Trends in
Cognitive Sciences

CellPress

Forum

The developing vocal self

Ana P. Pinheiro [1,*]

Developmental voice change exposes a gap in models of self-voice processing: these models explain predictive control and self-voice recognition but not how vocal signals integrate into self-representations. I propose a hierarchical framework in which recursive interactions across sensorimotor predictions, self-voice representations, and higher-order self-representations support the emergence of the vocal self.

When does a voice come to feel like 'my own'?

Current models of self-voice processing describe how individuals predict the sensory consequences of speaking and distinguish their own voice from others. Yet how these processes support the developmental emergence of the **vocal self** (see Glossary)—through which a voice is experienced as one's own and part of 'who I am'—remains less well understood.

Over the past decades, research has largely treated the voice as an auditory signal to be predicted or discriminated from others, rather than as an evolving signal progressively integrated into self-representations across development. Conceptual frameworks for speech motor control [1] and voice identity recognition [2] describe how **efference copies** and sensory feedback jointly enable the rapid detection and correction of mismatches between predicted and actual auditory feedback [1], and how internal **self-voice representations** support self-recognition even when the voice is not actively produced (e.g., during recorded playback) [2]. Together, these mechanisms underpin fluent speech and reliable self-other voice distinction. Implicit in these frameworks, however, is the assumption that the body producing the voice and the self it supports remain largely stable over time.

Developmental voice change exposes a conceptual gap. While predictive control mechanisms account for rapid adaptation to transient perturbations in auditory feedback, they do not fully specify how prolonged acoustic changes reshape self-voice representations. Similarly, prototype-based accounts posit relatively stable voice identity templates underlying self-voice recognition but remain less explicit about how and when these templates update and become integrated into **higher-order self-representations** as the body changes.

Puberty as a natural experiment

The voice bridges action, perception, and personal identity. Unlike most bodily signals, vocal output is both self-generated and externally perceivable, allowing speakers to monitor developmental changes directly through auditory feedback. It is also a primary medium through which individuals communicate and evaluate who they are. These features make developmental voice change a particularly informative model for studying how evolving bodily signals become integrated into self-representations over time.

But when and how does the vocal self emerge? Puberty provides a natural experiment (Figure 1). Hormone-driven growth of the vocal apparatus is a primary driver of long-term changes in fundamental frequency, formant structure, and timbre [3], transforming the voice into a distinctive acoustic signature. In contrast to the millisecond-scale perturbations typically modeled in predictive control research, these changes introduce sustained mismatches between established **sensorimotor predictions** and evolving auditory feedback. Amid pronounced hormonal and neural reorganization [4], the voice simultaneously becomes a

Glossary

Efference copy: an internal copy of outgoing motor commands used to generate predictions about the sensory consequences of one's own actions, including speaking, thereby enabling the rapid detection and correction of mismatches between expected and perceived vocal feedback.

Recalibration: the adaptive updating of predictive and representational systems in response to a persistent mismatch between expected and perceived vocal signals. During development, this occurs across sensorimotor predictions, self-voice representations, and higher-order self-representations.

Higher-order self-representations (self-concept): multimodal knowledge and beliefs about oneself, including identity, traits, and social categories. Throughout development, increasingly distinctive vocal signals become progressively integrated into these broader self-representations.

