Prenatal auditory experience and its sequelae

Marin Vogelsang[+,*,1,2], **Lukas Vogelsang**[+,1,2], **Sidney Diamond**[2], **Pawan Sinha**[2]

[1]Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract

Towards the end of the second trimester of gestation, a human fetus is able to register environmental sounds. This in-utero auditory experience is characterized by comprising strongly low-pass filtered versions of external sounds. Here, we present computational tests of the hypothesis that this early exposure to severely degraded auditory inputs serves an adaptive purpose – it may induce the neural development of extended temporal integration. Such integration can facilitate the detection of information carried by low-frequency variations in the auditory signal, including emotional or other prosodic content. To test this prediction, we characterized the impact of several training regimens, biomimetic and otherwise, on a computational model system trained and tested on the task of emotion recognition. We find that training with an auditory trajectory recapitulating that of a neurotypical infant in the pre-to-post-natal period results in temporally-extended receptive field structures and yields the best subsequent accuracy and generalization performance on the task of emotion recognition. This strongly suggests that the progression from low-pass-filtered to full-frequency inputs is likely an adaptive feature of our development, conferring significant benefits to later auditory processing abilities relying on temporally-extended analyses. Additionally, this finding can help explain some of the auditory impairments associated with preterm births, suggests guidelines for the design of auditory environments in neonatal care units, and points to enhanced training procedures for computational models.

1. Introduction

Expectant mothers often report that their unborn child exhibits movements in response to loud environmental sounds (1). Further attesting to these anecdotal reports, systematic developmental psychoacoustic studies have demonstrated that by around 20 weeks of gestational age, the fetus is equipped with a functioning auditory system, capable of registering the mother's soundscape (2) (see Figure 1A), and may even benefit from such auditory exposure, especially to the mother's vocalizations (3). However, the quality of this auditory experience is quite limited. Factors such as the fluid medium and surrounding tissues in the intrauterine environment, the impedance of the fetal skull, and the immaturity of the cochlea lead to a strong reduction of the audibility of high frequencies present in environmental sounds while affecting lower frequencies only marginally or even enhancing

[*]Corresponding author.
[+]Equal contribution

them (2, 4, 5). In other words, the intrauterine environment essentially acts as a low-pass filter on incident sounds, significantly limiting the fetus' exposure and sensitivity to high-frequency auditory stimuli. This is further illustrated in Figure 1A.

The question we consider in this paper is whether these degradations of incident sounds are epiphenomenal limitations imposed by biological processes, or if they may, in fact, have any salutary implications later in life. In other words, might the degradation in prenatal experience not limit a child's later auditory skills but possibly enhance them? This question has the potential to contribute to a broader and mostly unresolved debate in the field of developmental science, concerned with whether developmental constraints are to be considered as hindering, or rather helping, the acquisition of later perceptual abilities. In the traditional view, early limitations are typically seen as the outcome of constraints imposed by the physiological immaturities of the underlying neural and perceptual systems, and human development has to somehow surmount the challenges imposed by these early limitations to attain its later manifesting proficiencies. This perspective, despite its perhaps intuitive appeal, has been challenged in specific domains of investigation. The proposal put forward by (6) was among the first to promote the perspective that early limitations, rather than solely representing hurdles, may, in fact, be adaptive for later developmental proficiencies. Specifically, (6) proposed that by initially reducing the complexity of environmental stimuli, a learner's perceptual analysis would be rendered less overwhelming. Along similar lines, but focused specifically on learning, and using computational simulations to support their claims, other researchers have suggested that, for instance, developmental limitations in cognitive architectures can benefit language learning (7, 8) and that early limitations in an infant's visual system can benefit the acquisition of binocular disparity detection capabilities (9). Here, we wish to extend this ongoing investigation to the domain of auditory processing as well as to prenatal, rather than exclusively postnatal, development.

In the domain of auditory processing, the following logic makes a case for a way in which early degradations may be adaptive: By definition, low-pass-filtered audio streams, roughly mirroring the fetal experience in the uterus, contain markedly reduced fine-grained, high-frequency information. As a consequence of this reduction of higher-frequency content that is available to the auditory system at any moment in time, brief audio snippets are rendered relatively uninformative for inference upon information and events in the environment. In order to, nevertheless, derive meaningful inferences, a strategy that the system may learn to employ is to gather information across longer time scales, which would consequently induce the development of neural mechanisms capable of integrating information over extended time spans. Such temporal integration is known to, later in life, be useful for the processing of information carried by slow variations in the auditory signal (i.e., those relying on temporally-extended analysis), including the expression of emotions or other prosodic content (10, 11). Contrasting this perspective, the early availability of high frequencies in the auditory environment may render even short audio snippets sufficiently informative for the perceptual system and may thus preclude the development of extended temporal integration and, hence, compromise specific auditory skills that depend on it. Taken together, this leads us to hypothesize that degraded, approximately low-pass filtered, inputs play an adaptive role in (i) instantiating extended temporal integration mechanisms and, (ii)

as a consequence thereof, enable robust auditory analysis for tasks relying on temporally-extended information, such as the recognition of emotions or other prosodic content.

