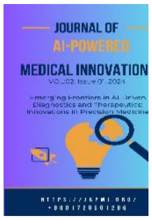
Journal of AI-Powered Medical Innovations
ISSN: 3078-1930 DOI: 10.60087
2026, Vol. 04, No. 1
Home page https://japmi.org/



Review Article

Hierarchy of Auditory Cortex Adaptation to Signal Degradation. From Acoustics to Prediction

Marie Setaire
Research group of the methods of science education and popularization

ABSTRACT

The brain does not process speech as a single stream. Instead, it operates across multiple levels at once — acoustic, voice-selective, prosodic, and multisensory — and each level has its own way of compensating when the signal degrades, whether through noise vocoding, disrupted binaural hearing, or peripheral deafness. This review brings together evidence showing that: (1) the bilateral superior temporal gyrus responds to acoustic degradation largely regardless of whether speech is intelligible; (2) the temporal voice areas (TVA) stay tuned to voices even under severe degradation, drawing on preserved temporal cues; (3) visual prosodic facial movements generate predictive auditory representations before the acoustic signal even arrives; and (4) bilateral cochlear implants normalize cortical activity patterns more effectively than a single implant. Drawing on fMRI, PET, and EEG data, we propose an integrative model of hierarchical cortical adaptation and consider what it means for auditory rehabilitation.

Keywords: auditory cortex; vocoding; cochlear implant; temporal voice areas; predictive coding; binaural hearing; cross-modal plasticity; superior temporal gyrus; speech perception; cortical adaptation

---

* Corresponding author: Marie Setaire Email: g.s.mariam@gmail.com

Received: 01-05-2026; **Accepted:** 25-05-2026; **Published:** 04-06-2026

1. 1. Introduction

Of all the auditory objects the human brain must process, speech is among the most demanding. Perceiving it calls on a distributed network of cortical and subcortical structures operating simultaneously across multiple levels of representation. Decades of neuroimaging work have sketched the broad outlines of this network, centred on the superior temporal gyrus (STG) and sulcus (STS) and extending into inferior frontal and parietal regions (Hickok & Poeppel, 2007; Scott & Johnsrude, 2003). What remains a live question — and the focus of this review — is how this network adapts when the input signal is degraded, whether by cochlear pathology, background noise, or the artefacts introduced by signal processing.

Auditory signal degradation can come from many sources. In everyday life the most common challenge is background noise; in clinical populations, the primary cause is peripheral hearing loss. Under experimental conditions, noise vocoding offers a precisely controllable model of spectral degradation that closely mimics what a cochlear implant (CI) processor does: the signal is broken into frequency bands in which the temporal amplitude envelope is preserved while fine spectral structure is discarded (Shannon et al., 1995). This controllability makes vocoding ideal for separating acoustic from linguistic contributions to the neural response to degraded speech.

A key theoretical question is whether the brain treats signal degradation as a single problem, or whether different levels of the auditory hierarchy each respond in their own specific way. The evidence reviewed here firmly supports the latter view. The bilateral STG/STS responds to acoustic degradation largely regardless of speech intelligibility (Strelnikov et al., 2011a; Obleser et al., 2008). The TVA hold on to voice-selectivity even under severe spectral degradation, apparently by drawing on preserved temporal cues (Strelnikov et al., 2011a; Belin et al., 2000). Visual prosodic facial movements trigger predictive auditory representations before the acoustic signal arrives (Strelnikov et al., 2015; van Wassenhove et al., 2005). And restoring binaural hearing through bilateral CI normalizes cortical activity patterns more effectively than monaural stimulation does (Strelnikov et al., 2011b; Litovsky et al., 2009).

This review brings these findings together within a unified hierarchical framework. We work through four processing levels in turn — acoustic, voice-selective, prosodic, and binaural — examining what each level does, how the levels interact, and what the whole picture means for hearing rehabilitation.

