Reverse-Engineering the Human Auditory Pathway

Lloyd Watts

Audience, Inc.
440 Clyde Avenue
Mountain View, CA, 94043
lwatts@audience.com

Abstract. The goal of reverse-engineering the human brain, starting with the auditory pathway, requires three essential ingredients: Neuroscience knowledge, a sufficiently capable computing platform, and a long-term funding source. By 2003, the neuroscience community had a good understanding of the characterization of sound which is carried out in the cochlea and auditory brainstem, and 1.4 GHz single-core computers with XGA displays were fast enough that it was possible to build computer models capable of running and visualizing these processes in isolation at near biological resolution in real-time, and it was possible to raise venture capital funding to begin the project. By 2008, these advances had permitted the development of products in the area of two-microphone noise reduction for mobile phones, leading to viable business by 2010, thus establishing a self-sustaining funding source to continue the work into the next decade 2010-2020. By 2011, advances in fMRI, multi-electrode, and behavioral studies have illuminated the cortical brain regions responsible for separating sounds in mixtures, understanding speech in quiet and in noisy environments, producing speech, recognizing speakers, and understanding and responding emotionally to music. 2GHz computers with 8 virtual cores and HD displays now permit models of these advanced auditory brain processes to be simulated and displayed simultaneously in real-time, giving a rich perspective on the concurrent and interacting representations of sound and meaning which are developed and maintained in the brain, and exposing a deeper generality to brain architecture than was evident a decade earlier. While there is much still to be discovered and implemented in the next decade, we can show demonstrable progress on the scientifically ambitious and commercially important goal of reverse-engineering the human auditory pathway.

As outlined in 2003 [1], the goal of reverse-engineering the human brain, starting with the auditory pathway, requires three essential ingredients: Neuroscience knowledge, a sufficiently capable computing platform, and a long-term funding source. In this paper, we will describe the first successful decade of this multi-decade project, and show progress and new directions leading into a promising second decade.

By 2003, the neuroscience community had a good understanding of the characterization of sound which is carried out in the cochlea and auditory brainstem, including the detection of inter-aural time and level differences (ITD and ILD)

J. Liu et al. (Eds.): IEEE WCCI 2012, LNCS 7311, pp. 47–59, 2012.

computed in the superior olivary complex (SOC, MSO, LSO) used for determining the azimuthal location of sound sources, and the essential brainstem foundations for extracting polyphonic pitch (delay lines needed for autocorrelation in the nucleus of the lateral lemniscus (NLL), and combination-sensitive cells in the inferior colliculus (IC)). While there was still significant uncertainty about the full role of the inferior colliculus, medial geniculate body (MGB) of the thalamus, and auditory cortical regions, there was sufficient clarity and consensus of the lower brainstem representations to begin a serious modeling effort [1].

In 2003, on a single-core 1.4 GHz processor, it was possible to build computer models capable of running these processes in isolation at near biological resolution in real-time, e.g., a 600-tap cochlea model spanning a frequency range 20Hz - 20kHz at 60 taps/octave with realistic critical bandwidths, efficient event-driven ITD and normalized ILD computations, and a plausible model of polyphonic pitch [1]. By 2008, these advances had permitted the development of products in the area of two-microphone noise reduction for mobile phones [2][3][4], leading to viable business by 2010, thus establishing a commercial foundation to continue the work into the next decade 2010-2020.

1 Neuroscience Advances in 2003-2010 Illuminate Cortical Architecture

During 2003-2010, new fMRI, multi-electrode, and behavioral studies have illuminated the cortical brain regions responsible for separating sounds in mixtures [5][6], understanding speech in quiet and in noisy environments [7], producing speech [7], recognizing speakers [8], and understanding music [9][10]. Similarly, there is greater clarity in the function of the hippocampus [11] and amygdala [12][13] in the limbic system, relating to long-term memory storage and retrieval, and emotional responses to auditory stimuli [12]. While there is still much to be discovered about the underlying representation of signals in the cortex, it is now possible to see an architectural organization begin to emerge within the auditory pathway, as shown in Figures 1 and 2.

These figures were created by starting with the auditory pathway diagram first published in [1], and then updating the cortical regions to show the speech recognition and production pathways from [7], speaker identification pathway from [8], music pathways inferred from the functional description in [9], and limbic system pathways from [10][11][12], with additional guidance from [14].

Based on Figures 1 and 2, we may make some observations about the human auditory pathway:

- The auditory pathway contains many different representations of sounds, at many different levels. The most fundamental representation is the cochlea representation carried on the auditory nerve, from which all other representations are derived. Any realistic computational model of the human hearing system will have to generate all of the representations and allow them to interact realistically, thus extracting and utilizing all of the information in the auditory signals.