To systematically probe these two predictions, and thereby evaluate our proposal's overall plausibility, we need to be able to actively manipulate experience and assess the consequences of these manipulations. While ethical and practical considerations render such deliberate manipulation infeasible in human participants, studies with computational model systems, in particular deep convolutional neural networks (DCNNs), which are capable of learning from experience, offer a powerful way forward. These networks, while not perfect models of the biological system, can serve as useful approximations of early sensory processing (12) and allow us to systematically assess the impact of the nature of the stimuli they have been presented. In other words, we can expose a deep convolutional network (see Figure 1B for a sketch of the network architecture used in this paper) to different types of inputs (some of which are designed to recapitulate biological development in terms of their temporal progression, while others are not) and examine the differences in outcomes.

The following illustrates how we can probe our two key predictions through an examination of these outcomes: In brief, a network such as the one we use takes short audio recordings of spoken utterances (represented as raw audio waveforms) as inputs and, in rough analogy to the biological system, processes these through a hierarchical cascade of temporal filters (using the convolutional operation). Due to this hierarchical organization, as one proceeds along the processing stages, more and more complex speech properties are extracted. These properties, in turn, are learned to be associated with classification labels (in our case, one of the seven emotions that a given utterance was spoken with). While these associations are initially random, as part of the training procedure, numerous audio clips are fed into the network along with the desired classification label (e.g., 'sad' or 'happy'), and the connections are iteratively refined so as to reduce classification error. Crucially, not only the links between higher-level features and final classification labels, but also the temporal filters in the early processing layers themselves (which are in rough analogy to temporal receptive fields found in the auditory cortex) are learned. The network starts out with randomly initialized receptive field structures across the different processing layers, but they, too, are subsequently sculpted to reduce classification error. Following the completion of training, the resulting receptive field structures can be examined, allowing us to probe the first key prediction we had stated earlier: that degraded, approximately low-pass filtered, inputs play an adaptive role in instantiating extended temporal integration mechanisms (i.e., receptive fields encoding lower frequencies and thereby encoding longer wavelengths). In addition, we can also probe our second key prediction, that as a consequence of instantiating extended temporal integration mechanisms, degraded, approximately low-pass filtered, inputs enable robust auditory analysis for tasks relying on temporally-extended information. To do so, we can directly examine the networks' test performances, and generalization behavior, on the emotion classification task – a task that bears great ecological significance for humans and also relies on temporally-extended auditory analysis.

Taken together, the main goal of our computational investigation is to expose our network to different temporal progressions of filtered speech signals (one of which is inspired by the developmental progression from low-pass-filtered to full-frequency inputs, and

three other progressions serving as non-developmental controls) while evaluating the two above-mentioned aspects of the resulting networks: (i) their temporal integration profiles (as is visible in the resulting receptive field structures), and (ii) their performance and generalization abilities on a task relying on temporally-extended analysis (specifically, emotion classification).

2. Methods

As introduced earlier, we utilized a deep convolutional neural network as our computational model system. More specifically, we used the model "M5" by (13), which is equipped with 4 subsequent convolutional layers (each involving convolution, batch normalization, application of the ReLU activation function, and max pooling), followed by global average pooling, and connected to the output nodes of the network, representing the 7 different emotion classes, through a single dense layer with softmax activation function.

As dataset for training and testing our network, we used the Toronto Emotional Speech Set (TESS). This sets contains 200 short spoken utterances (lasting 1–3 seconds) for each of 7 different emotions (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral) and 2 speakers (14). The dataset was divided into 10 equal-sized folds, yielding 10 different data splits with train/test set ratios of 90/10 each. The networks were trained and tested on all 10 folds, to investigate the stability of the findings.

To test our two key predictions, we trained different instances of the neural network on an emotion classification task, based on the Toronto Emotional Speech Set (14), using four qualitatively different regimens (of which one was developmentally-inspired and three were not), as illustrated in Figure 1C:

'Low-to-full', in which, inspired by the prenatal-to-postnatal progression in development, training commenced with low-pass-filtered and resumed with full-frequency inputs,

'Full-to-low', in which, as a control condition, training followed an inverse-developmental progression but had the same aggregate content as the previous ('low-to-full') regimen,

'Exclusively-full', in which, more simply, the entire training consisted of full-frequency input, as an additional control, and

'Exclusively-low', in which, similarly simply, the entire training was based on low-pass filtered input, as a final control.

