# Influence of pitch, loudness, and timbre on the perception of instrument dynamics

**Marco Fabiani[a]** and **Anders Friberg**
*KTH Royal Institute of Technology, Department of Speech, Music, and Hearing,*
*Lindstedtsv. 24, 100 44, Stockholm, Sweden*
*himork@kth.se, afriberg@kth.se*

**Abstract:**   The effect of variations in pitch, loudness, and timbre on the perception of the dynamics of isolated instrumental tones is investigated. A full factorial design was used in a listening experiment. The subjects were asked to indicate the perceived dynamics of each stimulus on a scale from *pianissimo* to *fortissimo*. Statistical analysis showed that for the instruments included (i.e., clarinet, flute, piano, trumpet, and violin) timbre and loudness had equally large effects, while pitch was relevant mostly for the first three. The results confirmed our hypothesis that loudness alone is not a reliable estimate of the dynamics of musical tones.
© 2011 Acoustical Society of America

## 1. Introduction

A composer writes a score, a musician performs it, and a listener enjoys it—this is the classic sequence of events behind music production and listening. Although the score contains a transcription of the way the piece is supposed to sound, the performer is free to interpret the suggestions of the composer. The musicians' goal is to convey their ideas to the listeners through the tones they produce. We make here a distinction between a note, i.e., a symbol in a score, and a tone, i.e., the corresponding sound played on a musical instrument. The notes in the score are mainly defined by three musical parameters: pitch, note value (i.e., duration), and dynamics. What we call dynamics is the musical attribute that is described by such adjectives as *piano, mezzo forte,* and *forte*. Four perceptual attributes are used instead to describe a tone: pitch, duration, loudness, and timbre. Pitch and duration are similar to both notes and tones. Timbre is often defined as the quality that allows us to distinguish between two instruments when all other attributes are the same. This definition implies that timbre is related to acoustical parameters, both in terms of spectral qualities (e.g., number of overtones) and dynamic qualities (e.g., attack speed).

In this study, we will focus on the communication of dynamics. It has been shown that musicians can successfully communicate dynamics to the listeners. In a previous experiment by Nakamura (1987), listeners were asked to identify the dynamics that three professional musicians attempted to convey, using their respective instruments (i.e., violin, oboe, and recorder), in several performances of the same piece of music. The results suggested that although dynamics variations such as *crescendo* appeared to be easier to convey, even absolute dynamics could be communicated fairly well. What are then the perceptual cues that the listener uses to identify the dynamics of a tone? It seems logical to associate dynamics with loudness, i.e., that attribute that allows us to order tones on a scale from quiet to loud, and which is mostly related to sound level. Nevertheless, Nakamura (1987) also observed that listeners could recognize dynamics even if there was no correspondence to a fixed sound level. This suggests that loudness only in part explains the perception of dynamics.

---

[a]Author to whom correspondence should be addressed.

Both common knowledge and measurements of acoustical parameters (see, e.g., Luce and Clark, 1967) tell us that timbre also varies within the same instrument, depending on its pitch and the dynamic at which it is played, as well as other playing techniques. Clark and Milner (1964) conducted an experiment where the loudness of isolated instrumental tones played at different dynamics was systematically varied. Seven subjects were asked to identify the intended dynamics, and the data were analyzed by counting the number of correct vs. incorrect identifications. The results showed that the dynamics were in general poorly identifiable, although there were instrument-dependent differences (e.g., better identification for trombone, worse for clarinet). According to the authors, the spectral qualities of timbre are weakly dependent on dynamics. Furthermore, the attack transients are not reliable cues for the identification of dynamic markings.

Melara and Marks (1990) studied the interaction between the auditory dimensions timbre and loudness and between timbre and pitch. Timbre was defined as the duty cycle of a variable pulse tone (high duty cycle creates a bright sound, low duty cycle a dull sound). Their results showed an interaction between loudness and timbre, i.e., bright sounds were perceived as being louder than dull sounds.

