ASYNCHRONY VERSUS INTENSITY AS CUES FOR MELODY PERCEPTION IN CHORDS AND REAL MUSIC

Werner Goebl[1), 2)] & *Richard Parncutt*[1)]

[1)]Department of Musicology, University of Graz
Mozartgasse 3, A-8010 Graz, Austria

[2)]Austrian Research Institute for Artificial Intelligence
Freyung 6/VI/7, A-1010 Vienna, Austria

werner.goebl@oefai.at; richard.parncutt@uni-graz.at

ABSTRACT

In expressive piano performance, the performer emphasises a melody by increasing its intensity and by anticipating it by some tens of milliseconds (*melody lead*). In this contribution, we continue previous research on the influence of asynchrony and intensity variation on the perceived salience of a particular tone or voice with three experiments. In Experiment I, three-tone piano chords are presented with each of the three tones simultaneously manipulated in timing and intensity by up to ±55 ms and +30/–22 MIDI velocity units. Loudness ratings depended mainly on relative intensity and relatively little on timing (e.g., anticipated tones were sometimes rated louder than delayed ones). The lower voice was generally rated louder than the middle voice. In Experiment II, a sequence of chords produced similar results; streaming enhanced the effect of asynchrony only marginally. In Experiment III, a short musical excerpt by Chopin was presented. Again, intensity was the dominating cue. In contrast to previous findings, a melody that was both delayed and louder in intensity was rated significantly louder than a melody that was simultaneous and louder.

1. INTRODUCTION

A melody in a multi-voiced musical context receives greater perceptual attention because it is played louder. Additionally, it is also played earlier than nominally simultaneous tones (*melody lead*, cf. Palmer, 1989, 1996; Repp, 1996b; Goebl, 2001). Here, we investigate how asynchrony enhances the perceptual salience of a single voice in relation to changes in loudness. In previous research on the perception of tone salience in dyads, we found that loudness is the dominating cue, asynchrony having only marginal influence (Goebl & Parncutt, 2002).

2. AIMS

In a multi-voiced context, we investigate the relative perceptual salience of individual voices that are shifted back and forth in time and varied in intensity simultaneously. We are interested in the relative contribution of each of these cues to the perceptual salience of a tone or voice. By comparison to previous work (Goebl & Parncutt, 2002), we extend the stimulus material to three-tone chords in order to study perceptual salience behaviour at different vertical positions within a chord (Experiment I). Studies on the detection of pianists' errors revealed that outer voices tend to receive greater perceptual attention than inner voices (Palmer & van de Sande, 1993; Palmer & Holleran, 1994; Repp, 1996a).

Alongside masking effects (Rasch, 1978), Bregman's theory of auditory scene analysis (Bregman & Pinker, 1978; Bregman, 1990) is generally invoked to explain the melody lead phenomenon. To test this hypothesis, the stimulus material of the present study is extended to sequences of chords (Experiment II) and to an excerpt of real music (Experiment III).

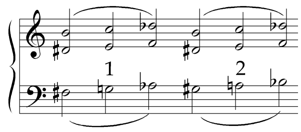


Figure 1. Experiment I & II. The two chords randomly transposed one semitone up and down.