Each training regimen thereby comprised a total of 100 training epochs (representing the number of times that the entire input data was fed into the network during training), with the two different stages in the 'low-to-full' and 'full-to-low' regimens consisting of 50 epochs each. Low-pass filtering was carried out at a cut-off frequency of 500 Hz, as inspired by previous recordings in the womb reported in (2) and illustrated in Figure 1A. The specifics of training, testing, and additional analyses are detailed in the Supplemental Methods section.

3. Results

To assess the impact of training regimen on our networks' early, learned representations, we examined the spectral profiles of their temporal receptive fields in the first convolutional layer. This analysis revealed that while the network trained on full-frequency inputs learned receptive fields with a broad range of peak frequencies, the network trained on low-pass filtered data acquired exclusively low-frequency receptive field structures (see Figures 2A&B). Further illustrating this point, Figure 2C shows spectral profiles of individual filters (those with sorting indices corresponding to the x-axis ticks in Figures 2A&B) for training following exclusively-low and exclusively-full regimens.

Further, training on low-frequency inputs resulted not only in more, but also purer, low-frequency filters. For filters with peak frequencies up to 0.5 kHz (126 filters following exclusively-low training, and 83 filters following exclusively-full training), there was a highly significant difference across the two training conditions in the contribution of frequencies up to 0.5 kHz to the filter responses ($t(198) = 13.91$, $p<0.001$ for two-tailed t-test; see Figure 3A). This correspondence, as evident in Figures 2 and 3A, attests to the idea that initial exposure to exclusively low-frequency content, as is characteristic of prenatal hearing, may enforce the development of extended temporal integration mechanisms.

Moving beyond homogenous training regimens, we next studied the impact of biomimetic and reverse-biomimetic protocols on the network's learned representations in the first convolutional layer. Specifically, we examined changes in receptive field properties as the network transitioned from one phase of training (low or full frequency inputs) to the other (full or low-frequency inputs). The results revealed that while only 13% of the receptive fields established during the first half of training on low-frequency inputs changed their peak frequencies upon transitioning to full-frequency inputs, this was the case for 45% of the receptive field structures when full-frequency training was later followed by low-frequency training (see Figure 3B). That these filters were almost exclusively enlarged temporally, further resulted in the full-to-low model approaching the spectral distributions of the exclusively-low and low-to-full models (see Figure 3C). These results suggest that temporally extended receptive fields, acquired through initial low-pass filtered training, may comprise more robustly useful processing units for the task, rendering the need for their later adjustment superfluous.

Complementing these results, we next examined the consequences of our four training regimens on the networks' later generalization performance by testing emotion classification on full-frequency and various low-pass filtered test data. The results depicted in Figure 4A reveal that while the model trained on full-frequency inputs exhibited poor generalization (red curve), the low-to-full model (green curve) yielded high generalization, performing better than all other regimens across nearly the entire range of test frequencies. This is particularly noteworthy given that the full-frequency inputs subsume the entire low-passed content, and considering that the exclusively-full network could have learned to discard high frequencies were they not required to achieve high performance levels on the task. Generalization curves from the remaining two conditions (black curve for the exclusively-low model; blue curve for the full-to-low model) are consistently lower than that of the

biomimetic one. Nevertheless, they indicate that including low-pass filtered inputs into some phase of the training enhances performance over the exclusively full-frequency training regimen.

To probe the relationship between performance scores and the receptive field characterizations described earlier, we assessed classification performance on full-frequency inputs while gradually ablating the individual networks' filters with the highest peak frequencies. While the network trained on exclusively full-frequency inputs exhibited the largest performance decrement, the low-to-full model was least affected by the removal of higher frequency units, which is noteworthy given its spectral similarity to the networks trained using the full-to-low and exclusively-low regimens (see Figure 4B). Moving beyond the first network layer, we also examined the response stability of the deeper networks' units when presented with full-frequency vs. low-frequency (500 Hz) test set inputs, operationalized as correlation score (r). As depicted in Figure 4C, across all four convolutional layers, activations in the network trained on full-frequency inputs were most varied, while the low-to-full network showed the least variation, further attesting to its invariance to high-frequency modulations.

