Chapter 10:

Getting involved

“My work in the Tribe program has really benefited me and has opened up my eyes to [voice] research and working with a startup.”

Alice Romanov,
(Tribe 1 Innovation Fellow, NeuroLex)

Chapter Overview

By now you know a lot about voice computing. But you may be looking for some practical experience. *It's hard to know where to start, especially when making a career shift.*

This chapter is about how to become more engaged in the voice developer community. Specifically, we'll cover:

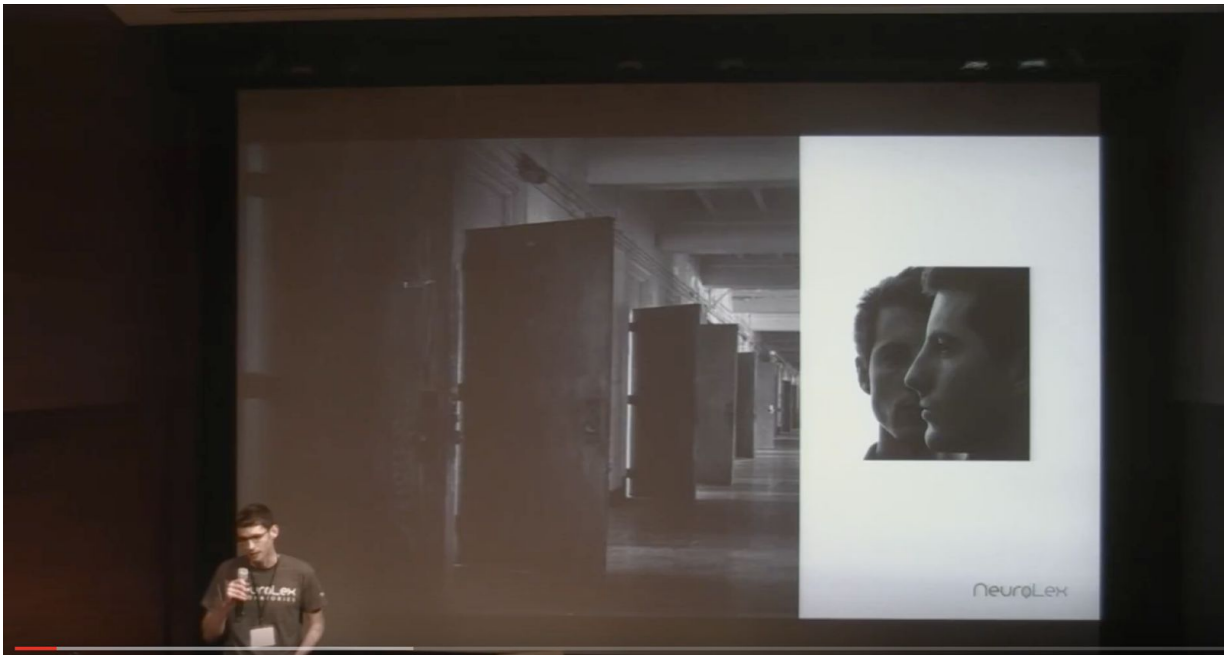
- 10.1 - NeuroLex's story
- 10.2 - The Innovation Fellows Program
- 10.3 - Building open source software
- 10.4 - Conferences
- 10.5 - Graduate schools
- 10.6 - Finding a job
- 10.7 - Launching a startup
- 10.8 - Teaching voice computing

In this way, you can help the voice community continue to grow and thrive!

10.1 NeuroLex's story

Eight years ago I got a call from my crying mother saying that my brother was in the mental hospital. He had a severe psychotic episode. He was not the person who I knew him to be (Figure 10.1.1).

Figure 10.1.1
YouTube video of NeuroLex's story: *Voicecamp demo day*



You can watch a video here.

Looking back, it was obvious something was off. He went to the primary care doctor 11 times complaining of the same vague symptoms - headaches, unclear thoughts, and anxiety - but never really got properly diagnosed. He even went to 6 different specialists, including a trained psychiatrist, and was misdiagnosed with an anxiety disorder. All this effort cost roughly $15,000 - all wasted money leading up to this hospitalization.[1]

The medical system failed my brother and is failing many other patients afflicted by schizophrenia. Many of these patients have family members abandon them and are left to be homeless on city streets.[2] I've reacted by embracing a deep empathy for such patients and have used this experience as motivation to improve, as opposed to react negatively, to the medical

[1] **There is a 80% relapse rate** within 2 years in patients with schizophrenia.
[2] **554,000 people** are homeless in the USA and there is an 80% relapse rate within 2 years within a first psychotic episode for schizophrenia patients.

system. As a biomedical engineer, this led me to lines of thought about how to better diagnose and manage the disease.

