

**JSLHR Research Note**

**“Common terminology and acoustic measures for human voice and birdsong.”**

**Areen Badwal<sup>a</sup>, JoHanna Poertner<sup>b</sup>, Robin A. Samlan<sup>\*b</sup>, and Julie E. Miller<sup>\*a,b</sup>.**

<sup>a</sup>Department of Neuroscience, <sup>b</sup>Department of Speech, Language and Hearing Sciences,  
University of Arizona, Tucson, 85721

\*denotes equal contributions

Corresponding Author:  
Julie E. Miller, Ph.D.  
Assistant Professor  
Dept. of Neuroscience  
University of Arizona  
1040 E 4<sup>th</sup> St. GS 423  
Tucson, AZ 85721  
Office Phone: (520) 626-0100  
E-mail: juliemiller@email.arizona.edu

## Abstract

**Purpose:** The zebra finch is used as a model to study the neural circuitry of auditory-guided human vocal production. The terminology of birdsong production and acoustic analysis, however, differs from human voice production, making it difficult for voice researchers of either species to navigate the literature from the other. The purpose of this research note is to identify common terminology and measures to better compare information across species.

**Methods:** Terminology used in the birdsong literature will be mapped onto terminology used in the human voice production literature. Measures typically used to quantify the percepts of pitch, loudness, and quality will be described. Measures common to the literature in both species will be made from the songs of three middle-age birds using Praat and Song Analysis Pro. Two measures, Cepstral Peak Prominence (CPP) and Wiener Entropy (WE), will be compared to determine if they provide similar information.

**Results:** Similarities and differences in terminology and acoustic analyses are presented. A core set of measures including frequency, frequency variability within a syllable, intensity, CPP, and WE are proposed for future studies. CPP and WE are related, yet provide unique information about the syllable structure.

**Conclusion:** Using a core set of measures familiar to both human voice and birdsong researchers, along with both CPP and WE, will allow characterization of similarities and differences among birds. Standard terminology and measures will improve accessibility of the birdsong literature to human voice researchers and vice versa.

**1. Introduction:** Songbirds have long been utilized as models for studying the neural circuitry for auditory-guided human vocal learning and production. Two songbird species, the zebra finch and Bengalese finch, are well-suited to these purposes due to their easy breeding and adaptability to captivity, preference for social housing and the abundance of literature on central brain mechanisms for vocal motor control. Similarities and differences in central and peripheral mechanisms of finch and human sound production are reviewed in the following paragraphs, followed by a comparison of their production mechanisms. It remains unclear how common acoustic measures of birdsong relate to human voice. The purpose of this research note is to propose a 'translational dictionary' to facilitate sharing results and knowledge across the human voice and birdsong literature including demonstrating the feasibility of applying acoustic measurements made in human voice to structural elements within the birdsong.

In the zebra finch species, the males sing and the females do not, a sexually dimorphic behavior established early on with the development and growth of the song control system in males but not in females. The finch is an excellent model for studying neural circuits for human vocal motor control (Brainard & Doupe, 2013). In finches, there are identifiable song-dedicated brain nuclei located within cortical and basal ganglia tissue. Evidence from anatomical and gene expression studies show a high degree of similarity between finch song control regions RA, LMAN, and Area X (abbreviations used as proper names) with human brain regions involved in speech motor production and planning. Specifically, RA is correlated with the primary motor cortex, LMAN with Broca's area, and Area X with the striatum (Pfenning et al., 2014). Recent electrophysiological evidence has identified a direct connection between the human primary motor cortex and the larynx in the control of pitch during speaking and singing, a finding that further supports the use of the finch model to study voice control (Dichter, Breshears, Leonard, & Chang, 2018).

Finches and humans also share a cortico-basal ganglia-thalamo-cortical loop that is under neuromodulatory control by dopamine (Simonyan, Horwitz, & Jarvis, 2012). In both

species, primary and supplementary motor cortices provide descending excitatory glutamatergic input to basal ganglia structures. The basal ganglia provide GABAergic inhibition of the thalamus which relays input back to higher cortical structures. In finches, song-dedicated cortical nucleus RA is the output nucleus of the song control circuit sending descending drive to the hypoglossal cranial motor neurons in the brainstem and then on to the syrinx, the main vocal organ (versus the larynx in humans).

Experimental studies in finches have provided needed insight into genetic and neuromodulatory control of birdsong as a model for human voice research. Notably, one widely used example of shared genetic function between finches and humans is the speech gene, *FOXP2*, which affects vocal learning and production in both species (Vargha-Khadem et al., 1998; Watkins, Gadian, & Vargha-Khadem, 1999; Lai, Fisher, Hurst, Vargha-Khadem, & Monaco, 2001; Haesler et al., 2007; Heston & White, 2015). In humans, abnormal expression of the *FOXP2* protein in human cortical and basal ganglia regions is associated with articulatory deficits including impaired sequencing of orofacial movements; whether *FOXP2* contributes to voice deficits (e.g. loudness, pitch or quality) has not been examined. Intriguingly, virally-driven genetic manipulation of *FoxP2* levels in Area X of adult male zebra finches alters one aspect of voice, pitch control, which is acoustically measured as changes in fundamental frequency ( $f_0$ ), of harmonic elements in both finch song and in human voice (Murugan, Harward, Scharff, & Mooney, 2013; Stemple, Roy, & Klaben, 2014). Changes in  $f_0$  are also detected with experimental manipulation of dopamine levels in finch Area X during auditory-driven vocal learning tasks in adult males and dependent on whether the male is singing alone or to a female (Hoffmann, Saravanan, Wood, He, & Sober, 2016; Leblois, Wendel, & Perkel, 2010; Leblois, & Perkel, 2012). However,  $f_0$  measurements made in birdsong are restricted to a small subset of harmonic elements within the bird's song, thereby requiring a large number of birds to achieve a sufficient sample size. To facilitate future comparisons between birdsong and voice research

studies, it is critical to identify acoustic voice measurements that can be reliably made from inharmonic aspects of the bird's song and have the potential to change with age or intervention.

Central mechanisms drive the peripheral mechanisms (e.g. song control nuclei) to produce the acoustic signal. Both central and peripheral mechanisms operate to determine  $f_0$  (pitch) and intensity (loudness), features that vary as the finch sings (Kao, Doupe, & Brainard, 2005; Brumm & Slater, 2006; Kao & Brainard, 2006; Sober, Wohlgemuth, & Brainard, 2008). Humans and zebra finches share several peripheral sound generation mechanisms and present with some key differences. As reviewed by Riede and Goller (2010), similarities and differences involve the respiratory system, oscillating masses, and supraglottal structures. Both species phonate primarily on exhalation. Muscular control regulates airflow to and from the lungs in both, though recoil forces do not appear to contribute to driving pressures in avian voice production (Riede & Goller, 2010). Both species have oscillating tissue masses that abduct and adduct, though the avian syrinx has two independently controlled sound generators, the medial and lateral labia. The most common vibratory modes present in human vocal fold vibration are also thought to occur with avian labial vibration. In songbirds, the medial-to-lateral vibratory mode has been observed and the rotational mode is hypothesized, though little is known about the histological composition of the labia, and songbirds might lack the layered structure of the lamina propria that underlies the rotational mode (Riede & Goller, 2010). The rate of vibration determines  $f_0$ . The supraglottal vocal tract contributes to tuning resonances in both species. Muscular control of the upper vocal tract in songbirds, consisting of the tracheal tube and the oropharyngeal-esophageal cavity, is used to enhance select components of the acoustic signal, much as humans use the pharynx, mouth, and nose to enhance specific components of the acoustic signal. In general, the songbird voice production systems appear to be adapted for high speed vocal output and control of timing, and human laryngeal muscles are adapted for fine control of tension (Riede & Goller, 2010).