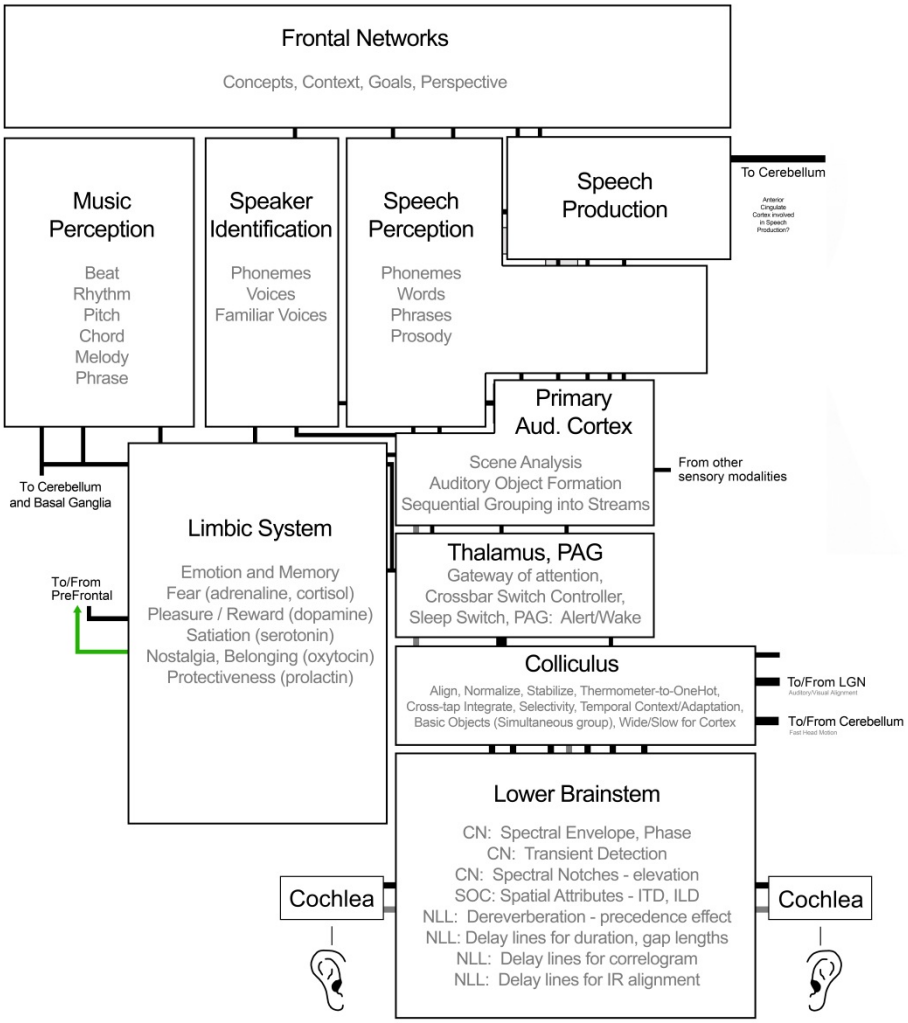

Fig. 1. Block diagram of the Human Auditory Pathway (high-level). Sounds enter the system through the two ears at the bottom of the diagram. The left and right cochleas create the spectro-temporal representation of sounds, which projects onto the cochlear nerve into the lower brainstem, beginning with the cochlear nucleus (CN), then projects to the superior olivary complex (SOC) and nucleus of the lateral lemniscus (NLL). From there, signals project to the inferior and superior Colliculus and Thalamus. The thalamus projects to the Limbic system (emotion and memory) and to primary auditory cortex, which then projects to the specialized pathways for speech recognition, production, speaker identification, and music perception.

