech Audiometry in the Young Child: *Ear and Hearing*,
978 *4*(1), 56-66. https://doi.org/10.1097/00003446-198301000-00010
979 Jerger, S., Jerger, J., & Lewis, S. (1981). Pediatric Speech Intelligibility test. II. Effect of receptive
980 language age and chronological age. *International Journal of Pediatric Otorhinolaryngology*,
981 *3*(2), 101-118. https://doi.org/10.1016/0165-5876(81)90026-4
982 Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence.
983 *Psychological Bulletin*, *109*(3), 490-501. https://doi.org/10.1037/0033-2909.109.3.490
984 Kidd, G., & Colburn, H. S. (2017). Informational Masking in Speech Recognition. In J. C. Middlebrooks,
985 J. Z. Simon, A. N. Popper, & R. R. Fay (Éds.), *The Auditory System at the Cocktail Party* (Vol.
986 60, p. 75-109). Springer International Publishing. https://doi.org/10.1007/978-3-319-51662-
987 2_4
988 Killion, M. C., Niquette, P. A., & Gudmundsen, G. I. (2004). Development of a quick speech-in-noise
989 test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired
990 listenersa). *J. Acoust. Soc. Am.*, *116*(4).
991 Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by
992 pediatric cochlear implant users. *Ear and Hearing*, *16*(5), 470-481.
993 Kolk, S. M., & Rakic, P. (2022). Development of prefrontal cortex. *Neuropsychopharmacology*, *47*(1),
994 41-57. https://doi.org/10.1038/s41386-021-01137-9
995 Kosky, C., & Boothroyd, A. (2003). Validation of an On-line Implementation of the Imitative Test of
996 Speech Pattern Contrast Perception (IMSPAC). *Journal of the American Academy of*
997 *Audiology*, *14*(02), 072-083. https://doi.org/10.3766/jaaa.14.2.3
998 Krull, V., Choi, S., Kirk, K. I., Prusick, L., & French, B. (2010). Lexical effects on spoken-word
999 recognition in children with normal hearing. *Ear and Hearing*, *31*(1), 102-114.
1000 https://doi.org/10.1097/AUD.0b013e3181b7892f

1001 Leibold, L. J., & Buss, E. (2013). Children's Identification of Consonants in a Speech-Shaped Noise or a
1002 Two-Talker Masker. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1144-1155.
1003 https://doi.org/10.1044/1092-4388(2012/12-0011)
1004 Leibold, L. J., & Buss, E. (2019). Masked Speech Recognition in School-Age Children. *Frontiers in*
1005 *Psychology*, *10*, 1981. https://doi.org/10.3389/fpsyg.2019.01981
1006 Leibold, L. J., Buss, E., & Calandruccio, L. (2018). Developmental Effects in Masking Release for
1007 Speech-in-Speech Perception Due to a Target/Masker Sex Mismatch. *Ear & Hearing*, *39*(5),
1008 935-945. https://doi.org/10.1097/AUD.0000000000000554
1009 Leibold, L. J., Buss, E., Miller, M. K., Cowan, T., McCreery, R. W., Oleson, J., Rodriguez, B., &
1010 Calandruccio, L. (2024). Development of the Children's English and Spanish Speech
1011 Recognition Test : Psychometric Properties, Feasibility, Reliability, and Normative Data. *Ear &*
1012 *Hearing*, *45*(4), 860-877. https://doi.org/10.1097/AUD.0000000000001480
1013 Lenth, R. V. (2021). *emmeans : Estimated Marginal Means, aka Least-Squares Means*.
1014 https://CRAN.R-project.org/package=emmeans
1015 Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in
1016 congenital amusia : Impaired recognition, but preserved emotional intensity.
1017 *Neuropsychology*, *32*(7), 880-894. https://doi.org/10.1037/neu0000461
1018 Lewis, H. D., Benignus, V. A., Muller, K. E., Malott, C. M., & Barton, C. N. (1988). Babble and Random-
1019 Noise Masking of Speech in High and Low Context Cue Conditions. *Journal of Speech,*
1020 *Language, and Hearing Research*, *31*(1), 108-114. https://doi.org/10.1044/jshr.3101.108
1021 Linares, D., & López-Moliner, J. (2016). quickpsy : An R Package to Fit Psychometric Functions for
1022 Multiple Groups. *The R Journal*, *8*(1), 122-131. https://doi.org/10.32614/RJ-2016-008
1023 Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The*
1024 *Journal of the Acoustical Society of America*, *117*(5), 3091-3099.
1025 https://doi.org/10.1121/1.1873913

