The Effect of Subtitles on Listeners’ Perceptions of Expressivity


<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DOI</td>
<td>10.1177/0305735612463951</td>
</tr>
<tr>
<td>Publisher</td>
<td>Sage Publications</td>
</tr>
<tr>
<td>Version</td>
<td>Accepted Manuscript</td>
</tr>
<tr>
<td>Terms of Use</td>
<td><a href="http://cdss.library.oregonstate.edu/sa-termsofuse">http://cdss.library.oregonstate.edu/sa-termsofuse</a></td>
</tr>
</tbody>
</table>
The Effect of Subtitles on Listeners’ Perceptions of Expressivity

Jason M. Silveira
Department of Music
Oregon State University
jason.silveira@oregonstate.edu

Frank M. Diaz
School of Music and Dance
University of Oregon
fdiaz@uoregon.edu

Abstract

The purpose of this study was to determine what effect, if any, subtitles would have on listeners’ perceptions of expressivity in an operatic performance. Specifically, this study addressed the following research questions: (1) Will there be differences in perceived expressivity among three listening conditions (audio only, audio + video, audio + video with subtitles)? (2) Will listening condition have an effect on listeners’ magnitude of response? (3) Will listening condition have an effect on listeners’ stated focus of attention during the listening task? (4) Where during the stimulus will listeners perceive moments of expressivity? A 13-minute excerpt from a live production of Puccini’s La Bohème was used as the music stimulus.

Participants (N = 103) were randomly assigned to the experimental (audio + video, audio + video with subtitles) and control (audio only) groups. Continuous data were collected via the Continuous Response Digital Interface (CRDI), and summative data were collected via a post hoc questionnaire. Results revealed significant differences in listeners’ continuous data among all three groups, with the audio condition evidencing the highest response magnitude, and the subtitles group receiving the lowest response magnitude. No significant differences were found among the groups with respect to summative perceptions of expressivity or focus of attention. Implications of these findings and suggestions for future research are discussed.

Keywords: Expressivity, Subtitles, Focus of Attention, CRDI, Perceptions
The Effect of Subtitles on Listeners’ Perceptions of Expressivity

Researchers have long been interested in how listeners attend to music, and what specific elements they attend to (Adams, 1994; Berlyne, 1974; Burnsed, 2001; Byrnes, 1997; Geringer & Madsen, 1995/1996; Madsen & Geringer, 1990; Madsen & Geringer, 2000/2001; Madsen, Geringer, & Fredrickson, 1997; Madsen & Kuhn, 1994; Price, 1983). This line of research has also included an examination of the effect of extra-musical factors (i.e., visual information) on performance evaluation (Madsen, 2009; McClaren, 1985; Morrison, Price, Geiger, & Cornacchio, 2009), listeners’ affective and cognitive responses (Geringer, Cassidy, & Byo, 1996, 1997), felt emotional response (Adams, 1994), perceived aesthetic response (Lychner, 2008), perceived musical tension (Frego, 1999), and perceived expressiveness (Broughton & Stevens, 2009; Davidson & Correia, 2002; Juchniewicz, 2008; Wöllner & Auhagen, 2008). Specifically, this research has examined the effect of a visual stimulus as an extra-musical factor on listeners’ perceptions of the listening experience.

Music performance carries with it a number of stimuli for both the aural and visual senses, and it has been suggested that this sensory integration can enhance the listening experience. While most live music performances facilitate this sensory integration, it is particularly evident in large operatic productions with their ornate scenery, stylized wardrobe, and the interweaving of the orchestra in the pit with the vocalists on the stage. Coordinating these elements to form a cohesive whole is no doubt a daunting task for the production designer. Copland (2002) poignantly stated the dilemma faced by many opera composers, namely, how to equalize and
balance the different elements in an opera (costumes, scenery, music, and text) to achieve a satisfying whole. This dilemma has also been a topic of discussion for many opera producers and composers, perhaps most notably by Richard Wagner, with his philosophy of unifying all works of art via the theater – gesamtkunstwerk (Tanner, 1997). Singing operatic music also carries with it additional complexities including diction and clarity of text, and these issues can be more troublesome when the language being sung is foreign to that of the listeners. For centuries, opera companies have dealt with the issue of staging operas in foreign languages, and the ensuing matter of text comprehension for audience members. As early as 1712 Addison (cited in Burton, 2010) crystallized the desire for some sort of translation of non-English operas with the following statement:

[O]ur great-grandchildren will be very curious to know the reason why their forefathers used to sit together like an audience of foreigners in their own country, and to hear whole plays acted before them in a tongue which they did not understand. (p. 179)