Particularly, there didn't seem to be a good way to diagnose and track schizophrenia,[3] so I sought some alternative ways to characterize his symptoms. One thing that seemed interesting was to look at his voicemails. It seemed like over this 5 year period voicemails that he left on my phone leading up to his first episode seemed more disordered, as he used new words and wasn't making much sense.[4] Therefore, I wondered if there were some features[5] in the transcript from voicemails that could be used to reliably detect when my brother was progressing towards or away from a psychotic episode.[6]

Then, I saw a paper was published in NPJ Schizophrenia that showed linguistic biomarkers could predict psychosis onset in high-risk youths.[7] In other words, you could predict who would or would not develop a psychotic episode from answering a very simple question like "how is your day today?".[8] This seemed quite remarkable and showed that some of the early ideas I had around early screening for psychosis were extensible to larger numbers of patients.[9]

A few weeks later I flew to New York City to meet with the authors of this study to see what could be done to translate this work to patients. As soon as I met with the authors, I felt as if my life calling were to commercialize this form of research. When I went there, I found out that voice biomarkers could characterize quite a few psychiatric and neurological conditions (Table 10.1.1).

Table 10.1.1
Summary of voice biomarker studies

Disease	Description
Early-stage psychosis	100% accuracy to detect who will convert using linguistic features (frequency use of determiners, maximum phrase length, and the first-order semantic coherence) and convex hull classifier with leave-one-out analysis. This was a proof-of-concept study with 35

[3] **Standard biomarkers do not work -** Blood, urine, and MRI images all were unreliable at diagnosing the disease.
[4] **Word salad -** The odd use of language associated with schizophrenia is known as 'word salad.'
[5] **Particularly features** that were a measure of disorder like lexical complexity and/or semantic coherence
[6] **It looked like** some of these features fed into anomaly detection methods (e.g. SVM models) could reliably predict when my brother was exhibiting more severe symptoms.
[7] **Features predicting psychosis** - the semantic coherence, the frequency use of determiners (e.g. 'this','that','whatever'), and the mean phrase length (e.g. length of each sentence)
[8] **After I saw this,** I immediately called Guillermo and flew to NYC to start NeuroLex to commercialize the technology - particularly, to build a universal voice test to screen for psychiatric and neurological symptoms early in primary care clinics much like Quest Diagnostics but for speech tests in the cloud.
[9] **Sample size** - the original sample size was very low (in the 30s), but now it's been tested on over 100 patients. The algorithm is roughly 80% accurate as measured by ROC curves.

	patients but has been followed on with 100+ patients at various sites (in talking with our collaborators).
Depression	a supervised learning support vector machine (SVM) approach was used to detect suicidal with mental illness, mental illness and not suicidal, and control (no mental illness) with 85% accuracy (N=379 patients).
Alzheimer's disease	The classification accuracy of automatic audio analyses were as follows: between HCs and those with MCI, 79% 6 5%; between HCs and those with AD, 87% 6 3%; and between those with MCI and those with AD, 80% 6 5%, demonstrating its assessment utility. Pause lengths in speech are most relevant here.
Parkinson's disease	Can classify PD patients or controls with 75% accuracy and could infer PD patients' level of motor impairment with 77% accuracy from semantic fields (via latent semantic analysis), grammatical choices (using part-of-speech tagging), and word-level repetitions (with graph embedding tools).