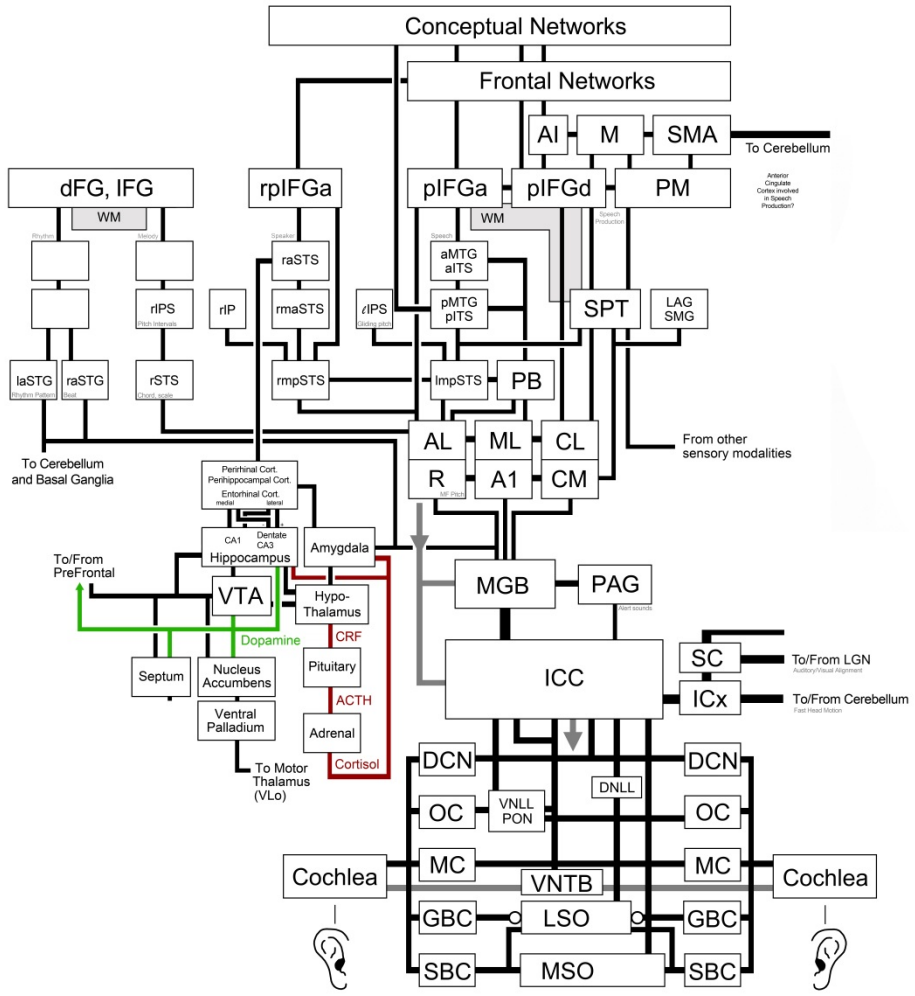

Fig. 2. Block diagram of the Human Auditory Pathway (detail). Original sources in [1][7][8][9][10][11][12]. Please see the Glossary of Terms for full list of abbreviations.

- It would appear that the functions of auditory scene analysis (breaking a mixture of sound up into its constituent sources [15]) must occur within the regions of inferior colliculus, thalamus (MGB), and primary auditory cortex, for the following reasons:
    - Below inferior colliculus, the lower brainstem is only extracting basic attributes of sound; it is too early in the system to have completed auditory scene analysis.

  - Above primary auditory cortex, we see regions specialized for deep analysis of isolated sounds (i.e. speech recognition, speaker identification, music perception). Thus, above auditory cortex, it appears that the auditory scene analysis has been largely completed.

- Thalamus (medial geniculate body (MGB)) functions largely as a wide, controllable cross-bar switch, to allow signals to be routed to cortex (selective attention) or cut off (not paying attention, or during sleep) [16]. However, some signals are capable of waking us up from sleep (i.e. baby cry), suggesting that some rudimentary signal classification is being done below the Thalamus, apparently in the inferior colliculus and peri-acqueductal gray (PAG) [17].
- The cortical speech recognition part of the human auditory pathway includes a phonological network (lmpSTS), lexical network (pMTG/pITS), and combinatorial network (aMTG/aITS) [7]. These elements are roughly analogous to the phoneme classifier, word recognizer, and language model of a conventional speech recognizer. However, as emphasized in [1], conventional modern speech recognizers do not include an auditory scene analysis engine to separate sounds in a mixture into their constituent sources prior to recognition. Instead, a conventional speech recognizer performs the front-end (Fast Fourier Transform and cepstrum) and projects them immediately to the back-end, which can only work well when the input signal is already isolated speech. The lack of an auditory scene analysis engine is the primary reason that modern speech recognizers exhibit poor noise robustness relative to human listeners, especially when the background noise consists of competing speech.
- There is a notably parallel structure between the speech recognition pathway [7] and the speaker identification pathway [8] – note that each has three major stages between primary auditory cortex and inferior frontal gyrus (IFG).
- Finally, the new block diagrams in Figures 1 and 2 indicate some important interactions between the auditory pathway and other parts of the brain. On the right side of both diagrams, there are additional connections:
  - To/From Cerebellum (at bottom right, from ICx): This connection is to trigger reflexive head movement in response to directional sound.
  - To/From LGN (at lower right, from SC): This is a bidirectional connection to allow a calibration and spatial alignment between the visual system and auditory system [18].
  - From other sensory modalities (middle right, to SPT (sylvian parietal-temporal junction)): This is the pathway by which lip-reading can assist the speech recognition pathway in the correct perception of spoken phoneme-level sounds [7], especially in noise where the auditory input may be corrupted [14].
  - To Cerebellum (upper right, from SMA): This is the motor output for speech production.