2. 2. Methodological Foundations: Vocoding as an Experimental Model

Noise vocoding was first established as a model of CI processing by Shannon et al. (1995), who showed that speech can be understood with as few as four frequency channels, as long as temporal envelope information is intact. Later work confirmed that both voice discrimination and speech recognition improve as channel number increases, and that varying the number of channels affects the brain in ways that closely parallel changes in electrode number and placement in actual CI users (Friesen et al., 2001; Lorenzi et al., 2006). The key point is what vocoding preserves and what it removes: it keeps the slow amplitude envelope (typically below ~50 Hz) that carries rhythm, syllabic rate, and broad prosodic contour, while stripping out the fine spectral structure needed to resolve formants and to tell voices apart by gender or identity.

Behaviourally, voice/non-voice discrimination holds up far better under vocoder degradation than lexical intelligibility does. At four and eight channels, listeners can still reliably distinguish voice from non-voice stimuli, while intelligibility has already fallen substantially (Massida et al., 2011). This behavioural split implies that there should be a corresponding neural split across levels of the auditory hierarchy — a prediction that the neuroimaging studies reviewed below put to the test.

It is worth noting that vocoding is not the same as simply adding broadband noise. The two manipulations differ in how they affect temporal fine structure, spectral contrast, and the acoustic cues that carry voice identity. Studies directly comparing the two confirmed qualitatively different patterns of STG activation (Obleser et al., 2008), which is why vocoding is such a useful and specific model for CI-related degradation.

3. 3. The Acoustic Level: STG/STS Responses to Spectral Degradation

3.1. 3.1. Bilateral Effects of Vocoding in the STG/STS

The most reliable neural marker of vocoder degradation is a drop in activity across the bilateral STG and STS. Early PET work showed that intelligibility-related BOLD responses tracked vocoder channel number in bilateral STG clusters extending into the left temporal pole (Scott et al., 2006). Later fMRI studies confirmed and sharpened this picture: channel number correlates with STS activity bilaterally, and sentence-level stimuli recruit an additional peak in the left inferior frontal gyrus (Obleser et al., 2007).

A key methodological step was independently manipulating acoustic degradation and linguistic intelligibility. In a factorial fMRI design crossing three stimulus categories — intelligible speech (IS), unintelligible speech in an unfamiliar language (US), and environmental sounds (ES) — with five vocoder degradation levels (Strelnikov et al., 2011a), regression analysis found a significant correlation between BOLD signal and degradation level across a large bilateral STG/STS cluster. Critically, this effect was equally strong for IS and US, with no significant Category × Degradation interaction in the whole-brain analysis. The main driver of STG deactivation under vocoding is therefore acoustic, not linguistic.

The only exception was the posterior STG, where a significant Category × Degradation interaction did emerge: deactivation was greater for IS than for US or ES. This region — distinct from the areas highlighted by Obleser et al. (2008) — appears to take on the extra work of lexico-semantic processing when intelligible speech must be decoded under poor acoustic conditions, consistent with the proposed role of posterior STG at the phonological–semantic interface (Hickok & Poeppel, 2007).

3.2. 3.2. Topography of Acoustic and Linguistic Effects

Together, these findings map out two functionally distinct zones within the superior temporal cortex. The anterior and mid-STG/STS respond mainly to acoustic degradation, and do so similarly for speech and non-speech alike. The posterior STG adds sensitivity to intelligibility, marking the point where acoustic processing gives way to linguistic-level adaptation. This anterior–posterior gradient fits well with hierarchical models of STG/STS processing (Scott & Johnsrude, 2003; Davis & Johnsrude, 2003) and with the dual-stream framework for auditory cortex organisation (Hickok & Poeppel, 2007).

Worth noting is that vocoded speech processing also recruits the cerebellum (Strelnikov et al., 2011a) — something not highlighted in earlier single-condition vocoding studies. Meta-analytic work confirms the cerebellum's involvement in both speech production and perception (Ackermann, 2008), and its engagement is particularly strong under challenging acoustic conditions such as noisy speech (Wong et al., 2008) and in CI users (Giraud et al., 2000). The bilateral cerebellar activation seen under vocoding is consistent with the cerebellum playing a predictive or forward-model role when auditory input is degraded.

