DOI: 10.61782/fa.2023.0536

DEVELOPMENT OF AUDITORY SENSITIVITY TO AMPLITUDE MODULATION CUES: SENSORY AND COGNITIVE DETERMINANTS AND RELATIONSHIP WITH SPEECH INTELLIGIBILITY

Irene Lorenzini[3]* **Léo Varnet**[1] **Christian Lorenzi**[2] **Laurianne Cabrera**[3]
[1] Université de Paris Cité, CNRS, Integrative Neuroscience and Cognition Center, F-75006 Paris, France
[2] Laboratoire des Systèmes Perceptifs (UMR 8248), CNRS, Ecole normale supérieure, Université Paris Sciences & Lettres (PSL), Paris, France

ABSTRACT

Speech sounds convey relatively slow Amplitude Modulation cues whose processing plays a crucial role for speech comprehension. However, the development of AM processing and its interaction with speech intelligibility remains unclear. Previous studies suggested that AM processing development relates to changes in the central filtering of AM cues or in 'processing efficiency' (i.e., a reduction in internal noise and/or improvements in the optimality of decision making). Here, we explored the contribution of (i) the ability to combine AM cues over time (temporal integration), (ii) response consistency for AM detection, on children's in-noise consonant discrimination. Temporal integration developed until 11 years. Response consistency in AM detection also increased with age. Temporal integration at higher AM rates and AM detection consistency were statistically related to identification thresholds in noise for a subset of the tested consonants. Children vocabulary was not a better predictor of speech intelligibility compared to the measures of AM processing. Overall, the development of AM processing and its interaction with speech intelligibility may result from changes in (central) processing efficiency for AM.

Keywords: *Auditory development, Speech perception development, Amplitude Modulation detection, Speech-in-noise perception development.*

1. INTRODUCTION

The human auditory system extracts and represents the audio-frequency components of complex sounds and their changes over time [1]. Slow and fast changes in amplitude – amplitude modulation (AM) or "temporal envelope" cues – convey information about syllabic and phonetic information, as well as voice pitch and supra-segmental information [2]. Both slow and fast AM cues are commonly assumed to be extracted by broadly tuned modulation filters located at central (*i.e.*, post-cochlear) stages of the human auditory system [3]. The specific tuning characteristics of these AM filters (*i.e.*, their bandwidth) explain masking or interference effects demonstrated repeatedly in the AM domain [3]. The ability to detect and combine these slow and fast AM cues over time (also called 'temporal integration') by human listeners is currently understood as resulting from the limitations introduced by internal variability – also called 'internal noise' – and the operation of (late) decision-making mechanisms based on template matching [4].

The relative importance of slow and fast AM for speech perception in adult listeners has been addressed in a wealth of studies. Overall, it has been shown that adults rely on the slowest AM cues (< 16 Hz) to identify speech sounds in

*__*Corresponding author*__: *first.author@email.ad.*

quiet. However, they do require the faster AM cues (> 16 Hz) to correctly perceive speech in complex backgrounds such as noise of competing voices [5]. More recently, speech perception in noise has been shown to be highly dependent on the ability to detect and use the AM cues of speech that interfere with the AM cues of a masking sound [6]. Indeed, Stone et al. (2012) showed that 'notionally' steady-state noise maskers with high levels of intrinsic random envelope fluctuations lead to a greater reduction in the intelligibility of speech compared to noise maskers with lower inherent fluctuations. In the same vein, speech intelligibility in noise for cochlear-implant adult users is significantly correlated with the basic ability to detect AM [7].

Interestingly, these two abilities – auditory temporal processing and speech perception in noise – both take a long time to reach adult-levels. A pioneering study compared children's sensitivity to AM at several rates (the 'temporal modulation transfer function' or TMTF) between 4 and 10 years of age using a broadband noise carrier [8]. This study indicated that AM detection thresholds (*i.e.*, AM sensitivity) improve with age and reach adult-like levels around 9 years. Moreover, the ability to track AM (measured using non-speech sounds) constrains children's capacity in perceiving speech in noise [9]. Thus, poor speech perception in noise in childhood may relate to poor auditory temporal processing. In a series of experimental and computational studies, we explored the development of AM processing from 5 to 11 years of age and investigated the development of sensory *vs* cognitive determinants in sensitivity to AM.