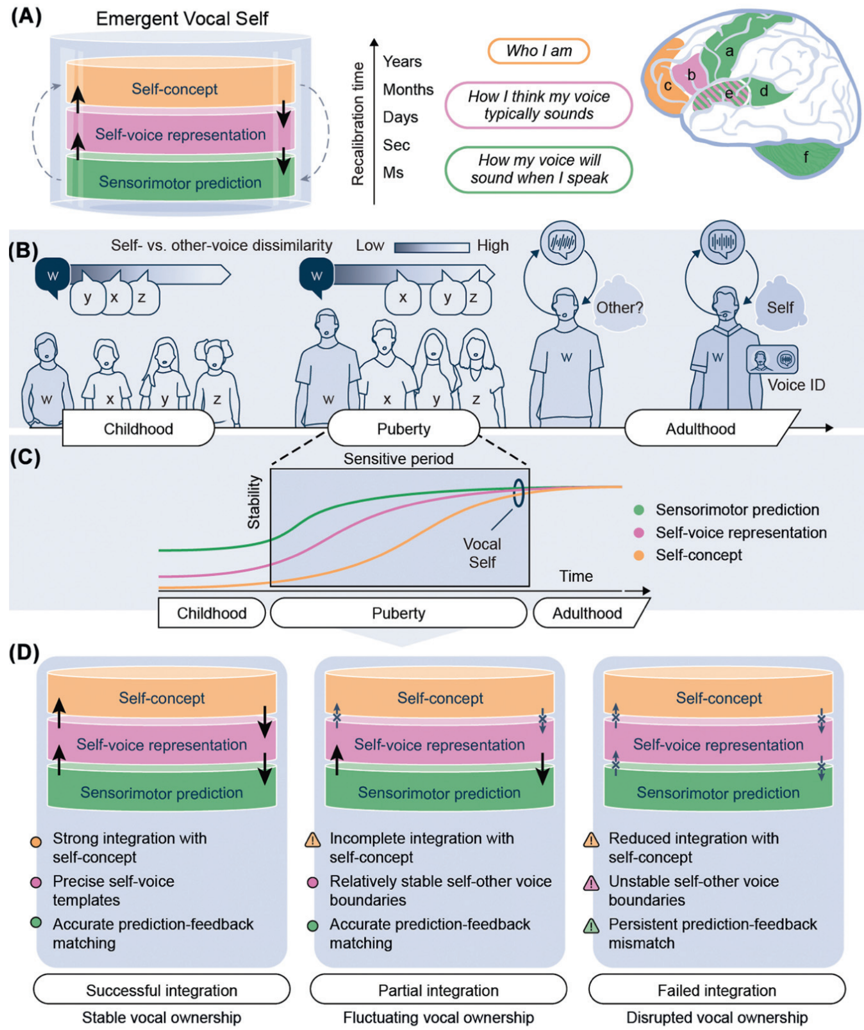
Self-voice representation: internal representations of the acoustic features of one's own voice across contexts and time. During development, these representations become progressively differentiated from other voices and integrated into higher-order self-representations. Although primarily grounded in the multisensory experience of active speech production, they may also incorporate acoustic features learned through the external playback of recorded speech.

Sensitive period (for the vocal self): a developmental window during which sensorimotor predictions, self-voice representations, and higher-order self-representations are especially plastic and sensitive to acoustic, neural, hormonal, and social influences.

Sensorimotor prediction: mechanisms that generate expectations about the sensory consequences of vocal motor commands, enabling rapid comparison between predicted and perceived auditory feedback, and supporting moment-to-moment control of vocal production.

Vocal ownership: the experience that a particular voice is one's own, emerging when sensorimotor predictions, self-voice representations, and higher-order self-representations remain sufficiently coordinated across development.

Vocal self: a developmentally emerging and progressively stabilized aspect of selfhood arising through recursive integration across sensorimotor predictions, self-voice representations, and higher-order self-representations, enabling the voice to be experienced as a meaningful and enduring aspect of the self.