4. 4. The Voice-Selective Level: Resilience of the Temporal Voice Areas

4.1. 4.1. The TVA: Organization and Normal Function

The temporal voice areas (TVA) are a set of regions distributed along the bilateral STS that respond more strongly to vocal sounds than to non-vocal ones. They were first identified using PET (Belin et al., 2000) and later characterised at finer spatial resolution with fMRI (Pernet et al., 2015). The TVA contain at least three functional subregions arranged along the anterior–posterior STS axis, each sensitive to different aspects of voice: speaker identity, voice gender, and voice quality (Belin et al., 2002; Latinus & Belin, 2011). Since their selectivity for vocal timbre seems to depend substantially on spectral information, the natural question is how TVA function survives when spectral detail is stripped away.

4.2. 4.2. TVA Voice-Selectivity Under Degradation

Evidence from factorial vocoding designs (Strelnikov et al., 2011a) shows that the TVA are largely resistant to spectral degradation. An ROI analysis using an independent TVA localizer (Belin et al., 2000) found no significant correlation between TVA activity and vocoder channel number, using either linear or logarithmic approaches ($p > 0.1$). Even more directly, comparing degraded with natural stimuli showed that the drop in right STS activity was significantly smaller for speech (IS + US combined) than for environmental sounds — confirming that the TVA retain a relative preference for vocal over non-vocal stimuli even under substantial degradation.

The most likely explanation is that the TVA fall back on preserved temporal cues. Both intelligible and unintelligible speech share distinctive amplitude modulation patterns — quasi-periodic syllabic envelope structure, characteristic onset and offset statistics — that set vocal sounds apart from non-vocal ones regardless of spectral content (Chandrasekaran et al., 2009). Reinforcing this, modulating broadband noise with speech-derived temporal envelopes is enough for listeners to classify sounds as speech even with no spectral information at all (Shannon et al., 1995). The TVA may therefore use this temporal signature as a degradation-resistant proxy for voiceness.

4.3. 4.3. Spatial Dissociation Between TVA and Degradation-Sensitive Regions

The activation peaks for the TVA (voice-selective response) sit spatially apart from the peaks of degradation-sensitive regions in the STG/STS. Euclidean peak-to-peak distances were 17.5 mm on the left and 7.3 mm on the right — well beyond the 6-mm smoothing kernel. At the same statistical threshold, only 3–16% of degradation-correlated voxels overlapped with the TVA localizer. Peripheral TVA zones did show greater engagement during degraded speech, forming a functional buffer between the voice-selective core and the broader acoustic-sensitivity regions. This spatial arrangement suggests that the mechanisms sustaining TVA voice-selectivity are, to a meaningful degree, insulated from global acoustic degradation effects.

5. 5. The Prosodic Level: Predictive Coding in Audiovisual Speech

5.1. 5.1. Visual Prosodic Cues as Predictors of Upcoming Acoustic Events

Speech prosody — the organisation of pitch, amplitude, and duration at the level above individual sounds — is communicated not just acoustically but also through systematic facial and head movements. Eyebrow raises, head nods, and chin movements co-vary with fundamental frequency and amplitude modulations across a wide range of languages and speaking styles (Munhall et al., 2004). Critically, these visual gestures arrive 100–500 ms before the corresponding acoustic events, making them potential predictive cues for the auditory system. Behavioural work has confirmed that visual prosodic information lowers auditory detection thresholds for amplitude changes — evidence of genuine cross-modal facilitation at the sensory level (Foxton et al., 2010).

At the neural level, it has long been established that visual speech — lip movements — modulates auditory cortex activity even when no sound is present, and shortens the latency of auditory cortical responses (van Wassenhove et al., 2005; Besle et al., 2004). These effects have been attributed to a predictive mechanism: because lip movements reliably precede sound, they act as bottom-up predictors that pre-activate auditory representations and ease acoustic processing (Arnal et al., 2009). Whether analogous predictive mechanisms work for prosody-related movements — which carry different information and operate on longer timescales — was, until recently, an open question.

A decisive EEG test using an audiovisual oddball paradigm with a three-word phrase answered this question (Strelnikov et al., 2015). Standard stimuli presented a neutral audiovisual phrase; deviants carried the identical acoustic track but added a visual prosodic accent — an eyebrow raise on the middle word — about 500 ms before the word’s acoustic position. An MMN-like negativity appeared at 360–400 ms, roughly 200 ms before the acoustic target word arrived. This component was absent in the visual-only condition (same facial movements, no sound), ruling out a purely visual MMN. Subtracting visual-only ERPs before comparing standard and deviant conditions confirmed that the response was genuinely audiovisual in nature (Strelnikov et al., 2015).