The importance of text comprehension in opera has been well documented throughout the history of opera performance. For almost 300 years, audience members brought copies of the libretti with them to enhance their experience at the opera (Germano, 2010). As the demand for “singable translations” of operas increased, so too did the philosophical and logistical arguments (Should the work of art be altered to suit non-native speakers? Will a translation distort the original meaning of the text? Will the translated text “fit” musically with the melodic line or rhyme scheme?). The debate over singable translations has continued since it began
in earnest at the turn of the 20th century. Advocates of singable translations acknowledged both the difficulty and importance of providing singable translations that conveyed the meaning of the source text (Spaeth, 1915). As technology has progressed, however, a new solution to the problem of text comprehension was proposed – surtitles. The act of creating surtitles has been described as a kind of hybrid form of translation and interpretation (Griesel, 2009). Those who create surtitles are charged with translating the words from a source text and creating an interpretation of that text, which becomes the target text for the audience. Surtitles are a unique translation of a source text, and do not present the same problems associated with a sung translation, where textual choices are constrained by the need to make the target text fit the notes and tempo of the original score (Mateo, 2007a).

The practice of projecting a text translation in an opera was first developed in 1983 by the Canadian Opera Company to aid in comprehension of the sung text. It was suggested that this translation would provide audience members with a better understanding of the opera and fuller enjoyment of the experience (Rich, 1984). Projecting text translations has also gained prominence in theatre as well. The process of providing subtitles to theatregoers first began in the 1980s in Scandinavia, and continues today (Griesel, 2009). Text projections were designed to be discrete, and were never meant to replace the source text but rather to compliment it. The initial reactions to the use of text projections in the opera house were mixed, and still continue today. While reactions have been predominantly
positive in Canada and Spain, text projections met with initial reluctance in New York, and have received mixed reactions in Great Britain (Mateo, 2007b).

The terminology for the practice of projecting translated text abounds (surtitles, subtitles, supratitles, supertitles). For the purposes of the present study, the term “subtitles” will be used for clarification and ease of reference. Some have argued against the use of subtitles in performance. Detractors have suggested that performers’ nonverbal information and the tonal sequences of the source language can convey the message of the production to the target language audience (Griesel, 2009). This view is rather poignantly expressed by music critic Andrew Clements (2000) with the following quote:

When opera ceases to be a medium in which the drama is presented through the symbiosis of the music and words, and instead is turned entirely into an exercise in reading, with the addition of some more or less engaging background music, then the medium is devalued, and it really will become museum art of a vacuous and purely decorative kind. (para. 6)

Time and effort expended reading subtitles has been cited as a possible detriment to the overall operatic experience (Virkkunen, 2004). However, given the difficulties surrounding the fidelity of singable translations, some have argued that subtitles provide an inexpensive solution to this problem (Orero & Matamala, 2007).

Supporters of subtitling have countered that subtitles are another method of communication and part of the Gesamtkunstwerk of opera that interacts with other symbolic modes used in the performance to create meaning (Virkkunen, 2004). Advocates posit that while reading subtitles may require more effort, they can
provide a more enriching, active, enjoyable experience, and perhaps even contribute to increased opera attendance (Mateo, 2007b).

While proponents of subtitling have suggested the translated text compliments the music, critics have argued that it distracts the audience from experiencing the gestalt nature of an operatic performance (Botstein, 1994; Low, 2001), and serves as a lexically weaker and pragmatically less effective translation of the text (Bruti & Perego, 2010). In either case, however, it appears that the importance of the text in vocal music is integral, and that the words are meant to accommodate the music and enhance the listening experience (Germano, 2010; Hackworth & Fredrickson, 2010; Peretz, Radeau, & Arguin, 2004; Skuggevik, 2010).

It has been suggested that formalized enhancement of a musical experience, such as program notes, structural descriptions, dramatic descriptions, and mental imagery can enhance the listening experience. The research in this area, however, is not without conflict. While some studies have suggested that formalized enhancements can have a positive effect on listeners’ value judgments of a piece of music (Bradley, 1972; Damon, 1933; Halpern, 1992; Zalanowski, 1986), other studies have not (Brown, 1978; LaBerge, 1995; Margulis, 2010; Prince, 1974).

In an early study, Damon (1933) found that program notes representing emotional or story-like programs resulted in greater enjoyment of listeners. In another study utilizing extra-musical information, Halpern (1992) examined nonmusicians’ value judgments of musical stimuli under three different conditions: analytical information with music, historical information with music, and music alone. Her results revealed that the historical information group evidenced
significantly higher value scores than the other two groups. A similar study by Zalanowski (1986) utilized both programmatic music and absolute music to determine listeners’ ratings for attention, enjoyment, and understanding across several different conditions: pay attention, form free mental imagery, follow a story program, follow an abstract verbal program, follow a concrete analytical program. Her results were somewhat mixed, however. The mental imagery condition led to the greatest enjoyment for both types of music (programmatic and absolute), while the abstract verbal program and concrete analytical program conditions did not evidence increased enjoyment.