Trends in Cognitive Sciences

Figure 1. The developing vocal self as a hierarchical system. (A) The vocal self comprises interacting levels operating over distinct timescales: sensorimotor prediction (milliseconds to seconds), self-voice representations (days to months), and higher-order self-representations (months to years). (B) Pre-pubertal voices are acoustically more similar and harder to distinguish. During puberty, voices become increasingly distinctive, forming individual 'voice IDs'. (C) Developmental trajectories (likely nonlinear) illustrate progressive coordination across levels. In childhood, simple motor-auditory mappings support basic self-other discrimination. During adolescence, hormone-driven changes in vocal tract anatomy demand recalibration of these mappings in parallel with self-concept development. By late puberty, self-voice representations become more tightly integrated with higher-order self-representations, yielding stable vocal ownership. Variability in the timing or tempo of integration across levels may produce distinct developmental profiles, including delayed or incomplete consolidation of the vocal self. (D) Successful cross-level integration sustains efficient vocal control, precise self-voice recognition, and stable vocal ownership. Partial integration may reflect delayed or incomplete integration of self-voice representations with higher-order self-representations, producing reduced certainty that the voice is a stable aspect of the self. Failed integration may reflect persistent misalignment across hierarchical levels, reducing the precision of vocal control, impairing self-voice recognition, and disrupting vocal ownership. Partial or failed integration may increase susceptibility to atypical self-voice processing across development. a: motor/premotor cortex; b: inferior frontal gyrus; c: medial prefrontal cortex; d: posterior superior temporal gyrus/sulcus; e: anterior superior temporal gyrus/sulcus; f: cerebellum.

salient social signal of gender, maturity, and identity.

During this period, adolescents may become temporarily 'out of sync' with their own voice as acoustic changes outpace the updating and integration of self-voice representations with higher-order self-representations. This developmental transition likely reflects heightened vocal plasticity, during which the emerging vocal self may be especially sensitive to neural, hormonal, perceptual, and social influences. In turn, atypical hormonal or neural trajectories and conditions involving altered self-experience may help reveal when and how the vocal self fails to consolidate.

From voice to vocal self

Although children can distinguish their own voices from others, self-voice representations in childhood may be less differentiated, less stable, and only weakly integrated into higher-order self-representations. During adolescence, acoustic changes transform the voice into an increasingly self-relevant signal. Amid ongoing bodily changes, a coherent vocal self must therefore be progressively constructed and stabilized.

This developmental perspective motivates a hierarchical framework (Figure 1) comprising interacting levels that operate over distinct timescales and must remain sufficiently coordinated for a stable vocal self to emerge across development. At the sensorimotor level, predictive mechanisms generate and recalibrate expectations about the sensory consequences of motor commands, enabling moment-to-moment control of vocal production [1]. At an intermediate voice identity level, prototype-like self-voice representations encode characteristic acoustic features of one's voice across contexts and time, supporting self-recognition even in the absence of action [2]. During development, these representations become progressively differentiated from other voices and increasingly stable across contexts. At a higher self-referential level, self-voice representations become integrated into multimodal self-representations supporting self-concept, including identity-relevant attributes and beliefs about oneself [5,6].

This framework is hierarchically organized across developmental and representational timescales, but interactions between levels are recursive rather than strictly feedforward. During development, lower-level sensorimotor **recalibration** continuously updates self-voice representations, supporting progressively more differentiated and stable representations of one's own voice. In turn, increasingly consolidated higher-order self-representations may indirectly influence lower-level processing through recursive interactions across hierarchical levels, stabilizing how self-voice representations are maintained and experienced as part of a coherent vocal self.

Emerging behavioral and neural evidence supports this multilevel organization. Voice ownership diminishes when vocal feedback conflicts with self-related beliefs [7] and stable acoustic voice features correlate with perceived and self-attributed personality traits [6], suggesting that vocal signals are evaluated both for correspondence with stored voice identity representations and for congruence with higher-order self-representations. Neuroimaging findings further indicate that voice-selective regions in the superior temporal cortex interact with cortical hubs involved in self-representation, such as the medial prefrontal cortex [8], forming a distributed network linking sensorimotor, identity-level, and conceptual aspects of the vocal self.

This framework does not replace predictive control and voice identity models but specifies how their underlying mechanisms may interact across development. Adaptation to a changing voice requires coordinated plasticity at multiple hierarchical levels: recalibration of sensorimotor predictions, updating of self-voice representations, and progressive integration into higher-order self-representations. A stable vocal self emerges when these levels remain sufficiently coordinated for a voice to be recognized as one's own and incorporated into the self-concept.