In a first study [10], we evaluated AM sensitivity at slow AM rates (4, 8, and 32 Hz) and masking effects produced by the intrinsic random AM fluctuations conveyed by a narrow-band noise carrier on auditory detection of a sinusoidal AM target. This was achieved to assess whether children have broader modulation filters than adults. In a second experiment [11], we evaluated temporal integration in the AM domain, that is the improvement of AM detection thresholds with increasing number of AM cycles, to assess whether children have less efficient central mechanisms involved in decision making than adults. In a third experiment [12], we measured response consistency in AM detection over trial-to-trial when listeners were tested at detection thresholds and when sinusoidal AM targets were presented with a narrow-band noise carrier. This final study was achieved to assess whether children have higher internal noise in the AM domain than adults. Overall, the results indicated that children of 5-6 years show the worst AM detection thresholds in all conditions and that children from 5-to-11 years show worse thresholds than young adults. Nevertheless, all age groups showed: 1) similar effects of AM rate and carrier, that is, better thresholds with increasing modulation rates and when presented with a deterministic tone carrier compared to narrow-band noise carriers, 2) similar temporal integration effects, that is, better AM detection thresholds with increasing number of AM cycles. Importantly, age-related changes were also observed on AM detection consistency.

To better understand these developmental trends in AM detection, we developed a series of computational models of human auditory processing based on the modulation-filterbank and template-matching concepts [3], [4], [13], [14], to simulate the AM detection thresholds of children and adults in each condition. Two sets of model parameters were manipulated to distinguish the role of sensory factors (*i.e.*, selectivity of modulation filters) and processing efficiency: 1) the effects of internal-noise sources modelled as additional gaussian noises at the output of modulation filters, and 2) sub-optimal decision strategies modelled as template matching operating on the noisy output of modulation filters. In these three studies, modelling revealed that age-related changes in AM detection were better simulated by changes in processing efficiency (maturation of internal noise and template matching strategy) rather than changes in sensory processing (*i.e.* maturation of modulation filtering).

In those studies, we also measured consonant-identification thresholds in a steady-state speech-shaped noise and performance for receptive vocabulary for all children. Consonant-identification thresholds improved in parallel with AM detection thresholds. Moreover, preliminary analyses showed that AM detection thresholds for 8-Hz target modulations, but not receptive vocabulary skills, were significantly predictive of consonant-identification thresholds in noise for place of articulation contrasts of fricative consonants. The goal of the present paper was to explore thoroughly the relationship between the different measures of AM processing and consonant-in-noise identification in children. In a first experiment, we investigated the relationship between temporal integration capacities and fricative consonant identification in noise. In a second experiment, we considered the relationship between AM detection consistency and identification thresholds of three phonetic features (voicing, place and manner of articulation) in noise. We hypothesized that measures of auditory AM processing related to processing efficiency, may be more strongly associated with speech-in-noise abilities compared to a measure of general linguistic development.

2. EXPERIMENT I

2.1 Methods

2.1.1 Participants

English-speaking children from 5 to 11 years were recruited in this experiment (see also [10], [11]). Consent forms were obtained from parents as approved by the university ethics committee (Project ID: 8107/001). All children included in these experiments had normal hearing (absolute thresholds for pure tones between 0.25 and 8 kHz < 20 dB HL) and typical cognitive development (reported by teachers and the block subtest from the Wechsler Intelligence Scale for Children evaluating non-verbal reasoning [15]). Twenty-one 5-6-year-old children (10 females; mean age = 5.7 years, *SD* = 0.4), 27 7–8-year-olds (13 females; mean age = 7.8 years, *SD* = 0.5) and 24 10–11-year-olds (12 females; mean age = 10.7, *SD* = 0.4) were included in the data analyses.

