Research title: Investigating a possible "musician advantage" for speech-in-speech perception: *The role of fundamental frequency (f0) separation*

INTRODUCTION: The purpose of this research is to investigate whether listeners' musical experience plays a role in their ability to perceive speech from a single speaker in the presence of a competing speaker (i.e., the "cocktail party" effect) on the basis of fundamental frequency (f0) differences. Musicians have had a unique form of intense auditory training involving fine-grained acoustic distinctions of musical sounds (e.g., pitch, amplitude, timing, etc.). Whether this experience can *transfer* to speech perception, however, is an unresolved question – some studies show musicians' enhanced speech- in-speech perception relative to nonmusicians (e.g., Parbery-Clark et al., 2009), while others report no significant difference between the groups (e.g., Ruggles et al., 2014). Yet, none of these studies have controlled for f0 separation and fluctuation – two acoustic cues that have been shown to lead to increased intelligibility in perceiving competing speakers (e.g., Summerfield & Assmann, 1991; Patel et al., 2010). As such, the present experiments investigated the effect of f0 separation on the identification of two concurrently presented steady-state artificial vowel sounds (Exp 1) and color/number from a target sentence presented with a competing talker (Exp2) – with musicians hypothesized to show greater accuracy than nonmusicians at smaller f0 differences based on their extensive training with pitch.

SUBJECTS: Musicians (EXP1: n=34; EXP2: n= 41) and nonmusicians (EXP1: n=34; EXP2: n = 41) were all native English speakers with no prior experience with a tonal language. Musicians were recruited based on having at least 10 years of musical training ($\bar{x}$ =23.26 yrs, sd=14.59) and being musically active ($\bar{x}$ =9.51 hrs weekly practice). Nonmusicians had extremely minimal (<1 year) to no musical training.

EXP 1: Five steady-state vowels (duration=260ms), /i, ɛ, æ, ɑ, u/, were synthesized at 6 different flat f0 levels (relative to 120 Hz in 0, 0.156, 0.306, 1, 2, & 3 semitone increases). On a given trial, listeners heard a double-vowel pair and were asked to identify the vowels they heard by means of two button presses (each button was labeled with a representative word for each vowel, e.g., 'boot', 'bat', etc.). In each pair, one vowel had an f0 of 120 Hz while the second vowel varied in f0 separation at one of the six possible levels. All possible vowel and f0 pairings were presented in 240 trials.

Results show that double vowel intelligibility is significantly higher in the musician group ($p<0.019$) and significantly improves with increases in pitch separation ($p<0.001$), decreasing listener age ($p<0.05$), and increasing Euclidean distance in F1/F2 frequency between the vowels ($p<0.001$).

EXP 2: Both target and masker sentences were selected from a single male speaker from the Coordinate Response Measure (CRM) database (Bolia et al., 2000). Target sentences, indicated by the call sign "baron" (e.g., "Ready baron go to blue one now"), were monotonized at six f0 levels relative to 100 Hz. Masker sentences, cued by different call signs (e.g., "Ready eagle..."), were monotonized at 100 Hz. In the 192 experimental trials, the subjects' task was to select the color/number combination from the target sentence.

Results suggest that increasing f0 separation ($p<0.001$) and decreasing age ($p<0.01$) significantly improve the identification of the color/number combination for the 'baron' target sentences, while musical training was not a significant main effect ($p<0.10$) in any of the models.

CONCLUSION: These results suggest that musicians' purported "advantage" in certain perceptually challenging situations – such as two concurrently presented 260ms vowels – may be rooted in their differential encoding of f0 cues to aid in speech stream segregation. However, when listeners have a longer integration time – as in sentence perception – no group level effects were observed.