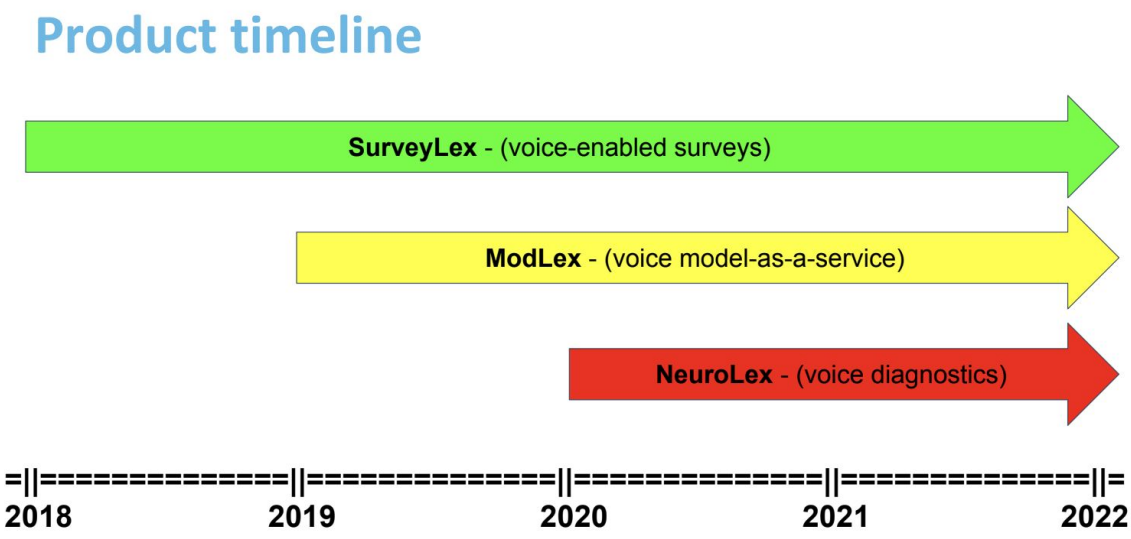
This led to our current company - NeuroLex - which aims to commercialize a universal voice test to refer patients to specialists faster; like a Quest Diagnostics but for speech tests in the cloud. In this way, patients like my brother could be diagnosed and treated earlier, leading to better outcomes (e.g. lower duration of untreated psychosis) and lower costs ($15k→ $3-5k).

However, since we started we've learned that healthcare is hard - *really hard.* It takes years of research to get new codes issues to reimburse new tests. We built one of the largest voice computing labs in the world. Headquartered at the University of Washington under our Chief Medical Officer (Reza Hosseini Ghomi), we have grown our lab to include over 80 research collaborators across the world. Despite this, we're still 3-5 years from launching a commercial healthcare product due to the evidence that health plans need to reimburse our tests (in terms of prospective studies / cost savings).

Meanwhile, we've found that one of the hardest things when we launched our research studies was that it was quite difficult to uniformly record audio across many clinical trial sites. This led to SurveyLex, a product where users speak, as opposed to type or click, their responses. Using machine learning, we can auto-label responses with emotions (e.g. happy/sad), age, gender, accents, and even race. In this way we can dramatically reduce the number of questions on surveys up to 500% and decrease dropout rates by up to 50%. In doing so, we create a better experience for the survey maker and the survey taker.

We still haven't given up on our healthcare journey. We see that fit into the same architecture that we've built going into years 3-5 on our roadmap (Figure 10.1.2). We'll get there eventually (~2021), just through non-traditional means.

Figure 10.1.2
NeuroLex product timeline



So why do I tell this story?

You may have an entrepreneurial itch. If you do, I highly encourage you to take the journey. *Nothing is more rewarding than building the future you want to live in and scaling that into the world.*