5.2. 5.2. Neural Dynamics of Audiovisual Prosodic Prediction

Spatiotemporal analysis revealed a clear pattern of neural propagation: early occipito-parietal activity (320–360 ms) spread to fronto-temporal sites (360–400 ms), then returned to occipital regions (400–440 ms). Source analysis identified generators in left frontal and left posterior temporal cortex at the peak of the response. This occipito-temporo-frontal loop — which completed itself before the acoustic target arrived — closely resembles the pathway previously identified for lip-reading prediction of phonemes (Arnal et al., 2009), but now operating at the prosodic rather than the segmental level.

The posterior temporal generator fits with the proposed role of this region in using visual information to anticipate upcoming auditory phonemes (Kilian-Hutten et al., 2011). The frontal component likely reflects the top-down transmission of prosodic predictions to auditory cortex via the dorsal stream. Together, the pattern is consistent with a two-pathway model of audiovisual speech prediction (Arnal et al., 2009): a fast, direct cortico-cortical route carrying visual motion parameters to auditory cortex, and a slower feedback pathway that signals prediction errors.

For auditory rehabilitation, this finding carries a concrete implication. CI processing severely disrupts the spectral cues that carry prosody — fundamental frequency and voice quality — and prosodic perception remains one of the most persistent deficits in CI users (Marx et al., 2015). If visual prosodic cues can pre-activate auditory prosodic representations before the acoustic signal even arrives, training CI users to make better use of these cues could offer a meaningful compensatory route to improved prosodic comprehension.

6. 6. The Binaural Level: Normalization of Brain Activity Through Bilateral CI

6.1. 6.1. Behavioral Advantages of Bilateral CI

Bilateral cochlear implantation is designed to restore binaural hearing for patients with profound bilateral deafness. Clinical evidence consistently shows that bilateral CI recipients outperform unilateral users on speech recognition in noise, sound localization, and listening effort (Litovsky et al., 2009; Mosnier et al., 2009; Brown & Balkany, 2007). These gains come from two main mechanisms: the head shadow effect (a better signal-to-noise ratio when the target reaches the better ear) and binaural summation (combining signals from both ears to improve overall sensitivity). What neuroimaging set out to address was whether these behavioural advantages also reflect genuine normalisation of cortical processing — not just a quantitative increase in auditory input.

6.2. 6.2. PET Evidence for Cortical Normalization

A PET study of word/nonword discrimination in five bilateral CI patients and five normal-hearing controls (Strelnikov et al., 2011b) took this question on directly. Under binaural stimulation, the group comparison showed no significant overactivation in patients; the only notable difference was moderate hypoactivation of the right superior temporal cortex, possibly reflecting a lack of inter-implant synchronisation at that threshold. Under monaural stimulation, however, patients showed bilateral overactivation of the posterior inferior temporal cortex and the cerebellum — regions associated with lexical interface processing and predictive compensatory mechanisms, respectively.

The interaction analysis confirmed that the monaural-minus-binaural contrast was significantly larger in patients than controls, specifically in the right cerebellum. This dissociation is theoretically important: it implies that bilateral CI imposes roughly the same computational demand as normal binaural hearing, while single-implant listening places substantially heavier demands on higher-order speech processing networks. The neural advantages of bilateral implantation therefore go beyond simply adding a second monaural signal.

6.3. 6.3. Implications of Cerebellar Involvement

The specific overactivation of the cerebellum under monaural CI — and its normalisation under binaural CI — connects the binaural level of the hierarchy to the predictive coding mechanisms described at the prosodic level. Meta-analytic evidence implicates the cerebellum in temporal prediction during speech processing (Ackermann, 2008), and its activation is disproportionately elevated under degraded speech conditions (Wong et al., 2008). The normalisation of cerebellar activity under binaural stimulation suggests that binaural integration reduces the system's reliance on compensatory prediction — in effect, the better-quality integrated signal produces fewer prediction errors, and the cerebellum has less corrective work to do.