Other studies have demonstrated conflicting results, suggesting that increased musical knowledge may not necessarily lead to increased enjoyment. Prince (1974) examined the effect of guided analytical listening on preference of junior high school students. He discovered that the analytic commentary did not increase preference ratings among the participants. Similarly, Brown (1978) ascertained that video instruction of musical concepts did not increase elementary school children’s preference for or attitudes toward school music, leading her to conclude that knowing about music does not necessarily lead to valuing that music.

In a more recent study, Margulis (2010) had participants rate their level of enjoyment of music stimuli under one of three conditions: listeners provided a dramatic description of the music, listeners provided a structural description of the music, and music alone. Results indicated that both types of descriptions had a significant negative effect on enjoyment, leading Margulis to conclude, “listening to music in terms of linguistic descriptions may in fact be less enjoyable” (p. 298).
Much research exists regarding listeners’ perceptions of audio and audio/video stimuli (Adams, 1994; Broughton & Stevens, 2009; Dahl & Friberg, 2007; Davidson, 1993; Di Carlo & Guaïtella, 2004; Gillespie, 1997; Madsen, 2009; McLaren, 1985, Vines, Wanderly, Krumhansl, Nuzzo, & Levitan, 2004). With the increased use of live operatic subtitles and the commercial availability of DVD recordings of live operatic performances, an examination of the effect of this visual component on listeners’ perceptions of the listening experience seems warranted. The present study attempted to determine what effect operatic subtitles might have on listeners’ perceptions of expressivity in the performance; would they enhance or detract from the overall listening experience?

Therefore, the following research questions were addressed:

1. Will there be differences in perceived expressivity among the three listening conditions (audio only, audio + video, audio + video with subtitles)?
2. Will listening condition have an effect on listeners’ magnitude of response?
3. Will listening condition have an effect on listeners’ stated focus of attention during the listening task?
4. Where during the stimulus did listeners perceive moments of expressivity?

Method

Participants

Participants for this study ($N = 103$) included undergraduate and graduate music students enrolled at two comprehensive universities in the Pacific Northwest. Participants included both instrumentalists ($n = 61$) and vocalists ($n = 42$). To remove text comprehension as a confounding variable, only those individuals who
did not speak the source language (Italian) were allowed to participate in the present study.

**Materials**

The audio/video stimulus for this study was a recording from the first act of Giacomo Puccini’s *La Bohème* (Zeffirelli, 2008). This particular musical excerpt was purposefully chosen for a number of reasons: (1) It has been used in a long line of research addressing constructs like “aesthetic response,” (e.g., Madsen, 1997) “musical tension,” (e.g., Fredrickson, 2000) and “expressivity” (e.g. Silveira & Madsen, in press); (2) Previous studies using this piece of music have provided a consistent data set for comparisons; (3) If differences among listening conditions were observed, then it is more likely that the differences can be attributed to the listening condition, and not the specific piece of music; (4) a plurality of participants in previous studies have evidenced a high magnitude of response with this piece (e.g., Byrnes, 1997; Madsen, 1997; Madsen & Geringer, 2008) and (5) It is sung in a language foreign to the listeners, thus allowing isolation of the text variable as a possible influencing factor on perceived expressivity (e.g., Hackworth & Fredrickson, 2010). The researchers asked two experts with over 50 years experience in operatic production and musicology to recommend a video recording of *La Bohème* that they considered to be of “highest artistic quality.” The production used in the present study (Zeffirelli, 2008) appeared on each expert’s list of “highest artistic quality” performances, and was chosen after considerable deliberation among the experts and the researchers.
The specific excerpt selected, which lasted 13 minutes, is the conclusion of Act 1 from *La Bohème*. It has been used in a long line of aesthetic response and musical tension research (Madsen, 1997; Madsen, Brittin, & Capperella-Sheldon, 1993; Madsen, Byrnes, Capperella-Sheldon, & Brittin, 1993; Madsen & Geringer, 2008; Madsen & Napoles, 2005; Southall, 2003). A professional digital recording engineer and videographer created the stimulus DVD to retain the DVD’s original video and audio quality.

A listener’s evaluation and enjoyment of music is a situation that unfolds over time. As a result, many of the static measurements available to the music researcher only offer a “snapshot” of the total listening experience. The Continuous Response Digital Interface (CRDI) is one type of measuring device designed to measure responses over time. Developed in the late 1980s, the CRDI was intended to allow listeners to respond non-verbally during ongoing music and/or during visual presentations. The CRDI allows listeners/viewers to respond differentially across time, which is a function not afforded to such static measurements as Likert scales, Osgood semantic differential scales, and behavioral checklists. As a tool to measure ongoing responses, the CRDI has been widely used to measure focus of attention, evaluation, aesthetic response, discrimination, and perception, since listeners’ affective or aesthetic responses change throughout the listening experience. Since this device has been used in numerous affective response and musical tension studies, and previous studies have illustrated differences in listeners’ continuous and summative responses to music (Brittin & Duke, 1997;
Duke & Colprit, 2001), it was decided to use this measuring device to measure ongoing perceptions of expressivity.

