

# Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity

Laurianne Cabrera,<sup>1,a)</sup> Léo Varnet,<sup>2</sup> Emily Buss,<sup>3</sup> Stuart Rosen,<sup>1</sup> and Christian Lorenzi<sup>2</sup>

<sup>1</sup>Department of Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, WC1N 1PF, London, United Kingdom

<sup>2</sup>Laboratoire des Systèmes Perceptifs, Ecole Normale Supérieure, Centre National de la Recherche Scientifique, Université Paris Sciences et Lettres, 29 Rue d'Ulm, 75005, Paris, France

<sup>3</sup>Department of Otolaryngology/Head and Neck Surgery, School of Medicine, University of North Carolina, Chapel Hill, North Carolina 27599, USA

(Received 25 April 2019; revised 28 August 2019; accepted 16 September 2019; published online 15 October 2019)

The ability to detect amplitude modulation (AM) is essential to distinguish the spectro-temporal features of speech from those of a competing masker. Previous work shows that AM sensitivity improves until 10 years of age. This may relate to the development of sensory factors (tuning of AM filters, susceptibility to AM masking) or to changes in processing efficiency (reduction in internal noise, optimization of decision strategies). To disentangle these hypotheses, three groups of children (5–11 years) and one of young adults completed psychophysical tasks measuring thresholds for detecting sinusoidal AM (with a rate of 4, 8, or 32 Hz) applied to carriers whose inherent modulations exerted different amounts of AM masking. Results showed that between 5 and 11 years, AM detection thresholds improved and that susceptibility to AM masking slightly increased. However, the effects of AM rate and carrier were not associated with age, suggesting that sensory factors are mature by 5 years. Subsequent modelling indicated that reducing internal noise by a factor 10 accounted for the observed developmental trends. Finally, children's consonant identification thresholds in noise related to some extent to AM sensitivity. Increased efficiency in AM detection may support better use of temporal information in speech during childhood.

© 2019 Acoustical Society of America. <https://doi.org/10.1121/1.5128324>

[JLL]

Pages: 2415–2429

## I. INTRODUCTION

Understanding speech in noisy backgrounds can be a challenge for all listeners and is particularly difficult for young children. The ability to identify speech in noise is relatively long to develop, as it continues to improve substantially until adolescence (Hall *et al.*, 2002; Johnson, 2000; Neuman and Hochberg, 1983; Stuart, 2008). This development may somewhat relate to the maturation of sensory auditory capacities, such as the ability to detect modulations in sound amplitude over time, that are essential to track the spectro-temporal features of speech sounds (Rosen, 1992; Stuart, 2005, 2008). Recent studies conducted with adults showed that the temporal modulations belonging to a background noise (e.g., a competing voice, a babble, a steady or a fluctuating noise) interfere with the temporal modulations of speech. The development of this phenomenon, called “modulation masking” (Houtgast, 1989; Bacon and Grantham, 1989; Dau *et al.*, 1997a; Biberger and Ewert, 2016, 2017; Jørgensen and Dau, 2011; Jørgensen *et al.*, 2013; Stone *et al.*, 2011), and its relationship with speech

intelligibility in noise has not been explored yet in childhood. The goal of the present study is to determine any changes in perceptual factors affecting modulation masking during development and clarify this relationship.

## A. Perception of speech in noise in childhood

Naturalistic environments are acoustically complex and rich in multiple sound sources. Young listeners face these “noisy” backgrounds in their everyday life, especially when at school. Children must focus and maintain attention to the speaker (e.g., the teacher) to learn new skills in these challenging listening conditions. Different types of degradation can alter speech intelligibility in a classroom, such as reverberation, steady or interrupted noise and multi-talker babble (Van Rietschote *et al.*, 1981). In that respect, it is worrying to note that the background sounds in classrooms may reach 70–77 dB LAeq (Jamieson *et al.*, 2004; Shield and Dockrell, 2003). When compared to adults, children show poorer word identification performance in noise at least until 10–11 years of age and more in some studies (Elliott, 1979; Elliott *et al.*, 1987; Hall *et al.*, 2002; Jacobi *et al.*, 2017; Johnson, 2000; Paps and Blood, 1989; Stuart, 2005, 2008). It has been proposed that the robustness of speech coding is late to develop because of immaturities in lexical processing, phonological representation of speech, and working memory (e.g., Hazan

<sup>a)</sup>Current address: CNRS (Integrative Neuroscience and Cognition Center, UMR 8002), Université Paris Descartes, Sorbonne Paris Cité, 45 Rue des Saints Pères, 75006, Paris, France. Electronic mail: laurianne.cabrera@gmail.com

and Barrett, 2000; McCreery *et al.*, 2017; Nitttrouer, 2004). Even when using easy phonetic discrimination tasks, typically developing 5- to 7-year-olds with normal hearing are still found to perform more poorly than adults or older children (Nishi *et al.*, 2010). Although it is likely that speech-in-noise performance is influenced by a wide variety of skills that develop in children, including memory, language and executive function, here we focus on the relationship between the development of basic auditory abilities and speech perception in noise. The present study explored the capacity to detect slow temporal variations in sounds, because this ability was found to be significantly correlated with speech identification in quiet and in noise for elderly people with normal hearing (Füllgrabe *et al.*, 2015) and hearing-impaired adults using cochlear implants (Cazals *et al.*, 1994; Fu, 2002; Gnansia *et al.*, 2014; Won *et al.*, 2011).

## B. Development of temporal auditory processing

Irrespective of language, the intelligibility of speech sounds relies on the accurate perception of patterns of slow (<5–10 Hz) modulations in amplitude (AM) (Drullman, 1995; Dubbelboer and Houtgast, 2007; Houtgast and Steeneken, 1985; Shannon *et al.*, 1995; Varnet *et al.*, 2017), so any development in the processing of such temporal features is likely to be important. Previous work shows that auditory AM sensitivity (as measured by detection thresholds for sinusoidal AM) improves with age between 5 and 10 years (Hall and Grose, 1994). Regarding this long development, two hypotheses have been put forward. First, the difference between children and adults may only relate to the maturation of sensory coding of AM information per se (e.g., the progressive tuning of mechanisms responsible for the extraction of AM information). Second, better thresholds may relate to the maturation of later stages of processing and reflect the efficiency with which the available AM information is used.

Very few studies have systematically measured AM detection thresholds in children at different AM rates. Overall, young listeners show worse auditory thresholds for AM detection at both slow and fast AM rates compared to adults. In a pioneering study, Hall and Grose (1994) compared adults' and children's AM detection thresholds using a broadband noise carrier that was modulated at five target AM rates (5, 20, 100, 150, and 200 Hz). The results indicated that sensitivity to AM was reduced in young listeners but reached adult-like levels at around 9 years of age. Nevertheless, the effect of changing AM rate on AM sensitivity was comparable for children and adults. That is, the higher the AM rate, the worse the AM detection threshold. This effect of AM rate on AM sensitivity suggests that factors constraining the temporal resolution of the human auditory system [the limit in the ability to follow (i.e., resolve) fast AM fluctuations] are probably mature early on and probably during infancy (Levi and Werner, 1996; Walker *et al.*, 2019; Werner, 1996). In a recent study, Buss *et al.* (2019) measured AM detection thresholds for 5-to-11-year-old children and adults using a 4300 Hz pure tone carrier modulated

at three different AM rates (16, 64, and 256 Hz). The results suggest that AM detection thresholds continue to improve with age at higher AM rates (64 and 256 Hz), but not at the lower AM rate tested (16 Hz). Altogether, these data indicate that at least some aspects of AM processing mature during childhood, but it is still unclear whether these observed changes reflect the maturation of sensory-processing mechanisms leading to better *temporal resolution* with age, or to improvement in *auditory efficiency*.

## C. Sensory factors involved in temporal processing

It has been recently suggested that the intrinsic random amplitude fluctuations conveyed by concurrent sounds such as stationary broadband noises may produce substantial perceptual masking of the amplitude modulations of the target speech signal (Stone *et al.*, 2011; Stone *et al.*, 2012; Stone and Moore, 2014). This phenomenon has been observed in adults, but is not yet fully characterized during development and could be related to worse AM auditory thresholds in children compared to adults. “AM masking” effects have been observed in adults using different psychophysical tasks and usually by using two simultaneous sounds, a target and a masker, that fluctuate at a similar AM rate (Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a; Sek *et al.*, 2015). AM masking effects can also be investigated when presenting a single sound such as a narrowband noise carrying a target sinusoidal AM as shown in Fig. 1. In this case, the random intrinsic amplitude fluctuations of the noise carrier may disrupt the detection of the (sinusoidal) target AM (Dau *et al.*, 1997a; Lorenzi *et al.*, 2001). Specifically, the amount of AM masking can be manipulated by changing the bandwidth of the narrowband noise carrier. One characteristic of a narrowband noise is that changes in its bandwidth result in changes in its AM spectrum, that is, the extent of modulation as a function of modulation rate. When the target sinusoidal AM fluctuates at a rate close to the modulations of the noise carrier, AM masking is maximal (Dau *et al.*, 1997a). These effects have been interpreted as revealing the existence of sensory *filters* tuned to a specific AM rate (Bacon and Grantham, 1989; Dau *et al.*, 1997a; Houtgast, 1989). Subsequent brain-imaging and electrophysiological studies conducted in humans revealed that these modulation filters are implemented at subcortical and cortical stages of the auditory pathway (Giraud *et al.*, 2000; Liégeois-Chauvel *et al.*, 2004). These studies on AM perception have provided support for the development of computational models of AM processing based on the concept of a modulation filterbank (Dau *et al.*, 1997a). These models are used to predict the perception of AM cues and have recently accounted for a wide range of detection, discrimination and identification data in adults (e.g., Biberger and Ewert, 2017).

The present study explored the development of AM masking. The AM carriers were either a pure tone or a single narrowband noise. Furthermore, computational modelling was used to predict the AM detection data. A recent study assessed the intelligibility of narrow bands of speech presented against spectrally overlapping pure tones or narrowband noises and showed no evidence of greater susceptibility

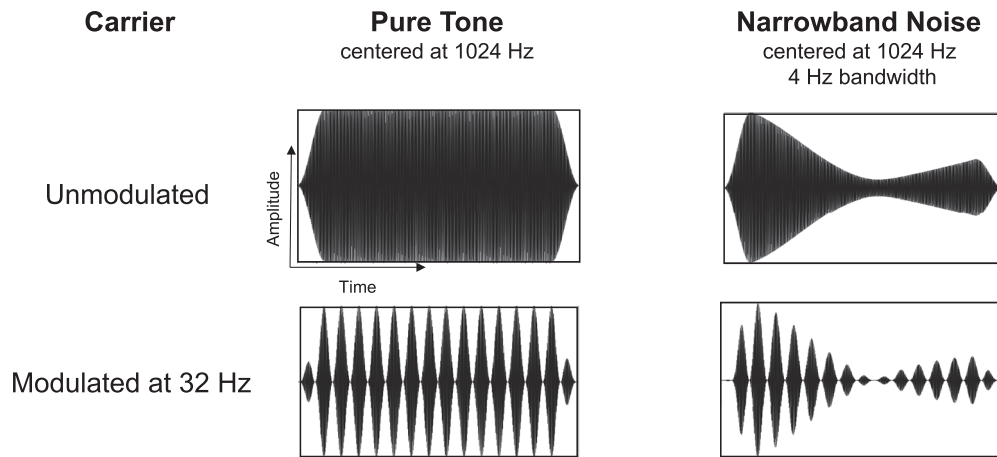


FIG. 1. Waveforms representing amplitude variations over time for a pure tone carrier (on the left) and a narrowband noise carrier (on the right). The top row shows the carrier without additional modulation applied, and the bottom row shows stimuli with 32-Hz AM.

to AM masking in children than adults (Buss *et al.*, 2018). Still, the necessity to interpret AM masking data in light of computer modelling results motivated the present study in its aim to tease apart the role of sensory vs efficiency factors involved in AM processing development.