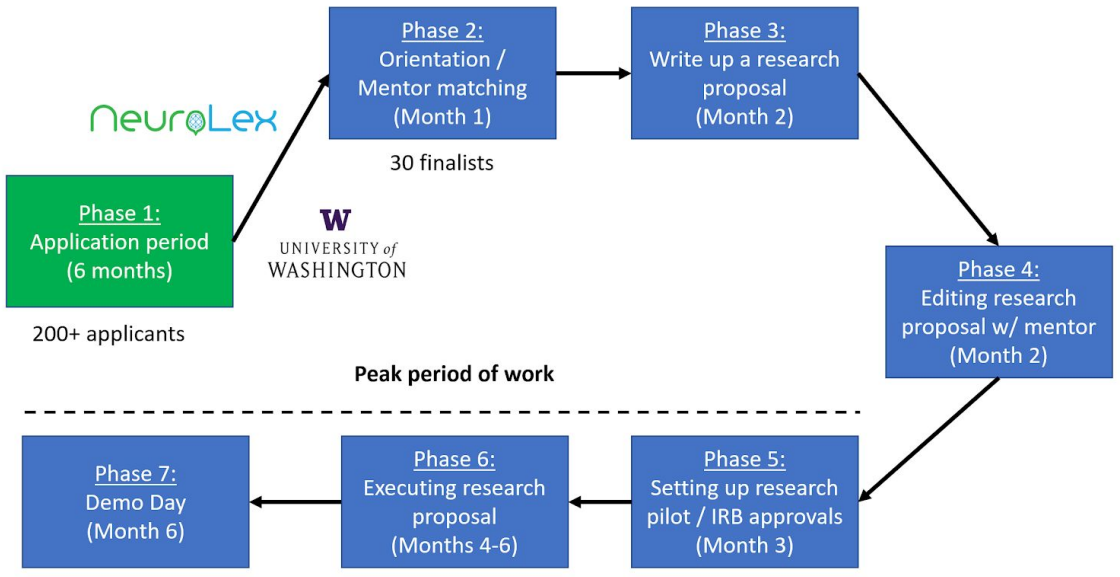
10.2 The Innovation Fellows Program

After telling NeuroLex's story many times, I kept running into many outstanding individuals that wanted to be involved in our vision. For a long time, I didn't really know how best to use our company to enable these individuals, as we obviously had limited resources as a startup company.[10] However, one day I woke up and wondered - "what if we could apply the Wikipedia model to startups, allowing for anyone to come into a company and contribute? What if someone could propose demo project - like proposing a new page or edit an existing page on Wikipedia?"

[10] Specifically, we had only raised $200,000 and had 10 part-time individuals contributing code to our architecture.

We thus invented the TRIBE model to engage outstanding individuals with our company in the form of a competition. These outstanding individuals apply here, and, if selected, pick a category to focus their demos (e.g. research, data science, or software). Fellows then are paired with mentors and propose and defend demo projects over the span of 6 months (Figure 10.2.1). The program culminates in a demo day where fellows present their work virtually to our team (at an internal corporate retreat).[11]

Figure 10.2.1
NeuroLex Innovation Fellows program timeline



So far, the program has been incredibly successful. We have received over >200 applicants and over 60 fellows have gone through the program and have had over 30 completed demo projects. To learn more, you can read this FAQ document. You can also apply to the program here.[12]

10.3 Building open source software

Another way that you can be involved is to contribute to the open source voice community. There are three main ways that you can contribute here:

1. Publish new datasets
2. Contribute to an existing open source project
3. Create a new open source project

[11] The winners (1st/2nd place) receive a small cash prize along and other perks (e.g. T-shirts).
[12] Contact me @ js@neurolex.co with any questions.

First, you can publish new datasets. Although large companies have begun open sourcing their data (e.g. Audioset / Google Research).[13] Here is a list of open source projects that have released audio data openly (Table 10.3.1). By contributing data to these projects and/or making new datasets you can greatly help speed up voice research initiatives.[14] [15] *NeuroLex is actively expanding many of the disease detection models through the train_diseases repository on GitHub. If this is of interest, let us know.*

Table 10.3.1
List of open source voice datasets

Dataset (link)	Description
Common Voice Dataset	150k recordings, open sourced by Mozilla Foundation. Useful for building models on age, accent, and gender.
Google Audioset	2 million recordings over 600+ classes of audio (contains over 1 million speech samples). Some example classes include: music, speech, vehicle, and musical instrument. Explore additional classes here.
Urban Sound dataset	1302 files of different urban related sounds. Specifically, each file is labeled with air_conditioner, car_horn, children_playing, dog_bark, drilling, enginge_idling, gun_shot, jackhammer, siren, and street_music.
NeuroLex Disease Dataset	This dataset contains over 500 YouTube samples labeled with addiction, adhd, als, anxiety, autism, cold, controls, depression, dyslexia, glioblastoma, graves disease, multiple_sclerosis, parkinsons, postpartum_depression, schizophrenia, sleep_apnea, and stressed. As of right now, it is the most diverse dataset for disease-related research across a range of diseases. We hope to expand it into the future.

[13] **Private datasets -** For a very long time, only a few companies have had training datasets for things like noise classification, text-to-speech models, and transcription models.
[14] **Kaggle -** Though there are multiple ways to publish your dataset, I'd highly recommend you to use Kaggle because I think it has the most robust community of data scientists to work with the data. I think you'd have much more impact there.
[15] **Project open data -** "In the United States, open and accessible data has been a standard since 2013. Three and a half years later, according to their open data dashboard, there are nearly 13,000 open datasets representing 20 US agencies. Along this journey, they have done an excellent job of outlining general principles of open data which you should expect to find anywhere open data lives. The principles, which you can find referenced at Project Open Data, include the following:

- Open
- Accessible
- Described
- Reusable
- Complete
- Timely
- Managed post-release"

NeuroLex Emotion Dataset	This dataset contains >2000 labeled emotions with 20 second voice files: happy, sad, neutral, angry, disgust, and fear. These were all extracted from YouTube videos.
Mivia audio events dataset.	6,000 recordings. Classes mostly include emergency related events like glass breaking, gunshots and screams. The 6000 events are divided into a training set (composed of 4200 events) and a test set (composed of 1800 events).
Karoldvl	2000 recordings of various animals sounds including dogs, rooster, pigs, cows, and frogs.