Lo, S., & Andrews, S. (2015). To transform or not to transform : Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. https://doi.org/10.3389/fpsyg.2015.01171

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words : The neighborhood activation model. *Ear and hearing*, *19*(1), 1.

Lüdecke, D. (2019). *esc : Effect Size Computation for Meta Analysis (Version 0.5.1)*. https://doi.org/10.5281/zenodo.1249218

Majerus, S., & Cowan, N. (2016). The Nature of Verbal Short-Term Impairment in Dyslexia : The Importance of Serial Order. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01522

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words : The importance of word onsets. *Journal of Experimental Psychology: Human perception and performance*, *15*(3), 576.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86.

McCreery, R. W., Miller, M. K., Buss, E., & Leibold, L. J. (2020). Cognitive and Linguistic Contributions to Masked Speech Recognition in Children. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3525-3538. https://doi.org/10.1044/2020_JSLHR-20-00030

Mendel, L. L. (2008). Current considerations in pediatric speech audiometry. *International Journal of Audiology*, *47*(9), 546-553. https://doi.org/10.1080/14992020802252261

Moulin, A., Bernard, A., Tordella, L., Vergne, J., Gisbert, A., Martin, C., & Richard, C. (2017). Variability of word discrimination scores in clinical practice and consequences on their sensitivity to hearing loss. *European Archives of Oto-Rhino-Laryngology: Official Journal of the European Federation of Oto-Rhino-Laryngological Societies (EUFOS): Affiliated with the German Society*

for Oto-Rhino-Laryngology - Head and Neck Surgery, *274*(5), 2117-2124. https://doi.org/10.1007/s00405-016-4439-x

Moulin, A., Favier, C., Richard, C., & Farget, V. (2013). French Babble Noises for Speech Perception Tests. Development and Spectro-temporal analysis. *Annales D'oto-Laryngologie Et De Chirurgie Cervico Faciale*.

Moulin, A., & Richard, C. (2015). Lexical Influences on Spoken Spondaic Word Recognition in Hearing-Impaired Patients. *Auditory Cognitive Neuroscience*, 476. https://doi.org/10.3389/fnins.2015.00476

Munro, K. J., & Lutman, M. E. (2005). The influence of visual feedback on closed-set word test performance over time. *International Journal of Audiology*, *44*(12), 701-705. https://doi.org/10.1080/14992020500271795

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, *28*(4), 661-677. https://doi.org/10.1017/S014271640707035X

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 516-524. https://doi.org/10.3758/BF03195598

Plaza, M., Cohen, H., & Chevrie-Muller, C. (2002). Oral Language Deficits in Dyslexic Children : Weaknesses in Working Memory and Verbal Planning. *Brain and Cognition*, *48*(2-3), 505-512. https://doi.org/10.1006/brcg.2001.1407

Pujol, J., Soriano-Mas, C., Ortiz, H., Sebastian-Galles, N., Losilla, J. M., & Deus, J. (2006). Myelination of language-related areas in the developing brain. *Neurology*, *66*(3), 339-343. https://doi.org/10.1212/01.wnl.0000201049.66073.8d

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., & McMurray, B. (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, *51*, 1690-1703. https://doi.org/10.1037/dev0000044

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers : Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431-2443. https://doi.org/10.1121/1.4794379

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, *118*(5), 2775-2778. https://doi.org/10.1121/1.2062650

Skarżyński, H., & Piotrowska, A. (2012). Screening for pre-school and school-age hearing problems : European Consensus Statement. *International Journal of Pediatric Otorhinolaryngology*, *76*(1), 120-121. https://doi.org/10.1016/j.ijporl.2011.10.016