#### D. Efficiency in temporal processing

Using a computational model of AM perception will help to assess whether the increase in AM sensitivity during childhood reflects changes in the characteristics of AM filters or changes in the ability to make efficient use of the available AM cues. The model implements two main components constraining processing efficiency in the AM domain: internal noise, the random component of efficiency, and the decision strategy, which is the systematic component of efficiency. It is possible that the extraction of AM cues (i.e., the sensory ability to resolve AM cues and decompose them via central AM filters) is adultlike in children, but that additional factors may limit the ability to use them optimally. Internal noise refers to variability in the decision statistic used by listeners (Green and Swets, 1966). In the case of AM processing, the decision statistic is thought to be related to the time course of envelope fluctuations, or envelope power, at the output of modulation filters (Strickland and Viemeister, 1996; Jepsen *et al.*, 2008). Internal noise then corresponds to the stochastic nature of neural responses to AM cues along the auditory pathway; however, it may also correspond to variability in the arousal of the organism and random fluctuations in auditory attention (e.g., Amitay *et al.*, 2013; Faisal *et al.*, 2008).

The hypothesis of larger internal noise in childhood has been explored for temporal masking tasks (Hartley and Moore, 2002; Hartley *et al.*, 2000; Hill *et al.*, 2004). The study of Hill *et al.* (2004) suggested that children's data in a backward masking task were better modelled when assuming a mature temporal resolution in children but poorer processing efficiency compared to adults. This result could be related to greater internal noise in children than adults, consistent with the suggestion made by Werner and Boike (2001) for infants. Buss *et al.* (2006) reached the same

conclusion by showing that the slope of psychometric functions for intensity discrimination (the function relating discrimination performance to stimulus intensity) is shallower in children compared to adults. Poorer thresholds for temporal processing may thus relate to a higher level of internal noise along the AM processing pathway in children compared to adults rather than poorer peripheral encoding.

It may also be the case that children do not use an optimal decision rule. For instance, children may not appropriately weight the AM information available in on- and off-frequency auditory channels, that is, channels tuned at or below and above the frequency region of the AM carrier, respectively (Jones *et al.*, 2015; Leibold and Buss, 2016). This would yield poorer detection performance because the presence of AM is better signaled by changes in neural excitation at the output of off-frequency channels than the channel centered on the signal frequency region (Moore and Sek, 1994; Zwicker, 1952; Zwicker, 1956). Such an “off-frequency” listening strategy is optimal, and presumably used by adult listeners with normal hearing (Moore and Sek, 1994), because cochlear compression applied by the active mechanism in the cochlea is maximal in the on-frequency channel and minimal or absent in off-frequency channels tuned above the carrier. Inconsistent with this hypothesis, a previous study indicated that children do use off-frequency cues at least for intensity discrimination (Buss *et al.*, 2013).

#### E. Goal of the study and main hypotheses

The main goal of the present study was to systematically assess the development of AM detection—and more specifically, AM masking—in childhood between 5 and 11 years of age. Here, the dependence of AM masking on the shape of the modulation spectrum of the AM masker was evaluated in a psychophysical task using non-linguistic sounds. To do so, children's auditory detection thresholds for sinusoidal AM were measured when that AM was applied to each of three different carriers varying in the modulation spectrum and strength of their inherent AM: tones, narrowband noises with small inherent AM fluctuations and noises with larger fluctuations. If younger children are more affected by the carrier



fluctuations than the older children, this would suggest that AM tuning improves with age. Such an outcome would be consistent with the idea that modulation masking plays a role in children's immature speech-in-noise recognition.

To better characterize AM sensitivity during development, we also measured AM detection using three target AM rates (4, 8, and 32 Hz). Younger children were expected to show worse (higher) AM detection thresholds (AMDTs) than older children but similar effects of target AM rate and modulation-masker strength on AMDTs (Hall and Grose, 1994). A computational model of AM processing based on the modulation filterbank concept and a template-matching decision strategy (Dau *et al.*, 1997a; Wallaert *et al.*, 2018) was also developed to test whether AM detection data in children were better simulated by changes in AM selectivity, internal noise and/or decision strategy (the appropriate integration of AM information from on- and off-frequency cochlear channels).

Finally, the ability of children to identify speech in the presence of noise was estimated by measuring consonant identification thresholds in a stationary noise using vowel-consonant-vowel stimuli. These were used instead of sentences or words in order to reduce the effect of lexical knowledge between 5 and 11 years on this task. It was expected that children would show similar consonant identification thresholds in noise by 7 years of age as observed in previous studies (Nishi *et al.*, 2010). The relationship between consonant perception abilities in noise and auditory temporal processes was then further explored.

## II. METHODS

### A. Participants

Eighty-two children were recruited at three different schools in London, and 32 adults were recruited through the university participant pool. For the children, parents and teachers reported typical language development and no special educational needs. Opt-out consent forms were distributed to the parents, and consent forms were signed by adult participants as approved by the university ethics committee. Adult participants filled out a questionnaire about speech and language history, and they received a monetary compensation for their time. The receptive vocabulary level of the children was measured using a standardized test [*The British Picture Vocabulary Scale* (Dunn and Dunn, 2009)]. For both children and adults, hearing sensitivity was assessed with pure tones at 250, 500, 1000, 2000, 4000, and 8000 Hz prior to testing.

Two participants (one child and one adult) were excluded because their hearing thresholds were above the normal range [ $>20$  dB hearing level]. An additional nine children were tested but not included in the final sample: three were at floor levels of performance for more than the half of the conditions (all aged 5–6 years), one withdrew from the study, four were not at school during testing, and one was excluded because of experimental error. Four adults were not included in the final sample because of missing data due to experimental errors.

The final sample included: 25 5–6-year-olds (11 females; mean age = 5.7 years, SD = 0.4), 29 7–8-year-olds (14 females; mean age = 7.8 years, SD = 0.4), 28 10–11-year-olds (13 females; mean age = 10.8, SD = 0.4) and 32 adults (24 females, mean age = 22.4 years, SD = 2.4). Experience with a second language was documented for all participants, as listeners learning English as a second language may show worse thresholds for English speech identification in noise (Lecumberri *et al.*, 2010). The proportion of bilingual participants was 7/25 for 5–6-year-olds, 13/29 for the 7–8-year-olds, 13/28 for the 10–11-year-olds, and 19/32 adults. This group of bilinguals included both sequential and simultaneous bilinguals. All but two children were born in the UK (both 10–11-year-olds). Of the adult bilinguals, only four were born in an English-speaking country.

### B. Stimuli used in the modulation-masking study

A total of nine conditions were used with three different AM carriers and three target AM rates ( $f_m$ ): 4, 8, and 32 Hz. In all the conditions, the stimuli were 500-ms long including 50-ms raised-cosine onset/offset ramps, and the inter-stimulus interval was 500 ms. Standard sounds were not modulated in amplitude, and target sounds were modulated at depths ranging from  $m = 100\%$  to  $m = 1\%$ , in 20 steps of 2 dB. The expression describing the target was

$$T(t) = [1 + m \sin(2\pi f_m t + \phi_m)]c(t), \quad (1)$$

where  $t$  is time,  $m$  is the modulation depth ( $0 \leq m \leq 1$ ),  $f_m$  is the modulation rate,  $\phi_m$  is the starting phase of the modulation, randomized on each trial, and  $c$  is the carrier signal.

In the tone conditions (TONE), the carriers were sine tones centered at 1027 Hz, generated with a random starting phase at a sampling frequency of 44.1 kHz. In the noise conditions, the carriers were 4-Hz wide narrow frequency bands. Fluctuations of the noise envelope were either relatively low in amplitude (NOISE-LOW) or high in amplitude (NOISE-HIGH). Each narrow-band noise carrier was generated by adding together five equal-amplitude sine tones with frequencies 1025, 1026, 1027, 1028, and 1029 Hz at a sampling frequency of 44.1 kHz. Each sine component had a random starting phase. The onset and offset of each sound was shaped with a 150-ms raised cosine ramp. There were three stages for selecting NOISE-LOW and NOISE-HIGH narrowband noise stimuli. In the first stage, noise samples were passed through a single 1-ERB wide, linear gammatone filter (Patterson *et al.*, 1995) with a center frequency of 1027 Hz. In the second stage, the envelope of the filter output was extracted via half-wave rectification and lowpass filtering (first order Butterworth; cutoff = 150 Hz) as in Strickland and Viemeister (1996). In the third and final stage, the standard deviation (SD) of the temporal envelope was computed, excluding the onset/offset ramps. From there, 200 sounds with “low” values of envelope SD were selected [SD ranging between 0.059 and 0.069 a.u. (arbitrary amplitude units), mean SD = 0.066 a.u.]. An additional 200 sounds with “high” values of envelope SD were selected (SD ranging between 0.19 and 0.21 a.u., mean SD = 0.19 a.u.).

Figure 2 shows the modulation spectra computed for the noise carriers in the NOISE-LOW and NOISE-HIGH conditions. Amplitude Modulation Index spectra (AMi) display the modulation index (the ratio between the temporal-envelope depth and the mean amplitude of the temporal envelope at the output of a cochlear filterbank) as a function of AM rate (see Varnet *et al.*, 2017, for details; Houtgast and Steeneken, 1985). These AM spectra were obtained by computing the long-term Fourier amplitude spectrum of the temporal envelopes across several audio-frequency bands. A 1/3-octave-band representation of AM power is usually used to emphasize modulation components in the high AM-rate region where the bandwidths of perceptual modulation filters is assumed to be larger (e.g., Dau *et al.*, 1997a). As expected, AM spectra are lowpass in shape, showing greater modulation energy below about 4 Hz (see Dau *et al.*, 1997a). It follows that higher AMDTs, and thus greater AM masking, should be observed for the lowest target AM rates ( $f_m = 4$  Hz). Any differences in AM masking across target AM rates between age groups might reflect differences in selectivity (i.e., the bandwidth of modulation filters, as estimated by their  $Q$  factor; see below). Finally, AM spectra also show that NOISE-HIGH stimuli show greater modulation energy than NOISE-LOW stimuli especially below 32 Hz. It follows that higher AMDTs, and thus, greater AM masking, should be observed for narrowband noise carriers in the NOISE-HIGH condition than in the NOISE-LOW condition, as observed previously for adults (e.g., Bacon and Grantham, 1989; Strickland and Viemeister, 1996).