7. 7. An Integrative Model of Hierarchical Cortical Adaptation

The evidence across all four levels points toward an integrative model in which auditory cortex adaptation to signal degradation is not a single unified response, but a hierarchically organised set of mechanisms — each with its own properties and its own limits.

At the acoustic level, the bilateral STG/STS responds to reductions in spectral detail with drops in BOLD activation that are largely unaffected by linguistic content. The posterior STG adds sensitivity to intelligibility, forming a transitional zone between acoustic and lexico-semantic processing. This level is directly shaped by CI channel number and electrode placement, and its activity closely mirrors intelligibility scores.

At the voice-selective level, the TVA maintain their preference for vocal sounds by drawing on temporal cues preserved by vocoding. Peripheral TVA zones show greater engagement as degradation increases, buffering the voice-selective core from the broader STG degradation response. This level is partially shielded from spectral degradation, but is vulnerable to long-term reorganisation under chronic deafness.

At the prosodic level, visual facial movements activate predictive audiovisual networks that pre-activate auditory representations on a 400–500 ms time horizon. This mechanism is shaped by both evolution and development to support natural face-to-face communication, and represents a high-level compensatory resource that becomes particularly valuable when the acoustic signal lacks prosodic cues — as in CI processing.

At the binaural level, restoring bilateral auditory input normalises cortical activity by reducing the computational demands that cause lexical interface regions and the cerebellum to overactivate under monaural degraded-signal processing. This normalisation suggests that the brain's binaural processing machinery remains functionally preserved in bilateral CI users, even after years of auditory deprivation.

The four levels are not independent of one another. When acoustic degradation is severe, the relative importance of the higher-level compensatory resources grows: temporal cues (voice-selective level), visual prosodic predictions (prosodic level), and binaural integration (binaural level) work together to determine how much of the speech signal can be recovered. The system is therefore functionally redundant — but only up to a point. No combination of higher-level mechanisms can fully replace the fine spectral information needed to discriminate voice attributes such as gender, emotion, and identity.

8. 8. Clinical Implications and Future Directions

8.1. 8.1. Bilateral Cochlear Implantation

The convergence of behavioural and neuroimaging evidence strongly supports the case for bilateral CI. The near-normalisation of cortical activity under binaural stimulation shows that the brain's binaural processing mechanisms can be brought back online with the right device provision. This is especially urgent for children: there is a sensitive period for binaural cortical organisation, and sequential bilateral implantation with a gap of more than roughly 1.5 years between implants is associated with persistent abnormal aural preference that does not resolve even with prolonged bilateral device use (Gordon et al., 2012). Where possible, the clinical goal should be early bilateral implantation — ideally simultaneous, or at least within this sensitive window.

8.2. 8.2. Prosodic and Audiovisual Rehabilitation

The evidence for visual prosodic predictive coding suggests that rehabilitation programmes could partially compensate for persistent prosodic deficits by systematically incorporating training on prosodically informative facial movements. Materials using natural face-to-face speech — rather than audio-only or degraded audiovisual stimuli — are likely to engage the predictive prosodic mechanism most effectively. This approach may also have something to offer patients with focal lesion-based aprosodia (Ross & Monnot, 2008), a clinically underrecognised condition where visual prosodic training has not yet been rigorously evaluated.

8.3. 8.3. Monitoring Cortical Reorganization in Children

One notable gap in the existing literature is that we have limited data on how hierarchical cortical adaptation unfolds during development, particularly in children with unilateral hearing loss. The fNIRS-based protocol developed by Calmels et al. (2022) for children aged 5–16 years with untreated single-sided deafness (SSD) is a methodologically well-suited step toward closing this gap. By measuring asymmetry indices (AI) and aural preference indices (API) during monaural and binaural stimulation, the protocol aims to detect the early cortical signs of maladaptive reorganisation — aural preference syndrome — before they become fixed. Connecting these cortical indices to behavioural measures of binaural hearing, language development, and quality of life will provide the clinical benchmarks needed to guide individualised rehabilitation planning.

