Interpreting tone of voice: Musical pitch relationships convey agreement in dyadic conversation

Brooke M. Okada and Lorin Lachs
Department of Psychology, California State University, Fresno, 2576 East San Ramon ST11, Fresno, California 93740
brookeokada@gmail.com, llachs@csufresno.edu

Benjamin Boone
Department of Music, California State University, Fresno, 2380 East Keats MB77, Fresno, California 93740
bboone@csufresno.edu

Abstract: Previous research has found that the musical intervals found in speech are associated with various emotions. Intervals can be classified by their level of consonance or dissonance—how pleasant or unpleasant the combined tones sound to the ear. Exploratory investigations have indicated that in an agreeable conversation, the pitches of the last word in an utterance and the first word of a conversation partner’s utterance are consonantly related; in a disagreeable conversation, the two pitches are dissonantly related. The present results showed that the intervals between the tonics of the utterances in a conversation corresponded to the agreement between interlocutors.

© 2012 Acoustical Society of America
PACS numbers: 43.71.Bp, 43.75.Cd [DD]
Date Received: May 9, 2012 Date Accepted: July 20, 2012

When two musical notes are played at the same time, the disparity between their pitches can be expressed in terms of a musical interval. Musical intervals can further be broadly classified as either consonant or dissonant. In musical terminology, the consonance or dissonance of an interval is loosely based on the way the sound waves comprising each pitch interact when perceived. The particular sounds and intervals considered pleasant or unpleasant varies among cultures; in Western music, consonant sounds are perceived as pleasant or stable, and dissonant sounds are perceived as unpleasant or unstable (Schmidt-Jones, 2010; Worrall, 2008).

In musical composition, pitch relationships are often used to convey emotion. For example, pieces in major keys are generally regarded as bright and happy, while pieces in minor keys are usually perceived as dark or somber. Thus when Chopin composed a prelude in the key of E minor, the perceived darkness of the key allowed him to express an intended perpetual sadness and isolation (Aitken, 1997). The consonance or dissonance of a pitch interval is similarly used to convey emotion. For example, dissonant intervals are used to build tension; Wagner used prolonged dissonances to heighten the emotional tension of the actors on stage and to display their drama in musical terms (Heisterman and Weinstock, 2007).

Music, like speech, is an acoustic signal that humans use to express emotion. This similarity appears to be more than just superficial; evidence has shown that the areas of the brain that process music and language overlap and that they are so deeply connected that music cognition is vital to language development in babies.
Cross (2003) asserts that music emerged from emotional speech and that music has had an evolutionary purpose as a means of creating and providing for communication within a group. Music can be used as a communication device in a variety of ways: It can provide a manner by which one can express emotions in a way that words cannot, it can cause physical responses like relaxation or stimulation, and it can even change our moods (Thayer, 1996). It has been claimed that, as it does for speech, the brain contains innate systems for music perception, which allow for nonverbal, musical communication (Gaston, 1968 as cited in Hallam, 2006).

Remarkably, musical relationships in the pitch contours of speech (which may be thought of as music-like) have been shown to communicate emotion. In an interesting explicit comparison of the way music and speech convey emotions, Curtis and Bharcha (2010) examined the pitch contours exhibited in emotional speech, predicting that observed speech contours would resemble those of the pitch changes used to express emotion in Western music. Participants were asked to read a story containing passages spoken by characters exhibiting a variety of different emotions. It was found that people reenacting a sad utterance used significantly more intervals of a minor third than any other interval. This use of the minor third in sad speech mimicked the use of the minor third in Western music to portray sadness (Curtis and Bharcha, 2010).

The connection between musical and spoken emotion expression has also been observed in the key of a spoken pitch contour. Schreuder, Van Eerten, and Gilbers (2005) analyzed the pitch contours of people reading a happy passage and a sad passage. They found that the pitch contours of those reading the happy, cheerful passages contained significantly more intervals of a major third, common to major modalities, and that the pitch contours of people reading sad passages contained intervals of a minor third, common to minor modalities (Schreuder et al., 2005).