Research studies using finches and other songbirds rely upon a standard set of acoustic measurements to describe birdsong. Several of these measures are often unfamiliar to researchers of human voice, and vice versa, limiting sharing of information across species. In order to interpret birdsong data in the context of human voice metrics, there needs to be a common ‘language’ between birdsong and human voice. Therefore, we had two aims in the current study: 1) Establish a “translational dictionary” of birdsong and human voice terminology and measures and, 2a) Determine the feasibility of using a set of measures familiar to human voice researchers that will characterize key features of birdsong, 2b) Identify the extent to which Wiener Entropy (WE) relates to Cepstral Peak Prominence (CPP). WE and CPP both provide information about harmonic and noise energy in the sound (Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000; Hillenbrand & Houde, 1996). Importantly, we hypothesize that WE or CPP can be used to provide similar information about the birdsong.

To our knowledge, these two aims have not been pursued previously in birdsong research and should facilitate future investigations using the songbird model to investigate the impact of aging and neurodegenerative diseases on vocal communication.

## **2. Aim 1: Establishing a translational dictionary**

There are two components to the translational dictionary: 1) perceptual terminology related to the sounds birds and humans produce, with emphasis on voice and 2) acoustic measures used to define voice quality, loudness, and pitch. Acoustic analysis of human voice and birdsong is not standardized in the birdsong or human voice literature. We therefore reviewed several examples of acoustic measurements made in adult birdsong with relevance to voice.

First, it is necessary to provide a description of the birdsong structure in terms familiar to human voice researchers. Both birdsong and human speech include time-varying acoustic products consisting of alternating harmonic and noisy units. The comparison is not intended to

suggest similarity in meaning (e.g., that a particular unit represents the smallest unit of meaning). We refer the reader to already existing literature that draws parallels between phonological and syntactical features of birdsong and human language structure during development and adulthood (see Chapters 9-12 in: Berwick & Chomsky, 2013; Lipkind et al., 2013).

The basic unit of birdsong is a sequence of repeated syllables known as a motif. Motifs are encoded by specific patterns of neuronal firing in song control regions and are separated in time by silent intervals (Hahnloser, Kozhevnikov, & Fee, 2002). The motif can vary over multiple renditions by the order of the syllables (syntax) or insertion of a new syllable type. The audio signals and narrow-band spectrograms for sample birdsongs from three zebra finches are shown in Fig. 1A-C. The basic motif differs across birds and the most complex motif for these three birds is in panel C. Fig. 1A shows two motifs. For comparison, Fig. 1D shows a sentence spoken twice (“Shhh, finches perch in trees”) by a middle-aged human female to illustrate similarities/differences in the spectral structure of elements, human words, vowels and consonants, to birdsong. The motifs and sentence show a generally similar pattern in that they are both comprised of several components and that the energy in some, but not all, components have an  $f_0$  and harmonics.

In birdsong, the motif is a series of “syllables” defined by their spectral profile and labeled in Fig. 1 with capital letters. Each syllable in the birdsong consists of one (e.g. syllables A,B,C,E in Fig. 1A) or more notes (e.g. two notes- syllable D, Fig. 1A). All acoustic analyses in this manuscript were completed on the syllable level. Birdsong syllables are referred to as ‘noisy’ or ‘harmonic,’ and the designation determines the type of analysis completed on the syllable. A well-defined harmonic syllable (e.g. syllables ‘B & E’ in Fig. 1A and ‘B’ in Fig. 1B) has an  $f_0$  and many harmonics, similar to a vowel such as the ‘ee’ in “trees.” Note that the  $f_0$  for the human sentence (Fig. 1D) is generally lower than for the bird motif. The associated human harmonics tend to fade by approximately 5000 Hz, whereas they continue to 10,000 Hz in the

bird's harmonic syllables. A noisy syllable (e.g., syllable A in Fig. 1A-B) has poorly defined harmonics with sound energy visible between them. This type of birdsong syllable is more similar to a human fricative or affricate (such as 'sh' or 'ch') than a vowel, yet the birdsong noisy syllable has more harmonic structure than a fricative or affricate. Some syllables are comprised of a harmonic and a noisy note (e.g., Fig. 1A, syllable 'D' and Fig. 1C, Syllable 'I'). In this manuscript, those syllables are considered "mixed." In our human sentence, the word 'finches' and 'trees' appear to resemble the 'mixed' type of birdsong syllables, in that there are harmonic and noisy elements within the words.

**Perceptual voice terminology:** Birdsong and human voice are both discussed using the perceptual terms "loudness," "pitch," and "quality." Working definitions are summarized in column two of Table 1 and their acoustic correlates in column three. The meanings of pitch and loudness are similar in the birdsong and human voice literature and are perceptually judged using terms such as "louder" or "quieter" (loudness) and "higher" or "lower" (pitch). Quality in birdsong is determined based on how similar one performance of a syllable or motif is to the next rendition; in early song development, quality is defined by the degree of song similarity (e.g. imitation) between the juvenile finch and his adult tutor (Tchernichovski & Nottebohm, 1998; Tchernichovski et al., 2000). Highly accurate and similar performances (scores) in adult males and reduced variability in  $f_0$  from one song rendition to the next indicate higher quality song and in social contexts, the higher quality song is preferred by a female finch (Kao et al., 2005; Woolley & Doupe, 2008). The accuracy and similarity of pitch and loudness from one rendition to the next in birdsong is enfolded into the concept of quality. In contrast, human voice "quality" is separated from pitch and loudness in human voice research. However, quality is difficult to define and typically described using a series of terms such as breathy, strained or rough (Bartholomew, 1934; Kempster, Gerratt, Abbott, Barkmeier-Kraemer, & Hillman, 2009). Together, the three aspects of voice provide considerable information about speaker identity, mood, physical health, and vitality.

**Acoustic Analysis:** When comparing standard acoustic measurements of birdsong to human voice, several similarities and differences emerge. In the birdsong field,  $f_0$  and aggregate acoustic measures, known as similarity scores, are used to describe the acoustic match of the juvenile pupil's song to that of its adult tutor and to compare the effects of pre versus post experimental treatments in finch song at different ages (Tchernichovski, Mitra, Lints, & Nottebohm, 2001; Haesler et al., 2007; Miller, Hilliard, & White, 2010; Heston & White, 2015). Similarity scores are calculated at the motif and syllable level using WE, frequency modulation (FM), pitch, pitch goodness and amplitude modulation (Tchernichovski et al., 2000). WE is a common measure used to describe the effects of experimental treatment on birdsong syllables. WE is defined as a measure of the width and uniformity of the power spectrum and is measured on a logarithmic scale where zero is white noise and negative infinity is complete order (Tchernichovski et al., 2000). Thus, syllables with harmonic structure have more negative (e.g. lower) WE scores than noisy syllables.

Birdsong analysis contrasts with human voice acoustic measurement, where several measures of intensity,  $f_0$ , and spectral correlates of quality are common together with composite measures of quality (e.g., cepstral spectral index of dysphonia (Awan, Roy, & Cohen, 2014). Current recommendations for standard acoustic evaluation include: speaking  $f_0$ , standard deviation of  $f_0$ , maximum phonational frequency range, speaking intensity, maximum intensity range, and CPP (Patel et al., 2018). CPP is a robust and frequently-used measure of overall voice quality that provides information about acoustic waveform periodicity (Hillenbrand, Cleveland, & Erickson, 1994; Samlan, Story, & Bunton, 2013; Fraile & Godino-Llorente, 2014). Many other measures of quality have been used over the past several decades, including long and short-term perturbation, ratios of harmonic to noise components and various measures of spectral slope (Buder, 2000; Kreiman, Gerratt, & Berke, 1994). To summarize, frequency and amplitude measures *are common* to analyses of both species, in spite of being used differently (i.e., as part of a composite score in birdsong and as individual measures in humans). Based on

these similarities, we determined that mean  $f_0$ , the variability of  $f_0$  within a syllable, and intensity will serve as part of a core set of measures familiar to researchers of both species. Measures of quality are defined differently for birdsong and human voice, though a measure of harmonic or noise structure is common to both (i.e., WE for birdsong and CPP for human voice). These two measures complete the core set used in Aim 2.