### C. Procedure used in the modulation-masking study

For each of the nine conditions, AMDTs were obtained using a three-interval, three-alternative forced-choice procedure (3I-3AFC), implemented on a touch-screen tablet. The sounds were played diotically through headphones (Sennheiser HD 25-SP II, Sennheiser electronic GmbH & Co. KG, Wedemark, Germany) at a level of 65 dB sound pressure level (SPL).

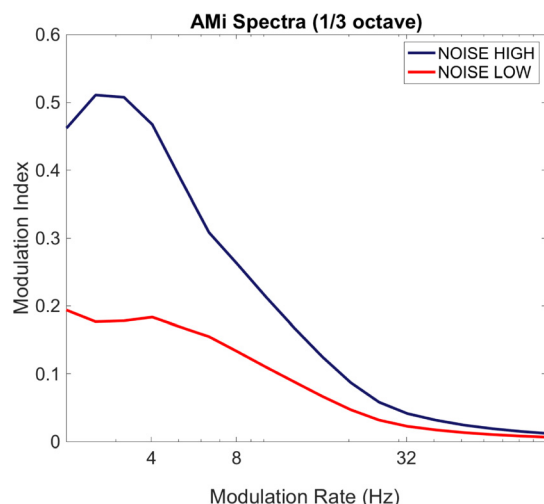


FIG. 2. (Color online) Amplitude-modulation index (AMi) spectra of the narrowband noise carriers in the NOISE-LOW (red lines) and NOISE-HIGH (blue lines) conditions.

For both children and adults, the experimenter sat next to the participants and explained that three animal characters would appear on the screen and produce a sound one after the other. Only one character would produce a different sound, and the task was to find the odd-one out. On each trial, participants had unlimited time to respond and received feedback (a tick or a red cross). The next trial started automatically 600 ms after the participant's response.

Children were tested at schools in a quiet room over a 3-week period. They collected stickers on a science certificate after each run to keep track of their progress and keep them motivated. After two or three runs, a small break was given, and each session lasted no more than 20 min. For the 5–6-year-olds, all the conditions were completed within five sessions, and for 7-to-11-year-olds only four sessions were necessary. Adults were tested at the university in a quiet room in a single 1.5 h session, with breaks following blocks of 3–4 runs. The picture background differed between conditions to keep the children interested in the game.

An initial one-down, one-up rule was used until the first reversal (Baker and Rosen, 2001), with the caveat that incorrect responses on the first trial were always ignored. This was followed by a 2-down 1-up adaptive procedure to track the 71% correct-point (Levitt, 1971). The first trial started at  $m = 100\%$ , with an initial step-size of 6 dB in modulation index, reduced to 4 dB after the first reversal, and 2 dB after the second reversal. The run stopped after the eighth reversal or after 32 trials. If participants were not able to detect the target at  $m = 100\%$ , this trial was repeated until correct detection (the limits of  $m$  were set to 100% and 1%). Only one estimate was collected for each condition. However, after a visual examination of the track, a second estimate was collected if fewer than five reversals were obtained, or if the track did not converge. The threshold in dB was the geometric mean of the last four reversals. Moreover, for the 5–6-year-olds, the first condition was always presented two times as they usually showed some difficulties in understanding the task the first time. The threshold was obtained only from the second run in this first condition. In addition to threshold, the standard deviation of the last four reversals was calculated. This measure of response variability was interpreted as an indicator of threshold reliability (Moore *et al.*, 2010). The easiest condition (TONE-carrier with 32-Hz target AM) was always completed first to make sure that the task was understood by all participants, followed by a randomized order for the remaining conditions.

### D. Modulation-masking data processing and analysis

The outlier labelling rule was applied to our data to identify potential outliers in all conditions (Hoaglin and Iglewicz, 1987). Four outliers were flagged in the 5–6-year-old group, two in the 7–8-year-olds, four in the 10–11-year-olds, and two in the adult group (11% of the data). When those outliers were removed, our data did not differ from normality (Kolmogorov-Smirnov test  $p_s > 0.05$ ) for all

conditions except in the NOISE-HIGH condition at 4 Hz (see below) and in the condition NOISE-HIGH at 32 Hz for the 11-year-olds, for which the data were a little skewed and kurtotic. However, for this last condition, comparison between the skewness value and its SD was considered as reasonable for a normal distribution; that is, the deviation from normality was not sufficiently large to preclude the assumption of normality (1.22 comprised between  $-/+ 2$ ; see Trochim and Donnelly, 2006). Thus, data were available from a total of 21 5–6-year-olds, 27 7–8-year-olds, and 24 10–11-year-olds. The analyses included only children’s data to focus on the developmental trends of AM processing during childhood with adult data being shown only for reference in the figures.

### E. Computational modelling of modulation masking

A computational model of AM processing was developed to test whether AM detection data in children were better captured by changes in AM frequency selectivity, internal noise and/or decision strategy. The general structure of the model of AM processing is similar to that used by Wallaert *et al.* (2018) and is fully described in Appendix A. Three main levels of processing were represented: peripheral processing (e.g., cochlear filtering), AM processing (extraction of AM information via AM filters), and template matching (decision making). Figure 3 represents the stages implemented at each level. Three parameters were manipulated in the model to fit the empirical data: (1) the quality factor  $Q$  reflecting the AM frequency selectivity at the envelope processing level where AM information is extracted and represented; (2) the variance of an additive internal noise affecting processing efficiency; (3) at the decision stage, the use of either an optimal strategy incorporating information from off-frequency channels (where cochlear compression is reduced), or a suboptimal strategy relying entirely on on-frequency channels, in a template matching procedure. Ninety thresholds were simulated using the same procedures implemented with human observers and averaged for each condition.

### F. Stimuli and procedure used in the speech-perception study

Syllables of the form /aCa/ were recorded by three native Southern British English speakers: one male ( $F_0 = 112$  Hz) and two females ( $F_0 = 153$  and 160 Hz). Only fricative consonants, known to be difficult to perceive in childhood (Ingram *et al.*, 1980; Moeller *et al.*, 2007), were used,  $C = /f/, /v/, /ʃ/, /ʒ/, /s/, /z/$ . Two exemplars of each syllable were selected for each speaker and equalized in intensity. A steady speech-shaped noise masker was designed to match the long-term spectrum of the female syllables. Two conditions were designed to assess consonant identification in noise, one presenting a minimal change in voicing  $/f/-/v/, /s/-/z/, /ʃ/-/ʒ/$  and the other one a change in place of articulation  $/f/-/s/, /v/-/z/, /s/-/ʃ/, /z/-/ʒ/$ .

The syllables were presented in a XAB task, where a first sound X was always a sample from the male speaker played in quiet, and the subsequent sounds (A and B) were

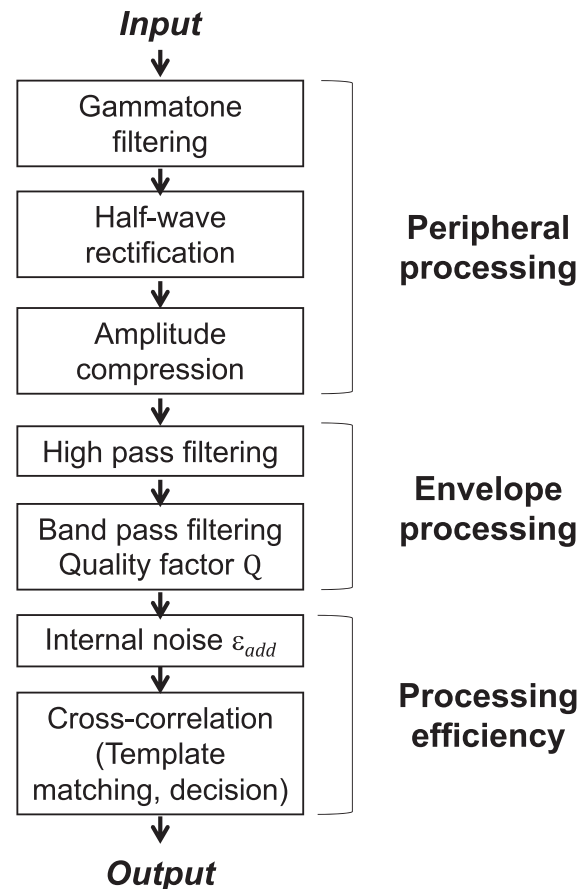


FIG. 3. Schematic representation of the computational model of AM processing.

sample from each of the two female speakers played in noise. The inter-stimulus interval was 500-ms. The same touch-screen and headphones as in the previous task were used. On each trial, the participant looked at one animated character on the top of the screen and two other characters on the bottom, next to each other, representing the three different speakers. The participants were asked to touch the character at the bottom who uttered the same sound as the one on the top. On each trial, participants had unlimited time to respond and received feedback (the character chosen displayed a “happy” or a “sad” face). The next trial started automatically  $\sim 600$  ms following the participant’s response. The noise level was varied following an adaptive 2-down 1-up procedure. The starting signal-to-noise ratio (SNR) was +20 dB. The first step size was 5 dB, reduced to 2 dB after two reversals. The run ended either when a total of six reversals was reached or after 32 trials. The stimuli were presented at 65 dB SPL, and the SNR was adjusted by fixing the masker level and adjusting the target level. The outcome measure was the averaged SNR level of the last four reversals corresponding to the consonant identification thresholds in noise of the participant. The outlier labelling rule was applied once again, and 4 outliers were flagged (three in the 10–11 years and one in the 5–6 years). Moreover, one child from the 7–8-year group was excluded from the following analyses because of experimental error.



### III. RESULTS

#### A. Sensitivity to AM in the presence of an AM masker

Figure 4 shows the AMDTs in dB obtained in each age group for the three carriers at the three modulation rates. Lower (better) AMDTs are observed with the TONE carrier and higher (worse) thresholds are observed with the NOISE HIGH carrier. Overall, each age group showed better thresholds with higher AM rate. The improvement in AMDTs with AM rate reflects the perceptual integration of AM cycles by listeners (Viemeister, 1979; Wallaert *et al.*, 2018), as the number of AM cycles increases when AM rate increases from 4 to 32 Hz (stimulus duration being constant at 500 ms in all conditions). A mixed-design analysis of variance (ANOVA) was run to assess the effect of the within-subject factors Carrier (TONE vs NOISE-LOW vs NOISE-HIGH) and Rate (4 vs 8 vs 32 Hz) and the between-subject factor Age Group (5–6 years vs 7–8 years vs 10–11 years) on the AM detection thresholds. This analysis revealed a main effect of Carrier [ $F(2,276) = 212.47$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.61$ ], Rate [ $F(2,276) = 899$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.87$ ], and Age [ $F(2,69) = 10.14$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.23$ ]. Overall, AMDTs significantly degraded between TONE, NOISE-LOW, and NOISE-HIGH conditions, and between 32, 8, and 4 Hz. Moreover, the group of 5–6-year-olds showed significantly worse thresholds than the 7–8- and 10–11-year-olds (that did not differ from each other). The analyses of the interactions revealed a significant interaction between Carrier and Rate [ $F(3,276) = 37.07$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.29$ ], and pairwise comparisons with Bonferroni corrections indicated that even though in all carrier conditions AMDTs significantly improved with AM rates, they did not differ significantly between the NOISE-LOW and TONE conditions at the 32 Hz modulation rate. Finally, the interaction between Carrier and Age Group was significant but with a small effect size [ $F(4,276) = 3.6$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.05$ ], and further pairwise comparisons with Bonferroni corrections indicated that only the 5–6-year-old group showed no significant difference between thresholds in the NOISE-LOW versus TONE conditions. Other effects were not significant [Rate  $\times$  Age:  $F(4, 276) = 0.64$ ,  $p = 0.63$ ,  $\eta_p^2 = 0.01$ ; Rate  $\times$  Age  $\times$  Carrier:

$F(8, 276) = 0.90$ ,  $p = 0.52$ ,  $\eta_p^2 = 0.03$ ]. A supplementary analysis showing that 5–6-year-olds' thresholds were not affected by specific attention issues in the TONE condition is available in Appendix B.