2.1.2 AM tasks

Stimuli

In order to assess temporal integration for AM detection, four experimental conditions were designed to manipulate AM rate (4 *vs* 32 Hz) and number of modulation cycles (2 *vs* 8 cycles) see [11]. These manipulations involved changes in sound duration such that the target sounds modulated at 32 Hz were 62.5- and 250-ms long in the 2 and 8-cycle conditions, respectively, and the target sounds modulated at 4 Hz were 500- and 2000-ms long in the 2 and 8-cycle conditions, respectively. The carriers were sinusoidal tones centred at 1027 Hz, generated with a random starting phase. All stimuli were generated at a sampling frequency of 44.1 kHz. In all conditions, the stimuli included 50-ms raised-cosine onset/offset ramps. Standard sounds were not modulated in amplitude, and target sounds were modulated at depths m ranging from m=100% to m=1% in 20 steps of 2 dB. The starting phase of the modulation was randomized on each trial.

Procedure

In each of the four AM conditions, AM detection thresholds were measured using a three-interval, three-alternative forced-choice (3I-3AFC) adaptive procedure. Stimuli were presented using headphones (Sennheiser HD 25-SP II) at 65 dB SPL. Responses were collected using a touch-screen tablet. On each trial, three animal characters appeared on the screen and produced a sound one after the other. The inter-stimulus interval was 500 ms and inter-trial intervals 600 ms. Children were asked to select the odd-one out, corresponding to the modulated target. Visual and corrective feedback was presented on each trial. A first one-down, one-up rule was used until the first reversal followed by a 2-down 1-up adaptive procedure [16]. The first trial was modulated at m = 100%. The first step size was 6 dB, reduced to 4 dB after the first reversal, and to 2 dB after the second reversal. The run stopped after the 8th reversal or after 32 trials. One estimate was collected for each participant in each condition and a second estimate was collected if fewer than 5 reversals were obtained, or if the track did not converge. The threshold in dB was the geometric mean of the last four reversals. Estimates of temporal integration were calculated as the difference between threshold obtained in the 8-cycle and 2-cycle condition at each modulation rate. The outcome variable analyzed was the average of these two scores for each participant standardized by age.

2.1.3 Speech-in-noise tasks

Stimuli

Three native Southern British English speakers were recorded: one male (F0 = 112 Hz) and two females (F0 = 153 and 160 Hz). /aCa/ syllables were selected, where C = /f/, /v/, /ʃ/, /ʒ/, /s/, /z/. A steady speech-shaped noise masker was generated with long-term spectrum similar to the spectrum of the syllables uttered by the female voices. Two phonetic conditions were presented to participants: one presenting a minimal change in voicing /f/-/v/, /s/-/z/, /ʃ/-/ʒ/ ('Voicing' condition), and one presenting a minimal change in place of articulation /f/-/s/, /v/-/z/, /s/-/ʃ/, /z/-/ʒ/ ('Place of articulation' condition) see [10]. Both conditions entailed fricative phonemes only.

Procedure

A forced-choice XAB task was used to present the syllables. It was implemented on the same tablet and used the same headphones calibrated at 65 dB SPL as above. On each trial, a first character on the top of the screen produced a sound X in quiet (always uttered by the male speaker), then, two characters on the bottom of the screen produced the sounds A and B, played in noise (each uttered by one of the two female speakers). The inter-stimulus and inter-trial intervals were the same as in the previous task and feedback was also provided. Children were asked to select the character at the bottom who pronounced the same sound as the one on the top. The noise level was varied on a trial-to-trial basis 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 stopped after the 6th reversal was reached or after 32 trials.

The consonant identification threshold in noise was calculated as the geometric mean of the last four reversals and was then standardized by age.

2.1.4 Receptive Vocabulary Assessment

Spoken language receptive vocabulary was assessed using the British Picture Vocabulary Scale (BPVS third edition [17]). Children were presented with four pictures on each trial and required to select the one that best illustrated the meaning of a word uttered by the experimenter. The raw scores were normalized by the age of the child.