4.   Discussion

Taken together, the computational results presented in this paper lend support to the proposal that commencing auditory development with degraded (i.e., exclusively low-frequency) stimuli may help set up temporally extended processing mechanisms that remain in place throughout subsequent experience with full-frequency inputs. The early instantiation of such mechanisms may, in turn, facilitate robust analyses of long-range modulations in the auditory stream, as indicated by generalized performances of the developmentally-inspired model when tested on the task of emotion recognition in full- and low-frequency conditions.

Beyond the superior generalization on the temporally-extended task of emotion recognition, the observed robustness of classification performance to low-pass filtering not only mimics an important feature of auditory recognition in humans (15) but might also facilitate auditory analyses in more challenging environments. For instance, as sound travels through air, its energy is more substantially reduced in the higher, compared to the lower, frequencies (16). One of the benefits that facility with using low-frequencies could thus confer to the mature auditory system is robust classification even over large distances. Along similar lines, such proficiency may also be responsible for the ability of patients suffering from age-related hearing loss, which first affects the sensitivity of higher-frequency bands, to continue to be able to identify speech sounds for part of the progression of their hearing loss (17).

Our results suggest that precluding early experience with low-frequency inputs leads to a reduced emphasis on receptive fields tuned to low frequencies, with corresponding decrements in prosodic classification performance, while not necessarily affecting the processing of informational content in speech that is based on higher frequencies. Interestingly, this computationally-derived result is corroborated by clinical data. Prematurely born infants, who did not get to experience exclusively low-frequency sounds for as long as normal-term infants but were almost immediately immersed into a full-

frequency environment (roughly akin to the exclusively-full regimen in the computational simulations presented here), have been reported to later exhibit impairments in the processing of low-frequency structure of sounds and show lower performance on prosody and emotion classification, even though their punctate acoustic detection thresholds are near normal (18–20). This result has implications for the design of auditory environments for pre-term babies in classical neonatal ICUs, which have been demonstrated to expose infants to frequencies above 500 Hz for the majority of the time (21). Auditory interventions in neonatal ICUs are, if applied, usually oriented towards silence (22, 23) or musical sounds (24–26) and have been reported to yield some benefits in terms of higher-level functioning. The findings presented in this paper, however, suggest that for pre-term babies, the auditory environment in a neonatal ICU, instead of, or in addition to, being controlled for the general type of audio signal, should be made more similar to the intrauterine one in terms of the specific spectral composition of sounds, thereby mirroring, and adding support to, previous perspectives on the impact of high-frequency noise in neonatal ICUs (21, 27). Specifically, based on the data we obtained in this paper, we argue that the environmental sounds allowed to reach the infant should be filtered with an approximation of the intra-uterine acoustic filtering characteristics.

It remains to be seen how the converging computational and empirical evidence on the benefits of initial exposure to exclusively low-frequency sounds early in development, as derived as part of the computational results presented in this paper in the context of emotion recognition, generalizes to other ecologically-relevant aspects and tasks of hearing such as the comprehension of other prosodic content. This could, in the future, be probed in pre-term vs. full-term babies and possibly be further examined as a function of the auditory intervention applied to the neonatal ICU environment.

Finally, it is worth noting that the data presented here are consistent with results previously reported in the visual domain (28). There it was found that initially low spatial acuity, that progressively improved over the first years of life, helped instantiate receptive fields capable of extended spatial integration, and more robust classification performance. The analogous results we have found in the auditory domain suggest that the adaptive benefits of initial sensory limitations may apply across different modalities. Thus, in addition to highlighting the potential benefits of sound degradation in the intrauterine environment, the results presented in this paper may also help explain the adaptive significance of some key human developmental progressions, prenatal or postnatal, and across sensory domains. The initially degraded inputs may provide a scaffold rather than act as hurdles. Taken together, this may not only help elucidate some of the mechanisms underlying our later auditory proficiencies, but also help us better understand why the choreography of developmental stages is structured in the way that it is. From the perspective of artificial intelligence, the salutary effects of these developmental trajectories upon later classification performance of biological systems, suggest that computational systems for analogous tasks may also benefit from incorporating biomimetic training regimens.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Data Availability Statement

The code supporting the findings of this study is openly available at https://github.com/marin-oz/HIA_Audition.

References

1. Murkoff H (2016). What to expect when you are expecting. Workman Publishing.

2. Hepper PG, Shahidullah BS (1994). Development of fetal hearing. Archives of Disease in Childhood, 71: F81–F87. [PubMed: 7979483]

3. Webb AR, Heller HT, Benson CB, & Lahav A (2015). Mother's voice and heartbeat sounds elicit auditory plasticity in the human brain before full gestation. Proceedings of the National Academy of Sciences, 112(10), 3152–3157.