As observed in Fig. 4, thresholds in the NOISE HIGH condition with 4 Hz AM rates were near 0 dB for all groups. We then calculated level of performance for all children in each condition to check whether in some conditions participants were at chance when detecting the AM target. The lowest performance was obtained in the NOISE-HIGH at 4 Hz (mean = 51% correct), while the performance was between 64% and 74% correct in the other conditions. As listeners were at chance in the 4 Hz condition for the NOISE HIGH, we repeated the analysis described above without the 4 Hz conditions. This ANOVA showed the same main effects of Carrier, Rate, and Age, and a significant interaction between Carrier and Rate ( $ps < 0.001$ ), but only a trend for an interaction between Carrier and Age [ $F(4,138) = 2.41$ ,  $p = 0.052$ ,  $\eta_p^2 = 0.07$ ]. The carrier by age interaction seems mainly driven by floor performance in the 4 Hz rate conditions. Without the 4 Hz condition, the 5–6-year-old group showed lower thresholds in both the TONE and NOISE-LOW conditions compared to the other two groups.

In summary, although younger children showed overall higher (worse) AMDTs, AM rate had the same effect on AMDTs over age, a pattern of results interpreted as indicating comparable “temporal resolution” (the ability to resolve temporal variations and thus, extract AM cues; Viemeister, 1979) and comparable “temporal integration” (the ability to combine AM cycles to improve AM detection performance; Viemeister, 1979; Wallaert *et al.*, 2018). Perhaps more interestingly, younger children did not show strong differences between the TONE and the NOISE-LOW conditions.

#### B. Modulation-masking effect

To further estimate the modulation-masking effect, difference scores were computed between NOISE-HIGH and TONE conditions, and between NOISE-LOW and TONE conditions. Figure 5 represents the masking effect in each condition at each modulation rate and for each age group. A repeated-measures ANOVA was run on the masking

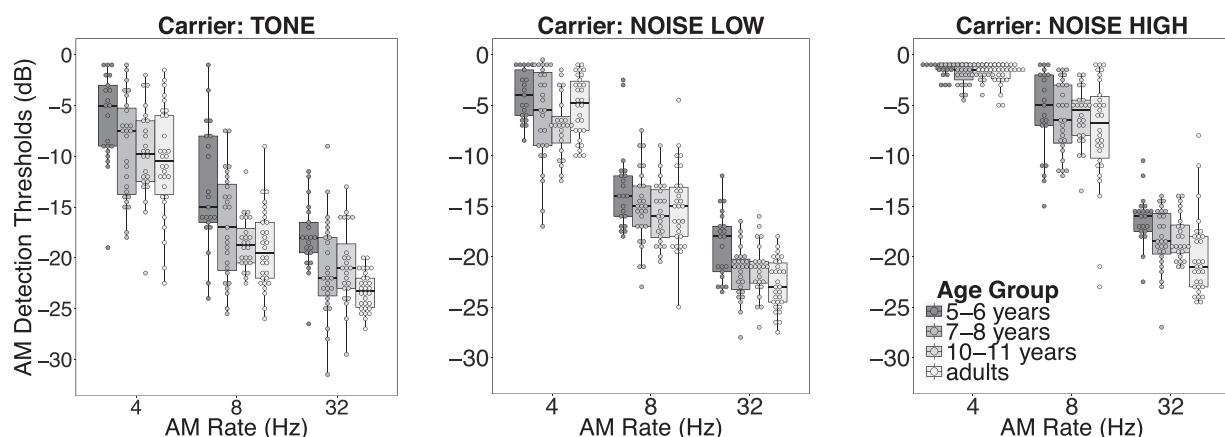


FIG. 4. AM detection thresholds [ $20\text{-log}_{10}(m/100)$ ; boxplots and individual data] for each age group, carrier condition (TONE, NOISE-LOW, and NOISE-HIGH) and modulation rate (Hz).

thresholds to assess the effect of Carrier (NOISE-LOW vs NOISE-HIGH) and Rate (8 vs 32 Hz, as participants were at chance at 4 Hz in the NOISE-HIGH condition) as within subject-factors and Age Group (5–6 years vs 7–8 years vs 10–11 years) as a between subject factor. This analysis revealed a significant effect of Carrier [ $F(1,69)=266.39$ ,  $p < 0.001$ ,  $\eta_p^2=0.79$ ], Rate [ $F(1,69)=59.14$ ,  $p < 0.001$ ,  $\eta_p^2=0.46$ ] and Age Group [ $F(2,69)=3.30$ ,  $p=0.04$ ,  $\eta_p^2=0.09$ ] as well as a significant interaction between Carrier and Rate [ $F(1, 69)=64.82$ ,  $p < 0.001$ ,  $\eta_p^2=0.48$ ]. No significant interaction between Carrier and Age Group [ $F(2,69)=1.18$ ,  $p=0.31$ ,  $\eta_p^2=0.03$ ], nor between Rate and Age Group [ $F(2,69)=2.43$ ,  $p=0.10$ ,  $\eta_p^2=0.07$ ]. Pairwise comparisons with Bonferroni corrections indicated that for all age groups, there was significantly more masking in the NOISE-HIGH condition (all  $ps < 0.001$ ) and that for both NOISE-LOW and NOISE-HIGH conditions, less AM masking was observed with increasing AM rate. Overall, the youngest group showed relatively *less* masking (for both noise conditions) compared to the oldest group of children ( $p=0.045$ ). Increased AM masking with age result from worse thresholds with tone carriers at 5 years, as if tone carriers were “noisier,” and especially in the 8 Hz condition (as observed in Fig. 5).

In summary, these results suggest that 5–6-year-olds show *poorer* AM sensitivity (irrespective of carrier) but *lower* AM masking than older children. However, they show the same *pattern* of AM masking as older children when the target AM rate increases from 8 to 32 Hz, suggesting that factors constraining AM masking (i.e., the shape and width of modulation filters) are mature by the age of 5.

### C. Modelling AM processing

A computational model of AM processing using the modulation filterbank concept and a template-matching decision strategy was run to explain these effects of age on AM detection and masking. Figure 6 shows the simulated AMDTs plotted along with the real data for each group and experimental condition. It is important to keep in mind that this model implemented adult-like modulation frequency selectivity and optimal listening strategy. The simulated data were obtained by

adjusting the SD of the internal noise of the model,  $\varepsilon_{add}$ , on the TONE, 8 Hz condition, for each group. Goodness of fit (the mean difference in absolute value), averaged across groups and experimental conditions, was 1.1, 1.1, and 1.5 dB for 4, 8, and 32 Hz, respectively. Overall, the SD of the internal noise yielding the best fits to the AMDTs was equal to  $2 \times 10^{-4}$  a.u. (arbitrary units) for adults,  $5 \times 10^{-4}$  a.u. for 10–11-year-olds,  $1 \times 10^{-3}$  a.u. for 7–8-year-olds, and  $2 \times 10^{-3}$  a.u. for 5–6-year-olds. It is noteworthy that the magnitude of the internal noise was a factor of 10 larger for the youngest children (5–6 years) compared to the oldest children (10–11 years) and adults. Moreover, the model correctly predicted that the overall amount of modulation masking is lower for the youngest children. For instance, between 5–6 years and 10–11 years, the amount of masking at 8 Hz increased from 1 to 6 dB for the NOISE-LOW carrier, and from 7 to 11 dB for the NOISE-HIGH carrier.

Another series of models was run to determine whether the worse AMDTs in the 5–6-year-olds could also be explained by poorer modulation-frequency selectivity or a sub-optimal listening strategy, whereby children apply more weight to the on-frequency channel. This suboptimal strategy could be adopted due to the perceptual salience of cues in this frequency region. Figure 7 shows three series of simulations for the data of 5–6-year-olds. Model A used five cochlear frequency channels (which is the optimal listening strategy), each with modulation filters with a  $Q$  value (Quality factor) of 1 (simulating adult-like modulation-frequency selectivity). Model B used five channels but with modulation filters of  $Q=0.5$  (poorer modulation-frequency selectivity). Model C used 1 channel (sub-optimal strategy) with modulation filters with  $Q=1$ . The SD of the internal noise was adjusted for each model and each group to fit model predictions to real data in the TONE, 8 Hz condition. With the NOISE-LOW and NOISE-HIGH carriers, the model best predicting the data of the 5–6-year-olds was model A, simulating adult-like modulation frequency selectivity and optimal listening strategy, but with an internal noise 10 times higher than the 10–11-year-olds’ data (goodness of fit = 1.3 dB). The model making the worst predictions was model B (goodness of fit = 2.3 dB).

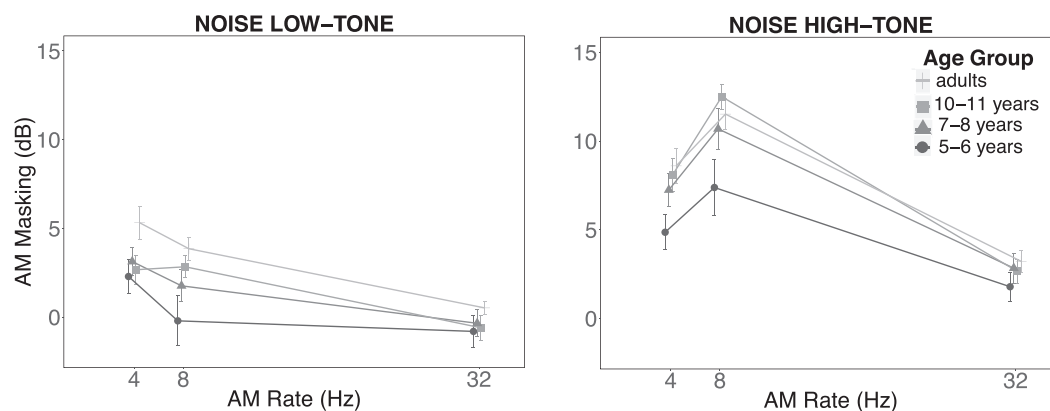


FIG. 5. Average differences in AMDTs (i.e., AM masking) between NOISE-LOW and TONE conditions (left panel) and between NOISE-HIGH and TONE conditions (right panel) at each modulation rate (Hz). Error bars show one standard deviation around the mean AM masking effect.