3. EXPERIMENTS I & II

3.1.1. Method

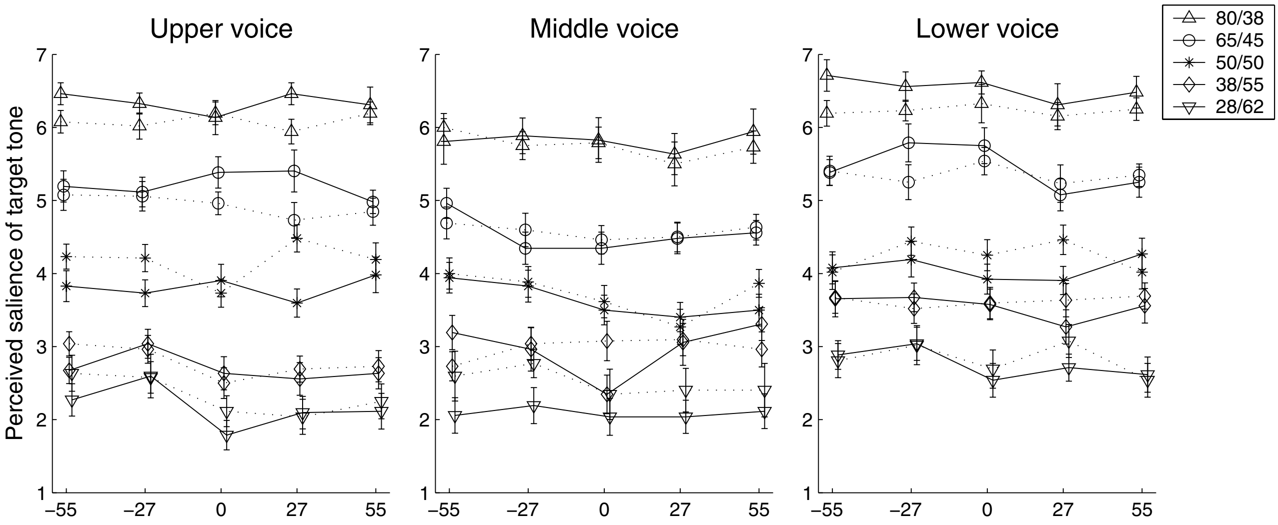
Participants. The three experiments were included in a single test session. The 26 musically trained participants comprised 17 pianists and 9 other instrumentalists (violin, violoncello, and acoustic guitar). They had been playing their instruments regularly for an average of 17.9 years (SD = 5.5 years). 23 of them had studied their instrument at post-secondary level, for an average of 7.2 years (SD = 4.1 years). The ages of all participants ranged from 19 to 35 years with an average of 26.5 years (SD = 4.5 years).

Stimuli. In the first experiment, three-tone piano chords, in which one tone (the *target*) was shifted in time and varied in intensity, were presented to the participants. The chords were a major triad in 2nd inversion and a minor triad in root position (see Figure 1). The pitch of each trial was randomly shifted one semitone up and down (Figure 1). The asynchronies of the target tone were –55, –27, 0, 27, or 55 ms in comparison to the other two chord tones. The five intensity combinations were [target tone/other two tones]: +30/–12, +15/–5, 0/0, –12/+5, –22/+12 MIDI velocity units relative to a medium intensity of 50 MIDI velocity units. These combinations were chosen so that the velocity differences implicated the above named asynchronies according to the *velocity artifact* (Repp, 1996b; Goebl, 2001) and that the overall loudness of the whole sonority remained approximately constant over the different intensity combinations. The target tone could occur in any of the three voices (upper, middle, lower). The listeners' attention was directed to the target by a priming tone which started 1200 ms before the tested chord and sounded for 600 ms. The intensity of the primer was always constant at the medium intensity of 50 MIDI velocity units. In the second experiment, the identical stimulus material was used, but with each chord repeated five times consecutively with an inter-onset interval of 300 ms.

Equipment. The piano sounds used in these experiments are taken from acoustic recordings of tones produced by the Bösendorfer SE290 playing back a computer-generated file.[1] The

[1] Each key was programmed to be depressed with MIDI velocities from 10 to 110 in steps of 2 units. The mapping between MIDI velocity and