The variations of timbre within the same instrument is an aspect that has been in part neglected in favor of studies concerned with the perceptual mechanisms involved in the discrimination between different instruments. Our experiment aims at closing this gap by investigating the perception of dynamic differences within the same instrument. We are in particular interested in the relative influence of the tone's perceptual attributes pitch, loudness, and timbre, as well as in their interactions. Our hypothesis is that both loudness and timbre substantially contribute to the perception of dynamics, while pitch is less important. Such a hypothesis is in line with the results by Melara and Marks (1990), i.e., brighter tones are perceived as louder, although we would like to verify it using real instrumental sounds instead of synthetic ones.

To verify our hypothesis, we designed an experiment where the participants were asked to rate the perceived dynamics of isolated tones on a six-step scale while pitch, loudness, and timbre were systematically varied. Isolated tones were used in order to remove the effect of musical context, in contrast with the study by Nakamura (1987) that used complete musical excerpts. Although our experiment is in principle similar to that conducted by Clark and Milner (1964), their conclusions appear to be in contrast with the findings by Nakamura (1987). These facts, together with the lack of a statistical analysis of the data and the small number of subjects, prompted us to conduct a new experiment. We did not try to separately study the spectral and dynamic attributes of timbre, since in normal situations we hear the complete tone and not only the attack or the sustain. Furthermore, Iverson and Krumhansl (1993) suggest that the effect of dynamic attributes extend throughout the entire tone.

## 2. Method

### 2.1 Stimuli

Five instruments were chosen for this study: clarinet, flute, trumpet, violin, and piano. Single tones were selected from a large corpus of samples recorded by the first author. The procedure for recording the clarinet, trumpet, flute, and violin samples is described in Fabiani (2009). The piano, a Steinway model C, was recorded at a later stage in a damped rehearsal room at the University College of Opera in Stockholm. The piano tones were produced by means of a striker pendulum with a constant weight in order to obtain consistent strike velocities (Askenfelt and Jansson, 1990). Two Bruel & Kjaer type 4003 microphones were placed inside the piano (direct sound), while a Studio Projects C3 microphone was place near the tail (diffuse sound). For this experiment, the diffuse sound samples were used. A calibration signal was also obtained for all the recording sessions in order to accurately estimate the loudness of the tones.

A full factorial design with the three factors pitch, timbre, and loudness was used in the experiment. Three pitches, one octave apart, were used for clarinet (B3, B4, B5), trumpet (F3, F4, F5), and flute (G4, G5, G6); four for piano, one octave apart (C3, C4, C5, C6); and four for violin, one per string (A3, E4, B4, G5).

Three dynamics-related timbres were used for all the instruments, by selecting the recorded *pianissimo* (*pp*, soft), *mezzo forte* (*mf*, intermediate), and *fortissimo* (*ff*, bright) samples for each pitch included in the study. Note that no changes were made to the spectral content of the sample: we relied entirely on the timbre differences produced by the musicians. Other playing techniques, such as the position of the bow on the string for violin, or different embouchures in brasses and woodwinds, influence the timbre. In this experiment we tried to reduce their effect by explicitly asking the musicians to produce neutral tones using consistent playing techniques throughout the recording session.

Three loudness levels were defined (low, mid, and high). To determine the loudness levels, the original loudness of the sustain part of all the previously selected samples was first estimated following the ITU-R recommendation BS-1770-1 (ITU-R, 2007). Piano tones were considered an exception, since they do not have a sustain part: in this case, the peak sound level in the attack part was used. The low, mid, and high loudness levels for each pitch were defined, respectively, as the original loudness of the *pianissimo, mezzo forte*, and *fortissimo* sample. Three versions of each sample were then obtained by scaling its amplitude so that it corresponded to the three original loudness levels. This means that there was a unique set of loudness levels for each pitch and instrument combination. Given a generic stimulus $S(p,l,t)$, where $p$, $l$, and $t$ represent, respectively, the pitch, loudness, and timbre levels, then for example the clarinet stimuli $S(B3,low,mf)$ and $S(B3,low,ff)$ had both the original loudness of $S(B3,low,pp)$, but not necessarily that of $S(B4,low,pp)$.