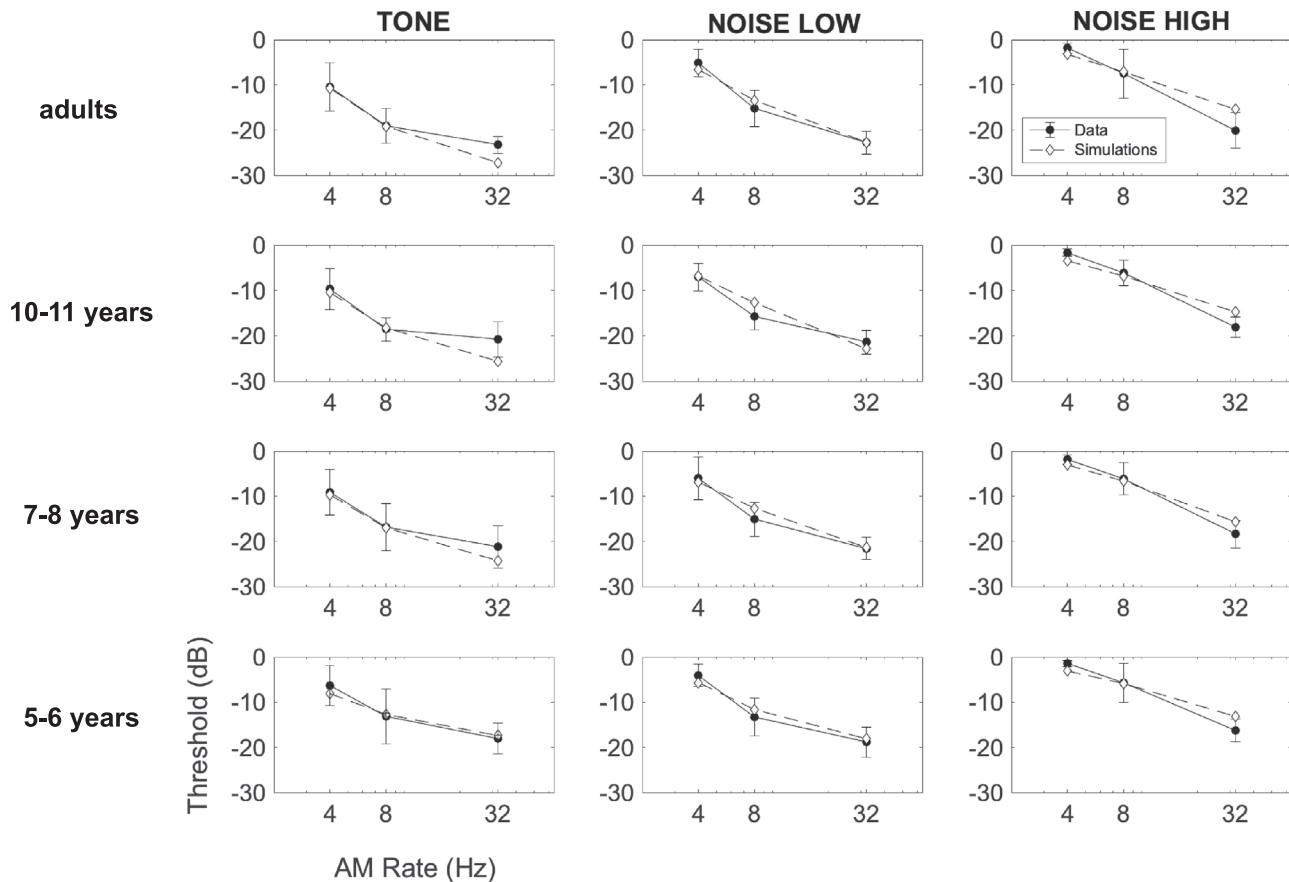


FIG. 6. Open diamonds with dashed lines represent simulated AM-detection data, while filled circles with continuous lines represent the averaged empirical observations. Data are shown as a function of AM rate (4, 8, and 32 Hz). The left column shows the data in the TONE conditions, the middle column the data for the NOISE-LOW conditions, and the right column the data for the NOISE-HIGH conditions. Each line represents one age group, from the adult group at the top to the youngest child group (5–6 years) at the bottom. In this model, the  $Q$  factor for modulation filters is set to 1 (adult-like modulation frequency selectivity). The decision device (the cross-correlator or “template-matching” process) uses temporal-envelope information at the output of off-frequency channels (the two simulated cochlear filters tuned below or above the carrier), which is an optimal listening strategy.

Thus, the data of younger children are better modelled assuming adult-like modulation-frequency selectivity ( $Q = 1$ ), optimal listening strategy (i.e., off-frequency listening) and higher internal noise than assuming poorer modulation-frequency selectivity or a sub-optimal listening strategy (on-frequency listening). However, it is possible that a combination of two factors determining efficiency (increased internal noise with sub-optimal listening) explains the worse AMDTs in younger children. Consistent with this idea, the goodness of fit for model C was 1.3 dB as for model A. However, predictions of model C were slightly worse than those of model A at 32 Hz in the NOISE-HIGH condition as observed in Fig. 7.

#### D. Speech-in-noise data

Finally, we also measured fricative consonant identification thresholds in noise for the same children in order to better understand the development of speech perception in noise and its relationship with temporal processing.

##### 1. Effect of age on consonant identification thresholds in noise

First, as shown in Fig. 8, consonant perception in noise appears to be adult-like around 7 years of age (Nishi *et al.*, 2010). A repeated-measures ANOVA was run on the thresholds

in noise to assess the effect of Speech Contrast (Place vs Voicing) as within subject factor, and Age Group (5–6 years vs 7–8 years vs 10–11 years) and Bilingualism (Monolinguals vs Bilinguals) as between subject factors. This analysis showed a significant main effect of Contrast [ $F(1,61) = 65.91, p < 0.001, \eta_p^2 = 0.52$ ], indicating worse (higher) thresholds for voicing contrasts than for place of articulation with fricative consonants. A main effect of Age Group was also observed [ $F(2, 61) = 9.36, p < 0.001, \eta_p^2 = 0.13$ ]. *Post hoc* analyses showed that the worst (highest) thresholds were obtained by the 5–6-year-olds. Their thresholds were significantly different from the two other groups ( $ps < 0.001$ ), but the two older groups did not differ from each other ( $p = 0.71$ ). No main effect of Bilingualism was observed [ $F(1,61) = 0.32, p = 0.57, \eta_p^2 < 0.01$ ], but there was a significant interaction Bilingualism  $\times$  Contrast [ $F(1,61) = 5.03, p = 0.03, \eta_p^2 = 0.08$ ] with a small effect size. Surprisingly, bilinguals tended to show lower (better) thresholds for voicing contrasts (2.5 dB more) compared to monolinguals, but not for place (1.2 dB less). However, when using Bonferroni corrections, these differences were not significant ( $p > 0.12$ ). No interaction was observed between Age and Contrast [ $F(2,61) = 0.21, p = 0.81, \eta_p^2 < 0.01$ ], Age and Bilingualism [ $F(2,61) = 0.52, p = 0.60, \eta_p^2 = 0.02$ ], or Age, Contrast, and Bilingualism [ $F(2,61) = 0.12, p = 0.89, \eta_p^2 < 0.01$ ].

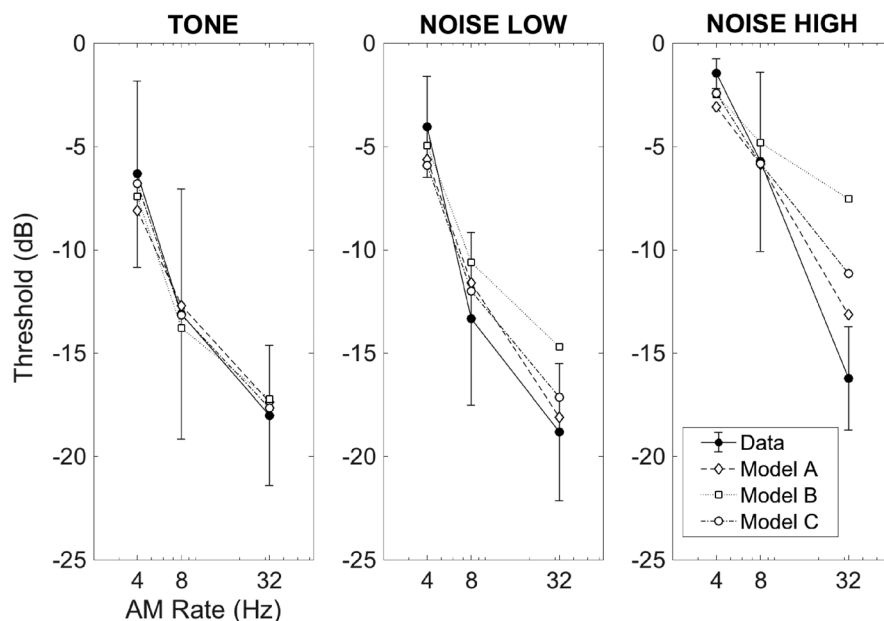


FIG. 7. Simulated thresholds for three models of AM processing are represented with the empirical data (filled circles with continuous lines) for the 5–6-year-olds in each carrier condition (TONE in the left, NOISE-LOW in the middle panel, and NOISE-HIGH in the right panel), plotted as a function of modulation rate. Open diamonds with dashed lines represent simulations of model A, with five channels (i.e., five cochlear filter outputs) and a Q value (i.e., AM-filter selectivity) of 1. Open squares with dotted lines represent simulations of model B, with five channels and a Q value of 5. Finally, open circles with dashed and dotted lines represent simulations of model C, with one channel and a Q value of 1. Models A and B allowed the decision device (the cross-correlator or “template-matching” process) to use temporal-envelope information at the output of all five channels, whereas model C restricted the operation of the decision device to the on-frequency channel where cochlear compression is maximal. Models A and B differed in terms of the degree of modulation-frequency selectivity; model B implements broader modulation filters.

## 2. Data reduction and regression analyses

The present study assessed two characteristics of temporal auditory processing: AM sensitivity and AM masking. To determine whether these temporal processing measures and vocabulary measures were predictive of children’s consonant identification in noise, backward regression analyses were conducted.

Before performing these regression analyses the number of predictors related to temporal processing was reduced. To explore the variability in AM masking data, a principal component analysis (PCA) with varimax rotation and Kaiser normalization was used. The thresholds for AM masking (i.e., the difference between TONE and NOISE-LOW at 32 and 8 Hz, and the difference between TONE and NOISE-HIGH at 32 and 8 Hz) were normalized by age (z scores) and entered in the PCA. This analysis resulted in the extraction of two principal components (PC) as shown in Table I. Together, they explained 78% of the variance in the AM masking data, with PC1 accounting for 45% and PC2 for 33%. The first PC was interpreted as an overall measure of AM masking for AM rate of 32 Hz, while the second PC reflected a measure of AM masking for AM rate of 8 Hz.