2.2 Results

2.2.1 Temporal integration for AM and speech-in-noise identification thresholds

As detailed in Cabrera et al. (2022) [11], statistical analyses showed no significant effect of Rate [$F(1,98) = 1.03$, $p = .312$], a significant effect of Age [$F(3,98) = 5.60$, $p = .001$, $\eta 2 = 0.15$] and no interaction between these factors on temporal-integration scores. The 5-6-year-olds showed higher integration scores (that is, better ability to combine AM cues) compared to older children due to their worst thresholds in the 2-cycle conditions.
For the two conditions of speech-in-noise, a significant main effect of Age and Condition on identification thresholds was observed, but no interaction between the two factors (see [10] for details). Higher (worse) thresholds for voicing contrasts as compared to place of articulation was observed overall, and the group of 5-6 years showed the worst thresholds.

2.2.2 Relationship between temporal integration for AM and speech-in-noise identification thresholds

Using backward regression analyses, we assessed whether temporal-integration scores, as averaged on the two AM rates (4 and 32 Hz), and receptive vocabulary scores could predict speech-in-noise identification thresholds in each phonetic condition.
None of the regression models predicted a significant part of the variance of the speech-in-noise thresholds (see Tab. 1).
We then ran regression models including temporal-integration scores at 4 and 32 Hz separately and receptive vocabulary scores. This model showed a significant contribution of the integration scores obtained in the 32-Hz condition in predicting identification thresholds for place of articulation [$R^2 = 9.4\%$, adj. $R^2 = 5\%$, $F(3,61) = 2.111$, $p = 0.108$, see Tab. 1 for more details].

Table 1. Summary of the backward regression models from Experiment I. β refers to the standardized regression coefficient. Bold indicates significance at α<0.05.

Outcome variable	Predictors	*β*	*t*	*p*	Status
Place	•Vocabulary • Integration	-0.085 0.060	-0.659 0.462	.512 .646	excluded excluded
Voicing	•Vocabulary • Integration	-0.009 0.044	-0.072 0.341	.943 .734	excluded excluded
Place	•Vocabulary • 4 Hz Integration **• 32 Hz Integration**	-0.092 0.205 **-0.274**	-0.749 1.610 **-2.152**	.456 .112 **.035**	excluded excluded ***included***
Voicing	•Vocabulary •4 Hz Integration •32 Hz Integration	-0.203 0.048 -0.182	-1.657 0.367 -1.426	0.102 0.708 0.159	excluded excluded excluded

3. EXPERIMENT II

3.1 Methods

3.1.1 Participants

French-speaking children from 6 to 9 years were recruited in this second experiment, see [12]. Consent forms were obtained from parents as approved by the university ethics committee (Numéro CER – Paris Descartes : 2018-41). All children included had normal hearing (absolute thresholds between 0.25 and 8 kHz < 20 dB HL) and typical cognitive development. Twenty-eight 6-to-7-year-olds (18 females; mean age = 6.6 years, *SD* = 0.3), 29 7-to-8-year-olds (17 females; mean age = 7.6 years, *SD* = 0.3), and 29 8-to-9-year-olds (17 females; mean age = 8.5 years, *SD* = 0.4) were included in the final sample.

Stimuli

Narrow-band noise carriers centered at 500 Hz were generated with a 4-Hz bandwidth. A total of 500 different noise carriers were generated by adding together five equal-amplitude sine tones with frequencies 498, 499, 500, 501 and 502 Hz. All sounds were 500-ms long, including a 14-ms raised-cosine onset/offset ramps and with a fixed starting phase. Sounds with "low" values of envelope *SD* were selected (ranging between 0.059 and 0.069 arbitrary units). AM spectra of these carriers show greater modulation energy below about 4 Hz (see [10]) and AM masking effects were shown previously in children using such carriers for AM detection of an 8-Hz modulation target.

Procedure

AM detection thresholds were first measured using a 2I-2AFC implemented on a touch-screen tablet. A transformed and weighted 1-up-1-down adaptive procedure was implemented [18] targeting 76 % response correct corresponding to a d' = 1. Then, a constant-stimuli procedure was implemented using a 2I-2AFC paradigm where the stimuli were repeated in two passes. Each noise carrier was only presented once in each pass. Between the two passes, the trials were exactly the same, that is, the same (physical) modulated and unmodulated carriers were presented in the same order within a trial. The order of trial presentation was the same in each pass. Each participant completed two AM-detection passes of 200 trials each, as divided into 10 blocks of 20 trials played at individual threshold. Once the 10 blocks of pass 1 were completed, the exact same blocks were repeated in pass 2.