**Procedures**

Participants were randomly divided into one of three groups. Participants in experimental group 1 ($n = 37$) viewed the audio + video stimulus with no subtitles. Experimental group 2 ($n = 35$) viewed the audio + video stimulus with subtitles. The control group ($n = 31$) listened to the audio portion of the recording while viewing a black screen. Participants were seated at a table in groups of three or fewer, and were partitioned from each other so that participants would be out of view from each other. Thus, each participant was isolated in his/her own listening and viewing station with a Continuous Response Digital Interface (CRDI) dial placed in front of him/her on the table. They were all instructed to listen to the recording while manipulating the CRDI dial to represent their perceptions of expressivity during the performance.

The graphic overlay (see Figure 1) placed on the CRDI was identical to the one originally created by Madsen and Fredrickson (1993) and subsequently used in a number of tension (Fredrickson, 1995; Fredrickson, 1997; Fredrickson, 1999; Fredrickson, 2000; Fredrickson, 2001; Fredrickson & Coggiola, 2003; Fredrickson & Johnson, 1996), flow (Diaz, 2011), and expressivity (Silveira & Madsen, in press) studies. The initial position of the CRDI dial was in the leftmost position, representing 0 degrees (from 0 to 255) at the “LESS” anchor. Upon entering the room, participants were given the following instructions:
You will now hear an excerpt from Giocomo Puccini’s Opera, *La Bohème*. The excerpt is from a live performance of the Metropolitan Opera with Nicola Luisotti conducting. The singers on the recording include soprano Angela Gheorghiu, and tenor Ramón Vargas. Music from the first act will be played in its original order, and include *Che gelida manina* (tenor aria), *Mi chiamano Mimi* (soprano aria), *Ehi Rodlofo!* (a short transitional interlude), and *O soave fanciulla* (soprano and tenor duet). As you listen to the music please move the dial to indicate your perception of expressivity. Do you have any questions?

As in previous studies related to aesthetic response, musical tension, and expressivity, participants were not given a specific definition of the construct “expressivity.” Rather, each listener brought with them their own interpretation of expressivity to the listening experience.

After listening to the excerpt, participants were asked to complete an exit questionnaire. This questionnaire contained questions asking whether they perceived expressivity while listening to the performance, their level of focus of attention, how long the expressivity lasted, and the magnitude of the musical experience. The questionnaire material was similar to that used in aesthetic response and musical tension research (Hackworth & Fredrickson, 2010; Madsen, Brittin et al, 1993; Madsen, Byrnes et al, 1993; Madsen & Geringer, 2008).

**Results**

The first research question concerned the effect of subtitles on listeners’ perceptions of expressivity. Participants’ self-report data indicated that all
participants perceived expressivity in the recording. Continuous data were also collected to examine listeners’ perceptions of expressivity in the performance. These data were collected in numerical format (ranging from 0 at the “LESS” anchor to 255 at the “MORE” anchor) using the Continuous Response Digital Interface (CRDI), which sampled participants’ responses every .5 seconds. Participants’ continuous responses were aggregated to create an overall response graph for each of the three conditions (see Figure 2). Visual inspection of the graphic data contours revealed that each group responded similarly regarding moments of high and low expressivity, represented by the graph’s peaks and nadirs respectively. As in previous studies utilizing this music as a stimulus, participants indicated three distinct moments of “high expressivity,” and two moments of “low expressivity” (Madsen, Brittin, et al., 1993; Madsen, Byrnes, et al., 1993; Madsen & Geringer, 2008; Madsen & Napoles, 2005; Silveira & Madsen, in press; Southall, 2003).

Based on an examination of Figure 2, the first distinct moment of “high expressivity” occurred in the last eight measures of the tenor aria (475 half seconds), which coincided with dense orchestration, a loud dynamic marking, and a high vocal tessitura. The second distinct moment of high expressivity occurred approximately two-thirds of the way through the soprano aria (1020 half seconds), which also coincided with dense orchestration, a loud dynamic, and a high vocal tessitura. It was also accompanied by a momentary slowing of the tempo. The third (and highest) moment of perceived expressivity occurred halfway through the tenor/soprano duet (1453 half seconds). As in the previous two highly expressive moments, this moment included a loud dynamic, dense orchestration, high vocal
tessitura, and also included octave unisons in both the vocal duet and orchestral accompaniment. The first moment of distinct “low expressivity” occurred at the end of the tenor aria immediately before the first applause break (538 half seconds), in which the orchestra is playing at a soft dynamic level on mostly sustained notes. The second moment of distinct low expressivity occurred during the interlude in which there is little orchestral accompaniment, and the vocalists are performing in a recitativo style. Additional moments of high and low expressivity are displayed in Tables 3 and 4.