Skarzynski, P. H., Wlodarczyk, A. W., Kochanek, K., Pilka, A., Jedrzejczak, W., Olszewski, L., Bruski, L., Niedzielski, A., & Skarzynski, H. (2015). Central auditory processing disorder (CAPD) tests in a school-age hearing screening programme – analysis of 76,429 children. *Annals of Agricultural and Environmental Medicine*, *22*(1), 90-95. https://doi.org/10.5604/12321966.1141375

Snowling, M. J., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *346*(1315), 21-27. https://doi.org/10.1098/rstb.1994.0124

Sommers, M. S., Kirk, K. I., & Pisoni, D. B. (1997). Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. *Ear and Hearing*, *18*(2), 89.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, *86*(2), 221-234. https://doi.org/10.1037/0022-0663.86.2.221

1101 Stekić, K., Ilić, O., Ković, V., & Savić, A. M. (2023). ERP Indicators of Phonological Awareness
1102 Development in Children : A Systematic Review. *Brain Sciences*, *13*(2), 290.
1103 https://doi.org/10.3390/brainsci13020290
1104 Thillay, A., Roux, S., Gissot, V., Carteau-Martin, I., Knight, R. T., Bonnet-Brilhault, F., & Bidet-Caulet, A.
1105 (2015). Sustained attention and prediction : Distinct brain maturation trajectories during
1106 adolescence. *Frontiers in Human Neuroscience*, *9*, 519.
1107 Tillmann, B., Albouy, P., Caclin, A., & Bigand, E. (2014). Musical familiarity in congenital amusia :
1108 Evidence from a gating paradigm. *Cortex*, *59*, 84-94.
1109 https://doi.org/10.1016/j.cortex.2014.07.012
1110 Vickers, D. A., Moore, B. C. J., Majeed, A., Stephenson, N., Alferaih, H., Baer, T., & Marriage, J. E.
1111 (2018). Closed-Set Speech Discrimination Tests for Assessing Young Children. *Ear & Hearing*,
1112 *39*(1), 32-41. https://doi.org/10.1097/AUD.0000000000000528
1113 Watson, B. U., & Miller, T. K. (1993). Auditory Perception, Phonological Processing, and Reading
1114 Ability/Disability. *Journal of Speech, Language, and Hearing Research*, *36*(4), 850-863.
1115 https://doi.org/10.1044/jshr.3604.850
1116 Werner, L., Fay, R. R., & Popper, A. N. (Éds.). (2012). *Human Auditory Development* (Vol. 42). Springer
1117 New York. https://doi.org/10.1007/978-1-4614-1421-6
1118 Wetzel, N., Widmann, A., & Schröger, E. (2009). The cognitive control of distraction by novelty in
1119 children aged 7–8 and adults. *Psychophysiology*, *46*(3), 607-616.
1120 Wightman, F., & Allen, P. (1992). Individual differences in auditory capability among preschool
1121 children. In *Developmental psychoacoustics* (p. 113-133). American Psychological
1122 Association. https://doi.org/10.1037/10119-004
1123 Wilson, R. H., Farmer, N. M., Gandhi, A., Shelburne, E., & Weaver, J. (2010). Normative Data for the
1124 Words-in-Noise Test for 6- to 12-Year-Old Children. *Journal of Speech, Language, and*
1125 *Hearing Research*, *53*(5), 1111-1121. https://doi.org/10.1044/1092-4388(2010/09-0270)

Wilson, W. J., & Arnott, W. (2013). Using Different Criteria to Diagnose (Central) Auditory Processing Disorder : How Big a Difference Does It Make? *Journal of Speech, Language, and Hearing Research*, *56*(1), 63-70. https://doi.org/10.1044/1092-4388(2012/11-0352)

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development : Typical developmental trends. *Journal of Child Psychology and Psychiatry*, *44*(4), 561-575.

Wróblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear and hearing*, *33*(6), 731-744. https://doi.org/10.1097/AUD.0b013e31825aecad

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F.-X., & Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences*, *102*(39), 14110-14115. https://doi.org/10.1073/pnas.0504446102

Ziegler, J. C., Pech-Georgel, C., George, F., & Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, *12*(5), 732-745. https://doi.org/10.1111/j.1467-7687.2009.00817.x