### 3. Aim 2: Methods

There are two sub-aims: a) Determine the feasibility of using a set of measures familiar to human voice researchers that will characterize key features of birdsong, b) Identify the extent to which WE relates to CPP.

**Subjects:** All animal use was approved by the Institutional Animal Care and Use Committee at the University of Arizona. For the song analyses, three adult male zebra finches at the mid-point of their lifespan (~ 865-898 days post-hatch) were used with the expectation that measurements made in these three birds would be feasible in young and elder adult finches. The finches were raised in different nest boxes within an aviary in which male and female finches can select their mates; therefore, the adult tutor was not identified but is likely different for each bird and may represent the influence of several tutors based on the distinct songs sung (Fig. 1). Bird identification codes refer to the leg band color (R-red; W-white and number; R1156, W35, R1157). Finches were moved to individual sound attenuation chambers (Eckel Noise Control Technologies, Cambridge, MA) and acclimated for two days under a 13:11 h light:dark cycle before beginning recordings.

**Song Recordings:** Methods followed those of Miller et al. - (Miller et al., 2008). Songs were recorded from males housed alone in sound attenuation chambers. Isolation of an individual male finch in these sound chambers is a routine experimental paradigm employed in songbird research. It enables recording and analysis of a single song in response to an experimental manipulation of the neural circuitry without the need to filter out competing noise

from a female, groups of other finches or human presence (selected references: Jarvis, Scharff, Grossman, Ramos, & Nottebohm, 1998; Miller et al., 2008). Secondly, we opted to collect songs from males singing in a solo context, also known as ‘undirected song’ (UD). UD song characteristics are advantageous for the current study because they include a greater degree of variability in acoustic features such as  $f_0$  compared to female-directed song performance (Kao et al., 2005). This is similar to human voice research, where it is common to evaluate the subject reading or imitating text and sustained vowels (Patel et al., 2018). Conversational speech is less frequently used as stimuli for laboratory measurement.

Two hours of UD song were collected from lights-on in the morning for all birds using Shure 93 lavalier condenser omnidirectional microphones connected to an audiobox (Audiobox: 44.1 kHz sampling rate/24bit depth). When singing UD song, male finches tend to stay stationary in their cage; our previous observations noted only a one decibel change in sound intensity detected by the microphone if the bird was at the far regions of the cage. Sounds were recorded and digitized using pre-set parameters for capturing zebra finch song in Sound Analysis Pro (SAP) (SAP, <http://soundanalysispro.com/>) (Tchernichovski et al., 2000), a freely available software platform. Birdsong researchers use SAP as well as custom written code in Matlab or in R for acoustic analysis (Miller et al., 2010; Burkett, Day, Penagarikano, Geschwind, & White, 2015).

**Song Analyses:** Motifs were identified as a sequence of repeated syllables separated by periods of silence. Acoustic features were analyzed for 25 renditions of each syllable within the bird’s motif immediately following lights-on in the morning. No appreciable increase in power has been previously observed in any statistical test when conducted on an  $n \geq 25$  syllables in a given behavioral condition sung in the two-hour recording period based on power calculations (Miller et al., 2010). Introductory notes and unlearned calls were excluded from the acoustic analyses. Syllables were identified as sound envelopes that could be separated from other

syllables by local minima and repeated across the 25 motifs. The motifs and syllables were segmented in Praat and then analyzed in SAP (WE) and Praat (Boersma & Van Heuven, 2001).

**Syllable Analysis in Praat:** Acoustic measurements of birdsong syllables that are correlates of loudness, pitch and quality were selected for analysis (Table 1). Spectrogram settings typically used for viewing speech and voice did not allow visualization of enough birdsong harmonics to make decisions about segmentation and confirm syllable type within the motifs. Settings were manipulated through trial and error so that the harmonic structure was clearly revealed. The following spectrogram settings were used: view range of 0 to 10000 Hz, window length of 0.025 s, and dynamic range of 70 dB. Mean intensity and smoothed CPP were computed for all analyzed syllables. Mean intensity was computed using the “get intensity” command (Maryn, 2017). The CPP was calculated using standard methodology as described in the Praat instruction manual included with the program. First, a power cepstrogram was generated using a pitch floor setting of 300 Hz, a maximum frequency of 20,000, a time step of 0.002 and pre-emphasis of 50. All settings are the Praat default settings except the pitch floor and maximum frequency, which were modified in order to accommodate the higher  $f_0$  of the birds. The “get CPPS” command was then used with a peak search range of 300 to 1500 Hz (also modified to accommodate the bird  $f_0$ ), time window of 0.0001, quefrequency averaging window of 0.00005, tolerance of 0.05, parabolic interpolation, tilt line quefrequency of 0.001 to 0.0, exponential decay and a robust fit method.

We followed the convention in birdsong analysis whereby  $f_0$  is only measured for harmonic syllables where it is relatively stable; that is, syllable types comprised of only one note in the form of flat harmonic stacks (Kao et al., 2005). The  $f_0$  was not calculated for noisy or mixed birdsong syllables because, by definition, they lack a consistent  $f_0$ . For frequency analysis, we specified a 75 to 1600 Hz range for the setting “pitch range” and selected the cross-correlation analysis option. Mean  $f_0$ , standard deviation, minimum and maximum  $f_0$ , were measured using the “voice report” function in Praat for each harmonic syllable (Praat manual).

The  $f_0$  range was computed in an Excel spreadsheet as the difference between maximum and minimum  $f_0$ , in Hz.

**Statistics:** Two undergraduate research assistants made all measurements and rated one bird in common to determine inter-rater reliability. Descriptive statistics include mean and standard deviation (SD) for each acoustic measure.

### 3. Aim 2, Results:

The Pearson correlation coefficient was used to calculate inter-rater reliability. Correlations were 0.87 or higher for each measure ( $r \geq 0.87$ ), confirming good reliability.

The mean and standard deviation (SD) of each acoustic measure is reported for all syllables within each bird (Table 2). The full dataset of mean and SD for each of the 25 copies of every syllable can be found in the online supplemental materials (Supplemental Table 1).

As presented in Fig. 2, CPP and WE are shown for three consecutive renditions of a harmonic syllable (top panel, Syllable 'E') vs. a noisy Syllable 'C' (bottom panel) from the bird's motif in Fig 1A. In this example, the harmonic syllables have low WE (closer to negative infinity) compared to noisy syllables which approach zero, closer to white noise (Fig. 2). Both syllable types are present in a song motif but there are fewer harmonic syllables produced making sample sizes small.

Fig. 3 shows scatterplots of CPP vs. WE for each bird. In Fig. 3a (R1156), the harmonic syllable 'E' shows high CPP and low WE (e.g, more negative). In contrast, noisy syllable 'A' has lower CPP and higher WE (e.g. less negative). Interestingly, CPP values for harmonic syllable 'B' are close to noisy syllable 'A' but WE is similar to harmonic syllable 'E'. Syllable 'D,' (mixed) comprised of two notes, a harmonic and a noisy, can be compared to the syllables with one harmonic or one noisy note. In comparison to harmonic Syllables 'B' and 'E,' Syllable D has a

lower CPP score and higher mean WE (Table 2). CPP and WE for Syllable 'D' are both lower than for the two noisy syllables, 'A' and 'C.'