Regarding the second research question, data were analyzed to determine if there were differences in response magnitude among the three listening conditions. The overall aggregated continuous data were compared using a one-way analysis of variance. While there continues to be some debate regarding the use of statistical analyses for continuous data points (Schubert, 2010), these types of analyses have been suggested as appropriate for general comparisons of the overall magnitude of related continuous data points (Levitin, Nuzzo, Vines, & Ramsay, 2007). Raw data consisted of mean responses for each group for each half second of recorded data (1566 half seconds). Results indicated a significant difference among the three groups $F(2, 4692) = 158.77, p < .001$. Pairwise comparisons were made using Tukey’s HSD post-hoc tests. Results revealed significant differences between the audio only and audio + video group ($p < .01$), the audio only group and the audio + video with subtitles group ($p < .001$), and the audio + video group and the audio + video with subtitles group ($p < .001$). Overall means (see Figure 3) for each group
were as follows: audio only ($M = 138.85, SD = 40.66$); audio + video ($M = 134.37, SD = 36.64$); and audio + video with subtitles ($M = 116.45, SD = 34.03$).

In addition to collecting continuous data, summative self-report data were also collected to address the second research question regarding differences in magnitude of response. Participants were asked to indicate the magnitude of this experience as compared to others they had experienced in the past. Raw data consisted of Likert-type responses to the question ranging from $1 = \text{Low}$ to $10 = \text{High}$. Means and standard deviations for listeners’ self-reported magnitude of response are displayed in Table 1. A one-way analysis of variance test was used to determine if there were significant differences among groups’ self-reported magnitude of response. While results revealed no significant differences among the groups $F(2, 100) = 1.85, p > .05$, the relative differences paralleled the continuous data in that the group reporting the highest magnitude of response was the audio only group, followed by the audio + video group, and finally the audio + video with subtitles group.

The third question guiding this research sought to determine if the listening conditions would have an effect on listeners’ stated focus of attention. An inspection of participants’ individual continuous data indicated that each individual was focused on the task as represented by differentiations in graphic contours from moment to moment. Participants’ self-report data were also analyzed to determine their stated level of focus of attention during the listening task. Raw data consisted of Likert-type responses to the question ranging from $1 = \text{Low}$ to $10 = \text{High}$. Means and standard deviations for listeners’ self-reported level of focus of attention are
displayed in Table 1. A one-way analysis of variance test was used to determine if there were significant differences among groups’ self-reported focus of attention. Results revealed no significant differences among the groups’ stated level of focus of attention, $F(2, 100) = 1.43, p > .05$.

In response to the fourth research question, listeners’ continuous and self-report responses were analyzed to determine where during the stimulus listeners’ perceived moments of expressivity occurred. Listeners’ continuous data were visually inspected (see Figure 2), and revealed that the moments of highest perceived expressivity occurred during parts of the arias, and for part of the duet. Those moments perceived as least expressive were during the applause and during the interlude (see Tables 3 and 4 for a more detailed account). Following the stimulus, listeners were asked to indicate how long the expressivity lasted (i.e., all of the act, parts of the act, arias, parts of the arias, other). Participants’ self-report data regarding moments of expressivity are included in Table 2. Both overall, and for each condition, frequency of category from most to least selected was: parts of the arias, parts of the act, all of the act, arias, and other.

**Discussion**

The purpose of the present research was to determine what effect a variety of listening conditions (audio, audio + video, audio + video with subtitles) would have on listeners’ perceptions of expressivity. Specifically, this study sought to examine the effect of these listening conditions on listeners’ magnitude of response regarding their perceptions of expressivity, their focus of attention, and their perceptions regarding the temporal location of expressive and non-expressive
moments in the stimulus. Based on the results obtained in the present study, while listening condition did not appear to have an effect on listeners’ self-reported summative ratings of magnitude of response, listening condition did seem to have an effect on continuous data recorded via the CRDI. The subtitles condition evidenced the lowest magnitude of response as compared to the audio and audio + video conditions for both continuous data (significant) and summative data (not significant). Additionally, statistical analysis revealed no significant differences among the three groups based on participants’ stated level of focus of attention. While not significantly higher, the subtitles group did evidence the highest level of focus of attention. Continuous and summative temporal results are similar to that of previous studies utilizing this piece of music, indicating that listeners perceived the highest moments of expressivity during “parts of the arias” and “parts of the act,” (see Table 2) and that there were three noticeable moments of “high expressivity,” and two moments of “low expressivity” during the excerpt (see Figure 2 and Tables 3 and 4).