These four external interfaces indicate that the auditory pathway does not act in isolation – it interacts with the visual and motor pathways to create a whole-brain system that can hear, see, move, and talk.

2 Compute Capacity in 2012 Is Capable of Comprehensive Simulation and Visualization of the Multi-representation System

In early 2012, high-end gaming notebook computers have 2.0 GHz microprocessors with 8 virtual cores, about 11.4 times the compute capability of the 1.4 GHz single-core machines of 2003. In 2003, it took the entire machine to compute any one of the basic brainstem representations of sound, by itself. In 2012, it is possible to compute all of the representations simultaneously, including new ones which had not been developed in 2003. In 2003, the highest resolution display on a notebook computer was XGA (1024x768 pixels), which was only enough to display a single representation at once. In 2012, with a 1080p HD display (1920x1080 pixels), it is possible to compute and display all of the existing representations simultaneously, as shown in Figure 4.

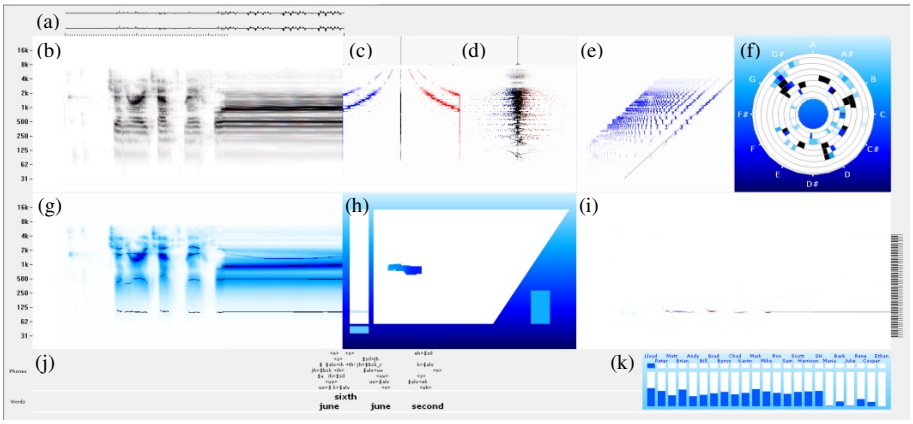

Fig. 3. Output of real-time, high-resolution functioning model of major auditory pathway elements. (a) Waveforms, at the level of the eardrums at the two ears. (b) Cochlea energy, as seen by the Multipolar Cells in the Cochlear Nucleus. (c) Inter-aural time difference (ITD), as computed by the medial superior olive (MSO). (d) Inter-aural level difference (ILD), as computed by the lateral superior olive (LSO) and normalized in the inferior colliculus (IC). (e) Correlogram, as computed in the nucleus of the lateral lemniscus (NLL) and inferior colliculus (IC). (f) Pitch Chroma Spiral (cortical pitch representation). (g) Pitch-adaptive spectral smoothing, with formant tracking (cortical speech representation). (h) Vocal Articulator mapping, in the sylvian parietal-temporal junction (SPT). (i) Polyphonic pitch. (j) Speech recognition. (i) Speaker identification.

There is still much to be done – in particular, the highest-level recognition functions (speech recognition, speaker ID) currently implemented are introductory placeholders based on fairly basic technologies. And currently, the representations are running simultaneously, but they are not yet interacting with each other. The true promise of integrating all of the representations together so that they can help each other is still to be done. But it is clear that we have sufficient neuroscience knowledge of a powerfully multi-representation system, and a sufficiently capable computing platform to be able to build the next level of the integrated system and visualize its output.

3 Next Steps in Neuroscience Research for the Next Decade 2010-2020

Neuroscientists are now beginning to explore the interactions between the scene analysis, speaker tracking, and speech recognition functions. One excellent example of this is the recent work by Dr. Eddie Chang at the University of California at San Francisco, in which subjects are asked to listen to a mixture of commands spoken by two different speakers (one male, one female), pick out a keyword spoken by one of them, and report the following command by the correct speaker [6], as shown in Figure 5.