4. Griffiths SK, Brown WS, Gerhardt KJ, Abrams RM, Morris RJ (1994). The perception of speech sounds recorded within the uterus of a pregnant sheep. J Acoust Soc of Am, 96(4), 2055–2063. [PubMed: 7963021]

5. Gerhardt KJ, & Abrams RM (1996, February). Fetal hearing: characterization of the stimulus and response. In Seminars in perinatology (Vol. 20, No. 1, pp. 11–20). WB Saunders.

6. Turkewitz G, & Kenny PA (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 15(4), 357–368.

7. Elman JL (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48(1), 71–99. [PubMed: 8403835]

8. Newport EL (1988). Constraints on learning and their role in language acquisition: Studies of the acquisition of American Sign Language. language Sciences, 10(1), 147–172.

9. Dominguez M, & Jacobs RA (2003). Developmental constraints aid the acquisition of binocular disparity sensitivities. Neural Computation, 15(1), 161–182. [PubMed: 12590824]

10. Ross M, Duffy RJ, Cooker HS, & Sargeant RL (1973). Contribution of the lower audible frequencies to the recognition of emotions. American Annals of the Deaf, 37–42. [PubMed: 4686794]

11. Snel J, & Cullen C (2013, September). Judging emotion from low-pass filtered naturalistic emotional speech. In 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 336–342). IEEE.

12. Norman-Haignere SV, McDermott JH (2018). Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. PLOS Biology, 16, 1–46.

13. Dai W, Dai C, Qu S, Li J, & Das S (2017, March). Very deep convolutional neural networks for raw waveforms. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 421–425). IEEE.

14. Dupuis K, & Pichora-Fuller MK (2010). Toronto emotional speech set (TESS). Link: https://tspace.library.utoronto.ca/handle/1807/24487

15. Bornstein SP, Wilson RH, Cambron NK (1994). Low- and high-pass filtered Northwestern University Auditory Test No. 6 for monaural and binaural Evaluation. Journal of American Academy of Audiology, 5, 259–264.

16. Little AD, Mershon DH, & Cox PH (1992). Spectral content as a cue to perceived auditory distance. Perception, 21(3), 405–416. [PubMed: 1437460]

17. Gates GA, & Mills JH (2005). Presbycusis. The lancet, 366(9491), 1111–1120.

18. Gonzalez-Gomez N, Nazzi T (2012). Phonotactic acquisition in healthy preterm infants. Developmental Science 15(6), 885–894. [PubMed: 23106742]

19. Ragó A, Honbolygó F, Róna Z, Beke A, Csépe V (2014). Effect of maturation on suprasegmental speech processing in full- and preterm infants: a mismatch negativity study. Res Dev Disabil. 35(1):192–202. [PubMed: 24171828]

20. Amin SB, Orlando M, Monczynski C, Tillery K (2015). Central auditory processing disorder profile in premature and term infants. Am J Perinatol. 32(4):399–404. [PubMed: 25261703]

21. Lahav A (2015). Questionable sound exposure outside of the womb: frequency analysis of environmental noise in the neonatal intensive care unit. Acta paediatrica, 104(1), e14–e19. [PubMed: 25264140]

22. Altuncu E, Akman I, Kulekci S, Akdas F, Bilgen HÜLYA, & Ozek E (2009). Noise levels in neonatal intensive care unit and use of sound absorbing panel in the isolette. International journal of pediatric otorhinolaryngology, 73(7), 951–953. [PubMed: 19406484]

23. Milette I (2010). Decreasing noise level in our NICU: the impact of a noise awareness educational program. Advances in Neonatal Care, 10(6), 343–351. [PubMed: 21102180]

24. Loewy J, Stewart K, Dassler AM, Telsey A, & Homel P (2013). The effects of music therapy on vital signs, feeding, and sleep in premature infants. Pediatrics, 131(5), 902–918. [PubMed: 23589814]

25. de Almeida JS, Lordier L, Zollinger B, Kunz N, Bastiani M, Gui L, … & Hüppi PS (2020). Music enhances structural maturation of emotional processing neural pathways in very preterm infants. NeuroImage, 207, 116391. [PubMed: 31765804]

26. Lordier L, Meskaldji DE, Grouiller F, Pittet MP, Vollenweider A, Vasung L, … & Hüppi PS (2019). Music in premature infants enhances high-level cognitive brain networks. Proceedings of the National Academy of Sciences, 116(24), 12103–12108.