Overall, several of the findings of the present study were consistent with previous research. All participants perceived moments of expressivity during the stimulus, and participants’ stated focus of attention was high, with group means ranging from 8.65 to 9.03 on a scale from 1 to 10. The temporal locations and graphic contours of the “high expressivity” and “low expressivity” moments are similar to that of previous research using this stimulus to explore listeners’ aesthetic response, felt emotional response, expressivity, and a more/less continuum (e.g., Lychner, 1998; Madsen & Geringer, 2008; Southall, 2003). This suggests that for this
piece of music, the relative relationships to magnitude might be the same across a variety of constructs (Lychner, 1998; Silveira & Madsen, in press). This finding also adds a measure of internal validity in that this audio/video recording of La Bohème, while different than the audio recording used in previous aesthetic response, musical tension, and expressivity research, yielded similar group responses for the audio only condition in the present study.

While the summative magnitude of response data revealed no significant differences among groups, examination of the continuous data yielded an interesting finding. Based on the continuous data, it appears that listeners’ magnitude of response was attenuated under the subtitles condition as compared to the other two conditions. It is interesting to speculate on possible reasons why this attenuation occurred. One possibility is that the subtitles may have served as a distraction inhibiting listeners’ perceptions of expressive moments. However, the subtitles condition did evidence the highest level of focus of attention, which could suggest that listeners were focused on the subtitles rather than the music being performed. In this regard, the subtitles may have actually served as a distraction from what some would call the intended focus of an opera – the music (Botstein, 1994; Low, 2001). Perhaps listeners’ attention was focused more on interpreting and contextualizing the translation rather than the music being performed. This supposition is not without some grounding in the research literature. As LaBerge (1995) suggested, when top-down controls are too strong and attentional shifts are inhibited (i.e., focusing on subtitles), features of the music may go unnoticed, or not fully be realized.
However, given that research has suggested that the mind cannot “multitask” (e.g., Buschman & Miller, 2010), but rather shifts rapidly back and forth between tasks, it could be that this shifting between cognitive (subtitles) and emotional (music) judgments had an effect on listeners’ continuous response measures. A previous study by Geringer and Madsen (1995/1996) addressed a similar issue when asking participants to identify to which elements of music they were attending. In their study, they used participants’ self-report data to address the focus of attention question. With recent advancements in eye tracking technology, it would be interesting if future studies like the present study tracked listeners’ eye movements to determine when the listener is attending to the subtitles, the action on stage, and the music.

Although the subtitles group evidenced the lowest magnitude of response for both summative and continuous measures, it may be that the translation of the text, and not the subtitles themselves that attenuated this group's responses. Subtitles act as one interpretation of the libretto; in essence, it is an edited version of the libretto for the audience. Some expressivity may have been lost with the subtitles because of what Bruti and Perego (2010) call semantic simplification of the foreign language text. It could be that some of the poetic nature of the original Italian text was lost or minimized when translated into English, and truncated to fit on the screen.

Previous research involving formalized enhancements of the listening experience (i.e., program notes, text translations) has produced mixed results. When examining the present study through the lens of previous research, an interesting
trend emerges. Two recent studies in the research literature revealed that text translation had no effect on participants’ perceptions of musical tension (Hackworth & Fredrickson, 2010) and program notes actually had a detrimental effect on listeners’ enjoyment of the music (Margulis, 2010). The present study seems to add to the growing body of research calling into question some of the previously held assumptions that program notes, text translations, and subtitles enhance the listening experience. It should be noted, however, that the aforementioned studies presented text translations and program notes before participants listened to the music stimulus. For the present study, listeners read the subtitles during the music stimulus, which may have served as a competition for focus of attention for the listeners in the subtitles condition.

Caution is warranted regarding the interpretation of the results from this study. While great care was taken in designing the study and choosing the stimulus, it is not without its limitations. While this excerpt of La Bohème has been extensively studied, different pieces of music would no doubt produce different results. To gain a fuller understanding of what effect subtitles might have on listeners’ perceptions of expressivity, other pieces of music will need to be examined. However, given the exploratory nature of this study, it seemed wise to choose a piece of music that has been well-studied in order to make comparisons against preexisting data regarding listeners’ affective responses. Another limitation is that only one live production was examined. Certainly other live productions of the same opera would yield different results. The results of the present study may be specific to this production of La Bohème. Opera libretti are not autonomous because subtitles are made for one
specific production, and they must be consistent with the performance on a
particular stage (Mateo, 2007a). Indeed varied translations of the source text will
produce different translated target texts. While this specific production was
purposefully chosen based on experts’ recommendations, examining other
productions of La Bohème would provide a deeper understanding of what effect the
production itself (in conjunction with the subtitles) might have on listeners’
perceptions. Additionally, translating a text from libretto to subtitles can be a
subjective experience, producing different results based on each translator’s
interpretation. The process of subtitling includes conveying the maximum amount
of information possible while utilizing a minimum amount of space (typically one or
two lines on screen). The use of alternative translations of this libretto, varying in
detail and length, might also have an effect on listeners’ perceptions.