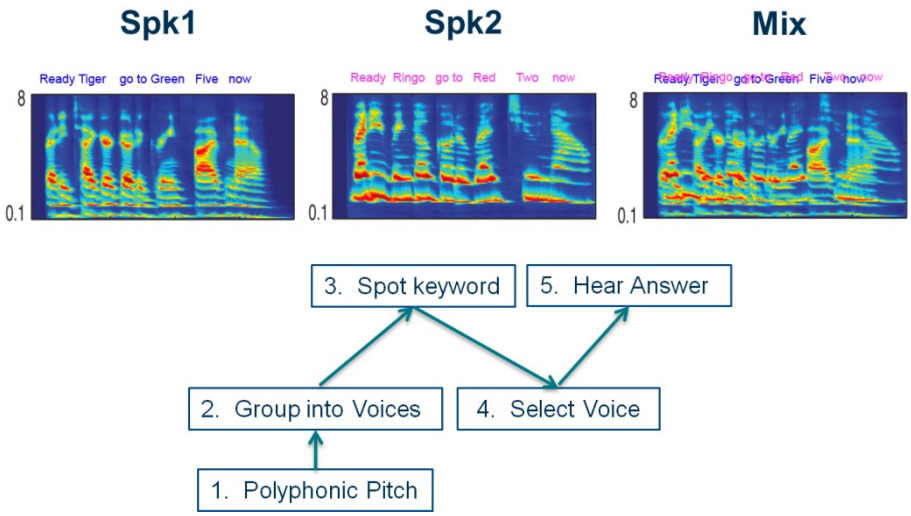

Fig. 4. Dr. Eddie Chang's task can be understood in the context of the whole auditory pathway. For the subjects to get the correct answer, they must separate the voices, presumably on the basis of polyphonic pitch, since the subjects are unable to reliably perform the task if there is not a clear pitch difference. Then they must spot the keyword, then track the voice that spoke the keyword, and then listen for the command in the chosen voice while ignoring the other voice, all while under time pressure.

Dr. Chang's task exercises the major elements of the auditory pathway – polyphonic pitch detection, grouping and separation into voices, word spotting, selective attention to the correct voice, and listening for the correct answer. And he is able to make direct multi-electrode recordings from the relevant brain regions of awake functioning human beings – his neurosurgery patients who have volunteered to participate in his study. This is a major recent advancement in auditory neuroscience, already shedding light on the detailed mechanisms of auditory attention, stream separation, and speech recognition, with much promise over the next decade 2010-2020.

While the architectural advances from the last decade's fMRI studies are very important and encouraging, a notable foundational weakness still remains: what is the general computational and learning strategy of the cortical substrate? In 2012, it is safe to say that there is no clear consensus, although there are many sophisticated models with persuasive proponents [19][20][21][22], including Hierarchical Bayesian Models [23], Hierarchical Temporal Memories [24], and Deep Belief Networks [25]. From my own work on modeling the human auditory pathway, it is apparent that the cortex must be capable of at least the following set of functions:

Table 1. Functions performed in cortex

Cortical Capability	Example
Finding patterns in sensory input	Recognizing sounds of speech
Recognizing temporal sequences	Recognizing speech and music
Memory Storage and Retrieval Creating new memories	Remembering and recalling a fact
Adding attributes to existing memories	Learning new meaning of a word
Associative Memory, Relational Database	Recalling a person by their voice, recalling all people with a similar voice
Organizing short-term and long-term memory (with hippocampus)	Memory updates during sleep
Learning	Learning a language or a song
Searching large spaces while maintaining multiple hypotheses	Understanding a sentence in which the last word changes the expected meaning. Getting a joke. Viterbi search in a modern speech recognizer.
Playing back sequences	Playing music, singing, speaking well-known phrases.
Predicting future, detecting prediction errors, re-evaluating assumptions	Motor control, getting a joke.
Tracking multiple moving targets	Polyphonic pitch perception
Separating multiple objects	Auditory Stream separation
Making decisions about what to pay attention to	Listening in a cocktail party
Local cross-correlations	Stereo Disparity in the visual system

Note that there are computer programs that can do each of the above things, *in isolation*, at some level of ability. For example, music sequencers can play back long and complicated sequences of musical notes. The Acoustic Model part of a modern speech recognizer has been trained to estimate the likelihood of phonemes, given speech input. Back-propagation and Deep Belief Networks are examples of programs

that learn. Google's web crawl and hash table updates are examples of organizing associative memories for fast recall. Creating new memories and adding new attributes to existing memories are routine operations on linked lists. Stereo disparity algorithms have been around since the early 1990's [26].