To summarize, we created five blocks of stimuli, one for each instrument, for a total of 27 ($3 \times 3 \times 3$) stimuli for clarinet, trumpet, and flute, and 36 ($4 \times 3 \times 3$) for violin and piano. The data for each instrument were analyzed separately.

## 2.2 Subjects

Twenty-one subjects (average age = 32, SD = 9) participated in the experiment, and were compensated with a free cinema ticket. Since they had to be familiar with the concept of dynamics, they all had some musical experience. Furthermore, it has been shown that in such experiments, musical expertise is a strong between-subjects factor (see, e.g., Pitt, 1994), the effect of which we wanted to reduce as much as possible. All the participants had played at least one musical instrument for an average of 19 yr (SD = 10), and listened to music for an average of 12 h a week.

## 2.3 Procedure and apparatus

The participants were asked to rate the perceived dynamics of each stimulus on a six level scale (*pp, p, mp, mf, f, ff*). Observe the slight difference between this task and the one described in Clark and Milner (1964), where the participants were asked to recognize the dynamics at which the tone was originally played.

The responses were collected using a simple computer-based interface created with MATLAB's GUI Builder (version R2010b). The subjects were first asked to fill in a questionnaire with some personal information, i.e., their age, years of musical training, instruments played, and how many hours per week they listen to music (classical and other genres). A description of the experiment's procedure was then shown on the screen.

The stimuli were played only once, in random order, and the participants were instructed to rate their dynamics as quickly as possible by clicking on one of the six buttons marked with the appropriate label. A single presentation and a fast response were chosen in order to prevent the subjects from over-analyzing the tones, something that does not happen in normal listening conditions. Clicking on the "Next" button

produced the next stimulus. To give an idea of the range of timbre and loudness varia-tions found in the main task, five additional samples were also selected from the sam-ples corpus. The subjects were not aware of these five training stimuli, which were inserted before the other stimuli but appeared to be part of the experiment. The train-ing stimuli were discarded in the data analysis.

The order of the instrument blocks was also randomized. At the end of each block the participants were asked to indicate the instrument-specific cues on which they based their ratings. Four options were available: timbre (e.g., brightness), attack speed, sound level, and others (with space for additional comments).

The experiment took place in the same semi-anechoic room used for recording the samples. The test took approximately 20 min to complete. The stimuli were played through one Genelec 1031-A studio monitor placed at a distance of about 1.5 m from the subject. The sound level of the stimuli was calibrated so that the *mezzo forte* sam-ple in the center octave would give a reading of approximately 78 dBA at 1 m on a sound level meter. This was deemed a comfortable level, considering there is an aver-age difference of 10 dB between *mf* and *ff*.

## 3. Results

We conducted a separate three-way analysis of variance for each instrument with repeated measures of pitch, loudness, and timbre as independent variables ($3 \times 3 \times 3$ factorial design for clarinet, flute, trumpet, violin; $4 \times 3 \times 3$ factorial design for piano). Although there were four pitch levels in the violin stimuli, a mistake in the data collec-tion prevented us from distinguish between the ratings for the stimuli with pitch E4 and B4. We therefore decided to compute their average and run the analysis of vari-ance with three pitch levels, i.e., octave 3, 4, and 5. Another technical problem caused the piano stimulus $S$(C4, low, *mp*) to be rated only by 9 out of 21 subjects.

Table 1 summarizes the results of the analysis of variance, including only the statistically significant factors and interactions. The assumption of sphericity was first verified with Mauchly's test. In all cases when the assumption was rejected, the Greenhouse-Geisser correction was applied to the corresponding degrees of freedom.

As shown in Table 1, the main effects of all three factors are significant for all instruments, suggesting as expected that pitch, loudness, and timbre all contribute to the perception of dynamics. What is more interesting here is the size of these main effects. The Generalized Eta-squared ($\eta_G^2$) as defined in Bakeman (2005) is used in Table 1 as a measure of the effect size. According to Bakeman, the Generalized Eta-squared should be preferred over the Eta-squared and Partial Eta-squared in repeated measures designs because it allows comparing the results from studies using different numbers of factors. As a rule of thumb, Bakeman suggests to consider $\eta_G^2$ values around 0.02 as small, around 0.13 as medium, and around 0.26 as large.