27. Lahav A, & Skoe E (2014). An acoustic gap between the NICU and womb: a potential risk for compromised neuroplasticity of the auditory system in preterm infants. Frontiers in Neuroscience, 8, 381. [PubMed: 25538543]

28. Vogelsang L, Gilad-Gutnick S, Ehrenberg E, Yonas A, Diamond S, Held R, Sinha P (2018). Potential downside of high initial visual acuity. Proc Natl Acad Sci U S A, 115(44):11333–11338. [PubMed: 30322940]

Research Highlights

- A human fetus' auditory experience comprises strongly low-pass-filtered versions of sounds in the external world. We examine the potential consequences of these degradations of incident sounds.

- Results of our computational simulations strongly suggest that, rather than being epiphenomenal limitations, these degradations are likely an adaptive feature of our development.

- These findings have implications for auditory impairments associated with preterm births, the design of auditory environments in neonatal care units, and enhanced computational training procedures.
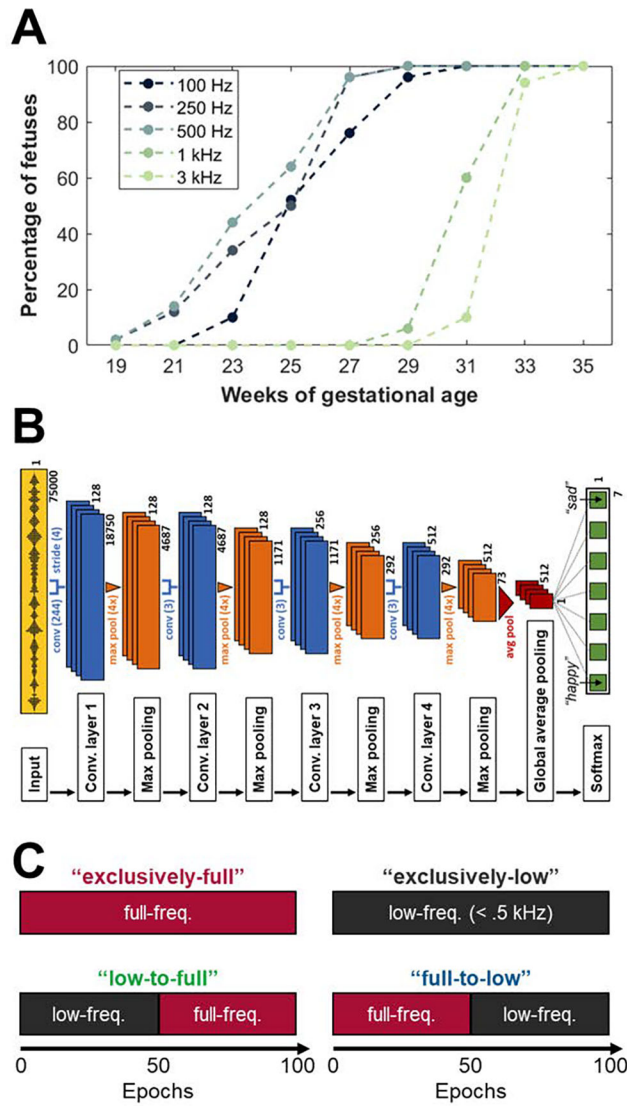
Figure 1.