The use of pitch as a mechanism for communicating emotion suggests that pitch can be used for communicating nonverbal relationships between interlocutors. Along these lines, Gregory, Dagan, and Webster (1997) found evidence of the use of musical relationships for conveying interlocutor relationships in dyadic conversation. For instance, the fundamental frequency of a speaker’s voice changes when he or she is speaking to someone of a higher social status. The speaking frequency of the lower status partner grows closer to that of the partner with the higher status. This phenomenon is so robust that pitch analysis of the spoken exchanges in recorded interviews can predict naïve listeners’ ability to predict the relative social status of the two conversation partners (Gregory and Webster, 1996).

Boone (2003) has proposed that the pitch relationships observed in conversational speech contain musical intervals for appropriately relaying interlocutor relationships in dyadic conversation. In an acoustical analysis of a dialogue between talk show hosts Regis and Kathie Lee, it was shown that the speech pitches observed in their conversations contained some of the harmonies found in Western music and that these harmonies were used to convey emotions similar to their use in Western music. For example, because Regis and Kathie Lee had many agreeable conversations, it was expected and observed that the pitches of their speech, when exchanging dialogue, often matched in unison or were consonantly related. In fact, this “pitch synchrony” was so great that when Regis finished an utterance and Kathie Lee continued it, she would often start on the same pitch. It was also observed that when Kathie Lee interrupted Regis, her chosen pitch was dissonantly related, conveying her disagreement.

In a further investigation of this phenomenon, Boone, Basiletti, Bissel, Olson, Peters, Sherrill, and Lachs (2007) conducted an acoustic analysis of speech pitch in these prerecorded televised interviews. To analyze the pitches, music graduate students took an ear training and interval identifying diagnostic test. From this set, only those students who consistently received exceptional scores on the diagnostic were asked to transcribe the pitches of the interviews. The students identified the intervals used in the dyadic exchanges, and these pitches were corroborated by computer analysis of the
pitches. A number of different intervallic relationships were found between the last pitch of one’s utterance and the first pitch of his or her conversation partner’s responding utterance. For example, many of the dyads showed a synchronization of pitches by using the three most consonant of intervals when exchanging utterances: a perfect unison, octave, and perfect fifth. Although these observations were made on naturalistic utterances, the evidence has suggested that musical relationships are used to convey emotions relevant to interlocutor agreement in spoken conversation.

In the present study, we observed the musical pitch relationships utilized by participants having agreeable or disagreeable conversations. To provoke these agreeable and disagreeable conversations, we used the Map Task Corpus from the Human Communication Research Centre at the University of Edinburgh (2010). In the Map Task, each participant is given a map of his or her own. One map has a drawn route interweaved within landmarks, and the other has no drawn route. The two participants are separated so they cannot see each other but can hear each other. The participant who receives the map with the drawn route is instructed to give directions to the other participant so that he or she can collaboratively trace the same route on the unmarked paper. In one condition, the maps given to both the “information giver” and the “information follower” contain the same landmarks in the same locations on the maps. We operationally defined conversations elicited in this condition as “agreeable.” However, in the other condition, the maps given to both participants mismatch: landmarks are either absent or in different locations on the different maps. We operationally defined conversations elicited in this condition as “disagreeable.”

The evoked conversations in these two conditions were examined for evidence of pitch synchronization—the use of musical intervals to convey interlocutor agreement in speech. We predicted that the agreeable conversations would contain more consonant musical intervals between participants, while the disagreeable conversations would contain more dissonant musical intervals.

1. Method
1.1 Participants
Twenty-four participants recorded speech samples in pairs to elicit dyadic conversation for a total of 12 dyads. All participants were undergraduate students recruited from the Introduction to Psychology pool at California State University, Fresno, and all reported that English was their native language.

1.2 Materials
For each pair, one participant was recorded with an AKG Emotion C900 condenser microphone while the other was recorded with a CAD GXL2200 large diameter condenser microphone. The participants were placed in separate rooms to allow the recordings for each participant to be contained on different audio tracks. The microphones were hooked up to an Apogee Duet, which allowed both vocal inputs to be simultaneously recorded to a MacBook Pro laptop using Logic Express. Two sets of Beyer Dynamic DT150 headphones were also connected to the Apogee Duet to allow the participants to hear each other.