Next, you can contribute to an existing open source project. Here are some open source projects currently in development (Table 10.3.2). Of course, this is a very limited list. You can do your own searches on GitHub and clone repos / make branches as necessary to improve existing repositories.

Table 10.3.2
List of some open source voice projects

Library	Description
Librosa	LibROSA is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems.
Pyaudioanalysis	pyAudioAnalysis is a Python library covering a wide range of audio analysis tasks. Through pyAudioAnalysis you can: ● Extract audio features and representations (e.g. mfccs, spectrogram, chromagram) ● Classify unknown sounds ● Train, parameter tune and evaluate classifiers of audio segments ● Detect audio events and exclude silence periods from long recordings ● Perform supervised segmentation (joint segmentation - classification) ● Perform unsupervised segmentation (e.g. speaker diarization) ● Extract audio thumbnails ● Train and use audio regression models (example application: emotion recognition) ● Apply dimensionality reduction to visualize audio data and content similarities

Soundfile	SoundFile can read and write sound files. File reading/writing is supported through libsndfile, which is a free, cross-platform, open-source (LGPL) library for reading and writing many different sampled sound file formats that runs on many platforms including Windows, OS X, and Unix. It is accessed through CFFI, which is a foreign function interface for Python calling C code. CFFI is supported for CPython 2.6+, 3.x and PyPy 2.0+. SoundFile represents audio data as NumPy arrays.
Sounddevice	This Python module provides bindings for the PortAudio library and a few convenience functions to play and record NumPy arrays containing audio signals.
SpeechRecognition	Library for performing speech recognition, with support for several engines and APIs, online and offline. Speech recognition engine/API support: ● CMU Sphinx (works offline) ● Google Speech Recognition ● Google Cloud Speech API ● Wit.ai ● Microsoft Bing Voice Recognition ● Houndify API ● IBM Speech to Text ● Snowboy Hotword Detection (works offline)
PocketSphinx	Pocketsphinx is a part of the CMU Sphinx Open Source Toolkit For Speech Recognition. This package provides a python interface to CMU Sphinxbase and Pocketsphinx libraries created with SWIGand Setuptools.
SIDEKIT	SIDEKIT is an open source package for Speaker and Language recognition. The aim of SIDEKIT is to provide an educational and efficient toolkit for speaker/language recognition including the whole chain of treatment that goes from the audio data to the analysis of the system performance.

Lastly, you can even make a new open source project. Some of the coolest things you can do is to take two or more open source voice projects and put them together in a new way to add value to the voice community. The sky is the limit with new open source repos. *I encourage you to release your code under an Apache 2.0 license to have the most impact and portability; otherwise, adding more restrictions may limit the use of the code by others.*

10.4 Conferences

Another way to get involved in the voice community is to attend a conference. The most popular conference seems to be the VOICE summit, which has Amazon Alexa as a diamond sponsor.

I'm still trying to navigate through all these conferences, but it's often a good way to see all the innovation going on in the voice computing space.

Table 10.4.1 **Voice computing conferences**[16]		
Conference	**Location**	**Dates**
Smart Voice Summit	Paris	February 1–2, 2018
Conversational Interaction Conference	San Jose, California	February 5–6, 2018
Bottish	Your Computer – free!	February 22 and April 26, 2018
Superbot	San Francisco	April 3, 2018
SpeechTEK	Washington, D.C.	April 9–11, 2018
Voice & Beyond	Berlin	Spring 2018
Conversational Commerce Conference	London	May 8–9, 2018
Business of Bots	San Francisco	May 15–17, 2018
Voicecon	New York City	May 22, 2018
Connections: Connected Home Conference	San Francisco	May 22–24, 2018
Voice Summit	Newark	July 24–26, 2018
ML Conf	Atlanta, SF, NYC, Seattle	November 14th, 2018
The Voice of Healthcare Summit	Boston	August 7, 2018

