Acta Psychologica 182 (2018) 189–193

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Determining the end of a musical turn: Effects of tonal cues

Lauren V. Hadley[a,*], Patrick Sturt[a], Nikki Moran[b], Martin J. Pickering[a]

[a] Department of Psychology, University of Edinburgh, United Kingdom
[b] Reid School or Music, University of Edinburgh, United Kingdom

A R T I C L E I N F O

Keywords:
Turn-taking
Musical coordination
Tonality
Prediction
Improvisation

A B S T R A C T

Successful duetting requires that musicians coordinate their performance with their partners. In the case of turn-taking in improvised performance they need to be able to predict their partner's turn-end in order to accurately time their own entries. Here we investigate the cues used for accurate turn-end prediction in musical improvisations, focusing on the role of tonal structure. In a response-time task, participants more accurately determined the endings of (tonal) jazz than (non-tonal) free improvisation turns. Moreover, for the jazz improvisations, removing low frequency information (< 2100 Hz) - and hence obscuring the pitch relationships conveying tonality - reduced response accuracy, but removing high frequency information (> 2100 Hz) had no effect. Neither form of filtering affected response accuracy in the free improvisation condition. We therefore argue that tonal cues aided prediction accuracy for the jazz improvisations compared to the free improvisations. We compare our results with those from related speech research (De Ruiter et al., 2006), to draw comparisons between the structural function of tonality and linguistic syntax.

1. Introduction

Accurate temporal coordination between members of a musical ensemble is essential for coherent performance (Keller, 2008), and such coordination requires performers to predict each other's behaviour (Pecenka & Keller, 2011; Phillips-Silver & Keller, 2012). The need to predict each other is particularly apparent in turn-taking contexts, as turns must be accurately timed between individuals when the primary communicative role switches. Musicians have access to auditory cues such as pitch, duration, and intensity, as well as visual cues such as gaze or body movement, to help them determine the end of a co-performer's turn. As the audio information is the primary musical signal, and visual information can be obscured in performance contexts (because of positioning, lighting etc.), we focus on auditory cues. These auditory cues are relevant both for performers predicting turn-ends when playing with a partner, and for non-performers predicting the music that they passively listen to. We therefore had listeners predict turn-ends for jazz and free improvisations under various conditions to determine the importance of cues relating to tonality in musical prediction.

Accurate turn-end prediction is critical not only for music performance but also for linguistic conversation (Stivers et al., 2009). In fact, research on conversation has investigated the importance of a variety of auditory cues. De Ruiter, Mitterer, and Enfield (2006) presented listeners with recorded utterances in their original form, lowpass filtered, or with a flattened contour. Lowpass filtering removed high frequencies commonly used to distinguish consonants, obscuring structural word-level cues including semantic and syntactic information. Flattening the contour set the pitch contour to horizontal (i.e. monotone), removing intonation cues. While listening, participants were told to 'press a button in front of them at the moment they thought the speaker would be finished speaking (Dutch: *is uitgesproken*).' The authors argued that 'The instruction encouraged the subjects to try to ANTICIPATE this moment, and not wait until the fragment stopped playing' (De Ruiter et al., 2006, p. 523). Listeners were able to predict the end of a conversational turn equally well when it was presented in its original form (M = 186 ms prior to offset) or when its contour was flattened (M ~ 160 ms prior to offset [estimated from figure]), but accuracy dropped when words were obscured through lowpass filtering (M ~ 470 ms prior to offset [estimated from figure]). This suggests that information conveyed in the upper frequencies of speech, including semantic and syntactic information, is vital for accurately predicting turn-ends of speech.

In the current paper, we will investigate the accuracy with which listeners predict the end of turns in a musical improvisation: a form of performance in which musicians generate coherent musical utterances in real time, without recourse to a pre-planned script. While certain aspects of an improvisation may nonetheless be predetermined (for example the harmonic progression of a jazz standard or an individually generated plan for an upcoming musical phrase), many decisions regarding content are made in real time (Ashley, 2016; Berkowitz, 2010).

* Corresponding author at: Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom.
E-mail address: lauren.hadley@cantab.net (L.V. Hadley).

https://doi.org/10.1016/j.actpsy.2017.11.001
Received 3 May 2017; Received in revised form 15 September 2017; Accepted 2 November 2017

We investigate how accuracy differs between musical genres (jazz and free improvisation), as well as the structural cues for prediction used within these genres, by evaluating the importance of high and low frequency information.

In music each note is a periodic waveform, and the number of times that this waveform repeats per second is its fundamental frequency. The fundamental frequency is the note's lowest frequency component, and is commonly understood as its pitch. However, each note's spectrum also includes multiples of this fundamental frequency, which make up the harmonics. Although it is possible to perceive pitch from these harmonics alone, inferring pitch without the fundamental frequency takes substantially longer than when the fundamental is present (Winkler, Tervaniemi, & Näätänen, 1997). Pitch provides structural information, with pitches organised into a hierarchy (in terms of their stability and probability) whose pattern may be perceived as tonality. Tonality thus describes the probability of different note-to-note progressions (and by extension, larger-scale progressions). Listeners do not require musical training beyond normal exposure to perceive tonal structure during music listening (see Huron, 2006 for a broad summary), and recognise anomalous progressions (Koelsch, Gunter, Friederici, & Schroger, 2000; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010). The lower frequencies of musical notes, therefore, convey their pitch and are critical for conveying tonality, which is commonly compared with syntax and semantics in language (Koelsch, 2005; Patel, 2003; Steinbeis & Koelsch, 2008).