Two sets of matching maps and two sets of mismatching maps were used across all participants. Each matching map had a corresponding mismatching map, containing the same route and landmarks on the map for the instruction giver.

1.3 Procedure
Participants were seated in two different rooms, each with a microphone and headphones. Each was also given his or her own map. The dyads were randomly assigned the roles of “information giver” and “information follower.” Before the recording started, the information giver was instructed to give the best directions to the information follower so that he or she could trace the route on his or her own map. The participants were told to communicate as much as possible and to use the names of the
landmarks to reproduce the route correctly. They were also asked to try not to talk over each other. Finally, they were told to “continue and talk it out” until getting to the finish point if there was any confusion. Once the “follower” reached the finish point, the recording was stopped.

Each dyad participated in both the agreeable and disagreeable conditions. One of the matching map sets was given to the dyad for the agreeable condition, and the other mismatching map set was given to the dyad for the disagreeable condition. Each participant was the “instruction giver” for one condition and the “instruction follower” for the other condition. A counterbalancing procedure was used to control for sequence effects.

2. Results
2.1 Scoring

When extracting the samples for data analysis, we chose the portion of the conversation concerning the route near two key landmarks in the mismatching maps and two analogous key landmarks in the matching maps. The key landmarks in the mismatching maps were the first two from the starting point that did not correspond with each other across versions of a dyad’s map set. Non-correspondences were defined as landmarks for which one participant’s landmark was either completely missing from the map or located in a different place on the map of the other participant in the dyad. In the corresponding matching maps, we analyzed utterances around the landmarks with the same serial positions as those chosen for the mismatching maps. For the conversational exchange surrounding each key landmark, the intervallic relationships between three utterances were examined. The three utterances analyzed were those that occurred when: (1) information giver said the name of the key landmark in instructions, (2) information receiver replied, and (3) information giver replied again. In other words, these three utterances immediately followed one another in the exchange, with no intervening words or sentences. This procedure resulted in 96 speech samples in total; however, there were two utterances for which the software could not determine a key, and these were excluded from all analyses.

Each of the utterances in a speech sample underwent separate pitch extraction in PRAAT, a software package for acoustical analysis (Boersma and Weenink, 2010). Thus pitch extraction for each utterance was carried out over a long interval containing multiple words. Because we used the default pitch extraction parameters in PRAAT, pitches were extracted at a rate of 100 pitches/s. Sinusoidal tones were then generated at the pitches and durations measured during pitch extraction. The tones associated with each utterance were submitted to MELODYNE EDITOR, a music editing software package, which rounded the pitches to the nearest semitone, produced a musical staff and key signature (or, tonicity) and converted the file into MIDI format (MELODYNE EDITOR, 2011). The MIDI was then imported into SIBELIUS, a music writing software package, which output musical staves and notes with rhythmic notation (SIBELIUS, 2008). After the musical key for each of the three utterances in an exchange was ascertained from MELODYNE EDITOR, the interval between the keys for each utterance in an exchange was determined.

Fisher’s exact test was used to analyze the frequency of occurrence of either consonant or dissonant intervals for matching and mismatching map conditions. Intervallic inversions were treated as equivalent, and the dividing line for classification as consonant or dissonant was between a minor 3rd/major 6th and a major 2nd/minor 7th. Table 1 shows the incidence of intervallic relationships for each type of map. Fisher’s exact test showed a significant relationship between interval classification and map type, one-tailed Fisher’s exact test, $P = 0.032$, indicating that the type of map influenced the number of exchanges that were consonantly or dissonantly related.

Figures 1 and 2 show the distribution of each type of intervallic relationship for the two map types. As seen in Fig. 1, there were no tritones observed in any
matching maps. Figure 2 shows that perfect fifths and perfect fourths were much more common in the matching maps than in the mismatching maps.

3. Discussion
3.1 Summary of findings

It was predicted that dyads conversing about matching maps would use more consonant intervals when communicating and that those conversing about mismatching maps would use more dissonant intervals when communicating. The results showed that there was a relationship between map type and intervallic classification when the intervals were classified as either consonant or dissonant. In addition, examination of the frequency of occurrence of each type of interval showed that tritones—considered the most dissonant interval possible—did not occur in the agreeable condition but did occur in the disagreeable condition. Furthermore, perfect fifths—the most consonant

Table 1. Frequency of occurrence of consonant and dissonant intervals for each map type.