[16] Taken from https://bespoken.io/blog/voicefirst-conferences/

Conversational Commerce Conference	San Francisco	September 1, 2018
VoiceCon Canada (no link)	Toronto	September 2018
All About Voice	Munich	October 12, 2018
AWS re:INVENT	Las Vegas	November 26–30, 2018
Voice Conference	Berlin	December 5–7, 2018
CES	Las Vegas	January 8–11, 2019
The Alexa Conference	Chattanooga, Tennessee	January 15–17, 2019
Lingofest	Ogden, Utah	2019

10.5 Graduate schools

It's a bit unclear to me which schools are the best for voice computing. Generally, they seem to fall in line with programs that have great graduate schools in computational linguistics. Here's a short list of 8 schools that are known for their speech processing and linguistics programs, which give strong foundations for voice computing:

1. Stanford University (MS/PhD in linguistics)
2. Carnegie Mellon University (MS/PhD in Language / Information Technology)
3. Johns Hopkins University[17] (MA, PhD)
4. Georgia Tech (MS/PhD in machine learning/signals)
5. University of Pennsylvania (MS, PhD in linguistics)[18]
6. Massachusetts Institute of Technology (MS in machine learning, PhD in linguistics)
7. University of Washington (MS, PhD in computational linguistics)
8. UC Berkeley (MS, PhD in signal processing)

In general, though, the field of voice computing seems to be emerging, with few universities actually carving out training for the field. Your best bet seems like you should pick a focus in

[17] Johns Hopkins University's Whiting School of Engineering is home to the Center for Language and Speech Processing (CLSP). The CLSP is one of the most well-known language research groups in the world

[18] Mark Liberman is the Director of the Linguistic Data Consortium, a well-known data repository for speech.

affective computing, machine learning, and/or linguistics and then tailor your research in the voice area.

If you are in grad school doing research, there are also quite a few conferences to present your research (Table 10.4.1). In general, the most important voice-related research conferences tend to be ICASSP, AVEC, and Interspeech each year.

Table 10.4.1
List of voice-related research conferences

Event date	Event name	Location	Deadeline	Camera Ready
Apr 22-27, 2018	ICASSP2018 Int'l Conf. on Acoustics	Seoul	Oct/27/2017	Feb 09, 2018
Aug 20-24, 2018	ICPR2018 24rd Int'l Conf. on Pattern Recognition	Beijing	Nov/05/2017	May 20, 2018
Sep 02-06, 2018	Interspeech2018 19th Int'l Conf. on Speech Communication and Technology	Hyderabad	Mar/23/2018	Jun 17, 2018
Oct 22-26, 2018	AVEC2018 8th Audio/Visual Emotion Challenge and Workshop	Seoul	Jun/30/2018	Aug 14, 2018
Nov 05 - Dec 08, 2018	IVA2018 18th Int'l Conf. on Intelligent Virtual Agents	Sydney	May/15/2018	Aug 00, 2018
May 12-17, 2019	ICASSP2019 Int'l Conf. on Acoustics	Brighton	Nov/26/2018	Mar 18, 2019
May 14-18, 2019	FG2019 14th IEEE Int'l Conf. on Automatic Face and Gesture Recognition	Lille	Sep/21/2018	Feb 15, 2019
Aug 04-10, 2019	ICPhS2019 9th Int'l Congress of Phonetic Sciences	Melbourne	Feb/00/2019	May 00, 2019
Sep 00-00, 2019	ACII2019 The 8th Int'l Conf. on Affective Computing and Intelligent Interaction	Cambridge	May/00/2019	Aug 00, 2019
Sep 15-19, 2019	Interspeech2019 20th Int'l Conf. on Speech Communication and Technology	Graz	Mar/00/2019	May 00, 2019

Sep 22-25, 2019	ICDAR2019 15th Int'l Conf. on Document Analysys and Recognition	Brisbane	Mar/15/2019	-
Apr 04-09, 2020	ICASSP2020 Int'l Conf. on Acoustics	Barcelona	Nov/00/2019	Mar 00, 2019
May 25-29, 2020	Eurographics 2020 The 41th annual Conf. of the European Association for Computer Graphics	Norrköping	Oct/00/2019	Feb 00, 2020
Sep 08-10, 2020	ICFHR-2020 17th Int'l Conf. on Frontiers in Handwriting Recognition	Dortmund	Mar/00/2020	Jun 00, 2020
Sep 14-18, 2020	Interspeech2020 21st Int'l Conf. on Speech Communication and Technology	Shanghai	Mar/00/2020	May 00, 2020
Apr 25-30, 2021	ICASSP2021 Int'l Conf. on Acoustics	Toronto	Nov/00/2020	Mar 00, 2020
Apr 24-28, 2022	ICASSP2022 Int'l Conf. on Acoustics	Singapore	Nov/00/2021	Mar 00, 2022