As the main analyses of AMDTs (see Sec. III A) revealed an interaction between AM rate and Carrier (no difference between NOISE LOW and TONE carrier at 32 Hz indicating less masking effect at 32 Hz), and a floor effect at 4 Hz in the NOISE-HIGH condition, we chose to select only PC2 related to 8 Hz modulation rate for the AM masking data for the following regression analyses. For the predictor reflecting more basic AM sensitivity, we computed the

average AMDTs obtained at 8 Hz with the three carriers, but age-normalized to make them independent of age.

Thus, in the following backward regression analyses, one measure reflecting AM masking at 8 Hz, one measure reflecting AM sensitivity at 8 Hz, and the vocabulary scores of the children were entered as potential predictors (main effects only) of consonant identification thresholds in noise. As the phonetic contrast was a significant factor in the previous analyses (thresholds for identification of place being better than those for voicing), regression models were performed for voicing and for place separately (using age-normalized z scores as the outcome variables).

For the voicing contrasts, identification thresholds in noise were not predicted by any of the predictors (see Table II). For place of articulation, which was easier to identify compared to voicing in this task, identification thresholds in noise were best predicted by a model including the average age-normalized AMDTs at 8 Hz thought to reflect AM sensitivity [ $R^2 = 9.0\%$ ,  $\text{adj } R^2 = 7.6\%$ ,  $F(1, 66) = 5.77$ ,  $p = 0.01$ , see Table II].

## IV. DISCUSSION

The main goal of the present study was to systematically assess the development of temporal auditory processing to better understand the development of auditory mechanisms involved in speech identification in noise. More precisely, this study measured AM detection abilities between 5 and 11 years of age by exploring the development of selective processing of AM features as well as the development of efficiency in using the available AM information.

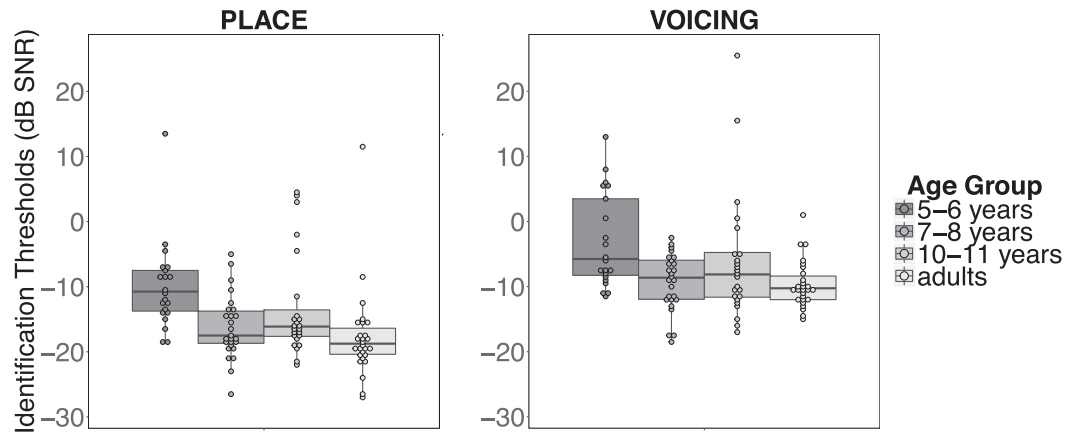


FIG. 8. Identification thresholds (in dB SNR) of consonants varying in (a) place of articulation (b) or voicing, plotted by age group (boxplots and individual data).

The coding of AM is constrained by sensory factors such as the temporal resolution of the auditory system (the ability to detect fast AM changes) and the selectivity of AM filters (tuning of modulation filters to specific AM rates). We measured sensitivity to AM cues using three target AM rates (4, 8, and 32 Hz). As expected based on previous data (Hall and Grose, 1994), 5–6-year-olds showed worse AMDTs in all conditions than older children and adults, but all age groups showed a similar effect of AM rate on AM detection. For the AM rates under study here (AM rates  $\leq 32$  Hz), this result may reflect mature temporal resolution and mature temporal integration of AM cues by 5 years of age, as all the listeners show better coding of AM cues with higher modulation rates.

The characteristics of AM filters involved in the processing of temporal-envelope information in sounds were further measured by looking at the dependence of AM masking on the shape of the modulation spectrum of the AM masker. Results indicated that young children (5–6 years) showed a relatively smaller *amount* of AM masking than older children (although that claim may only be fully supported by the results at 8 Hz). However, those young children clearly showed the same *pattern* of AM masking as older children. Children’s ability to detect AM in the presence of the (low-pass) modulation masker changed similarly as a function of AM rate irrespective of age; in addition, their ability to detect AM decreased irrespective of age when the strength of the modulation masker increased (from NOISE-LOW to NOISE HIGH). This indicates that the selective processing of AM information is mature by 5 years of age. Our results

are consistent with previous studies regarding the development of temporal resolution (e.g., Hall and Grose, 1994; Levi and Werner, 1996; Stuart, 2005) and also revealed mature AM selectivity by 5 years of age. Thus, the overall worse AMDTs and smaller amount of AM masking in younger children may not only be related to the development of sensory mechanisms involved in the extraction of AM information. It has been proposed that temporal processing is not only constrained by sensory coding of AM cues, but also by the ability to make efficient use of the extracted AM cues. Higher efficiency in processing AM cues may then result in better AMDTs.

In order to differentiate effects of sensory processing and processing efficiency on AM detection capacity in childhood, we used a computational model of AM processing to predict AMDTs. The model confirmed that the worse thresholds of 5–6-year-olds cannot be explained by changes in AM sensitivity (i.e., sensory processing), but are more likely to result from higher levels of internal noise (related to processing efficiency) in this youngest group. Furthermore, the model correctly predicted the more modest AM masking at this age when higher levels of internal noise were computed for this group.

We also assessed the role of decision and listening strategies in the children’s data. Modelling indicates that reduced sensitivity to AM might also result from sub-optimal decision strategies (i.e., inappropriate weighting of AM information across cochlear frequency channels). Nevertheless, comparisons across models suggest that reducing internal

TABLE I. PCA for AM masking data (i.e., difference between the thresholds in the NOISE conditions and the TONE condition). Factor loadings  $>0.4$  are highlighted in bold font.

	Principal Component 1	Principal Component 2
Masking NOISE HIGH 32 Hz	<b>0.90</b>	−0.01
Masking NOISE LOW 32 Hz	<b>0.89</b>	0.11
Masking NOISE HIGH 8 Hz	−0.12	<b>0.87</b>
Masking NOISE LOW 8 Hz	0.24	<b>0.83</b>

TABLE II. Summary of the backward regression models.  $\beta$  refers to the standardized regression coefficient. Bold indicates significant at  $\alpha < 0.05$ . The variance inflation factor (VIF) values for all predictors are  $<3.5$ .

Outcome variable	Predictors	$\beta$	$t$	$p$	Status
Voicing	Vocabulary	−0.02	−0.19	0.85	<i>excluded</i>
	AM sensitivity	−0.14	1.13	0.26	<i>excluded</i>
	AM masking	−0.03	−0.22	0.83	<i>excluded</i>
Place	Vocabulary	−0.12	−0.96	0.34	<i>excluded</i>
	<b>AM sensitivity</b>	<b>0.44</b>	<b>2.54</b>	<b>0.01</b>	<b><i>included</i></b>
	AM masking	−0.10	−0.78	0.44	<i>excluded</i>



noise by a factor 10 between 5 and 11 years may better account for the developmental trends in AM sensitivity and amount of AM masking than changes in the decision rule or listening strategy. This result is consistent with previous modelling studies on backward masking, suggesting that younger children show greater internal noise than older children and adults, resulting in poorer detection thresholds (Hill *et al.*, 2004).

Higher levels of internal noise in younger children may result in higher within-individual variability in listeners' responses. This higher variability may be related to the stochastic nature of neural responses, which may reduce with neural maturation. This reduction with age might arise because of progressive increase in neural inhibition and decrease in spontaneous neural activity in auditory brain structures over development (Babola *et al.*, 2018; Cai *et al.*, 2017; Wang and Bergles, 2015). Decrease of internal noise levels with age could also relate to progressive decrease in random fluctuations of selective auditory attention capacities (Jones *et al.*, 2015). This last notion is also consistent with the present data, where three children were excluded from the youngest group due to inability to complete the task (performing at floor for more than the half of the conditions).

In our study, we also evaluated children's ability to identify speech in noise by using a consonant identification task in a stationary speech-shaped noise masker. Consistent with previous data on children's consonant identification using a stationary white noise masker (Nishi *et al.*, 2010), our results showed that fricative identification thresholds in steady speech-shaped noise are mature by 7 years of age. The population tested in the present study included participants who were exposed to numerous languages. We expected bilingual listeners to show worse identification thresholds for English consonants in noise compared to monolinguals (Lecumberri *et al.*, 2010). However, no main effect of language background was observed in the present speech in noise task, even though bilinguals showed slightly better thresholds for voicing contrasts compared to monolinguals. This interaction between language and speech contrasts is congruent with the fact that different cues indicate voicing between languages, and less so for place of articulation (Lisker and Abramson, 1964). Furthermore, this weak effect is consistent with previous studies showing that after entering school (at age 5 in the UK), sequential bilinguals show similar performance as English monolingual children when detecting English voicing contrasts (McCarthy *et al.*, 2014). Future studies using more complex speech sounds, such as words and sentences, may help to explore the effects of bilingualism on speech in noise perception.

In the aim to better understand the development of consonant identification in noise, regression models were used to explore the relationship between temporal auditory processing, vocabulary level and consonant identification in noise in children. For the place of articulation contrasts, which were easier to identify in noise than voicing, differences in children's thresholds were predicted by a measure of AM sensitivity at 8 Hz. Data from our AM tasks suggest that capacity to extract AM cues via AM filters is mature

by 5 years, but that the level of internal noise of the AM-processing pathway reduces markedly with age. According to our model of AM processing, AM sensitivity at 8 Hz is determined by three factors: the signal-to-noise ratio in the AM domain (as controlled by stimulus fluctuation and the frequency selectivity of AM filters); the variance of internal noise in the AM domain; and the optimality of the decision strategy. Thus, it seems that poor speech in noise performance for some phonetic contrasts relates to some extent to factors responsible for the efficient processing of slow AM cues (the variance of internal noise and the decision strategy). It is possible that higher efficiency in using relatively slow modulations ( $\leq 8$  Hz), corresponding to the most salient modulation components of speech (Varnet *et al.*, 2017), relates to some extent to the improvement in speech intelligibility between 5 and 7 years of age.

The present psychophysical tasks do not allow us to evaluate specifically and individually the level of internal noise for each child, to assess directly whether children who have higher levels of internal noise for AM coding also show poorer speech intelligibility. Future studies implementing new psychophysical tasks estimating an individual listener's processing efficiency for AM cues (internal noise and decision strategies) and using more complex speech sounds (i.e., words) will help to better understand the relationship between temporal processing and the development of speech in noise abilities.