In Fig. 3b (W35), harmonic syllable 'B' has high CPP and low WE whereas noisy syllables 'A' and 'E' have lower CPP and higher WE (see also mean scores, Table 2). Syllables C and D are composed of mixed notes so their CPP scores are lower than harmonic Syllable B but greater than pure noisy Syllables 'A' and 'E'. Mixed syllable 'C' consists of mostly harmonic notes and therefore, its WE score is lower, similar to purely harmonic syllable 'B'. By contrast, mixed Syllable 'D' has a noisy note with harmonic note which may explain its high WE score in 3b.

In Fig. 3c (R1157), syllable 'H' has the highest CPP value compared to mixed syllables, 'B', 'D', 'E', 'F' and 'I' as well as noisy Syllables 'A', 'C', 'G'. Syllable 'H' has a harmonic appearance but its  $f_0$  varies over the length of the syllable and therefore, its  $f_0$  was not analyzed. The WE scores are not consistent among the syllables. Mixed Syllable 'D' has the lowest WE whereas noisy Syllable 'C' has the highest scores.

#### **4. Discussion**

The current study establishes a “translational dictionary” of birdsong and human voice terminology/measures that can be used as a tool to interpret how manipulations of neural circuitry in birdsong can be applied to a better understanding of voice disorders. Furthermore, we identified a set of acoustic measures common to evaluation of human voice and birdsong. In Praat, measurements of birdsong were feasible following some adjustments to the settings typically used for human voice analysis. Praat presents a platform that is more familiar to human voice researchers and may facilitate collaborations between voice and songbird researchers.

We hypothesized that the human voice measure CPP and the birdsong measure WE would provide the same information about the harmonic and noise components of the birdsong syllable. Contrary to our hypothesis, the relationship between CPP and WE is complex. For

syllables that were assigned to clear categories of harmonic or noisy, their CPP and WE scores showed a fairly consistent inverse relationship. Harmonic syllables have high CPP values and low WE scores compared to noisy syllables, suggesting that both metrics provide similar information. However, in the case of 'mixed' bird syllables consisting of harmonic and noisy elements, the relationship was not preserved because WE scores were too variable. For mixed syllables, CPP scores appear to be a more reliable measure. The CPP was generally mid-range when the mixed syllable had a long harmonic component (i.e., syllables C and D in Fig. 3B and I in Fig. 3C) and low when the harmonic component was short (i.e., syllable D in Fig. 3A and syllables B, D, E, F in Fig. 3C). Because calculation of CPP does not rely on previous determination of  $f_0$  (Hillenbrand et al., 1994), it is one of the measures commonly used in human voice analysis that is reliable when the acoustic signal is noisy (Awan, Roy, & Dromey, 2009). This is important to assessment of disordered voice and also when the assessment stimuli (e.g., sustained sounds, words, phrases) contain fricatives or affricates, an acoustic situation that mirrors the noisy and mixed syllables of birdsong. To our knowledge, limitations of WE for noisy and mixed syllables are not reported in the birdsong literature.

The CPP was not always the more reliable determinant of noisy versus harmonic energy, however. There were also some instances where CPP was similar for noisy and harmonic syllables (e.g. Fig. 3A- Syllables A vs. B) and the WE score differentiated the syllables. In this particular case, the noisy syllables had some low frequency harmonics which appear reflected in the CPP value. Thus, WE could also occasionally distinguish fine detail between two syllables better than CPP. Our results suggest that the measures provide complementary information and utilization of both CPP and WE scores is warranted to reliably characterize fine differences in acoustic properties between syllables.

Given the ability to make successful comparisons between birdsong and human voice measures, our analyses can be extended in future work to include more harmonic syllables in birdsong and to make comparisons at the motif level with human speech measurements

including speech and articulation rate. Here, we only studied males singing undirected song (e.g. vocally practicing alone), as a model for human voice research, but it would be informative for future studies to analyze female-directed song as a model for conversational speech.

One limitation of using zebra finches as a model for human voice is that the males sing and the females do not. Therefore, future studies will need to incorporate other songbird species (e.g. Northern cardinals, house wrens) in which both the females and males sing (Odom & Benedict, 2018) in order to make comparisons of pitch, loudness and vocal quality with human male and female subjects. For example, adult human males have a lower  $f_0$  compared to females (Baken & Orlikoff, 2000) whether that holds true for songbird species in which both the females and males sing requires determination. In addition to differences based on gender, changes in human voice are impacted by age – children have a higher  $f_0$ , than adults (Baken & Orlikoff, 2000), men's  $f_0$  might increase as they age, and women's  $f_0$  might decrease (Hollien & Shipp, 1972; Dehqan, Scherer, Dashti, Ansari-Moghaddam, & Fanaie, 2012; Goy, Fernandes, Pichora-Fuller, & van Lieshout, 2013). Whether young male zebra finches learning their song have a higher  $f_0$  compared to adult finches is not known. In a small study conducted in a related species, Bengalese finch, reductions in syllable pitch and intensity were detected in adulthood as the birds aged (Cooper et al., 2012).

Because male zebra finches are known as close-ended learners, meaning they retain a similar motif structure from development into adulthood, it would be useful to conduct vocal analyses on canaries (open-ended learners) that change their songs each breeding season (Nottebohm, Nottebohm, & Crane, 1986). By applying our human voice analyses to canary songs, we can obtain additional insight into the neural plasticity mechanisms that drive vocal quality based on environmental needs.

The further development of comparative analyses between voice/speech measures in normal human populations with birdsong will facilitate future comparisons examining age and neurodegenerative disease related changes on vocal output.

## 5. Acknowledgements

We would like to thank Stephanie Munger (U.Arizona) for birdsong collection, Dr. Nancy Day (UCLA) for Matlab code and University of Arizona Animal Care. Data collection was supported by University of Arizona start up funds to JEM and financial support to A. Badwal from the University of Arizona Undergraduate Biological Research Program (UBRP).

## 6. References

- Awan, S. N., Roy, N., & Cohen, S. M. (2014). Exploring the relationship between spectral and cepstral measures of voice and the Voice Handicap Index (VHI). *J Voice*, 28(4), 430-439. doi:10.1016/j.jvoice.2013.12.008
- Awan, S. N., Roy, N., & Dromey, C. (2009). Estimating dysphonia severity in continuous speech: application of a multi-parameter spectral/cepstral model. *Clinical linguistics & phonetics*, 23(11), 825-841.
- Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*: Cengage Learning.
- Bartholomew, W. T. (1934). A physical definition of "good voice-quality" in the male voice. *J Acoust Soc Am*, 6(1), 25-33.
- Berwick, R. C., & Chomsky, N. (2013). *Birdsong, speech, and language: exploring the evolution of mind and brain*: MIT press.
- Boersma, P., & Van Heuven, V. (2001). Speak and unSpeak with PRAAT. *Glott International*, 5(9/10), 341-347.
- Brainard, M. S., & Doupe, A. J. (2013). Translating birdsong: songbirds as a model for basic and applied medical research. *Annu Rev Neurosci*, 36, 489-517. doi:10.1146/annurev-neuro-060909-152826