(a) Experiment Ia/b



(b) Experiment II

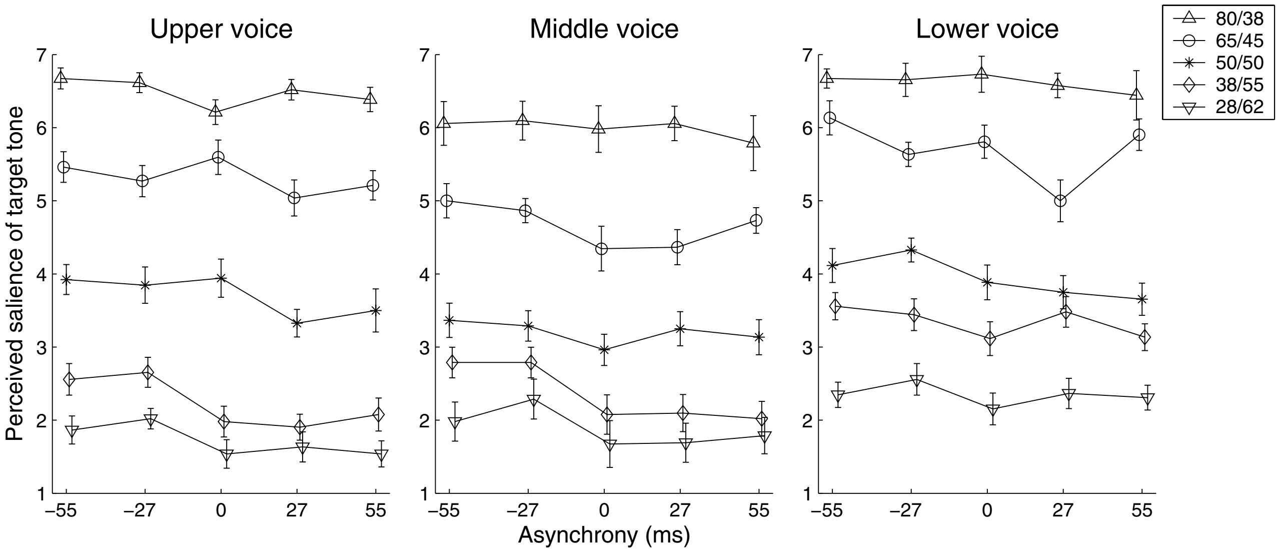


Figure 2: Mean ratings over 26 participants, separately for different voices (panels), intensity combinations (different markers), and asynchronies (*x* axes). **(a)** The upper panels depict results from the two blocks of Experiment I: the first block is plotted in solid lines, the second in dashed lines. **(b)** The lower panels show results of Experiment II. Error bars denote 95% confidence intervals of the means.

intensity of the recorded piano samples is referred to in terms of peak sound level in decibels and not in terms of MIDI velocity units.[2] The sound samples are added up on hard disk to avoid sympathetic resonances.

Procedure. A graphical user interface presented the stimuli in random order and guided the participants through the experiments. Each stimulus was preceded by a priming tone indicating which tone to rate. At the same time, the chord was presented in musical notation with an arrow pointing at the target tone. The stimuli were presented diotically (same signal in each ear) via headphones (Sennheiser 25–1). The participants rated the target tone relative to the other tones in the chord on a 7-point scale from 1 "very much softer" to 7 "very much louder" with 4 "equally loud". A short training session familiarised the participants with the stimulus material. The participants completed two blocks of Experiment I, one before and one after Experiment II; the order of experiments within the session was thus Ia, II, Ib, III. The session lasted between one hour and one hour and a half. The participants were paid 15 Euros for their services.

3.1.2. Results & Discussion

The mean loudness ratings for the two blocks of Experiment I are plotted separately for the three voices, the 5 intensity combinations, and the 5 asynchronies in Figure 2a. The ratings of the five intensity combinations showed the same order in all voices and both blocks. The ratings were generally higher

hammer velocity was $MIDI_{vel} = 52 + 25 \cdot \log_2(\text{hammer velocity})$. The sound was recorded with a TASCAM digital audio tape recorder (DA–P1) and AKG (CK91) stereo microphones. The recorded signal was transferred digitally to computer hard disk and stored in WAV format (16-bit, 44100 Hz sampling frequency).

[2] The relationship between MIDI velocity units and dB peak sound level of the recorded tones from the Bösendorfer was approximated by $pSL = -77.19 + 26.11 \cdot \log_{10}(MIDI_{vel}) + 5.277 \cdot \log_{10}(MIDI_{vel})^2$.

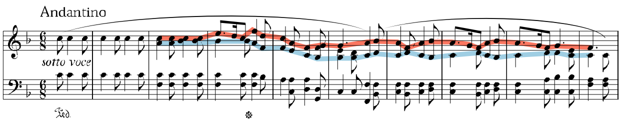


Figure 3. The first nine bars of Chopin's Ballade, op. 38 with the two selected voices marked (in different colours).

(louder) in the lower voice and lowest in the middle voice.

Repeated-measures ANOVAs were conducted on the ratings of the two blocks of Experiment I separately, with voice (3), asynchrony (5), and intensity (5) as within-subject factors and instrument (pianist, non-pianists) as between-subject factor. There were significant 3-way interactions of voice, asynchrony, and intensity [$F(32,768) = 3.82$, $\varepsilon_{G.G.} = 0.386$, $p_{adj} < 0.001$; $F(32,768) = 3.72$, $\varepsilon_{G.G.} = 0.329$, $p_{adj} < 0.001$, respectively][3]. There was no significant difference between the ratings of the pianists and the ratings of performers of other instruments.

The effects of asynchrony were small and partly unpredictable. The target tone was rated slightly louder when it was earlier and softer than the other tones. Although this tendency was barely significant (according to Bonferroni corrected post-hoc comparisons), it might be explained by spectral masking.

The ratings of Experiment II are plotted in Figure 2b. A repeated-measures ANOVA again revealed a significant interaction of voice, asynchrony, and intensity [$F(32,768) = 3.56$, $\varepsilon_{G.G.} = 0.352$, $p_{adj} < 0.001$] and no effect or interaction with participant's instrument. Again, an anticipated voice received slightly louder ratings when it was softer in comparison to the condition without time shift. The middle voice was generally rated softer than the lower voice. In general, the results of this experiment were similar to those of Experiment I.