A. Prenatal auditory sensitivity: Percentage of fetuses that responded to each of the given frequencies at different gestational ages, adapted from (2). **B.** Architecture of the network 'M5' by (13), equipped with four convolutional layers (each involving convolution, batch normalization, application of the ReLU activation function, and max pooling) which are followed by global average pooling and a softmax layer containing seven output nodes of the network, representing the seven different emotion classes to be predicted (anger, disgust, fear, happiness, pleasant surprise, sadness, and neutral). **C.** Schematics of the four training regimens used, each comprising 100 epochs of training.
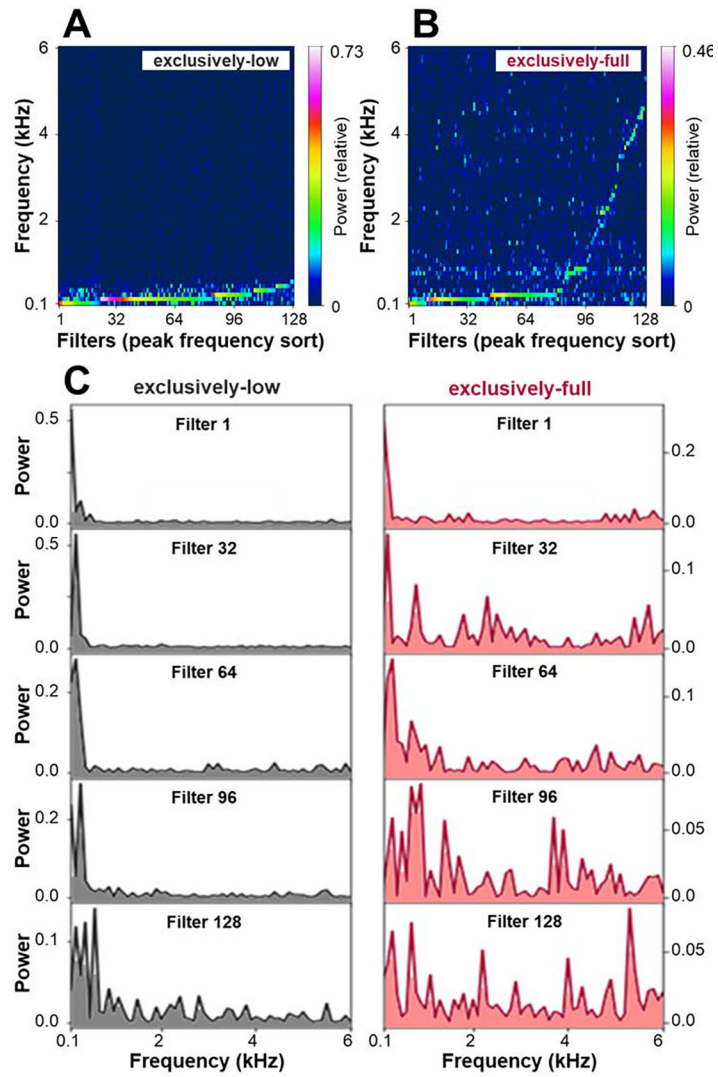
Figure 2.
Receptive field analysis. **A&B.** Spectral distribution of first-layer filters in the network trained on full-frequency (A) and low-frequency input (B). Colors code for normalized power obtained through frequency decomposition, with the sum of values up to 6 kHz normalized to 1. **C.** Close-up of spectral profiles of individual filters (filters with the 1st, 32nd, 64th, 96th, and 128th lowest peak frequency), when training followed exclusively-full and exclusively-low frequency regimens, respectively.
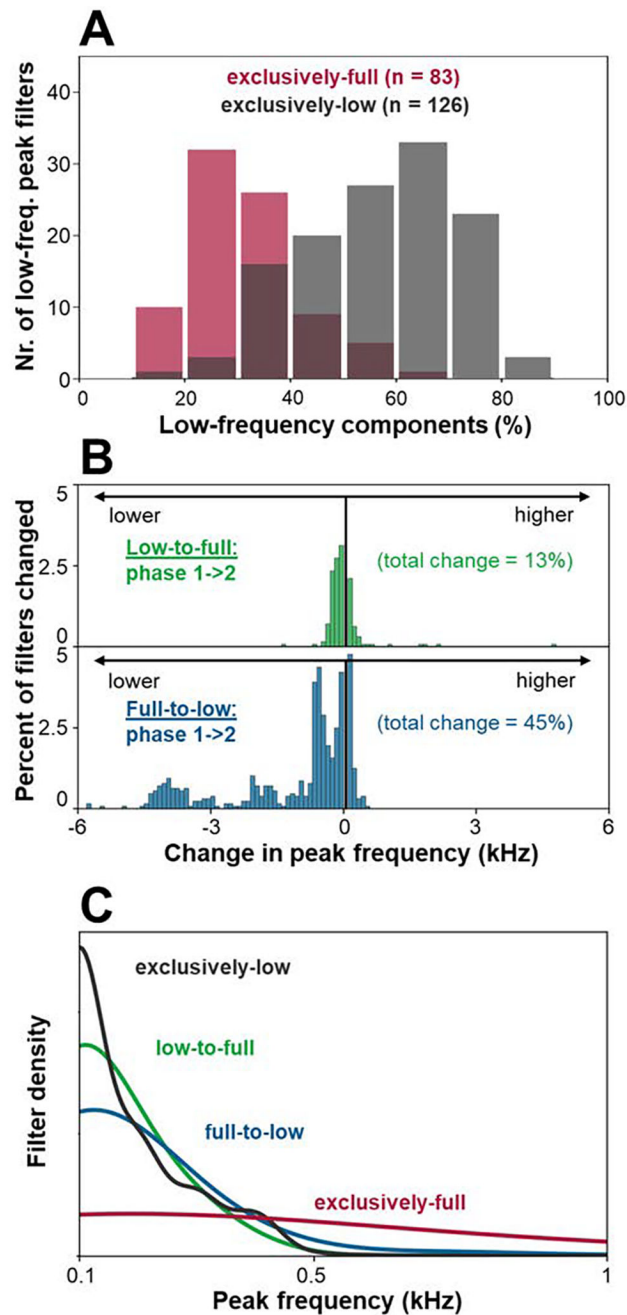