- Brumm, H., & Slater, P. J. (2006). Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Animal Behaviour*, 72(3), 699-705.
- Buder, E. H. (2000). Acoustic analysis of voice quality: A tabulation of algorithms 1902–1990. *Voice quality measurement*, 119-244.
- Burkett, Z. D., Day, N. F., Penagarikano, O., Geschwind, D. H., & White, S. A. (2015). VoICE: A semi-automated pipeline for standardizing vocal analysis across models. *Sci Rep*, 5, 10237. doi:10.1038/srep10237
- Cooper, B. G., Méndez, J. M., Saar, S., Whetstone, A. G., Meyers, R., & Goller, F. (2012). Age-related changes in the Bengalese finch song motor program. *Neurobiol Aging*, 33(3), 564-568.
- Dehqan, A., Scherer, R. C., Dashti, G., Ansari-Moghaddam, A., & Fanaie, S. (2012). The effects of aging on acoustic parameters of voice. *Folia Phoniatica et Logopaedica*, 64(6), 265-270.
- Dichter, B.K., Breshears, J.D., Leonard, M.K., & Chang, E.F. (2018). The control of vocal pitch in human laryngeal motor cortex. *Cell*, 174, 21-31.
- Frailé, R., Godino-Llorente, J.I (2014). Cepstral peak prominence: A comprehensive analysis. . *Biomedical Signal Processing and Control*, 14, 42-54.
- Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice*, 27(5), 545-555.
- Haesler, S., Rochefort, C., Georgi, B., Licznarski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS Biol*, 5(12), e321.
- Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, 419(6902), 65-70.
- Heston, J. B., & White, S. A. (2015). Behavior-linked FoxP2 regulation enables zebra finch vocal learning. *J Neurosci*, 35(7), 2885-2894. doi:10.1523/JNEUROSCI.3715-14.2015

- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *J Speech Hear Res*, 37(4), 769-778.
- Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2), 311-321.
- Hoffmann, L. A., Saravanan, V., Wood, A. N., He, L., & Sober, S. J. (2016). Dopaminergic contributions to vocal learning. *J Neurosci*, 36(7), 2176-2189.  
doi:10.1523/JNEUROSCI.3883-15.2016
- Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech, Language, and Hearing Research*, 15(1), 155-159.
- Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., & Nottebohm, F. (1998). For whom the bird sings: context-dependent gene expression. *Neuron*, 21(4), 775-788.
- Kao, M. H., & Brainard, M. S. (2006). Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J Neurophysiol*, 96(3), 1441-1455.
- Kao, M. H., Doupe, A. J., & Brainard, M. S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, 433(7026), 638-643.
- Kempster, G. B., Gerratt, B. R., Abbott, K. V., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124-132.
- Kent, R. D., & Ball, M. J. (2000). *Voice quality measurement*: Singular.
- Kreiman, J., Gerratt, B. R., & Berke, G. S. (1994). The multidimensional nature of pathologic vocal quality. *J Acoust Soc Am*, 96(3), 1291-1302.
- Kreiman, J., Gerratt, B. R., Garellek, M., Samlan, R., & Zhang, Z. (2014). Toward a unified theory of voice production and perception. *Loquens*, 1(1).

- Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, 413(6855), 519-523.
- Leblois, A., Wendel, B. J., & Perkel, D. J. (2010). Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *J Neurosci*, 30(16), 5730-5743.
- Leblois, A. P., D.J. (2012). Striatal dopamine modulates song spectral but not temporal features through D1 receptors. *European Journal of Neuroscience*, 35, 1771-1781.
- Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., . . . Okanoya, K. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, 498(7452), 104.
- Maryn, Y. (2017). Practical acoustics in clinical voice assessment: a Praat primer. *Perspectives of the ASHA Special Interest Groups*, 2(3), 14-32.
- Miller, J. E., Hilliard, A. T., & White, S. A. (2010). Song practice promotes acute vocal variability at a key stage of sensorimotor learning. *PLoS One*, Jan 6, 5(1), e8592.
- Miller, J. E., Spiteri, E., Condro, M. C., Dosumu-Johnson, R. T., Geschwind, D. H., & White, S. A. (2008). Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J Neurophysiol*, 100(4), 2015-2025.
- Murugan, M., Harward, S., Scharff, C., & Mooney, R. (2013). Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. *Neuron*, 80(6), 1464-1476. doi:10.1016/j.neuron.2013.09.021
- Nottebohm, F., Nottebohm, M. E., & Crane, L. (1986). Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and neural biology*, 46(3), 445-471.
- Odom, K. J., & Benedict, L. (2018). A call to document female bird songs: Applications for diverse fields. *The Auk*, 135(2), 314-325.

- Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyiski, D., Eadie, T., . . . Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association Expert Panel to Develop a Protocol for Instrumental Assessment of Vocal Function. *American Journal of Speech-Language Pathology*, 1-19.
- Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., . . . Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, 346(6215), 1256846. doi:10.1126/science.1256846
- Riede, T., & Goller, F. (2010). Peripheral mechanisms for vocal production in birds - differences and similarities to human speech and singing. *Brain Lang*, 115(1), 69-80.
- Samlan, R. A., Story, B. H., & Bunton, K. (2013). Relation of perceived breathiness to laryngeal kinematics and acoustic measures based on computational modeling. *J Speech Lang Hear Res*, 56(4), 1209-1223. doi:10.1044/1092-4388(2012/12-0194)
- Simonyan, K., Horwitz, B., & Jarvis, E. D. (2012). Dopamine regulation of human speech and bird song: A critical review. *Brain Lang*, 122(3), 142-150.
- Sober, S. J., Wohlgemuth, M. J., & Brainard, M. S. (2008). Central contributions to acoustic variation in birdsong. *J Neurosci*, 28(41), 10370-10379.
- Stemple, J. C., Roy, N., & Klaben, B. K. (2014). *Clinical voice pathology: Theory and management*: Plural Publishing.
- Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, V291(N5513), 2564-2569.
- Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, 59, 1167-1176.
- Tchernichovski, O., & Nottebohm, F. (1998). Social inhibition of song imitation among sibling male zebra finches. *Proc. Natl. Acad. Sci. USA*, 95, 8951-8956.

- Vargha-Khadem, F., Watkins, K. E., Price, C. J., Ashburner, J., Alcock, K. J., Connelly, A., . . .  
Passingham, R. E. (1998). Neural basis of an inherited speech and language disorder.  
*Proc Natl Acad Sci U S A*, *95*(21), 12695-12700.
- Watkins, K. E., Gadian, D. G., & Vargha-Khadem, F. (1999). Functional and structural brain  
abnormalities associated with a genetic disorder of speech and language. *Am J Hum  
Genet*, *65*(5), 1215-1221.
- Woolley, S. C., & Doupe, A. J. (2008). Social context-induced song variation affects female  
behavior and gene expression. *PLoS Biol*, *6*(3), e62.

## 7. Figure and Tables list:

**Fig. 1. Birdsong motifs vs. Human Sentence.** Audio waveform (top) and spectrogram (bottom). Spectrogram is depicted as time (in seconds, x-axis) vs. frequency (Hz, y-axis). Color in the spectrogram represents the amplitude of the spectral energy at that frequency with red being the most intense followed by orange, yellow and blue. **A)** Two motifs from bird R1156 shown as a sequence of repeated syllables (ABCDE) separated by a pause (red line). Syllables are assigned unique letters based on their visual appearance and confirmed by acoustic measurements. **B)** Motif exemplar for bird W35 comprised of syllables ABCDE. **C)** Motif exemplar for bird R1157, consisting of syllables A-I. **D)** Human sentence 'SHHH, FINCHES PERCH IN TREES,' repeated twice.

**Fig. 2. Harmonic vs. Noisy Syllable Exemplars, CPP and WE.** Three consecutive renditions of R1156 harmonic Syllable 'E' (**top**) vs. Syllable 'C' (**bottom**) are shown. Cepstral Peak Prominence (CPP) and Wiener Entropy (WE) scores are reported below each syllable. CPP is higher for harmonic Syllable E compared to noisy Syllable C. WE entropy scores are lower (more negative) for Syllable E than for Syllable C.