The outcome measures in this task were: Percent Correct (PC) of AM detection in each pass (calculated on the 200 trials) and across the two passes (average PC), which signals accuracy in the perceptual decision, and Percent of Agreement (PA) between the response given for the same stimuli in each pass. PA signals response consistency in the double-pass, a proxy of internal noise for AM detection, and was calculated on a trial-to-trial basis between pass 1 and 2.

3.1.2 Speech-in-noise tasks

Stimuli

Three native French speakers were recorded: one male (F0 = 110 Hz) and two females (F0 = 195 and 205 Hz). /aCa/ syllables were selected such as C = /f/, /v/, /ʃ/, /ʒ/, /s/, /z/ (fricatives), /b/, /p/, /d/, /t/, /g/, /k/ (stops). A steady speech-shaped noise masker was generated based on the female tokens. Six phonetic conditions were presented with a minimal change in: 1) voicing, for either fricative or stop consonants (conditions 'Voicing fricatives' and 'Voicing stop'; 2) place of articulation for either fricative or stop consonants (conditions 'Place fricatives' and 'Place stop'); 5) manner of articulation for either voiceless or voiced consonants (conditions 'Manner voiced' or 'Manner voiceless').

Procedure

The exact same XAB procedure as in Exp. I was used to present the French consonants in the six phonetic conditions. An identification threshold in noise was obtained in each condition, calculated as the geometric mean of the last four reversals.

3.1.3 Receptive Vocabulary Assessment

Receptive vocabulary was measured using the standardized French version of Picture Vocabulary Scale tests (Échelle de Vocabulaire en Images Peabody).

3.2 Results

3.2.1 Precision and consistency in AM detection

Percentage Correct (PC) and Percentage of Agreement (PA) for AM detection were analyzed in linear mixed-effects models including the fixed effects of Age, Pass (1, 2) and Block (from 1 to 10) and participants as the random factor. This analyses showed a significant Age effect on both PC and PA [for PC and PA respectively: $F(1,86) = 4.466$; $p = 0.037$; η2 = 0.049; $F(1,86) = 4.780$; $p = 0.031$; $\eta^2 = 0.052$]

3.2.2 Speech-in-noise identification thresholds

A linear mixed-effect model was run to assess the fixed effects of Age and Phonetic Condition ('Voicing fricative', 'Voicing stop', 'Place fricative', 'Place stop', 'Place fricative', 'Manner Voiced' and 'Manner Voiceless') on identification thresholds. The fixed effect of Phonetic Condition proved significant [$F(5, 415.79) = 63.950$; $p < .001$; $\eta^2 = 0.435$]. Post-hoc Tukey comparisons highlighted a pattern of differences in performance in the various conditions signaling growing difficulty going from the condition Place fricative, yielding the lowest (best) thresholds to the condition Place stop, yielding the highest (worse) thresholds. Precisely, thresholds in the condition Place Fricative were significantly lower (*i.e*., better) than in the conditions Manner (Voiced and Voiceless) and Voicing (Fricative and Stop), yielding similar thresholds. Thresholds obtained in condition Place Stop, in turn, were significantly higher (*i.e.*, worse) than all other conditions. Note that an analogous pattern was identified in a group of 15 French adults performing the same experiment. Fig. 1 reports identification thresholds in the six conditions. The main effect of Age was marginally significant [$F(1, 82.15) = 3.911$; p = .051; $\eta^2 = 0.045$] and no significant interaction was detected [$F(5, 415.79) = 0.517$; $p = 0.763$; $\eta^2 = 0.006$].

3.2.3 Relationship between precision and consistency in AM detection and speech-in-noise

We assessed whether precision and consistency in AM detection (PC and PA, respectively) and receptive vocabulary scores were predictive of speech-in-noise identification thresholds. We tested this hypothesis using backward regression analyses in independent models for PC and PA as the two variables were highly correlated (Pearson's $r = .851$; $p < 0.001$).