Figure 3.
Additional receptive field analysis. **A.** Histogram of the proportion that frequencies up to 0.5 kHz contribute, relative to all frequencies up to 6 kHz, to the response of individual filters with low peak frequencies (up to 0.5 kHz), following exclusively-low and exclusively-full frequency training. **B.** Histogram of units whose peak frequency changed from the first to the second half of training, separately for low-to-full and full-to-low training (pooled across 10-fold variations; see Methods section). **C.** Kernel density estimation plot of the distribution of filters' peak frequencies for all networks (pooled across 10-fold variations).
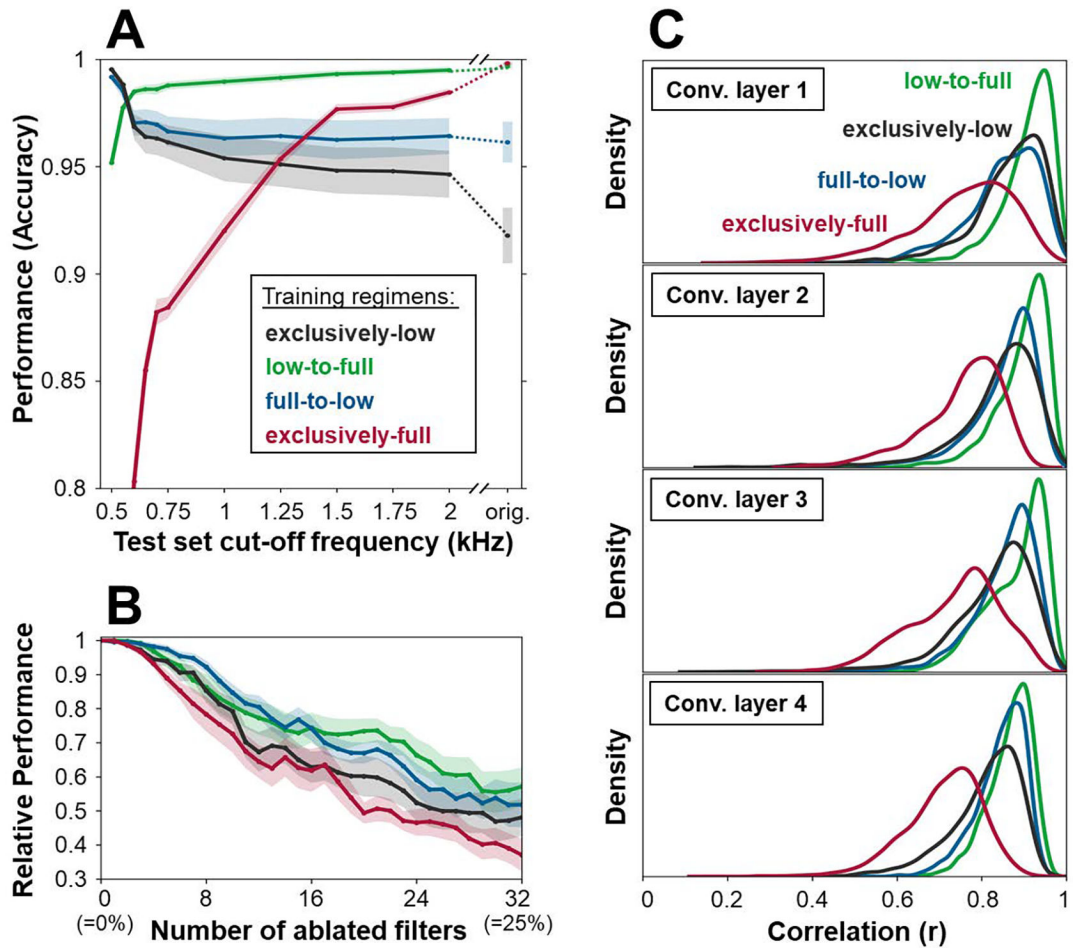
Figure 4.
A. Mean and standard error of 10-fold cross-validated emotion recognition performances on full-frequency and various low-frequency test sets. **B.** 10-fold cross-validated classification performances (baseline-normalized, for visualization), when gradually removing the units with highest peak frequency. **C.** Distribution of correlations of units' activities between low-pass filtered and full-frequency inputs, across layers (10-fold pooled).