We investigated turn-end prediction for two forms of musical improvisation, namely jazz and free, across a range of listeners using a paradigm similar to De Ruiter et al. (2006). Broadly speaking, jazz improvisation uses a tonal framework to constrain both note-to-note and larger-scale structure (Barrett & Peplowski, 1998), whereas free improvisation does not use a tonal framework, but uses clustering to constrain large-scale structure (Dean, Bailes, & Drummond, 2014). We first compared turn-end accuracy for original jazz and free improvisations. We then compared turn-end accuracy for jazz and free improvisations that had been highpass filtered, obscuring pitch information (and hence in the jazz improvisations, tonal information). Such filtering has been used to investigate the salience of specific spectral cues in music and speech research (Moore & Tan, 2003; Schellenberg, Iverson, & Mckinnon, 1999). We expected listeners' turn-end accuracy to be higher for the (tonal) jazz than the (non-tonal) free improvisations. We further hypothesised that if tonal cues are critical for predicting the end of a musical turn, removing low frequencies should impair accuracy for jazz improvisations more than free improvisations. We included a control manipulation (lowpass filtering) to ensure that effects were not due to general spectral depletion.

2. Methods

2.1. Participants

Forty-seven participants took part in the study.[1] We excluded participants if they responded to < 70% of stimuli (1 participant), or produced responses that fell > 2.5SD from the participants' overall mean (1 participant). Of the remaining 45 participants, 29 were female, and ages ranged from 18 to 65 (M = 27, SD = 11). We recruited participants by email, either within the university or through musician networks, to ensure a range of musical training. A standard musical questionnaire was used to record extent of musical training on a 7-point scale (0/0.5/1/2/3–5/6–9/10 or more years; Müllensiefen, Gingras, Stewart, & Musil, 2011), revealing that participants had a broad range of musical training experience. Nine participants had no training, one had 1 year, four had 2 years, 11 had 3–5 years, 11 had 6–9 years, and nine had 10 or more years.

2.2. Materials

The experimental items were extracted from the recordings by Moran and Keller (2016), which included idiomatic duets from six pairs of jazz improvisers and six pairs of free improvisers. The jazz improvisers followed the harmonic progression of a jazz standard (*Autumn Leaves*, J. Kosma, 1945), and played with a regular pulse in the range of 400–500 ms (equivalent to a range of 120–150 beats per minute). Meanwhile, the material generated by the six pairs of free improvisers was characterised by the absence of consistent and predictable tonal structure, and the absence of a regular and reliable pulse (and hence metre). All improvisers specialised in the style they performed and had at least 7 years of performance experience. The final phrases of 90 solo improvisation turns within these recordings were determined (47 jazz, 43 free) with phrases defined as perceptually complete musical units as judged by the first author (a music graduate). These phrases were taken from a point of silence until the end of the final note (endpoint determined in Logic Pro), and did not include any sounds other than those of the instrument.

The maximum f0 in these stimuli was 1846 Hz and the mean was 336 Hz. Ritsma (1967) found that frequencies of 500–1500 Hz are most important to determine the pitch of complex tones with an f0 of up to 400 Hz, and hence we generated a *highpass* condition excluding all f0 frequencies (threshold 2100 Hz), and a *lowpass* condition incorporating all f0 frequencies (threshold 2100 Hz). Highpass filtering obscured the pitch of the jazz and free improvisations, and, as tonality is based on pitch, the tonal framework of the jazz improvisations. Lowpass filtering, on the other hand, did not affect the pitch of the jazz and free improvisations but provided a spectrally reduced control. We additionally included the *normal* condition (original recordings). Stimuli in each condition were equated in root mean square amplitude to level the extracts in loudness.

Each participant heard one version of each item. Items were divided into three groups (approximately matched in distribution of improvisation styles, durations, and instruments), and each participant heard one group in each condition (i.e., 30 normal, 30 highpass, and 30 lowpass stimuli), in an individually randomised order (see Table 1).

2.3. Procedure

Each participant was randomly assigned to one of three stimulus lists (15 participants per list), and instructed: ‘Press the button at the moment you think the musician will have finished their solo. Try to anticipate the end of their turn.’ Participants were warned that some stimuli would sound unusual, but to give their best guess as to when the solo would be finished. The experiment began with a practice block, in which one example of each spectral manipulation was presented, followed by the main experiment. On each trial, the participant pressed a button to indicate readiness, after which a fixation cross was presented for 500 ms, followed by the stimulus. Participants responded by pressing the button again, which cut the music and ended the trial. (The music did not continue after the button press, as we judged that participants might then wait for the end of the performances before responding.) If participants did not press the button during a trial, a time out occurred 2 s after the end of the stimulus. The procedure then repeated. Participants were told that they could take breaks between trials. Button press timings were recorded relative to the start of the stimulus. The lack of a button press was recorded as a missed trial.

2.4. Design and data analysis

The experiment involved a 2 Improvisation (free, jazz) × 3 Spectral Manipulation (normal, highpass, lowpass) design. After exploring the

[1] As there was no way to estimate effect size a priori, sample size was determined through reference to studies using similar paradigms in language (cf. Magyari & de Ruiter, 2012).