In principle, I see nothing in the brain that could not be implemented on a sufficiently fast computer with enough memory, although matching the low power consumption of the brain will favor a parallel/slow architecture over the conventional fast/serial architecture. It is common to regard the cortical columns as basic units of computation [19][21][22][24], and in principle, I see no reason why these columns (or groups of columns) could not be reasonably modeled by a sufficiently capable microprocessor running a suitable program, provided the microprocessors can communicate adequately with each other. But the key question is:

In such a model, should each cortical processor be running the same program?

I believe the answer is *No*. The highly differentiated functions performed in the different cortical regions shown in Figure 2 and listed in Table 3, suggest that, while the cortical structure (hardware) may be quite uniform across the cortex, the functions performed in the mature brain in each region (software) must be quite specialized for each region. For example, the functions of stream separation performed in auditory cortex are extremely different than the functions of phoneme recognition performed in the left medial posterior Superior Temporal Sulcus (lmpSTS), which in turn are extremely different from the functions of working memory for extracting the meaning of sentences in and near the posterior Inferior Frontal Gyrus (pIFG). And all of these are fundamentally different from the functions of controlling movement in motor cortex or computing the cross-correlations for determining stereo disparity in visual cortex.

It is not clear whether the functional specialization in the mature cortex is the result of a uniform cortical structure in which different regions learn their specialized function solely because of their unique inputs (i.e., wiring determines function), or if there is some other additional way that the specialized functions in each region are determined during development – perhaps genetic [27][28][29]. For example, recent evidence from 2001-2009 points to mutations in the FOXP2 gene as causing severe speech and language disorders [30][31][32][33][34], including defects in processing words according to grammatical rules, understanding of more complex sentence structure such as sentences with embedded relative clauses, and inability to form intelligible speech [35].

I am emphasizing this point because the observation that the cellular structure of cortex appears uniform has led to a widely accepted hypothesis that there must be a single learning or computational strategy that will describe the development and operation of all of cortex. For this hypothesis to be true, the learning or computational strategy would have to be capable of developing, from a generic substrate, a wide variety of very different functional specialties, including functions well-modeled as correlators, hierarchical temporal memories, deep belief networks, associative memories, relational databases, pitch-adaptive formant trackers, object trackers, stream separators, phoneme detectors, Viterbi search engines, playback sequencers, etc.

In any case, so far, to even come close to matching the functions that are observed by my neuroscience collaborators working in mature brains, I have found it necessary to write very specialized programs to model each functional area of the mature brain.

4 Non-technical Issues: Collaboration and Funding for 2010-2020

In 2003 [1] and 2007 [2], I outlined the importance of collaboration with leading neuroscientists, and of finding a funding model that would sustain the multi-decade project of reverse-engineering the brain, beginning with the auditory pathway. The basic science work and early prototypes were done in 1998-2000 at Interval Research, and in 2000-2003 in the early days of Audience. From 2004-2010, the focus was on building a viable business to commercialize the practical applications of research into the auditory pathway. In 2010-2011, we revisited the neuroscience community and found substantial progress had been made in the cortical architecture of the auditory pathway, and Moore's Law has ensured that compute capacity has grown ten-fold as expected. It remains to be seen what new insights and products will emerge from the next phase of scientific exploration over the next few years, but we can at least say that after the first decade, the neuroscience, compute capacity and funding aspects of the project have all advanced in sync with each other, as hoped in [1], and the multi-decade project is still on track.

5 Conclusions

The goal of reverse-engineering the human brain, starting with the auditory pathway, requires three essential ingredients: Neuroscience knowledge, a sufficiently capable computing platform, and a long-term funding source to sustain a multi-decade project. All of these were available on a small scale at the time of founding Audience in 2000, enough to begin the project in earnest. By 2010, neuroscience knowledge had advanced dramatically, giving major insights into cortical architecture and function, compute capacity had grown ten-fold, and a commercial foundation had been established to allow the project to continue into the next decade 2010-2020. While there is still much work to do, and many risks remain, the multi-decade project still appears to be on track.

