A Real-time Bol to MIDI Converter

/ta/   /ba/   /do/   /tee/
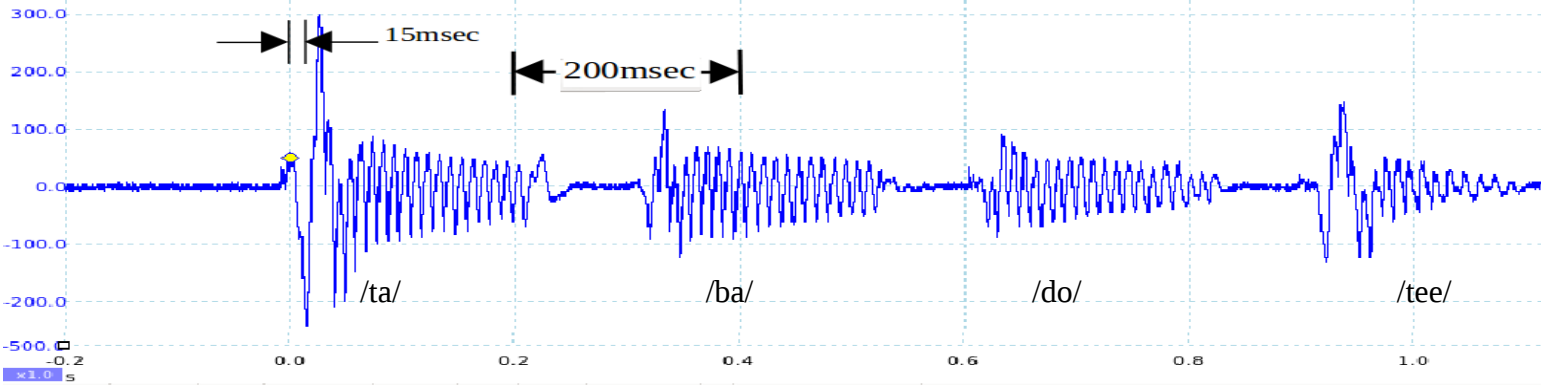
 This page is only a statement of the problem to be solved.
 The Problem: Convert vocal percussion sounds, like ta, da, & tee, into MIDI data in real-time. The vocal percussion sound set was first planned to be the collection of bols from the Indian tradition, but it can be anything really, even beatboxing. The percussionist would assign each vocalization to a different synthesizer patch. Vocalizations get converted into MIDI note on/off commands. It looks like a straightforward DSP problem, except for the bit about real-time.
 A delay from an action to hearing a sound of 35msec is noticeable to most people. Due to unavoidable delays elsewhere, the actual processing budget may be under 15msec. There is not going to be a lot of signal to analyze in 15msec.

 The top graph shows ta, ba, do, tee spoken in succession. Each sound is about 200 msec long. Percussionists are faster. This guy puts an upper and perhaps unattainable limit on the speed of bol recognition:
https://www.youtube.com/watch?v=TXS6UByE_y8
 (Amazing Indian Percussion 3 (Vocal Percussion Lesson Konokkol))
 and these two give one an idea of the difficulty of separating individual sounds in the beatbox genre:
https://www.youtube.com/watch?v=pg1qLJ_6-LE
 (father vs. daughter beatboxing)

 Zooming in on the first 15msec of each sound we get the plots to the right. The graph below is more generous with time and shows the spectrum of the first 25msec of each sound. At first glance is appears that there *may* be enough signal differentiation there for that set of 4 sounds. But in the end there may be more than a dozen sounds and they will arrive in a real-world noisy environment. I don't know yet if I can simply do a spectral analysis to discriminate the sounds.

 There is more analysis to be done regarding the perception of simultaneity and I'm building a small device to test that with musicians and ordinary folks. Using real microphones and A/D sampling and a MIDI connection and synthesizer delay and 4 feet of sound propagation, I should be able to determine exactly what my time budget for signal analysis is. Stay tuned (as they say).

ta
ba
do
tee

Dan Downs   OstensiblyMyID@gmail.com   Scottsdale, AZ, November 2016