## V. CONCLUSIONS

Speech is associated with slow and salient amplitude modulation (AM) cues. The present study indicated that for AM rates below 32 Hz, 5–6-year-olds show overall poorer thresholds for detecting AM and relatively less AM masking compared to older children (resulting from worse thresholds with tone carriers at 5 years as if tone carriers were “noisier”). However, characteristics of AM sensitivity (the changes in the ability to detect AM as a function of AM rate) and AM masking (the changes in the ability to detect a masked target AM as a function of the separation between target and masker rates) are comparable between 5–6 and 10–11 years of age for rates of 4–32 Hz. Computational modelling indicated that changes in processing efficiency (e.g., a reduction in internal noise in the AM domain) better explain the improvement of AM sensitivity and the increase in AM masking with age than changes in sensory coding of AM information. The present study also suggests a relationship between noise-masked consonant identification and AM sensitivity. Specifically, the ability to discriminate place of articulation contrasts for English fricatives was correlated with sensitivity to 8-Hz AM, but not with AM masking. This suggests that the development of speech intelligibility in childhood may be related in part to the efficiency with which children make use of the relatively slow temporal information of speech.

## ACKNOWLEDGMENTS

This work was supported by the European Union's Horizon 2020 research and innovation programme under

the Marie Skłodowska-Curie Grant Agreement No. 659204 held at the Department of Speech, Hearing and Phonetic Sciences, UCL and by Grant No. ANR-17-CE28-008 awarded to L.C. C.L. was also supported by Grants Nos. ANR-11-0001-02 PSL and ANR-10-LABX-0087. The authors wish to thank Gianna Li and Lucía Sanz for help with testing at schools, as well as Lorna Halliday for helpful feedback. The authors declare no conflict of interest.

## APPENDIX A: DESCRIPTION OF THE MODEL SIMULATING THE AM DETECTION DATA

The first stage was a set of five linear gammatone filters (Patterson *et al.*, 1995) that simulated the bandpass filtering of the basilar membrane in the cochlea. One gammatone filter (called the on-frequency filter) was centered at 1024 Hz. The remaining four filters (the off-frequency filters) were centered one and two Cams (units of the ERBN-number scale; Glasberg and Moore, 1990) above and below the carrier frequency (780.75, 895.82, 1166.79, and 1325.86 Hz). In the second stage, the outputs of the five gammatone filters were half-wave rectified. The rectified output of the on-frequency filter was processed by a broken-stick input-output function, which applied compression for levels above 40 dB SPL (see Moore, 2012). Above 40 dB SPL, the compression was implemented by raising the amplitude to the power 0.3. The rectified signals at the outputs of the four off-frequency filters were not compressed. This crudely simulates the physiological finding that compression in the cochlea occurs mainly for input frequencies close to the characteristic frequency of the place whose response is being measured (Robles and Ruggero, 2001).

The signals derived from the outputs of the five auditory filters were highpass filtered (cutoff = 2.5 Hz; rolloff = 20 dB/decade) to simulate the effects of adaptation in the AM domain (Tchorz and Kollmeier, 1999). The resulting signals were passed through a set of 10 bandpass (modulation) filters, one set for each band, to simulate frequency-selective processing in the AM domain (Dau *et al.*, 1997a,b). The quality factor ( $Q$ ) quantifies the degree of selectivity of the modulation filter, and  $Q$  increasing with frequency selectivity because  $Q = \text{center frequency}/\text{bandwidth}$ . The modulation filters had a  $Q$  value of 1 and a rolloff of  $\pm 20$  dB/decade (Ewert and Dau, 2000; Ewert *et al.*, 2002; Lorenzi *et al.*, 2001; Sek and Moore, 2002). The center frequencies of the modulation bandpass filters ranged between 2 and 120 Hz (Moore *et al.*, 2009) and were spaced logarithmically. For each modulation filter centered below 10 Hz, the waveform at the output of the filter was passed on for further processing, while for each filter centered at and above 10 Hz only the Hilbert envelope of the output was passed on. This was done to simulate the loss of sensitivity to envelope phase for rates above 10 Hz (Dau *et al.*, 1997a,b). Each envelope of the envelope (the so-called “venvelope”; Ewert *et al.*, 2002) was scaled so that the root-mean-square value at the output of modulation filters was the same before and after the Hilbert transformation.

The outputs of the modulation filters (the envelope for  $fm < 10$  Hz; the venvelope for  $fm \geq 10$  Hz) were down-sampled by a factor 10 and then “degraded” by a source of internal noise, which were introduced to limit the performance of the model (Dau *et al.*, 1997a,b). This noise, named  $\varepsilon_{add}$ , was additive. It was modeled as a Gaussian noise with zero mean and with magnitude specified by its SD. Independent noise samples were added on a sample-by-sample basis to each modulation filter output. For each group of listeners, the SD of  $\varepsilon_{add}$  was initially adjusted to fit AMDTs measured with an 8-Hz AM rate and a tonal carrier. Goodness of fit was estimated by computing the difference between the observed and simulated AMDTs. Goodness of fit was typically below 2 dB in the simulations presented, except for simulations using modulation filters with a  $Q$  value of 0.5 (cf. Fig. 7).

The decision stage was realized as a simplified version of the optimal detector described by Dau *et al.* (1997a,b). The output of the model was computed for the two stimuli in each trial. A template was generated at the beginning of each simulated threshold measurement (that is, for the first trial of each adaptive staircase) as the difference between the model output in response to the target and standard stimuli. This was done with the internal noise source set to zero. Note that this largely cancelled out the effects of the added AM on the internal representation, and therefore it represented an idealized version of what listeners do. For each subsequent trial, the template was cross-correlated with the model output for each interval in that trial. The lags used by the cross-correlation device were restricted to  $\pm 1$  modulation cycle. The normalized cross-correlation was computed separately for each band and each modulation filter. The correlation functions were then summed across all bands to estimate the lag corresponding to the best match. The interval leading to the largest cross-correlation was taken as the target interval identified by the model for that trial. Stimuli were generated as for the behavioral experiment including the roving of overall level.

## APPENDIX B: SUPPLEMENTARY ANALYSES OF CHILDREN'S THRESHOLDS

The group of 5–6-year-olds showed a specific pattern of results compared to other age groups. This pattern of results could be related to specific attention issues in the TONE condition. Variable attention would lead to wider swings in the adaptive track and hence increased SDs for threshold estimates. To evaluate whether 5–6-year-olds' thresholds were differentially affected by attention compared to the other age groups, the SD for the last four reversals were analysed using a  $3 \times 3 \times 3$  repeated-measures ANOVA (Age Group  $\times$  Carrier  $\times$  Rate). This analysis showed that in each age group, values of SD were affected by the Carrier [ $F(2,276) = 33.52$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.195$ ], and that the NOISE-HIGH carrier led to higher SDs than the two other carriers, which did not differ between each other. The SDs were also significantly affected by the Rate [ $F(2, 276) = 230.31$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.625$ ], with higher rates associated with lower SDs. The SDs were also affected by Age [ $F(2, 69) = 9.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.218$ ], with 5–6-year-

olds showing overall higher SDs than the 7–8- and 10–11-year-olds, which did not differ between each other. This analysis also revealed a significant interaction between Rate and Carrier [ $F(4, 276) = 15.67$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.185$ ], but no interaction involving the factor Age Group. Thus, it seems that the thresholds of the 5–6-year-olds were generally somewhat less reliable than for the older children. Nevertheless, their SDs were not affected differently by Rate or Carrier, thus the small difference in thresholds between TONE and NOISE-LOW carriers at 5–6 years cannot be explained by specific attentional issues in the TONE conditions at this age.