**Fig. 3. Scatterplots of CPP vs. WE Scores for all Syllables.** CPP (x-axis) vs. WE (y-axis) scores are plotted for 25 renditions (copies) of each syllable within a bird. Filled circles indicate harmonic syllables whereas open circles represent noisy syllables. Mixed syllables that consist of two or more notes are depicted as plus signs. Colors are used to represent each syllable. R1156 (**A**) and W35 (**B**) have five syllables each in their motifs whereas R1157 (**C**) has nine syllables. The '?' signifies that the syllable type is not clear but contains some harmonic elements.

Table 1. Definition in Human Voice and Birdsong

Perceptual component of voice	Meaning	Acoustic correlates in human and finch
Loudness	The percept of the amount or intensity of the sound; volume. Influenced by intensity, $f_o$ , and spectral profile	Intensity level or sound pressure level; reported in decibels, sound pressure level (dB SPL)
Pitch	The percept of how high or low a sound is. Related in part to $f_o$ (of human vocal folds, finch syringeal labial folds, or acoustic waveform), and also to intensity and spectral profile	$f_o$ , in Hertz (Hz)
Quality	<p>Human voice: The interaction of the acoustic signal with the listener. Perceived as an overall pattern, related in part to harmonics and spectral slope, inharmonic energy, time-varying frequency and intensity, and characteristics of the vocal tract</p> <p>Birdsong: A similarity score that indicates goodness of match of the song motif across multiple renditions in an adult bird or how similar the juvenile finch's song is to his adult tutor's song</p>	<p>Human voice: CPP, in dB, is the currently-recommended measure of overall voice quality. CPP describes the prominence of harmonic energy in the acoustic waveform.</p> <p>Birdsong: Composite acoustic measures of similarity and accuracy (scale of 0-100) that include WE, pitch, frequency modulation and spectral continuity taken from 50ms (similarity) or 7ms (accuracy) sampling windows at either the motif or syllable level</p> <p>Birdsong WE: a measure of the periodic versus aperiodic energy in a birdsong syllable. Measured on a logarithmic scale from zero to minus infinity. White noise <math>\log(1) = 0</math> and complete order <math>\log(0) = \text{minus infinity}</math></p>

**Table 1.** Definitions are compiled from selected references: (Tchernichovski et al., 2000);

Hillenbrand & Houde, 1996; Baken & Orlikoff, 2000; Kreiman & Gerratt in: Kent & Ball,

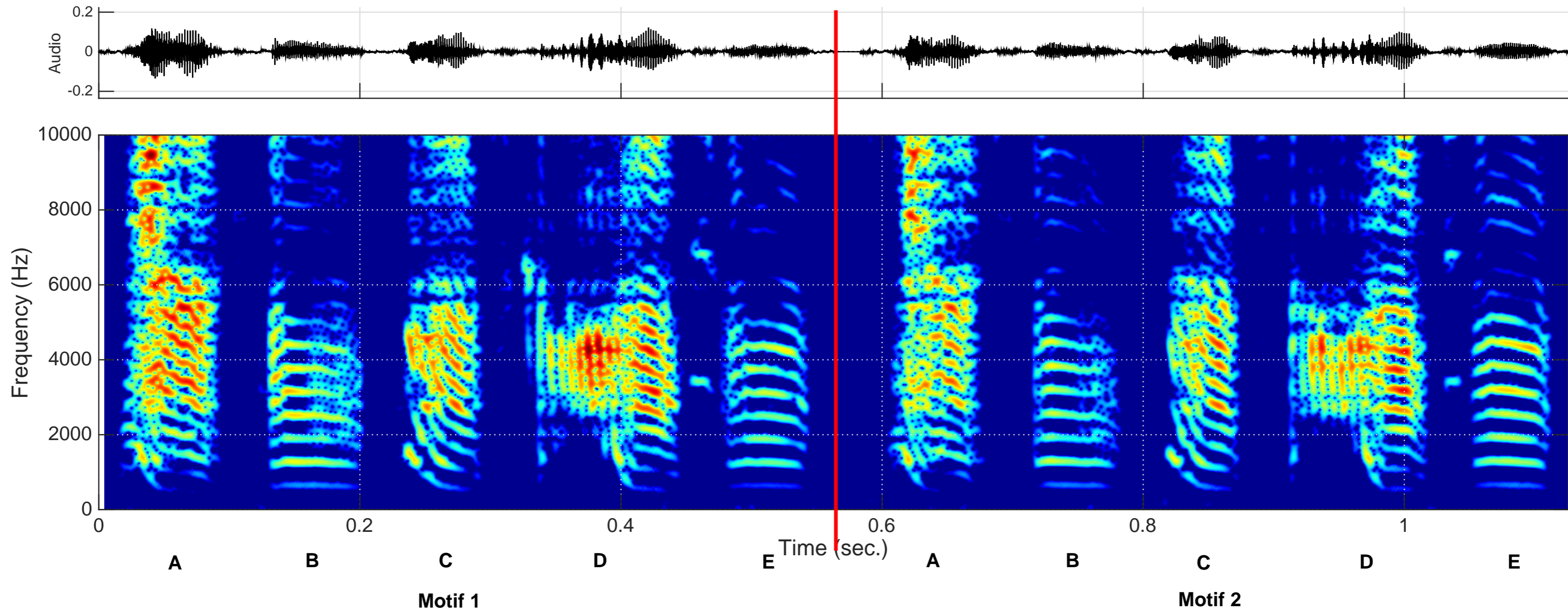
2000;Kreiman, Gerratt, Garellek, Samlan, & Zhang, 2014;Patel et al., 2018). Human measurements in voice research are obtained using either sustained vowel phonation and/or connected speech during a standard reading passage. Fundamental frequency =  $f_0$  Cepstral Peak Prominence = CPP

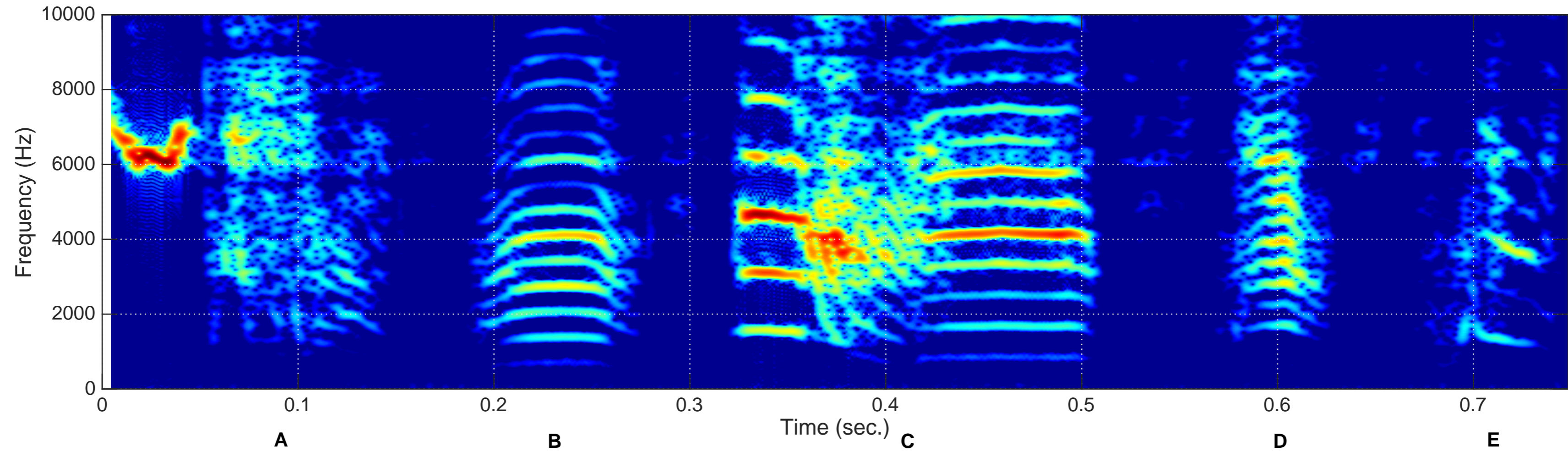
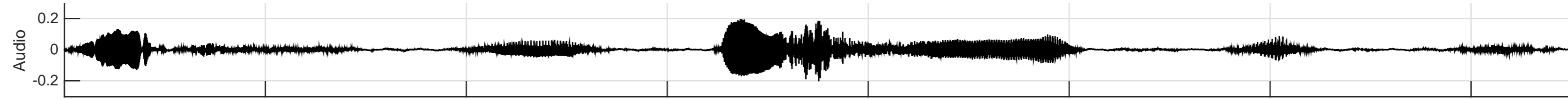
**Table 2. Mean and Standard Deviation Scores for Acoustic Measurements of Birdsong**