8.4. 8.4. Outstanding Questions

Several open questions call for further investigation. First, it remains unknown whether TVA voice-selectivity holds up in long-term CI users in the same way it does under the short-term acute degradation modelled by vocoding in normal listeners. Chronic deafness may produce cross-modal reorganisation of the TVA, potentially undermining their temporal-cue-based resilience. Second, the conditions under which visual prosodic predictive mechanisms interact with CI-processed acoustic input in real time are uncharted territory; EEG studies using naturalistic audiovisual speech with CI simulation would be a natural first step. Third, the considerable individual variability in hierarchical compensation strategies calls for systematic characterisation of the neurobiological predictors — structural connectivity, resting-state network topology — that determine which compensatory mechanisms a given patient can draw on.

9. 9. Conclusion

The evidence presented here makes the case that adaptation to auditory signal degradation is not a single unified process, but a collection of partially independent mechanisms distributed across a hierarchy of cortical levels. The acoustic level (STG/STS) governs overall sensitivity to spectral content, responding to degradation largely regardless of linguistic context. The voice-selective level (TVA) keeps vocal identity processing alive by drawing on preserved temporal cues. The prosodic level uses visual facial movements to build predictive auditory representations before the acoustic signal arrives. And the binaural level determines the overall computational burden on the system — bilateral input substantially reducing the higher-order compensatory demands that degraded monaural signals impose.

Understanding this hierarchy has direct implications for rehabilitation practice. It provides neurofunctional support for early bilateral cochlear implantation, motivates the development of audiovisual prosodic training protocols, and points to functional near-infrared spectroscopy as a practical tool for monitoring cortical adaptation in children. The key message from the hierarchical framework is that optimal rehabilitation needs to operate at all levels simultaneously: device optimisation addresses the acoustic level, but prosodic prediction and binaural integration require targeted training and — where clinically indicated — bilateral device provision.

References

1. Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. Trends in Neurosciences, 31(6), 265–272. https://doi.org/10.1016/j.tins.2008.02.011
2. Arnal, L. H., Morillon, B., Kell, C. A., & Giraud, A. L. (2009). Dual neural routing of visual facilitation in speech processing. Journal of Neuroscience, 29(43), 13445–13453. https://doi.org/10.1523/JNEUROSCI.3194-09.2009
3. Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. Cognitive Brain Research, 13(1), 17–26. https://doi.org/10.1016/S0926-6410(01)00084-2
4. Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. Nature, 403(6767), 309–312. https://doi.org/10.1038/35002078
5. Besle, J., Fort, A., Delpuech, C., & Giard, M. H. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. European Journal of Neuroscience, 20(8), 2225–2234. https://doi.org/10.1111/j.1460-9568.2004.03670.x
6. Calmels, M.-N., Gallois, Y., Marx, M., Deguine, O., Taoui, S., Arnaud, E., Strelnikov, K., & Barone, P. (2022). Functional reorganization of the central auditory system in children with single-sided deafness: A protocol using fNIRS. Brain Sciences, 12(4), 423. https://doi.org/10.3390/brainsci12040423
7. Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., & Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. PLoS Computational Biology, 5(7), e1000436. https://doi.org/10.1371/journal.pcbi.1000436
8. Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. Journal of Neuroscience, 23(8), 3423–3431. https://doi.org/10.1523/JNEUROSCI.23-08-03423.2003
9. Foxton, J. M., Rivière, L. D., & Barone, P. (2010). Cross-modal facilitation in speech prosody. Cognition, 115(1), 71–78. https://doi.org/10.1016/j.cognition.2009.11.009
10. Friesen, L. M., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants.

Journal of the Acoustical Society of America, 110(2), 1150–1163. https://doi.org/10.1121/1.1381538