Developmentally, this framework motivates concrete hypotheses about **sensitive period** dynamics [9]: when recalibration and consolidation are maximal, and how factors such as pubertal timing, pubertal tempo, and social context shape the emergence of a vocal self (Box 1). The framework predicts that lower-level sensorimotor recalibration may stabilize more rapidly than self-voice and higher-order self-representations, producing transient asymmetries between vocal control, self-recognition, and **vocal ownership** during periods of rapid vocal change. For example, persistent misalignment between ongoing acoustic change and the updating of self-voice representations may result in broader, less precise self-other perceptual boundaries and greater uncertainty in self-voice attribution, potentially increasing vulnerability to atypical self-voice processing.

Toward a developmental neuroscience of the vocal self

The developing voice, being both measurable and experimentally manipulable, offers a tractable model for probing how the brain constructs, stabilizes, and maintains a vocal self amid prolonged bodily change. Acoustic trajectories can be precisely quantified and linked to neural, hormonal, perceptual, and self-report measures using longitudinal designs and computational methods. For example, longitudinal studies could determine how changes in vocal acoustics relate to neural and hormonal measures, self-voice recognition accuracy, and self-concept development, and whether these domains

Box 1. Testing the developmental dynamics of the vocal self

Voice research lacks developmental models linking acoustic, behavioral, neural, and hormonal dynamics to the emergence of the vocal self. Longitudinal, multimodal designs are particularly well suited for studying these processes. Repeated assessment of vocal acoustics, self-voice perception, hormonal maturation, brain development, and self-concept can characterize how vocal individuality emerges and stabilizes across puberty. Computational modeling and multimodal data fusion [10] can further estimate person-specific developmental trajectories, nonlinear change, and cross-domain coupling between acoustic, neural, hormonal, and behavioral measures.

Longitudinal studies should ideally begin before pubertal voice change to establish individual baselines and continue into late puberty to capture consolidation of self-voice and higher-order self-representations. Annual or biannual assessments may characterize slower developmental trajectories, whereas dense sampling precision approaches [11], including weekly or monthly assessment 'bursts' during periods of rapid vocal change, may better resolve short timescale acoustic change and perceptual recalibration.

This framework generates several testable predictions. Periods of rapid vocal change are predicted to precede transient broadening or reduced precision of self-other voice categorization boundaries. This may be reflected in altered performance in self-other voice discrimination paradigms across pubertal stages, potentially following a nonlinear trajectory as changes in vocal acoustics temporarily destabilize self-voice representations before their consolidation in later puberty. In addition, rapid vocal change may transiently increase the mismatch between predicted and perceived vocal signals during talk-listen tasks, resulting in reduced suppression of self-generated vocal feedback. Such effects may be especially pronounced in male adolescents, whose voices typically undergo larger pubertal changes compared to females [3]. Changes in self-perceived attributes (e.g., maturity, masculinity/femininity, or dominance) are predicted to lag behind vocal change, reflecting the slower updating of higher-order self-representations.

By formalizing the developmental dynamics of the vocal self, these approaches advance the field from descriptive accounts toward mechanistic understanding, opening new avenues for developmental and translational research.

follow distinct developmental trajectories. Precision approaches that collect dense within-person data can further refine these analyses, improving the reliability of individual developmental trajectory estimates [11]. Integrating these multimodal data yields testable predictions about how the vocal self emerges, stabilizes, and consolidates from childhood into adulthood, as well as the conditions under which consolidation may be delayed or disrupted (Box 1).