There are large instrument-dependent differences in the size of the main effects. For trumpet and violin, pitch ($\eta_G^2 = 0.07$ and $\eta_G^2 = 0.05$, respectively) seems to play a minor role compared to loudness and timbre. Pitch ($\eta_G^2 = 0.74$) has the same importance as timbre ($\eta_G^2 = 0.72$) in clarinet, although loudness has a large effect as well ($\eta_G^2 = 0.59$). For flute and piano, loudness seems to be the most important factor ($\eta_G^2 = 0.60$ and $\eta_G^2 = 0.71$, respectively), with pitch and timbre on the same level, albeit lower for flute ($\eta_G^2 = 0.25$ and $\eta_G^2 = 0.17$, respectively, for pitch and timbre) than for piano ($\eta_G^2 = 0.42$ and $\eta_G^2 = 0.41$). These results confirm our hypothesis that both loud-ness and timbre play an important role in the perception of dynamics. They also sug-gest that for three of the five instruments we studied (i.e., clarinet, flute, and piano), pitch plays a major role as well.

As an example, Fig. 1 shows mean and confidence intervals of the rate dynam-ics for trumpet and flute, grouped according to the three main factors. Observe how all the means increase with increasing values of loudness and timbre, a trend found for all instruments. Figure 1(a) clearly shows the small effect of pitch and the large effect of timbre and loudness for trumpet. For example, the stimuli $S$(F3,low,*ff*) and

Table 1. Summary of the data analysis results. Only statistically significant effects and interactions are included. Legend: P = Pitch, L = Loudness, T = Timbre. df* = Greenhouse-Geisser correction.

| | Clarinet | | | | | Piano | | | |
|---|---|---|---|---|---|---|---|---|---|
| Effect | df | F | Sig. | $\eta_G^2$ | Effect | df | F | Sig. | $\eta_G^2$ |
| P | 1.5,29.5* | 12.91 | 0.000 | 0.74 | P | 3,24 | 61.62 | 0.000 | 0.42 |
| L | 1.2,23.4* | 82.08 | 0.000 | 0.59 | L | 1.2,9.2* | 128.17 | 0.000 | 0.71 |
| T | 1.4,28.8* | 161.78 | 0.000 | 0.72 | T | 1.2,9.8* | 30.11 | 0.000 | 0.41 |
| P × T | 4,80 | 10.62 | 0.000 | 0.06 | P × L | 6,48 | 7.41 | 0.000 | 0.13 |
| P × L × T | 8,160 | 2.17 | 0.032 | 0.02 | P × T | 6,48 | 7.14 | 0.000 | 0.09 |
| | Flute | | | | | Violin | | | |
| P | 1.4,29.9* | 36.60 | 0.000 | 0.25 | P | 1.2,24.8* | 13.89 | 0.001 | 0.05 |
| L | 1.2,23.1* | 140.34 | 0.000 | 0.60 | L | 1.5,30.3* | 212.70 | 0.000 | 0.59 |
| T | 2,40 | 32.60 | 0.000 | 0.17 | T | 1.3,25.1* | 83.45 | 0.000 | 0.57 |
| P × L | 4,80 | 10.06 | 0.000 | 0.07 | P × L | 2.7,54.5* | 9.28 | 0.000 | 0.05 |
| P × T | 4,80 | 5.97 | 0.000 | 0.03 | P × T | 4,80 | 8.40 | 0.000 | 0.03 |
| P × L × T | 8,160 | 3.74 | 0.000 | 0.03 | L × T | 4,80 | 4.38 | 0.003 | 0.10 |
| | Trumpet | | | | | | | | |
| P | 1.2,23.8* | 9.52 | 0.004 | 0.07 | | | | | |
| L | 1.2,23.5* | 68.49 | 0.000 | 0.30 | | | | | |
| T | 1.1,22.5* | 66.61 | 0.000 | 0.49 | | | | | |
| P × L | 2.9,58.5* | 5.14 | 0.003 | 0.02 | | | | | |

*S*(F3,high,*ff*) were in average rated *mezzo piano* and *forte*, respectively, although they had the same original timbre (i.e., *ff*). Observe also the different behavior for the flute [Fig. 1(b)], where loudness is instead the most important factor, while timbre plays a minor role, especially in octave 3.