Additional limitations regarding the methodology deserve mentioning. The
design of the present study was to investigate the effects of subtitles on ratings of
expressivity in the most ecologically valid manner possible. Listening to music,
listening to and watching a music performance, and listening to and watching a
music performance with subtitles are tasks with which most people are familiar. As
a result, these tasks were the basis of determining the experimental and control
groups. Future studies of this nature might also consider adding three more groups
for more control of the variables: audio + subtitles (no video); video + subtitles (no
audio); video only (no audio or subtitles). Another possible limitation could have
been the dependent variable itself (expressivity). Based on research proposed by
Lychner (1998), it appears that listeners perceive the constructs of aesthetic
response, felt emotional response, and a more/less continuum similarly. The term “aesthetic” has specific connotation for musicians, and previous evidence suggests that emotional response can be interpreted as either a construct felt or perceived in relation to the stimulus (i.e., cognitivist/emotivist theory). As a result, for the present study, the word “expressivity” was purposefully chosen in the hopes that it would be less susceptible to misinterpretation by the musically trained participants. However, caution is warranted in interpreting the results of this study, given the vagaries regarding participants’ conceptualization of the term “expressivity,” and the sometimes arbitrary nature in distinguishing between categorical and dimensional models of emotion (Schubert, 2010). Future studies might consider a similar design, but utilizing different constructs for listeners to rate/evaluate.

In addition to the methodological limitations mentioned above, there are potential complications regarding procedures in analyzing the data. A number of different techniques have been proposed regarding the analysis of continuous data points. There still remains a debate regarding the use of parametric statistics in analyzing data obtained via the CRDI. Specifically, the issue of serial correlation has been cited as a possible problem in using parametric statistics to analyze related continuous data points (Schubert, 2010). The rationale behind this issue is that as participants manipulate a continuous measuring device (e.g., CRDI), each data point must “pass through” the adjacent points on the scale. As such, they do not fluctuate randomly from time point to time point, and serial correlation may violate the assumptions made in parametric statistics. Future studies of this kind might include different methods of data analysis to address the issue of serial correlation (i.e.,
autocorrelation function, second-order standard deviation threshold, functional $F$ tests, etc.).

While the present study attempted to better understand the effect of subtitles on listeners’ perceptions of expressivity, it did not address the issue of comprehension. While emotional attachment during the listening experience has been described as the *sine qua non* of meaningful listening (e.g., Madsen & Geringer, 2000/2001), subtitles do allow for an understanding of the text being sung on stage. The purpose of the present study was not to examine listeners’ comprehension of the story line during the excerpt, rather their perceptions of expressivity occurring in the music. This was not meant to devalue the importance of text comprehension, rather it was meant to isolate one aspect of the listening experience. Although the data obtained in this study seem to suggest that subtitles may have impeded listeners’ affective responses, more research is needed to determine the effect of subtitles as a comprehension tool.

Based on the results of the present study, a number of different areas for future research are recommended aimed at understanding the relationship among subtitles, expressivity, focus of attention, and comprehension. Although the participants in the present study were highly trained musicians (college music majors), it would be interesting to see how musically untrained participants would respond to this stimulus. It is possible that musical sophistication could have an effect of how listeners attend to both the cognitive portion (subtitles) and affective portion (music) of the stimulus. Previous research has revealed similarities and differences in how musicians and nonmusicians respond to a music stimulus (e.g.,
Madsen, Byrnes et al., 1993) and how they attend to a music stimulus (e.g., Geringer & Madsen, 1995/1996). Therefore, it seems warranted for the present study to continue to explore the effect of subtitles on the listening experience of nonmusicians. Additionally, since supporters of subtitles have suggested that subtitles can provide a more enjoyable experience at the opera, perhaps future studies might examine participants’ satisfaction and enjoyment of the listening experience.