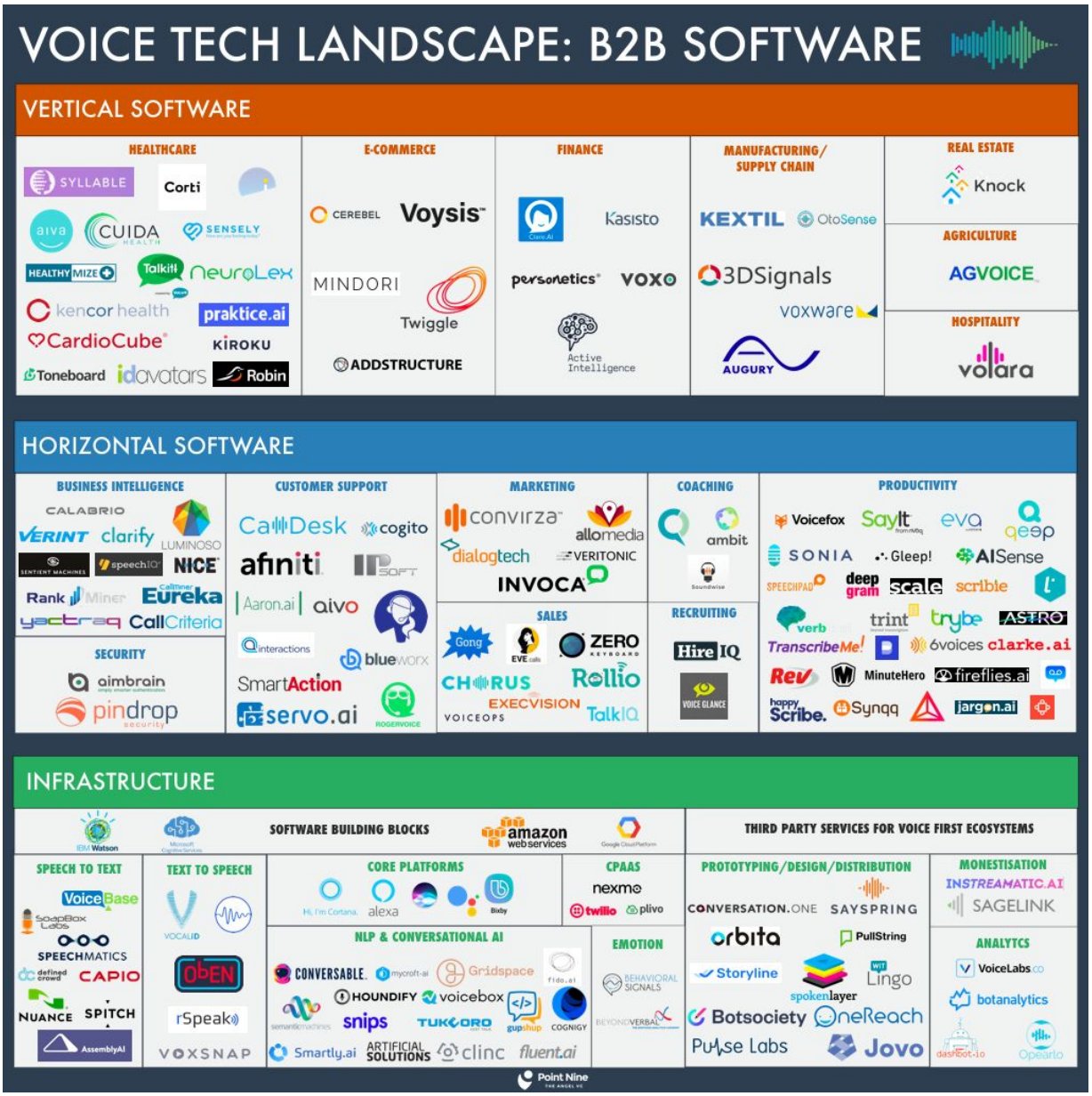
10.6 Finding a job

If you have gotten this far, perhaps you feel proficient enough to interview for an entry-level or mid-level job at a Tech company. Note that many of these companies require that you have at least a Master's (MS) or PhD degree to do cool stuff, though there are job for entry-level developers.