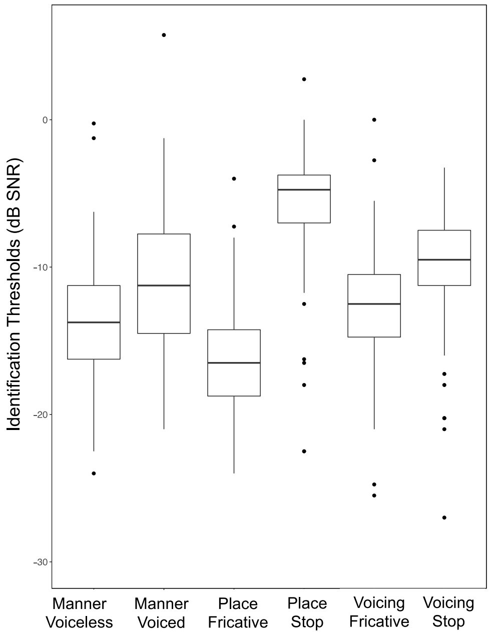

Figure 1. Average Speech-in-Noise identification thresholds (in dB SNR) in the six conditions of Experiment II.

As thresholds were comparable in the conditions Manner Voiced and Manner Voiceless on one hand, and in the conditions Voicing Stop and Voicing Fricative on the other hand, thresholds were averaged for "Manner" and for "Voicing" (note, however, that, as a marginal effect was observed in the model targeting PA and Vocabulary as predictors of the Voicing condition, for this model only, separate analyses were then run for Voicing Fricative and Voicing Stop).

A first series of models assessed the effects of PC and vocabulary scores. Only one model significantly explained the variance in the Place Fricative condition [$R^2 = 9.3\%$, adj $R^2 = 7\%$, $F(2,77) = 3.966$, $p = .23$] where Vocabulary scores significantly predicted the thresholds while PC scores were marginally predictive (see **Tab. 2**). A second series of models then assessed PA and vocabulary scores as predictors and only revealed a marginal effect of PA in the Voicing condition. Subsequent analyses were run on the conditions Voicing Stop and Voicing Fricative separately. Thresholds obtained in the Voicing Stop condition were best predicted by a model including PA scores [$R^2 = 13\%$, adj. $R^2 = 11.9\%$, $F(1,76) = 11.534$, $p = 0.001$, see Tab. 2].

Fig. 2 shows that better PA scores were related to lower (better) identification thresholds in the Voicing Stop condition.

Table 2. Summary of the backward regression models from Experiment II. β refers to the standardized regression coefficient. Bold indicates significant at α<0.05.

Outcome variable	Predictors	*β*	*t*	*p*	Status
Place Stop	•Vocabulary •PC	-0.177 0.049	-0.156 0.428	.876 .669	excluded excluded
Place Fricative	**•Vocabulary** •PC	**-0.244** -0.181	**-2.244** -1.670	**.028** .099	**included** excluded
Manner	•Vocabulary •PC	-0.205 0.092	-1.864 0.840	.066 .403	excluded excluded
Voicing	•Vocabulary •PC	-0.101 -0.125	-0.881 -1.091	.381 .279	excluded excluded
Place Stop	•Vocabulary •PA	-0.047 0.089	-0.412 0.768	.682 .434	excluded excluded
Place Fricative	Vocabulary •PA	-0.188 -0.092	-1.682 -0.828	.096 .410	excluded excluded
Manner	•Vocabulary •PA	-0.211 -0.030	-1.905 -0.273	.060 .785	excluded excluded
Voicing	•Vocabulary •PA	-0.047 -0.213	-0.423 -1.903	.674 .061	excluded excluded
Voicing Stop	•Vocabulary **•PA**	-0.358 **-0.028**	-3.326 **-0.264**	.793 **.001**	excluded ***included***
Voicing Fric	•Vocabulary •PA	-0.074 -0.051	-0.662 -0.453	.510 .652	excluded excluded

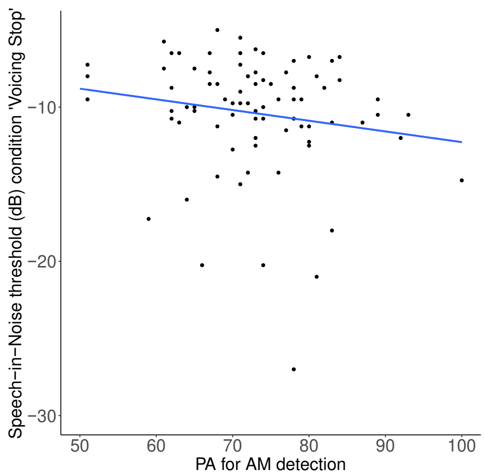

Figure 2. Identification thresholds in the Voicing Stop condition (in dB SNR) plotted as a function of Percentage of Agreement for AM detection obtained in the double-pass procedure, Experiment II.