11. Giraud, A. L., Truy, E., Frackowiak, R. S., Grégoire, M. C., Pujol, J. F., & Collet, L. (2000). Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. Brain, 123(7), 1391–1402. https://doi.org/10.1093/brain/123.7.1391
12. Gordon, K. A., Salloum, C., Toor, G. S., van Hoesel, R., & Papsin, B. C. (2012). Binaural interactions develop in the auditory brainstem of children who are deaf: Effects of place and level of bilateral electrical stimulation. Journal of Neuroscience, 32(12), 4212–4223. https://doi.org/10.1523/JNEUROSCI.5741-11.2012
13. Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. Nature Reviews Neuroscience, 8(5), 393–402. https://doi.org/10.1038/nrn2113
14. Kilian-Hutten, N., Vroomen, J., & Formisano, E. (2011). Brain activation during audiovisual exposure anticipates future perception of ambiguous speech. NeuroImage, 57(4), 1601–1607. https://doi.org/10.1016/j.neuroimage.2011.05.043
15. Latinus, M., & Belin, P. (2011). Human voice perception. Current Biology, 21(4), R143–R145. https://doi.org/10.1016/j.cub.2010.12.033
16. Litovsky, R. Y., Parkinson, A., & Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. Ear and Hearing, 30(4), 419–431. https://doi.org/10.1097/AUD.0b013e3181a165be
17. Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proceedings of the National Academy of Sciences, 103(49), 18866–18869. https://doi.org/10.1073/pnas.0607364103
18. Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., & Deguine, O. (2015). Speech prosody perception in cochlear implant users with and without residual hearing. Ear and Hearing, 36(2), 239–248. https://doi.org/10.1097/AUD.0000000000000105
19. Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., & Deguine, O. (2011). Voice discrimination in cochlear-implanted deaf subjects. Hearing Research, 275(1–2), 120–129. https://doi.org/10.1016/j.heares.2010.12.013
20. Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., Fraysse, B., Bordure, P., Mondain, M., Bouccara, D., Bozorg-Grayeli, A., Borel, S., & Ambert-Dahan, E. (2009). Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients. Audiology & Neuro-Otology, 14(2), 106–114. https://doi.org/10.1159/000159121
21. Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. Psychological Science, 15(2), 133–137. https://doi.org/10.1111/j.0963-7214.2004.01502010.x
22. Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. Journal of Neuroscience, 28(32), 8116–8123. https://doi.org/10.1523/JNEUROSCI.1290-08.2008

23. Obleser, J., Wise, R. J., Alex Dresner, M., & Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. Journal of Neuroscience, 27(9), 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007
24. Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., & Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. NeuroImage, 119, 164–174. https://doi.org/10.1016/j.neuroimage.2015.06.050
25. Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. Brain and Language, 104(1), 51–74. https://doi.org/10.1016/j.bandl.2007.04.007
26. Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. Trends in Neurosciences, 26(2), 100–107. https://doi.org/10.1016/S0166-2236(02)00037-1
27. Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech — a positron emission tomography study. Journal of the Acoustical Society of America, 120(2), 1075–1083. https://doi.org/10.1121/1.2216565
28. Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. Science, 270(5234), 303–304. https://doi.org/10.1126/science.270.5234.303
29. Strelnikov, K., Massida, Z., Rouger, J., Belin, P., & Barone, P. (2011a). Effects of vocoding and intelligibility on the cerebral response to speech. BMC Neuroscience, 12, 122. https://doi.org/10.1186/1471-2202-12-122
30. Strelnikov, K., Foxton, J., Marx, M., & Barone, P. (2015). Brain prediction of auditory emphasis by facial expressions during audiovisual continuous speech. Brain Topography, 28(3–4), 494–505. https://doi.org/10.1007/s10548-013-0338-2
31. Strelnikov, K., Rouger, J., Eter, E., Lagleyre, S., Fraysse, B., Demonet, J.-F., Barone, P., & Deguine, O. (2011b). Binaural stimulation through cochlear implants in postlingual deafness: A positron emission tomographic study of word recognition. Otology & Neurotology, 32(8), 1210–1217. https://doi.org/10.1097/MAO.0b013e31822e5bd6
32. Brown, K. D., & Balkany, T. J. (2007). Benefits of bilateral cochlear implantation: a review. Current Opinion in Otolaryngology & Head and Neck Surgery, 15(5), 315–318. https://doi.org/10.1097/MOO.0b013e3282ef3d3e
33. van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. Proceedings of the National Academy of Sciences, 102(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102
34. Wong, P. C., Uppunda, A. K., Parrish, T. B., & Dhar, S. (2008). Cortical mechanisms of speech perception in noise. Journal of Speech, Language, and Hearing Research, 51(4), 1026–1041. https://doi.org/10.1044/1092-4388(2008/075)