You can check out the voice tech landscape to help guide you applying for jobs (Figure 10.6.1). In general, voice companies are separated into ***verticals***,[19] ***horizontals***,[20] or ***infrastructure***.[21] Often, teams within large companies have different teams and divisions based on these categories. Some teams are more product-driven, others are more research driven.

[19] **Vertical categories -** healthcare, e-commerce, finance, manufacturing/supply chain, and real-estate
[20] **Horizontal categories -** business intelligence, customer support, marketing/sales, coaching, recruiting, productivity
[21] **Infrastructure categories -** cloud platforms, speech-to-text, text-to-speech, core platforms, NLP/conversational AI, CPAAS, Emotion detection, prototyping voice apps, voice data monetization, and analytics

Figure 10.6.1
The voice tech landscape



If you're looking for a more corporate job, you can check out any of the large tech companies as well. In general, the two most active companies in the voice computing space seems to be Amazon and Google because of their market dominance with smart speakers. *Boston and New York City also seem to be emerging hubs for voice computing with many of the voice and speech understanding teams being strong represented in these areas.*

- Amazon Alexa
- Facebook
- Google Assistant
- Microsoft Cortana
- IBM Watson
- Sonos
- Bose
- Twilio

10.7 Launching a startup

If you've built something cool, there is no better time to start a company. Here are some early-stage accelerator funds that can help get you off the ground, often investing roughly $100k for 7% equity (Table 10.7.1).[22]

Table 10.7.1
List of accelerators to launch your voice startup[23]

Accelerator	Description
Y-Combinator	The most well-known accelerator that has funded over 900 startups. In the request for startups YC has stated they are excited about the voice computing space.
Techstars	Various accelerators across the USA. Have a strong interest in machine learning and voice computing startups.
Betaworks	Betaworks is a great leader in the voice computing space. They were one of the first VCs to explicitly host a voice-based accelerator program called **VoiceCamp** (NYC, NY).
Alexa Accelerator	Also, the Amazon Alexa Fund (Accelerator) exists to seed voice-based startups out of Seattle, WA.

10.8 Teaching opportunities

We don't have enough teachers of voice computing! You can definitely use this book to teach others about your craft and get them up-to-speed on the technology. *I find teaching is the best*

[22] **Accelerator definition -** If you are unfamiliar with the term 'accelerator' it's just an investing group that invests in your company in exchange for equity, often through SAFE notes or convertible debt. They often help you for a constrained period of time (3-6 months) before moving onto a new batch. During the time they help you they often connect you to many mentors and customers to help speed up your growth.
[23] **Investors -** If you're looking to finance a company, I highly recommend you connect with other VCs in this space to build a thesis.

way to learn something. You always learn something new from others learning as well because they provide a fresh perspective and look on problems you are currently facing.

10.9 Conclusion

This chapter summarized how you could get involved in the voice computing community. This is the last chapter of the book; thanks for taking this journey with me. Check out any of the following links for additional information!

- NeuroLex
    - NeuroLex's story (video)
    - The Innovation Fellows Program
    - TRIBE FAQ document
    - TRIBE 2 Demo Day
- Open source software
    - Librosa
    - Pyaudioanalysis
    - Soundfile
    - Sounddevice
    - SpeechRecognition
    - PocketSphinx
    - SIDEKIT
- Conferences
    - Smart Voice Summit
    - Conversational Interaction Conference
    - Bottish
    - Superbot
    - SpeechTEK
    - Voice & Beyond
    - Conversational Commerce Conference
    - Business of Bots
    - Voicecon
    - Connections: Connected Home Conference
    - Voice Summit
    - ML Conf
    - The Voice of Healthcare Summit
    - Conversational Commerce Conference
    - All About Voice
    - Voice Conference
    - AWS re:INVENT
    - CES
    - The Alexa Conference
    - Lingofest
- Graduate schools
    - Stanford University

    - Carnegie Mellon University (Language technologies institute)
    - Johns Hopkins University (Center for Speech Language Technology)
    - Georgia Tech
    - University of Pennsylvania (Department of Linguistics)
    - Massachusetts Institute of Technology (EECS)
    - University of Washington
    - UC Berkeley (Computational Linguistics)
- Corporate jobs
    - Voice tech landscape
    - Amazon Alexa
    - Facebook
    - Google Assistant
    - Microsoft Cortana
    - IBM Watson
    - Sonos
    - Bose
    - Twilio
- Investment
    - Y-Combinator
    - Techstars
    - Betaworks Voicecamp
    - Alexa Accelerator

If I can help you with anything, let me know. I'm always around. You can reach me @ js@neurolex.co