communications, forensic screening, and real-time risk assessment in environments affected by emerging voice manipulation technologies.

2pSC5. Using machine learning modeling to analyze neural development of music processing in children with different language backgrounds. Elliot S. Blin (The Bronx High School of Sci., 2077 5th Ave. Apt. 5C, New York City, NY 10035, elliotblin2509@gmail.com), Anita Hyseni, Michael Yushavaiev, Benjamin Lee, Elizabeth Dickey (St. John's Univ., New York City, NY), and Yan H. Yu (Commun. Sci. & Disord., St. John's Univ., Bayside, NY)

Both music experience and language experience influence auditory processing in the developing brain. However, it remains unclear how bilingual experience specifically affects music processing and its underlying neural mechanisms. The primary objective of this study was to examine the neural network involved in music processing in children and to explore how music and language experience interact at the cortical level during development. Children between the ages of five and ten from Mandarin, Spanish, French, Standard American English, and African American English speaking households participated in a passive listening oddball task that included six types of musical changes, including intensity, pitch, rhythm, timbre, slide, and location. Neural responses were recorded using event-related potentials from 65 scalp sites. Brain connectivity was analyzed across frequency bands between 1 and 40 Hz. Functional connectivity measures were used as input features for machine learning classification models to distinguish children by language background. Support vector machine, random forest, and elastic net methods were applied, and model performance was compared across acoustic features. The results showed that machine learning approaches successfully captured differences in neural connectivity patterns, providing insight into how language background shapes music processing networks in the developing brain.

2pSC6. Neural networks for speech separation in infants. Ashlyn Hall (Emory Univ., 17 Forest Cir, APT 606, Newnan, GA 30265, ashlynhall1204@gmail.com) and Gordon Ramsay (Dept. of Pediatrics, Emory Univ., Atlanta, GA)

Speech separation, isolating individual vocal sources from a mixed audio signal, remains one of the hardest challenges in artificial intelligence. Despite progress through deep learning and signal processing, achieving robust separation in real-world conditions continues to fall short, particularly for infant vocalizations that may provide insights into early identification of autism spectrum disorder (ASD). Unlike machines, the human auditory system can filter relevant voices from chaotic noise using a combination of spatial cues, frequency sensitivity, and can go for contextual awareness. This is known as the "cocktail party effect," so we seek to explore new ways to model sound separation in this study to see if it can be potential use for applications in infant monitoring ASD. First is to use existing architectures such as DCCRNet, ConvTasNet, DP-RNN, and DPTNet to find which is better in separate overlapping sources and suppress background noise. These models were trained on recordings of infants from Marcus Autism Center and seeing if these models can get a high positive scale-invariant signal-to-distortion ratio (SI-SDR) of the accuracy of distinct babbling or more speech-like. With the next work on improving performance by building on the ConvTasNet architecture, which had the best results and may help reveal meaningful acoustic differences between typical and disordered speech.

2pSC7. Effect of formant tuning on the voice source and vocal fold contact pressure in a computational model of voice production. Zhaoyan Zhang (Dept. of Head and Neck Surgery, Univ. of California Los Angeles, 1000 Veteran Ave. 31-24 Rehab Ctr., Los Angeles, CA 90095, zyzhang@ucla.edu)

Source–filter interaction, in general, has small effect on vocal fold vibration and vocal fold contact pressure in the frequency range typical of speech. However, this effect may become noticeable during formant tuning when a vocal tract resonance becomes close in frequency to either the first or second harmonic of the voice source spectrum, as often in classical or music theater singing. In this study, the effect of formant tuning on the voice source was investigated in a three-dimensional computational model of voice production. Simulations were performed with parametric variations in both vocal fold and vocal tract configurations. Conditions of strong formant tuning to the first or second harmonic were identified. Acoustic, aerodynamic, and vibratory measures of the voice source and peak vocal fold contact pressure at these conditions were then compared to those at conditions of weak formant tuning. The results are discussed with implications to vocal health and voice training. [Work supported by NIH.]

2pSC8. Optimizing linear prediction order for formant analysis of speech. Joshua Sims (Linguist, Indiana Univ., 2461 S Woolery Mill Dr., Bloomington, IN 47403, jodasims@iu.edu)

Linear predictive coding (LPC) is used in phonetics to measure formant frequencies of speech, but these formant readings are prone to error. Measured formants are attracted to F0 harmonics (Klatt, 1986; Monsen and Engebretson, 1983). Furthermore, different speakers or segments may require different analysis parameters (Escudero *et al*., 2009), due in part to changes in vocal tract length owing to physiology, larynx height, and lip rounding. Common methods for minimizing error in formant readings include adjusting parameters ad hoc, discarding outliers, finding parameters which minimize variation (Escudero *et al*., 2009), produce the smoothest contours (Barreda, 2021), or realistic formant and bandwidth values (Nearey *et al*., 2002). In autoregressive methods, an LPC of order *n* is computed from a filter of order *n-1* to minimize prediction error (see Vallabha and Tuller 2002). The trend in this error over successive orders yields a heuristic for LPC order fit. In this paper, I propose that the optimal LPC order for a given frame is that which optimally reduces prediction error over its inferior order. I evaluate this proposal with formant measurements of vowels produced by Klatt synthesis (Klatt & Klatt, 1990).

2pSC9. Implementing the UCLA voice synthesizer as a jsPsych plugin for web-based voice perception research. June M. Contreras (Speech Lang. and Hearing Sci., Univ. of Texas at El Paso, 3532 Keltner, El Paso, TX 79904, jmcontreras2@miners.utep.edu)

The UCLA Voice Synthesizer serves as an important tool in voice research, offering precise control over acoustic parameters to model both normal and pathological voices. The synthesizer's architecture enables precise manipulation of acoustic parameters—including glottal source, noise, pitch, loudness, and vocal tract resonances—supporting both research and clinical applications such as gender-affirming voice therapy. However, its current implementation is constrained by limited flexibility for browser-based experimentation on online platforms such as Prolific. This abstract proposes a port of the UCLA Voice Synthesizer as a jsPsych plugin that loads the UCLA Voice Synthesizer modules compiled as WebAssembly. jsPsych's plugin framework allows researchers to design custom plugins for use in online experiments; this project compiles the original C/C++ UCLA Voice Synthesizer code to WebAssembly using Emscripten, loading it with jsPsych as a plugin. The poster will include a QR code to a web demo of the plugin for participants to experiment with.

2pSC10. Exploring (psycho)linguistically informed approaches to ASR dataset development for minoritized language varieties. Rachel Hayes-Harb (Linguist, Univ. of Utah, Salt Lake City, UT), Shannon Barrios (Kahlert School of Computing, Univ. of Utah, 255 S. Central Campus Dr., LNCO Bldg., Ste. 2300, Salt Lake City, UT 84112, s.barrios@utah.edu), Neal Patwari, and Maneesha Rani Saha (Kahlert School of Computing, Univ. of Utah, Salt Lake City, UT)

Automatic speech recognition (ASR) systems have become ubiquitous, and they provide an increasingly unavoidable means for human–computer interaction. Unfortunately, they typically perform better for speakers of historically favored languages (Sprugnoli *et al*., 2017) and accents (Markl, 2022), with particularly poor performance for so-called "non-native" accents (DiChristofano *et al*. 2024). Our goal is to mitigate the biases that permeate ASR by exploring psycholinguisticallyinformed methods for dataset construction by examining the generalizability of listener and task factors that are known to modulate transcription accuracy in laboratory studies in contexts designed to resemble crowdsourced ASR dataset development. Over several experiments we have found mixed support for the hypotheses