References

1. Watts, L.: Visualizing Complexity in the Brain. In: Fogel, D., Robinson, C. (eds.) Computational Intelligence: The Experts Speak, pp. 45–56. IEEE Press/Wiley (2003)
2. Watts, L.: Commercializing Audi-tory Neuroscience. In: Frontiers of Engineering: Reports on Leading-Edge Engineering from the 2006 Symposium, pp. 5–14. National Academy of Engineering (2007)
3. Watts, L.: Advanced Noise Reduction for Mobile Telephony. IEEE Computer 41(8), 90–92 (2008)
4. Watts, L., Massie, D., Sansano, A., Huey, J.: Voice Processors Based on the Human Hearing System. IEEE Micro, 54–61 (March/April 2009)
5. Mesgarani, N., David, S.V., Fritz, J.B., Shamma, S.A.: Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. J. Neurophysiol. 102(6), 3329–3339 (2009)

6. Mesgarani, N., Chang, E.: Robust cortical representation of attended speaker in multitalker speech perception. Submitted to Nature (2011)
7. Hickok, G., Poeppel, D.: The cortical organization of speech processing. Nature Reviews Neuroscience 8(5), 393–402 (2007)
8. von Kriegstein, K., Giraud, A.L.: Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NeuroImage 22, 948–955 (2004)
9. Peretz, I., Zatorre, R.: Brain organization for music processing. Annual Review of Psychology 56, 89–114 (2005)
10. Levitin, D.: This is Your Brain on Music. Dutton Adult (2006)
11. Andersen, P., Morris, R., Amaral, D., Bliss, T., O'Keefe, J.: The Hippocampus Book. Oxford University Press (2007)
12. LeDoux, J.: The Emotional Brain. Simon & Schuster (1998)
13. Whalen, P., Phelps, E.: The Human Amygdala. The Guilford Press (2009)
14. Hervais-Adelman, A.: Personal communication (2011)
15. Bregman, A.: Auditory Scene Analysis. MIT Press (1994)
16. http://en.wikipedia.org/wiki/Thalamus
17. Parsons, C., Young, K., Joensson, M., Brattico, E., Hyam, J., Stein, A., Green, A., Aziz, T., Kringelbach, M.: Ready for action: A role for the brainstem in responding to infant vocalisations. Society For Neurosciences, Poster WW23 299.03 (2011)
18. Hyde, P., Knudsen, E.: Topographic projection from the optic tectum to the auditory space map in the inferior colliculus of the barn owl. J. Comp. Neurol. 421(2), 146–160 (2000)
19. Calvin, W.: The Cerebral Code. MIT Press (1998)
20. Douglas, R., Martin, K.: In: Shepherd, G. (ed.) The Synaptic Organization of the Brain, 4th edn., pp. 459–510. Oxford University Press (1998)
21. Mountcastle, V.B.: Introduction to the special issue on computation in cortical columns. Cerebral Cortex 13(1), 2–4 (2003)
22. Dean, T.: A computational model of the cerebral cortex. In: The Proceedings of Twentieth National Conference on Artificial Intelligence (AAAI 2005), pp. 938–943. MIT Press, Cambridge (2005)
23. George, D., Hawkins, J.: A Hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex. In: Proceedings of the International Joint Conference on Neural Networks (2005)
24. Hawkins, J., Blakeslee, S.: On Intelligence. St. Martin's Griffin (2005)
25. Hinton, G.E., Osindero, S., Teh, Y.: A fast learning algorithm for deep belief nets. Neural Computation 18, 1527–1554 (2006)
26. Okutomi, M., Kanade, T.: A Multiple-Baseline Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence 15(4), 353–363 (1993)
27. Dawson, G., Webb, S., Wijsman, E., Schellenberg, G., Estes, A., Munson, J., Faja, S.: Neurocognitive and electrophysiological evidence of altered face processing in parents of children with autism: implications for a model of abnormal development of social brain circuitry in autism. Dev. Psychopathol. 17(3), 679–697 (2005), http://www.ncbi.nlm.nih.gov/pubmed?term=%22Dawson%20G%22%5BAuthor%5D
28. Dubois, J., Benders, M., Cachia, A., Lazeyras, F., Leuchter, R., Sizonenko, S., Borradori-Tolsa, C., Mangin, J., Hu, P.S.: Mapping the Early Cortical Folding Process in the Preterm Newborn Brain. Cerebral Cortex 18, 1444–1454 (2008)
29. Kanwisher, N.: Functional specificity in the human brain: A window into the functional architecture of the mind. Proc. Natl. Acad. Sci. USA (2010)
30. Lai, C., Fisher, S., Hurst, J., Vargha-Khadem, F., Monaco, A.: A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413(6855), 519–523 (2001)

31. MacDermot, K., Bonora, E., Sykes, N., Coupe, A., Lai, C., Vernes, S., Vargha-Khadem, F., McKenzie, F., Smith, R., Monaco, A., Fisher, S.: Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am. J. Hum. Genet. 76(6), 1074–1080 (2005)
32. http://www.nytimes.com/2009/11/12/science/12gene.html
33. Konopka, G., Bomar, J., Winden, K., Coppola, G., Jonsson, Z., Gao, F., Peng, S., Preuss, T., Wohlschlegel, J., Geschwind, D.: Human-specific transcriptional regulation of CNS devel-opment genes by FOXP2. Nature 462, 213–217 (2009)
34. http://www.evolutionpages.com/FOXP2_language.htm
35. Vargha-Khadem, et al.: Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. Proc. Nat. Acad. Sci. USA 92, 930–933 (1995)