Several significant interactions between factors emerged from the analysis. Apart from some exceptions, their effect size is mostly very small compared to that of the main effects. The interaction Pitch × Loudness in flute has a moderate effect ($\eta_G^2 = 0.07$), especially if compared to that of timbre: note in Fig. 1(b) how the effect of
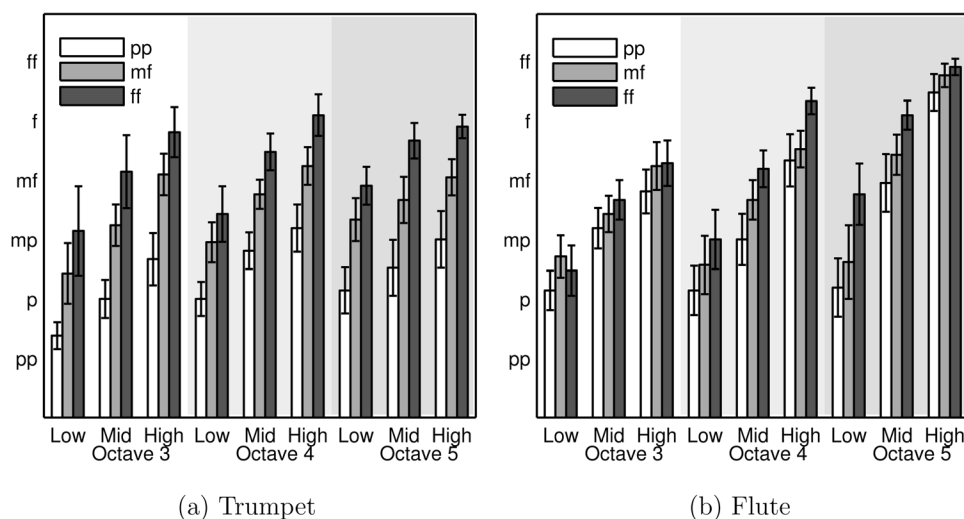


(a) Trumpet                    (b) Flute

Fig. 1. Average ratings across subjects of each stimuli for (a) trumpet and (b) flute, grouped first by loudness (low, mid, high), and then by octave. Error bars indicate confidence intervals (p < 0.05).

loudness increases with pitch. The same interaction has also a relatively larger effect size for piano ($\eta_G^2 = 0.13$). The Pitch $\times$ Timbre interaction in piano is also appreciable ($\eta_G^2 = 0.09$), as the Loudness $\times$ Timbre for violin ($\eta_G^2 = 0.10$, larger than the main effect of pitch).

Finally, *post hoc* pairwise comparisons of the means of the main effects were computed using a confidence interval corrected with the Bonferroni method. For loudness and timbre, all pairwise comparisons for all instruments were significant. For pitch, there was a significant difference between the two highest octaves for clarinet (4 and 5), flute (5 and 6), and violin (4 and 5). For trumpet, a significant difference between the two lowest octaves (3 and 4) was found. Finally, all pairwise comparisons except that between the two lowest octaves (3 and 4) were significant for piano.

A general preference for timbre emerged from the answers to the question regarding the cues used to rate the dynamics of the stimuli. An exception were piano tones, for which attack speed and loudness were considered more important. This was expected since piano is the only percussive instrument. Other freely suggested cues include bow noises for violin, breath noise for flute and clarinet, and mechanical noises for the piano, pitch stability, and direct knowledge of the instrument.