- Amitay, S., Guiraud, J., Sohoglu, E., Zobay, O., Edmonds, B. A., Zhang, Y.-X., and Moore, D. R. (2013). "Human decision making based on variations in internal noise: An EEG study," *PLoS One* 8(7), e68928.
- Babola, T. A., Li, S., Gribizis, A., Lee, B. J., Issa, J. B., Wang, H. C., Crair, M. C., and Bergles, D. E. (2018). "Homeostatic control of spontaneous activity in the developing auditory system," *Neuron* 99(3), 511–524.
- Bacon, S. P., and Grantham, D. W. (1989). "Modulation masking: Effects of modulation frequency, depth, and phase," *J. Acoust. Soc. Am.* 85(6), 2575–2580.
- Baker, R. J., and Rosen, S. (2001). "Evaluation of maximum-likelihood threshold estimation with tone-in-noise masking," *Brit. J. Audiol.* 35(1), 43–52.
- Biberger, T., and Ewert, S. D. (2016). "Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility," *J. Acoust. Soc. Am.* 140(2), 1023–1038.
- Biberger, T., and Ewert, S. D. (2017). "The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking," *J. Acoust. Soc. Am.* 142(2), 1098–1111.
- Buss, E., Hall, J. W., III, and Grose, J. H. (2006). "Development and the role of internal noise in detection and discrimination thresholds with narrow band stimuli," *J. Acoust. Soc. Am.* 120(5), 2777–2788.
- Buss, E., Hall, J. W., III, and Grose, J. H. (2013). "Factors affecting the processing of intensity in school-aged children," *J. Speech Lang. Hear. Res.* 56(1), 71–80.
- Buss, E., Leibold, L. J., and Lorenzi, C. (2018). "Speech recognition for school-age children and adults tested in multi-tone vs multi-noise-band maskers," *J. Acoust. Soc. Am.* 143(3), 1458–1466.
- Buss, E., Lorenzi, C., Cabrera, L., Leibold, L. J., and Grose, J. H. (2019). "Amplitude modulation detection and modulation masking in school-age children and adults," *J. Acoust. Soc. Am.* 145(4), 2565–2575.
- Cai, D., Han, R., Liu, M., Xie, F., You, L., Zheng, Y., Zhao, L., Yao, J., Wang, Y., Yue, Y., and Schreiner, C. E. (2017). "A critical role of inhibition in temporal processing maturation in the primary auditory cortex," *Cerebral Cortex* 28(5), 1610–1624.
- Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (1994). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," *J. Acoust. Soc. Am.* 96, 2048–2054.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* 102(5), 2892–2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* 102(5), 2906–2919.
- Drullman, R. (1995). "Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level," *J. Acoust. Soc. Am.* 98(3), 1796–1798.
- Dubbelboer, F., and Houtgast, T. (2007). "A detailed study on the effects of noise on speech intelligibility," *J. Acoust. Soc. Am.* 122(5), 2865–2871.
- Dunn, L. M., and Dunn, D. M. (2009). *The British Picture Vocabulary Scale* (GL Assessment, London).
- Elliott, L. L. (1979). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," *J. Acoust. Soc. Am.* 66(3), 651–653.
- Elliott, L. L., Hammer, M. A., and Evan, K. E. (1987). "Perception of gated, highly familiar spoken monosyllabic nouns by children, teenagers, and older adults," *Percept. Psychophys.* 42(2), 150–157.
- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* 108(3), 1181–1196.
- Ewert, S. D., Verhey, J. L., and Dau, T. (2002). "Spectro-temporal processing in the envelope-frequency domain," *J. Acoust. Soc. Am.* 112(6), 2921–2931.
- Faisal, A. A., Selen, L. P., and Wolpert, D. M. (2008). "Noise in the nervous system," *Nat. Rev. Neurosci.* 9(4), 292.
- Fu, Q.-J. (2002). "Temporal processing and speech recognition in cochlear implant users," *Neuroreport* 13(13), 1635–1639.
- Füllgrabe, C., Moore, B. C., and Stone, M. A. (2015). "Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition," *Front. Aging Neurosci.* 6, 347.
- Giraud, A. L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., and Kleinschmidt, A. (2000). "Representation of the temporal envelope of sounds in the human brain," *J. Neurophysiol.* 84(3), 1588–1598.
- Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* 47(1–2), 103–138.
- Gnansia, D., Lazard, D. S., Léger, A. C., Fugain, C., Lancelin, D., Meyer, B., and Lorenzi, C. (2014). "Role of slow temporal modulations in speech identification for cochlear implant users," *Int. J. Audiol.* 53(1), 48–54.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- Hall, J. W., 3rd, and Grose, J. H. (1994). "Development of temporal resolution in children as measured by the temporal modulation transfer function," *J. Acoust. Soc. Am.* 96(1), 150–154.
- Hall, J. W., III, Grose, J. H., Buss, E., and Dev, M. B. (2002). "Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children," *Ear Hear.* 23(2), 159–165.
- Hartley, D. E., and Moore, D. R. (2002). "Auditory processing efficiency deficits in children with developmental language impairments," *J. Acoust. Soc. Am.* 112(6), 2962–2966.
- Hartley, D. E., Wright, B. A., Hogan, S. C., and Moore, D. R. (2000). "Age-related improvements in auditory backward and simultaneous masking in 6- to 10-year-old children," *J. Speech Lang. Hear. Res.* 43(6), 1402–1415.
- Hazan, V., and Barrett, S. (2000). "The development of phonemic categorization in children aged 6–12," *J. Phon.* 28(4), 377–396.
- Hill, P. R., Hartley, D. E., Glasberg, B. R., Moore, B. C., and Moore, D. R. (2004). "Auditory processing efficiency and temporal resolution in children and adults," *J. Speech Lang. Hear. Res.* 47(5), 1022–1029.
- Hoaglin, D. C., and Iglewicz, B. (1987). "Fine-tuning some resistant rules for outlier labeling," *J. Am. Stat. Assoc.* 82(400), 1147–1149.
- Houtgast, T. (1989). "Frequency selectivity in amplitude-modulation detection," *J. Acoust. Soc. Am.* 85(4), 1676–1680.
- Houtgast, T., and Steeneken, H. J. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* 77(3), 1069–1077.
- Ingram, D., Christensen, L., Veach, S., and Webster, B. (1980). "The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years," in *Child Phonology* (Elsevier, Amsterdam), pp. 169–192.
- Jacobi, I., Sheikh Rashid, M., de Laat, J. A. M., and Dreschler, W. A. (2017). "Age dependence of thresholds for speech in noise in normal-hearing adolescents," *Trends Hear* 21, 1–9.
- Jamieson, D. G., Kranjc, G., Yu, K., and Hodgetts, W. E. (2004). "Speech intelligibility of young school-aged children in the presence of real-life classroom noise," *J. Am. Acad. Audiol.* 15(7), 508–517.
- Jepsen, M. L., Ewert, S. D., and Dau, T. (2008). "A computational model of human auditory signal processing and perception," *J. Acoust. Soc. Am.* 124(1), 422–438.
- Johnson, C. E. (2000). "Children's phoneme identification in reverberation and noise," *J. Speech Lang. Hear. Res.* 43(1), 144–157.
- Jones, P. R., Moore, D. R., and Amitay, S. (2015). "Development of auditory selective attention: Why children struggle to hear in noisy environments," *Dev. Psychol.* 51(3), 353–369.
- Jørgensen, S., and Dau, T. (2011). "Predicting speech intelligibility based on the signal-to-noise envelope power ratio after modulation-frequency selective processing," *J. Acoust. Soc. Am.* 130(3), 1475–1487.
- Jørgensen, S., Ewert, S. D., and Dau, T. (2013). "A multi-resolution envelope-power based model for speech intelligibility," *J. Acoust. Soc. Am.* 134, 436–446.
- Lecumberri, M. L. G., Cooke, M., and Cutler, A. (2010). "Non-native speech perception in adverse conditions: A review," *Speech Commun.* 52(11), 864–886.



- Leibold, L. J., and Buss, E. (2016). "Factors responsible for remote-frequency masking in children and adults," *J. Acoust. Soc. Am.* **140**(6), 4367–4377.
- Levi, E. C., and Werner, L. A. (1996). "Amplitude modulation detection in infancy: Update on 3-month-olds," *Assoc. Res. Otolaryngol.* **19**, 142–150.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**(2B), 467–477.
- Liégeois-Chauvel, C., Lorenzi, C., Trébuchon, A., Régis, J., and Chauvel, P. (2004). "Temporal envelope processing in the human left and right auditory cortices," *Cerebral Cortex* **14**(7), 731–740.
- Lisker, L., and Abramson, A. S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," *Word* **20**(3), 384–422.
- Lorenzi, C., Simpson, M. I., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. (2001). "Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers," *J. Acoust. Soc. Am.* **110**(5), 2470–2478.
- McCarthy, K. M., Mahon, M., Rosen, S., and Evans, B. G. (2014). "Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition," *Child Dev.* **85**(5), 1965–1980.
- McCreery, R. W., Spratford, M., Kirby, B., and Brennan, M. (2017). "Individual differences in language and working memory affect children's speech recognition in noise," *Int. J. Audiol.* **56**(5), 306–315.
- Moeller, M. P., Hoover, B., Putman, C., Arbataitis, K., Bohnenkamp, G., Peterson, B., Wood, S., Lewis, D., Pittman, A., and Stelmachowicz, P. (2007). "Vocalizations of infants with hearing loss compared with infants with normal hearing: Part I—phonetic development," *Ear Hear.* **28**(5), 605–627.
- Moore, B. C. (2012). *An Introduction to the Psychology of Hearing* (Brill, Leiden).
- Moore, B. C., Füllgrabe, C., and Sek, A. (2009). "Estimation of the center frequency of the highest modulation filter," *J. Acoust. Soc. Am.* **125**(2), 1075–1081.
- Moore, B. C., and Sek, A. (1994). "Effects of carrier frequency and background noise on the detection of mixed modulation," *J. Acoust. Soc. Am.* **96**(2), 741–751.
- Moore, D. R., Ferguson, M. A., Edmondson-Jones, A. M., Ratib, S., and Riley, A. (2010). "Nature of auditory processing disorder in children," *Pediatrics* **126**(2), e382–e390.
- Neuman, A. C., and Hochberg, I. (1983). "Children's perception of speech in reverberation," *J. Acoust. Soc. Am.* **73**(6), 2145–2149.
- Nishi, K., Lewis, D. E., Hoover, B. M., Choi, S., and Stelmachowicz, P. G. (2010). "Children's recognition of American English consonants in noise," *J. Acoust. Soc. Am.* **127**(5), 3177–3188.
- Nittrouer, S. (2004). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," *J. Acoust. Soc. Am.* **115**(4), 1777–1790.
- Papso, C. F., and Blood, I. M. (1989). "Word recognition skills of children and adults in background noise," *Ear Hear.* **10**(4), 235–236.
- Patterson, R. D., Allerhand, M. H., and Giguere, C. (1995). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," *J. Acoust. Soc. Am.* **98**(4), 1890–1894.
- Robles, L., and Ruggero, M. A. (2001). "Mechanics of the mammalian cochlea," *Physiol. Rev.* **81**(3), 1305–1352.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory and linguistic aspects," *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **336**(1278), 367–373.
- Sek, A., Baer, T., Crinnion, W., Springgag, A., and Moore, B. C. (2015). "Modulation masking within and across carriers for subjects with normal and impaired hearing," *J. Acoust. Soc. Am.* **138**(2), 1143–1153.
- Sek, A., and Moore, B. C. (2002). "Mechanisms of modulation gap detection," *J. Acoust. Soc. Am.* **111**(6), 2783–2792.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**(5234), 303–304.
- Shield, B. M., and Dockrell, J. E. (2003). "The effects of noise on children at school: A review," *Build. Acoust.* **10**(2), 97–116.
- Stone, M. A., Füllgrabe, C., Mackinnon, R. C., and Moore, B. C. (2011). "The importance for speech intelligibility of random fluctuations in 'steady' background noise," *J. Acoust. Soc. Am.* **130**(5), 2874–2881.
- Stone, M. A., Füllgrabe, C., and Moore, B. C. (2012). "Notionally steady background noise acts primarily as a modulation masker of speech," *J. Acoust. Soc. Am.* **132**(1), 317–326.
- Stone, M. A., and Moore, B. C. (2014). "On the near non-existence of 'pure' energetic masking release for speech," *J. Acoust. Soc. Am.* **135**(4), 1967–1977.
- Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**(6), 3638–3646.
- Stuart, A. (2005). "Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise," *Ear Hear.* **26**(1), 78–88.
- Stuart, A. (2008). "Reception thresholds for sentences in quiet, continuous noise, and interrupted noise in school-age children," *J. Am. Acad. Audiol.* **19**(2), 135–146.
- Tchorz, J., and Kollmeier, B. (1999). "A model of auditory perception as front end for automatic speech recognition," *J. Acoust. Soc. Am.* **106**(4), 2040–2050.
- Trochim, W., and Donnelly, J. P. (2006). *Research Methods Knowledge Base* (Atomic Dog Publishers, Cincinnati, OH).
- Van Rietschote, H. F., Houtgast, T., and Steeneken, H. J. M. (1981). "Predicting speech intelligibility in rooms from the modulation transfer function IV: A ray-tracing computer model," *Acta Acust. Acust.* **49**(3), 245–252.
- Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., and Lorenzi, C. (2017). "A cross-linguistic study of speech modulation spectra," *J. Acoust. Soc. Am.* **142**(4), 1976–1989.
- Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," *J. Acoust. Soc. Am.* **66**, 1364–1380.
- Walker, B. A., Gerhards, C. M., Werner, L. A., and Hom, D. L. (2019). "Amplitude modulation detection and temporal modulation cutoff frequency in normal hearing infants," *J. Acoust. Soc. Am.* **145**(6), 3667–3674.
- Wallaert, N., Varnet, L., Moore, B. C., and Lorenzi, C. (2018). "Sensorineural hearing loss impairs sensitivity but spares temporal integration for detection of frequency modulation," *J. Acoust. Soc. Am.* **144**(2), 720–733.
- Wang, H. C., and Bergles, D. E. (2015). "Spontaneous activity in the developing auditory system," *Cell Tissue Res.* **361**(1), 65–75.
- Werner, L. A. (1996). "The development of auditory behavior (or what the anatomists and physiologists have to explain)," *Ear Hear.* **17**(5), 438–446.
- Werner, L. A., and Boike, K. (2001). "Infants' sensitivity to broadband noise," *J. Acoust. Soc. Am.* **109**, 2103–2111.
- Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., and Rubinstein, J. T. (2011). "Acoustic temporal modulation detection and speech perception in cochlear implant listeners," *J. Acoust. Soc. Am.* **130**(1), 376–388.
- Zwicker, E. (1952). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones," *Acta Acust. Acust.* **2**(3), 125–133.
- Zwicker, E. (1956). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs," *Acta Acust. Acust.* **6**(4), 365–381.