4. DISCUSSION

With these experiments, we sought to study the relationship between AM processing skills and speech-in-noise perception during childhood. Precisely, we respectively targeted the links between speech-in-noise identification thresholds and, in a first study, temporal integration (the improvement of AM detection thresholds with increasing number of AM cycles) while, in a second study, precision and consistency in AM detection (*i.e.*, percentage correct and percentage of agreement in a double-pass procedure).

Our first experiment did not show strong links between general temporal integration capacities for AM detection and fricative consonants in-noise identification thresholds during childhood. In this group of English-speaking children aged from 5 to 11 years of age, only the integration scores in the 32-Hz condition representing how much children' AM detection thresholds improve from 2 vs 8 cycles, were predictive of the identification thresholds in noise obtained in a condition contrasting fricative consonants on the place contrast (adj $R^2 = 5\%$). Temporal integration for AM has been interpreted as reflecting the properties of the correlation operation of the decisional mechanism based on template matching [4, 11, 15]. Further work is warranted to explain why a relationship was specifically observed for relatively fast (32 Hz) AM cues and place of articulation identification of fricative consonants.

In our second experiment, we then found that identification thresholds in the same phonetic condition (Place Fricative) were best predicted by a combination of children's vocabulary scores and their precision in an 8Hz-AM detection, PC scores (adj $R^2 = 7\%$). Interestingly, this condition contrasting fricative consonants on place of articulation was the easiest condition for both English (Exp. I) and French (Exp. II) children. Our results showed that the identification thresholds in this specific condition are somewhat related to AM detection abilities. Fricative consonants and place of articulation contrasts are known to be the more difficult to process by children with sensorineural hearing loss using cochlear implants. It therefore seems very important to measure the identification thresholds of children using cochlear implants for this phonetic contrast and to assess whether their thresholds can be predicted by their processing of AM information, that is conveyed by their implants.

Finally, we also observed a strong and very specific link between PA (measuring consistency in AM detection at 8 Hz) and consonant identification thresholds in noise for minimal phonetic contrasts of voicing on stop consonants. Voicing cues for stops are conveyed by rather slow AM variations < 16 Hz and in light of our previous modeling studies, PA can be considered as a proxy of internal noise for AM processing [14]. Thus, the present results suggest that, to some extent, internal noise (and, thus, processing efficiency for AM) plays a role in the development of speech-in-noise identification skills, supporting at least the ability to track minimal differences in voicing while listening in noisy conditions. This result can be interpreted as the sign of a developmental relationship between the progressive reduction of neural variability along the auditory pathway – such as the stochastic nature of neuronal firing, the internal state of the listeners' organism, or random fluctuations in attention – impacting perceptual decisions and more efficient perception of the linguistic signal. Further work combining electrophysiological measures (to measure efficiency in neural temporal processing during development) and psychophysical tasks is required to test this hypothesis.

The present experiments revealed some links between AM processing and consonant-identification in noise for children with normal hearing. Vocabulary levels, overall, were not strongly predictive of identification thresholds in noise.

5. ACKNOWLEDGMENTS

This work was supported by the ANR-17-CE28-0008 (Agence National de la Recherche) to Laurianne Cabrera and by the MSCA-INCODE (Grant Agreement n. 895484) to Irene Lorenzini. Leo Varnet and Christian Lorenzi were supported by the FrontCog ANR-17-EURE-0017 grant. The authors wish to thank Emmanuel Ponsot for valuable insights on the analysis and modelling of the data; all the children that took part into the project and all the teachers that helped realizing the testing sessions; Paula Perrineau and Alicia Lefebvre for help with testing; Sarah Attia and Andrew King for help with setting up the experimental design.