Previous research related to focus of attention has suggested that total involvement while listening to music is necessary to elicit an intense emotional attachment. Interestingly, the present study revealed that, while not significant, the subtitles group evidenced the highest level of focus of attention, but the lowest magnitude of response. What is unclear, however, was whether listeners’ attention was focused on the music or the subtitles, and during which moments their attention shifted from one to the other. Future research in this area might seek to make this distinction clearer in reporting results. Another avenue of research might also utilize a mixed methods design to gain a richer context from which to draw implications. While this study was primarily quantitative in nature, future studies might incorporate qualitative data in the form of participants’ free responses. This might add another dimension regarding what participants perceived in each of the listening conditions (i.e., Did they want the subtitles, or were they distracting?). An examination of the continuous response literature over the past two decades reveals that a qualitative component to these studies has been absent. Asking participants to explain their thought process and reasoning behind their responses might yield
previously unexplored variables in this line of research. Finally, it appears that more research is needed that measures both cognitive (comprehension) and affective (expressivity) outcomes while reading subtitles during a listening experience to determine to what degree subtitles can enhance or detract from a performance.
References


Table 1

Means and Standard Deviations for Participants’ Self-Reported Summative Magnitude of Response and Focus of Attention

<table>
<thead>
<tr>
<th>Listening Condition</th>
<th>Magnitude</th>
<th>Focus of Attention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Audio only</td>
<td>6.90</td>
<td>1.19</td>
</tr>
<tr>
<td>Audio + Video</td>
<td>6.42</td>
<td>1.85</td>
</tr>
<tr>
<td>Audio + Video with Subtitles</td>
<td>6.11</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Table 2

Frequency of Selected Categories for Listeners’ Self-report Data Regarding Moments of Expressivity

<table>
<thead>
<tr>
<th>Condition</th>
<th>All of Act</th>
<th>Parts of Act</th>
<th>Arias</th>
<th>Parts of Arias</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>Audio + Video</td>
<td>11</td>
<td>13</td>
<td>7</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Audio + Video with Subtitles</td>
<td>7</td>
<td>17</td>
<td>7</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>44</td>
<td>24</td>
<td>54</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 3

*Selected Moments of Perceived High Expressivity*

<table>
<thead>
<tr>
<th>Time (in half seconds)</th>
<th>Textural elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>156</td>
<td>Tenor aria, long note values, high vocal tessitura, soft dynamic</td>
</tr>
<tr>
<td>310</td>
<td>Tenor aria, increased rhythmic activity in orchestra, sustained note for tenor, loud dynamic</td>
</tr>
<tr>
<td>475</td>
<td>Tenor aria, dense orchestration, high vocal tessitura, loud dynamic, slowing of tempo</td>
</tr>
<tr>
<td>758</td>
<td>Soprano aria, high vocal tessitura, slowing of tempo, crescendo</td>
</tr>
<tr>
<td>853</td>
<td>Soprano aria, sparse orchestration, slowing of tempo, short note values</td>
</tr>
<tr>
<td>1020</td>
<td>Soprano aria, dense orchestration, loud dynamic, high vocal tessitura</td>
</tr>
<tr>
<td>1128</td>
<td>Soprano aria, dense orchestration, high vocal tessitura, slowing of tempo, crescendo</td>
</tr>
<tr>
<td>1340 (slight increase)</td>
<td>Interlude, soft dynamic, sparse orchestration, innuendo suggested</td>
</tr>
<tr>
<td>1453</td>
<td>Duet, loud dynamic, dense orchestration, high vocal tessitura, octave unisons</td>
</tr>
</tbody>
</table>
**Table 4**

*Selected Moments of Perceived Low Expressivity*

<table>
<thead>
<tr>
<th>Time (in half seconds)</th>
<th>Textural elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>205</td>
<td>Tenor aria, slowing of tempo, low vocal tessitura, minimal orchestral accompaniment (recitative-like)</td>
</tr>
<tr>
<td>244</td>
<td>Tenor solo, tenor is not singing at this point in time, sparse orchestral accompaniment, soft dynamic</td>
</tr>
<tr>
<td>386</td>
<td>Tenor aria, vocal/orchestral octave doubling (minimal accompaniment), soft dynamic</td>
</tr>
<tr>
<td>538</td>
<td>End of tenor aria, soft dynamic, sustained notes</td>
</tr>
<tr>
<td>840</td>
<td>Soprano aria, soprano is not singing at this point in time, slowing of tempo, soft dynamic, minimal orchestration</td>
</tr>
<tr>
<td>948</td>
<td>Soprano aria, cellos are the only ones playing here, slowing of tempo, soft dynamic</td>
</tr>
<tr>
<td>1091</td>
<td>Soprano aria, slower tempo, soft dynamic, sparse orchestration</td>
</tr>
<tr>
<td>1288–1370</td>
<td>Interlude, little orchestral accompaniment, recitative</td>
</tr>
</tbody>
</table>
Figure 1

CRDI Overlay Used for the Present Study
Figure 2

*Composite Response Graph for the Three Listening Conditions*
Figure 3

Means Plots of Overall Response Magnitudes for the Audio, Audio + Video (A/V), and Audio + Video with Subtitles (A/V + S) Conditions.