## 4. Conclusions and discussion

In this study we investigated the effect of pitch, loudness, and timbre on the perception of the dynamics of isolated tones produced by different musical instruments. We focused our attention on the perception of dynamics in the musical and music performance sense, as opposed to the perception of loudness.

The results showed that, as hypothesized, loudness and timbre played in general about equally important roles, with the exception of flute, where timbre was found to be less influential. Observe that our definition of timbre includes both spectral (e.g., brightness) and dynamic (e.g., attack speed) attributes. The use of real recordings limited the possibility of manipulating such factors separately without introducing disturbing artifacts. Furthermore, removing for example the tone's attack would have made little difference since, according to Iverson and Krumhansl (1993), the dynamic attributes extend past the onset. These general results are in line with those obtained by Melara and Marks (1990) for synthetic instruments. In contrast to what Clark and Milner (1964) argues, and in accordance to Nakamura (1987), the subjects in our experiment could determine fairly accurately the dynamics of isolated tones.

The perception of dynamics appears to be less dependent on pitch. For most instruments, there was a significant difference between the two highest octaves. This could be explained by the register changes occurring in such instruments as clarinet and trumpet, which cause important timbre differences. Interactions between the three factors had in general a much smaller effect size compared to the main effects.

There were clear differences in judgment strategy for different instruments, as indicated by the different main effects' sizes. This suggests that there is a learning effect that is specific for each instrument. Other cues that contribute to the dynamics ratings such as mechanical noises and breath were indicated by the participants, especially by those playing that particular instrument. This suggests that, even though we tried to reduce the effect of musical background on the ratings by employing expert musicians, it might still have influenced the results. Nevertheless, although these noises are easy to detect in such experimental conditions, we expect these cues to become less relevant in a complete piece of music.

We think that our results are important from the point of view of music recording and listening. Modern recording techniques, where instruments are recorded in separate tracks and successively mixed, introduce a lot of modifications to the relative sound level of tones (e.g., different gains, compressors, limiters). If, as we have shown, loudness variations influence the perceived dynamics of a tone, the intentions of the musician are in part lost. From a different perspective, if modifying a music

performance is our primary goal, manipulating the dynamics of tones should involve not only sound level but also timbre modifications (Fabiani, 2009, 2011).

### References and links

Askenfelt, A., and Jansson, E. V. (**1990**). "From touch to string vibrations. I: Timing in the grand piano action," J. Acoust. Soc. Am. **88**, 52–63.

Bakeman, R. (**2005**). "Recommended effect size statistics for repeated measures designs," Behav. Res. Methods **37**(3), 379–384.

Clark, M., and Milner, P. (**1964**). "Dependence of timbre on the tonal loudness produced by musical instruments," J. Audio Eng. Soc. **12**, 28–31.

Fabiani, M. (**2009**). "A method for the modification of acoustic instrument tone dynamics," in *Proceedings of the 12th International Conference on Digital Audio Effects*, Como, Italy.

Fabiani, M. (**2011**). "Interactive computer-aided expressive music performance—Analysis, modification, and synthesis methods," Ph.d. thesis, KTH Royal Institute of Technology, Stockholm, Sweden.

ITU-R (**2007**). "Recommendation ITU-R BS. 1770-1 algorithms to measure audio programme loudness and true-peak audio level," Technical Report, International Telecommunication Union–-Radiocommunication Sector, Geneva.

Iverson, P., and Krumhansl, C. L. (**1993**). "Isolating the dynamic attributes of musical timbre," J. Acoust. Soc. Am. **94**, 2595–2603.

Luce, D., and Clark, M. J. (**1967**). "Physical correlates of brass-instrument tones," J. Acoust. Soc. Am. **42**, 1232–1243.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness.," Percept. Psychophys. **48**, 169–178.

Nakamura, T. (**1987**). "The communication of dynamics between musicians and listeners through musical performance," Atten. Percept. Psychol. **41**, 525–533.

Pitt, M. A. (**1994**). "Perception of pitch and timbre by musically trained and untrained listeners," J. Exp. Psychol. Human. **20**, 976–986.