6. REFERENCES

[1] B. C. Moore, *An introduction to the psychology of hearing*, vol. 4. Academic press San Diego, 2004. Accessed: Jul. 17, 2013. [Online]. Available: http://www.emeraldinsight.com/products/books/info/download/jan_mar_12review.pdf

[2] S. Rosen, “Temporal information in speech: acoustic, auditory and linguistic aspects,” *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 336, no. 1278, pp. 367–373, Jun. 1992, doi: 10.1098/rstb.1992.0070.

[3] T. Dau, B. Kollmeier, and A. Kohlrausch, “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,” *J. Acoust. Soc. Am.*, vol. 102, no. 5 Pt 1, pp. 2892–2905, Nov. 1997.

[4] T. Dau, B. Kollmeier, and A. Kohlrausch, “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,” *J. Acoust. Soc. Am.*, vol. 102, no. 5 Pt 1, pp. 2906–2919, Nov. 1997.

[5] J. M. Eaves, A. Q. Summerfield, and P. T. Kitterick, “Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes,” *J. Acoust. Soc. Am.*, vol. 130, no. 1, pp. 501–507, Jul. 2011, doi: 10.1121/1.3592237.

[6] M. A. Stone, C. Füllgrabe, and B. C. Moore, “Notionally steady background noise acts primarily as a modulation masker of speech,” *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 317–326, 2012.

[7] J. H. Won, W. R. Drennan, K. Nie, E. M. Jameyson, and J. T. Rubinstein, “Acoustic temporal modulation detection and speech perception in cochlear implant listeners,” *J. Acoust. Soc. Am.*, vol. 130, no. 1, pp. 376–388, Jul. 2011, doi: 10.1121/1.3592521.

[8] J. W. Hall 3rd and J. H. Grose, “Development of temporal resolution in children as measured by the temporal modulation transfer function,” *J. Acoust. Soc. Am.*, vol. 96, no. 1, pp. 150–154, Jul. 1994.

[9] A. Stuart, “Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise,” *Ear Hear.*, vol. 26, no. 1, pp. 78–88, 2005.

[10] L. Cabrera, L. Varnet, E. Buss, S. Rosen, and C. Lorenzi, “Development of temporal auditory processing in childhood: Changes in efficiency rather than temporal-modulation selectivity,” *J. Acoust. Soc. Am.*, vol. 146, no. 4, pp. 2415–2429, 2019.

[11] L. Cabrera, I. Lorenzini, S. Rosen, L. Varnet, and C. Lorenzi, “Temporal integration for amplitude modulation in childhood: Interaction between internal noise and memory,” *Hear. Res.*, vol. 415, p. 108403, 2022.

[12] I. Lorenzini, L. Varnet, L. Christian, and L. Cabrera, “Reduced processing efficiency impacts auditory detection of amplitude modulation in children: evidence from an experimental and modelling study.” PsyArXiv, Dec. 01, 2022. doi: 10.31234/osf.io/jwr4e.

[13] N. Wallaert, B. C. Moore, S. D. Ewert, and C. Lorenzi, “Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation,” *J. Acoust. Soc. Am.*, vol. 141, no. 2, pp. 971–980, 2017.

[14] S. Attia, A. King, L. Varnet, E. Ponsot, and C. Lorenzi, “Double-pass consistency for amplitude-and frequency-modulation detection in normal-hearing listeners,” *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3631–3647, 2021.

[15] D. Wechsler, “Wechsler preschool and primary scale of intelligence—fourth edition,” *Psychol. Corp. San Antonio TX*, 2012.

[16] H. Levitt, “Transformed up-down methods in psychoacoustics,” *J. Acoust. Soc. Am.*, vol. 49, no. 2B, pp. 467–477, 1971.

[17] L. M. Dunn and D. M. Dunn, *The British picture vocabulary scale*. GL Assessment Limited, 2009.

[18] M. A. Garcıa-Pérez, “Forced-choice staircases with fixed step sizes: asymptotic and small-sample properties,” *Vision Res.*, vol. 38, no. 12, pp. 1861–1881, 1998.