Glossary of Terms

Abbreviation	Full Name	Function
SBC	Spherical Bushy Cell	Sharpen timing, phase locking for ITD comparison
GBC	Globular Bushy Cell	Condition for ILD amplitude comparison
MC	Multipolar Cell	Detect amplitude independent of phase
OC	Octopus Cell	Broadband transient detection
DCN	Dorsal Cochlear Nucleus	Elevation processing
MSO	Medial Superior Olive	ITD comparison
LSO	Lateral Superior Olive	ILD comparison
VNTB	Ventral Nucleus of the Trapezoid Body	Control efferent signals to cochlea OHCs (top-down gain control loop)
MNTB	Medial Nucleus of the Trapezoid Body	Inverter between GBC and LSO to allow amplitude subtraction operation
VNLL	Ventral Nucleus of the Lateral Lemniscus	Prepare for broad system-wide reset in ICC (triggered temporal integration?)
PON	Peri-Olivary Nuclei	
DNLL	Dorsal Nucleus of the Lateral Lemniscus	Precedence effect processing of spatial information, compensate for reverberation
ICC	Inferior Colliculus (Central)	Scaling, normalizing (L-R)/(L+R), align data structure, selectivity
ICx	Inferior Colliculus (Exterior)	Audio visual alignment
SC	Superior Colliculus	Audio visual alignment
MGB	Medial Geniculate Body (Thalamus)	Attentional relay, sleep switch
PAG	Peri-acqueductal Gray	Wake from sleep from sounds like baby cry
LS	Limbic System (includes Amygdala, Hippocampus, hypothalamus, Pituitary gland, adrenal gland)	Fast-acting fear pathway, memory controller, hash table generator
A1	Primary Auditory Cortex	Primary area of Auditory cortex
R	Rostral part of Auditory Cortex	
CM	Caudal Medial part of AC	
AL	Anterior Lateral part of AC	Extraction of spectral shape – pitch-adaptive spectral smoothing, or preparations for it
ML	Medial Lateral part of AC	
CL	Caudal Lateral part of AC	
STS	Superior Temporal Sulcus	Phonological network (phonemes, speech components). Possible site of pitch-adaptive spectral smoothing and formant detection

PB	ParaBelt region	Pitch, noise
pMTG	Posterior Medial Temporal Gyrus	Lexical network (words, vocabulary, HMM)
pITS	Posterior Inferior Temporal Sulcus	Lexical network (words, vocabulary, HMM)
aMTG, aITS	Anterior Medial Temporal Gyrus, Anterior Inferior Temporal Sulcus	Combinatoric network (sentences, grammar, HMM)
SPT	Sylvian Parietal-Temporal junction	Sensori-motor interface
LAG, SMG	Left Angular Gyrus Super Modular Gyrus	Activated in degraded/challenging speech conditions
rmpSTS	Right medial posterior Superior Temporal Sulcus	Voice recognition
rmaSTS	Right medial anterior Superior Temporal Sulcus	Non-familiar voices
raSTS	Right anterior Superior Temporal Sulcus	Familiar voices
IP	Inferior Parietal	
pIFGa	Posterior Inferior Frontal Gyrus (anterior part)	Syntax and Semantics in speech comprehension, working memory for speech
pIFGd	Posterior Inferior Frontal Gyrus (dorsal part)	Phonemes in speech production
PM	Pre-Motor Cortex	
AI	Anterior Insula	Modulation of speech production (disgust)
M	Motor Cortex	
SMA	Supplemental Motor Area	Interface between Motor Cortex and Cerebellum, subvocalization, rhythm perception and production
rSTS	Right Superior Temporal Sulcus	Chord and scale in music
rIPS	Right Inferior Parietal Sulcus	Pitch intervals in music
lIPS	Left Inferior Parietal Sulcus	Gliding pitch in speech
raSTG	Right anterior Superior Temporal Gyrus	Beat in music
laSTG	Left anterior Superior Temporal Gyrus	Rhythm pattern in music
dFG, IFG	Dorsal Frontal Gyrus, Inferior Frontal Gyrus	Working memory for pitch, tones harmonic expectations/violations
	Cerebellum, basal ganglia	Auditory intervals (lateral cerebellum, basal ganglia), Motor timing (medial cerebellum, basal ganglia)