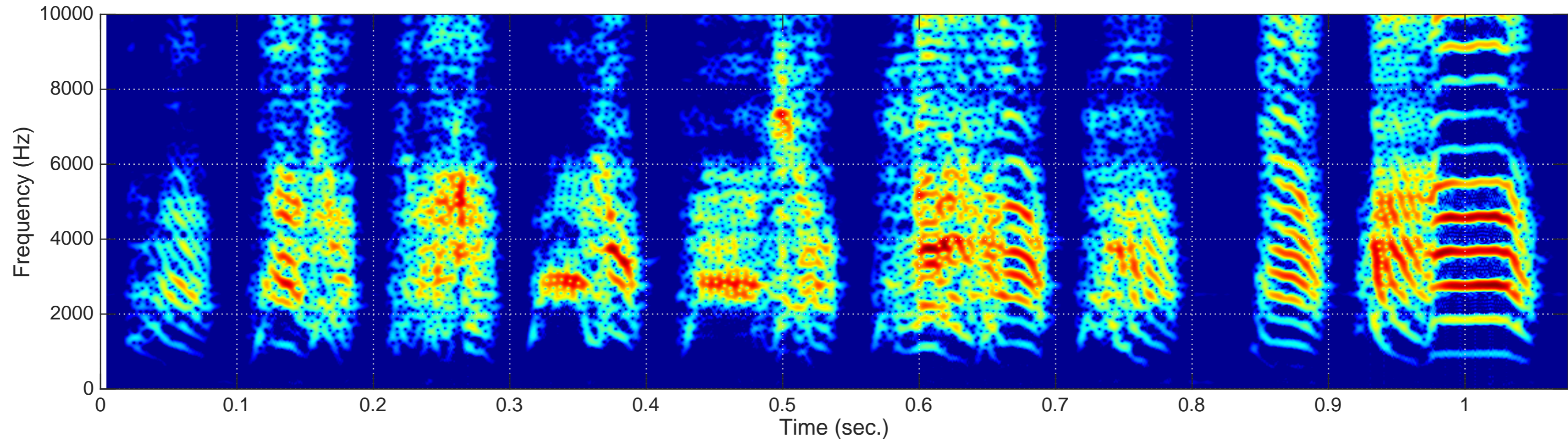
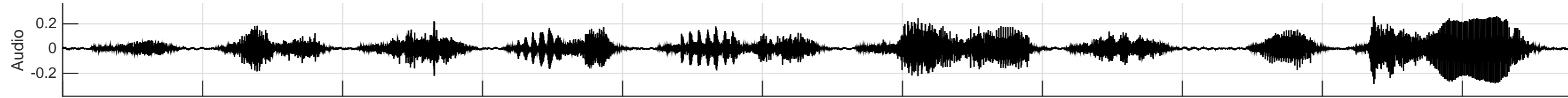
Bird & Syllable ID	Intensity		CPP		WE		$F_0$	$F_0$	$F_0$ .range	$F_0$ .range
	db SPL		db				Hz		Hz	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
W35 Syll A	61.82	4.89	9.68	0.21	-2.45	0.18	-	-	-	-
W35 Syll B	56.39	1.70	19.70	0.60	-3.08	0.14	644.02	5.30	118.43	15.56
W35 Syll C	65.25	1.81	15.20	0.37	-2.88	0.22	-	-	-	-
W35 Syll D	57.43	1.12	16.20	0.88	-1.55	0.15	-	-	-	-
W35 Syll E	51.26	0.84	10.10	0.41	-2.75	0.20				
R1156 Syll A	62.89	1.19	16.66	0.92	-1.17	0.08	-	-	-	-
R1156 Syll B	56.57	1.37	16.40	1.35	-2.97	0.18	615.54	7.81	55.10	13.68
R1156 Syll C	60.22	0.83	15.32	0.68	-1.91	0.12	-	-	-	-
R1156 Syll D	61.26	1.46	12.00	0.80	-2.73	0.13	-	-	-	-
R1156 Syll E	54.38	0.95	20.77	0.90	-2.87	0.20	612.50	6.78	53.23	12.99
R1157 Syll A	57.30	3.71	11.42	1.25	-2.86	0.19	-	-	-	-
R1157 Syll B	64.80	2.26	10.72	0.45	-2.06	0.10	-	-	-	-
R1157 Syll C	65.86	0.78	9.52	0.37	-1.92	0.32	-	-	-	-
R1157 Syll D	66.04	0.88	9.30	0.40	-3.34	0.26	-	-	-	-
R1157 Syll E	66.22	1.28	9.67	0.52	-2.47	0.18	-	-	-	-
R1157 Syll F	69.51	0.79	11.90	0.31	-2.14	0.12	-	-	-	-
R1157 Syll G	64.41	0.68	10.43	0.47	-2.31	0.13	-	-	-	-
R1157 Syll H	66.42	1.15	19.30	0.99	-2.19	0.14	-	-	-	-
R1157 Syll I	71.26	1.63	16.98	0.65	-2.88	0.15	-	-	-	-

**Table 2.** Mean and Standard Deviation (SD) are reported for each syllable across three birds. The  $f_0$  measurements were only made for syllables with a clear one note harmonic structure. The  $f_0$  range represents the difference between the lowest and highest  $f_0$  during a sustained phonation. Units are reported for the mean scores except for WE which is a pure number and unitless (Tchernichovski et al., 2000). Dashes indicate data not applicable. db = decibels; Hz=hertz

**Supplemental Table 1.** Excel file contains raw data from 25 copies of each syllable across the three birds. Copy number (Column C) refers to the motif in which that syllable was present. Mean and SD were obtained from these data and are reported in Table 2.