<table>
<thead>
<tr>
<th>Interval classification</th>
<th>Consonant</th>
<th>Dissonant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching</td>
<td>27</td>
<td>19</td>
<td>46</td>
</tr>
<tr>
<td>Mismatching</td>
<td>18</td>
<td>30</td>
<td>48</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>49</td>
<td>94</td>
</tr>
</tbody>
</table>

*Incidence of relationships in matching and mismatching maps

Fig. 1. Distribution of intervallic relationships for matching maps.
interval that does not involve pitches that are separated by octaves—were much more common in the agreeable condition than they were in the disagreeable condition.

The present results add to the literature concerning interactive speech alignment (e.g., Dias and Rosenblum, 2011; Pardo et al., 2012) in which the speech of interlocutors converges along acoustic-phonetic dimensions. For example, Pardo (2006) used a map task analogous to the one we used to elicit speech samples (although the agreement of the maps was not a manipulated variable used for analysis). Rather the map task was a vehicle for the elicitation of shared words that could later be compared for the presence of phonetic alignment. The results showed that the utterances of one member of a dyad were judged to be more perceptually similar to those of the other member when those utterances were taken from the map task conversation rather than from pre- and post-map task utterances. Pardo (2006) proposes that alignment may serve a social function in that it might contribute to comprehension, rapport, or group cohesion.

Such an explanation is compatible with our results: After all, it is plausible that group cohesion is strengthened when interlocutors agree with one another or share some common frame of reference. However, it should be noted that the results presented here differ from the existing work on phonetic alignment in the dimensions along which the alignment is occurring. While phonetic alignment shows a tendency for phonemes to become more perceptually similar, our results point to a tendency for alignment for pitch to occur on the musically relevant dimension of tonality. In other words, interlocutors are not always merely imitating each other’s pitches (although that does happen, as in the occurrence of unison relationships) but rather appear to be relating to each other in ways that are relevant to a socially relevant higher-order variable of pitch relationships: consonance. Our results support the hypothesis that Californian members of a conversing dyad use musical intervals common in the Western musical tradition to convey agreement during speech.

More sensitive analyses of the particular musical intervals used by interlocutors remain to be conducted. In addition, there are reasons to suspect that the results presented here underestimate the prevalence of the effect. For example, because consonance and dissonance fall on a continuum, the musical relationships in the middle of the continuum like a major sixth and a minor seventh are the most harmonically similar. Thus when we categorized the intervals as either consonant or dissonant, these similar intervals were separated and may have affected our analyses. Another problem with the present results may have resulted from our operational definition of agreeable
and disagreeable utterances. During our analyses, we observed that there were convers-
sations that seemed superficially to be agreeable when there were missing landmarks
and some conversations that seemed superficially to be disagreeable on the matching
maps. A more fine-grained analysis, involving coding of the semantic content of the
various utterances made in each exchange, might reveal a more robust effect. However,
with these caveats in mind, it appears that musical intervals play a role in conveying
interlocutor agreement in spoken communication.

References and links
Westport, CT).
program].” http://www.praat.org (Last viewed November 1, 2010).
International Conference on Arts and Humanities, http://www.hichumanities.org/AHProceedings/ (Last
viewed September 9, 2011).
acoustic analysis of musical patterns of pitch variation in dyadic conversation,” Poster presented at the
28th Annual Central California Research Symposium, Fresno, CA.
Lond. 40(12), 1457.
in conversation partners’ fundamental frequencies to perceptions of communication quality,” J. Nonverbal
Behav. 21(1), 23–43.
Gregory, S. W., Jr., and Webster, S. (1996). “A nonverbal signal in voices of interview partners
effectively predicts communication accommodation and social status perceptions,” J. Pers. Soc. Psychol.
70(6), 1231–1240.
(Celemony).
119, 2382–2393.
roommates,” J. Phone. 40(1), 190–197.
latest/ (Last viewed October 31, 2010).
dissertation (University of Groningen, Groningen, Netherlands).
University Press, Oxford).
PPoM/consonance/percept.html.