This framework also has translational relevance. Mapping the interaction of sensorimotor predictions, self-voice representations, and higher-order self-representations may illuminate conditions marked by atypical self-voice processing, such as psychosis or vulnerability to auditory hallucinations [12]. It may also inform interventions involving intentional modification of vocal characteristics. For example, periods of heightened vocal plasticity may influence how readily new vocal patterns become integrated into self-representations in contexts such as gender-affirming interventions. A functional and neural model of the vocal self across development is therefore essential.

Concluding remarks

Developmental change reveals that self-voice processing involves more than prediction and discrimination. As vocal signals become increasingly distinctive, a central challenge is to construct and maintain a stable vocal self through recursive coordination across sensorimotor predictions, self-voice representations, and higher-order self-representations. Rather than being a fixed system, the vocal self emerges as a developmental achievement through which a voice becomes a meaningful and enduring aspect of the self. Development, in this sense, becomes central to self-voice theory by revealing whether hierarchical levels remain sufficiently coordinated to support vocal individuality, self-continuity, and stable vocal ownership beyond puberty. By positioning the voice as a tractable model of selfhood, this developmental framework opens new avenues through which the link between voice and self can be mechanistically understood, measured, and ultimately leveraged in clinical and technological contexts.

Acknowledgments

I thank the members of the VoicES Lab for their insightful discussions on self-voice processing across development and its neurofunctional mechanisms. I am also grateful to the Editor and anonymous reviewers for their thoughtful comments and suggestions, and to Gil Costa for assistance with figure generation. This work was supported by Fundação para a Ciência e Tecnologia and the BIAL Foundation (grant numbers 2023.00041.RESTART and BIAL 146/2020).

Declaration of interests

The author declares no competing interests.

[1]Faculdade de Psicologia, CICPSI, Universidade de Lisboa, Alameda da Universidade, Lisboa 1649-013, Portugal

*Correspondence:
appinheiro@psicologia.ulisboa.pt (A.P. Pinheiro).

https://doi.org/10.1016/j.tics.2026.06.005

© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

References

1. Todorović, S. *et al.* (2024) Cortico-cerebellar monitoring of speech sequence production. *Neurobiol. Lang. (Camb.)* 5, 701–721
2. Orepic, P. *et al.* (2023) Bone conduction facilitates self-other voice discrimination. *R. Soc. Open Sci.* 10, 221561
3. Markova, D. *et al.* (2016) Age- and sex-related variations in vocal-tract morphology and voice acoustics during adolescence. *Horm. Behav.* 81, 84–96
4. Curtis, M. *et al.* (2024) Disentangling the unique contributions of age, pubertal stage, and pubertal hormones to brain structure in childhood and adolescence. *Dev. Cogn. Neurosci.* 70, 101473
5. Tucker, A.E. *et al.* (2024) How does our voice reflect who we are? Connecting the voice and the self using implicit association tests. *J. Voice* https://doi.org/10.1016/j.jvoice.2024.10.018
6. Stern, J. *et al.* (2021) Do voices carry valid information about a speaker's personality? *J. Res. Personal.* 92, 104092
7. da Cruz Martinho, D.H. and Constantini, A.C. (2025) Gender presentation and voice satisfaction: self-perception of voice versus external perception. *J. Voice* https://doi.org/10.1016/j.jvoice.2024.11.043
8. Guldner, S. *et al.* (2020) Vocomotor and social brain networks work together to express social traits in voices. *Cereb. Cortex* 30, 6004–6020
9. Frankenhuis, W.E. and Walasek, N. (2020) Modeling the evolution of sensitive periods. *Dev. Cogn. Neurosci.* 41, 100715
10. Kraft, D. *et al.* (2023) Domain adapted brain network fusion captures variance related to pubertal brain development and mental health. *Nat. Commun.* 14, 6698
11. Lee, H.J. *et al.* (2025) Using precision approaches to improve brain-behavior prediction. *Trends Cogn. Sci.* 29, 170–183
12. Duggirala, S.X. *et al.* (2024) Hallucination proneness alters sensory feedback processing in self-voice production. *Schizophr. Bull.* 50, 1147–1158