**A**

**B**

**C**

**D**

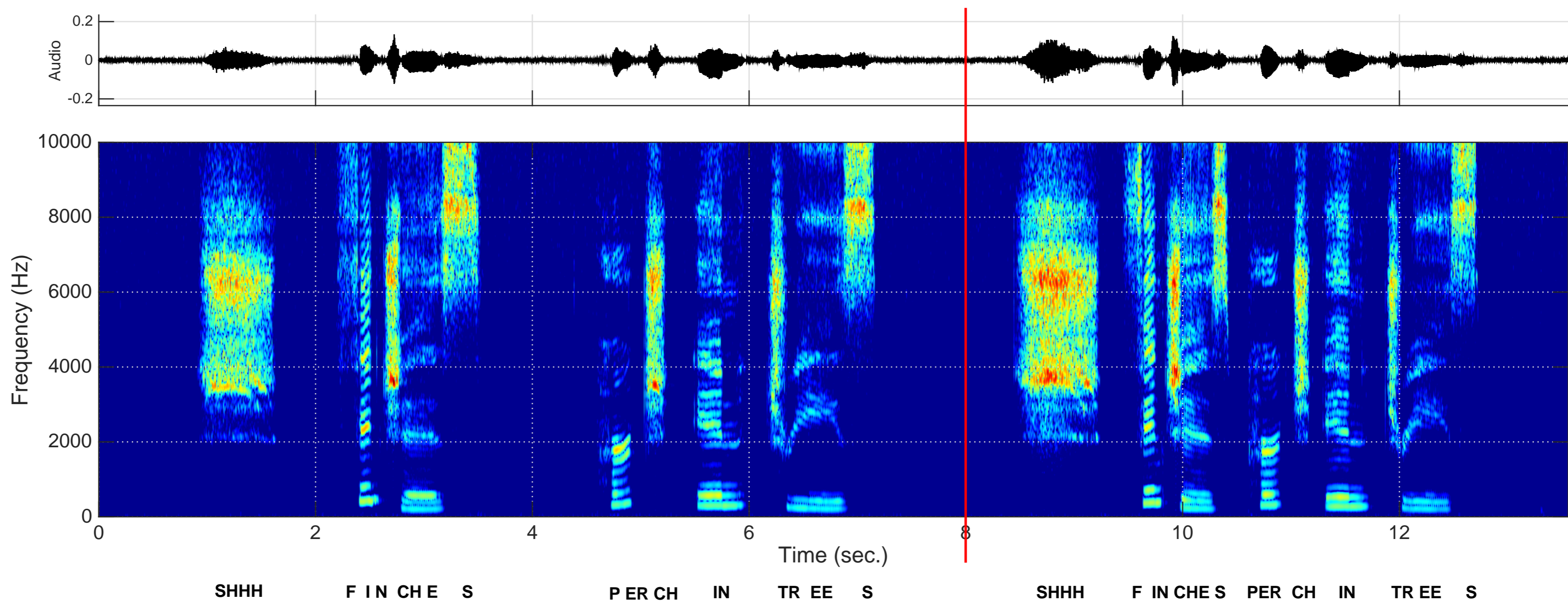
**E**

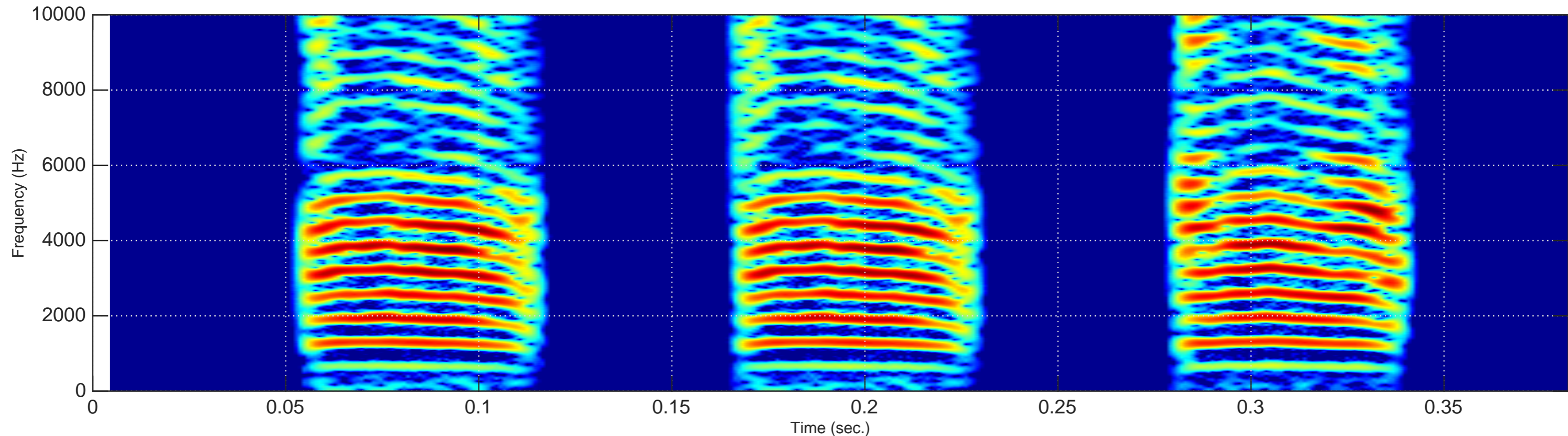
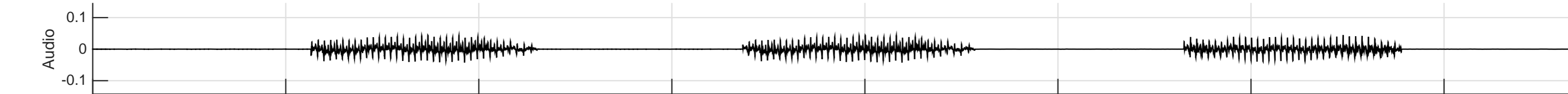
**F**

**G**

**H**

**I**

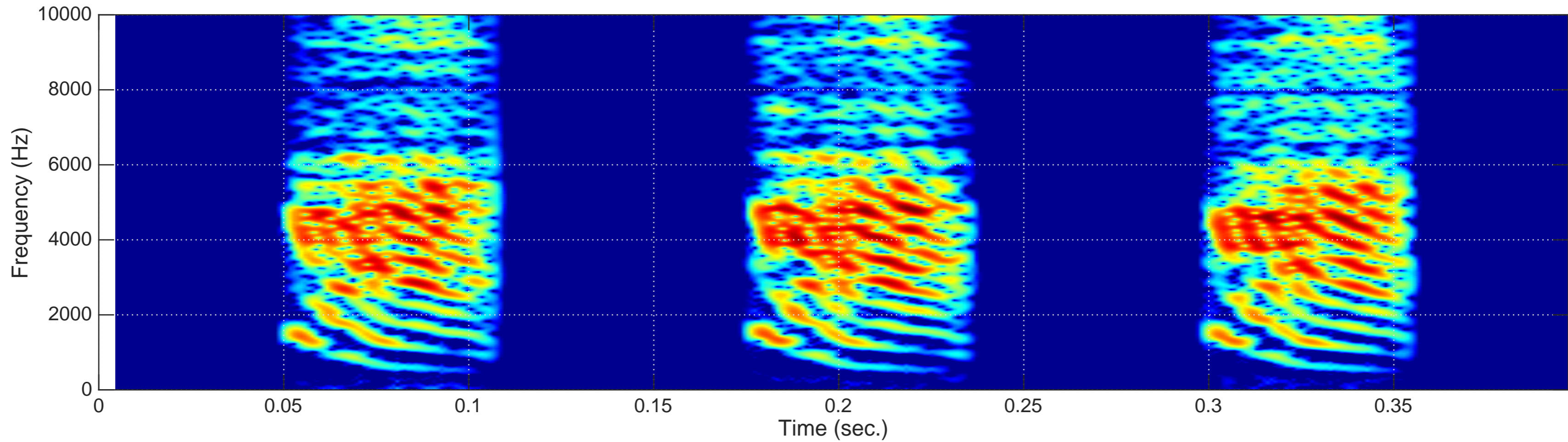
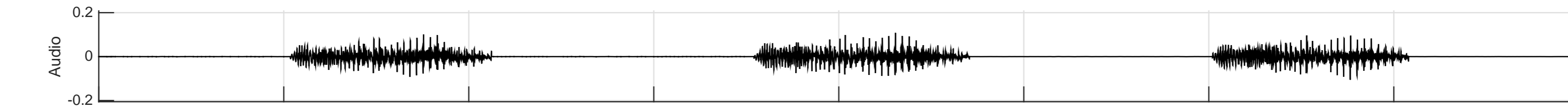




WE = -3.04  
CPP = 21.7

WE = -3.04  
CPP = 21.9

WE = -2.90  
CPP = 21.3



WE = -1.94  
CPP = 13.7

WE = -1.97  
CPP = 14.8

WE = -1.85  
CPP = 15.1