In order to test whether anticipation (–55 and –27 ms) changed rating in comparison to delay (+27 and +55 ms), these asynchrony conditions were linearly contrasted to each other (asynchrony: +1, +1, 0, –1, –1), separately for each intensity combination in each voice and each block of the two experiments ($5 \times 3 \times 3 = 45$ contrasts). In Experiment I, 5 or 6 of these contrasts were significant; in Experiment II, this number increased to 10. Thus, delayed tones tended to sound softer than anticipated tones of equal asynchrony and intensity. This trend was stronger in the streaming experiment (II). However, the effect was quite inconsistent: in Experiment Ia, middle voice (38/55) and in Experiment Ib, upper voice (50/50), the +27 ms condition was rated significantly *louder* than the corresponding simultaneous condition.

To conclude, the main cue for the loudness of a tone or voice was intensity; the effect of temporal shifting was relatively small. Asynchrony became relevant when intensity was absent as a cue (voices equally loud) or when the target tone/voice was very soft. In the latter case, anticipation helped to overcome the spectral masking that occurred when the tones were simultaneous.

4. EXPERIMENT III

The third experiment investigated the influence of asynchrony and intensity on the perceived salience of a voice within a real music excerpt. Participants and procedure were the same as in the previous experiments.

4.1.1. Method

Stimuli. This experiment involved the first nine bars of Chopin's *Ballade* op. 38 as the test piece with the upper and upper middle voices as possible melodies (Figure 3). As before, these two voices were varied both in asynchrony (–55, –27, 0, 27, 55 ms) and intensity (upper voice 0, +12, +24 MIDI velocity units; middle voice 0, +10, +20 units). The increments in MIDI velocity were obtained from measurements of 22 expressive performances of that piece (Goebl, 2001). The extent of asynchrony corresponded to melody leads resulting from above named velocity combinations according to the velocity artifact (Repp, 1996b; Goebl, 2001). MIDI files were generated by imposing the variation of the stimulus material (2 voices $\times$ 5 asynchronies $\times$ 3 intensities) onto a prototypical timing and intensity curve from an expressive performance by one performer. Manipulations began in the middle of bar 2 (the first one and a half bars were not manipulated). An artificial pedal track was added to the MIDI files. The performance files were played back on the Bösendorfer SE290 and recorded onto DAT. The recordings were transferred digitally on a computer hard disk and stored in WAV files (44.1 kHz, 16-bit, stereo). In order to reduce the number of stimuli, only orthogonal and diagonal combinations of asynchrony and intensity (22) were tested (see results in Figure 4).

Procedure. The participants rated which of the two voices attracted their attention more, again on a 7-point scale from 1 "very much the lower one" over 4 "they sound equally loud" to 7 "very much the upper one". The participants were guided by the same graphical user interface as before with a short familiarisation phase before the actual experiment. They saw a score with the two designated melodies marked in colour (as in Figure 3) and heard the recordings in stereo via headphones.

4.1.2. Results & Discussion

The mean loudness ratings separately for the two voices (panels), three intensity combinations (lines), and five asynchronies (*x* axes) are plotted in Figure 4. When the two voices were equally loud, both voices were rated significantly quieter than a rating of 4 when the upper was simultaneous or late by comparison to the upper voice (according to *t*-tests single means). It is unclear whether the middle voice was in fact louder than the upper voice (even though the two were equally loud in terms of MIDI velocity units) or whether musically trained participants expected the upper voice (melody) to be louder in this musical context and therefore considered the middle voice to be louder when this expectation was violated.

Repeated-measures ANOVAs on the ratings with asynchrony and voice as within-subject factors (separately for the 3 intensity conditions) revealed effects of asynchrony in the equal intensity conditions and in the upper voice emphasised conditions. The results of post-hoc comparisons are sketched in Figure 4.

The effects of asynchrony were small compared to the effects of intensity. When intensity was missing as a cue, anticipation could lead to a slightly enhanced perception of a voice (in our data more in the middle than in the upper voice), but when the voices were played louder, asynchrony became a minor cue (especially in the middle voice). Surprisingly, a melodic delay could increase the rated loudness of the upper voice, when it was already louder than the middle voice; but no analogous effect was observed in the middle voice.

[3] The *p*-values are corrected according to Greenhouse-Geisser, the corrected degrees of freedom are not printed.

Experiment III

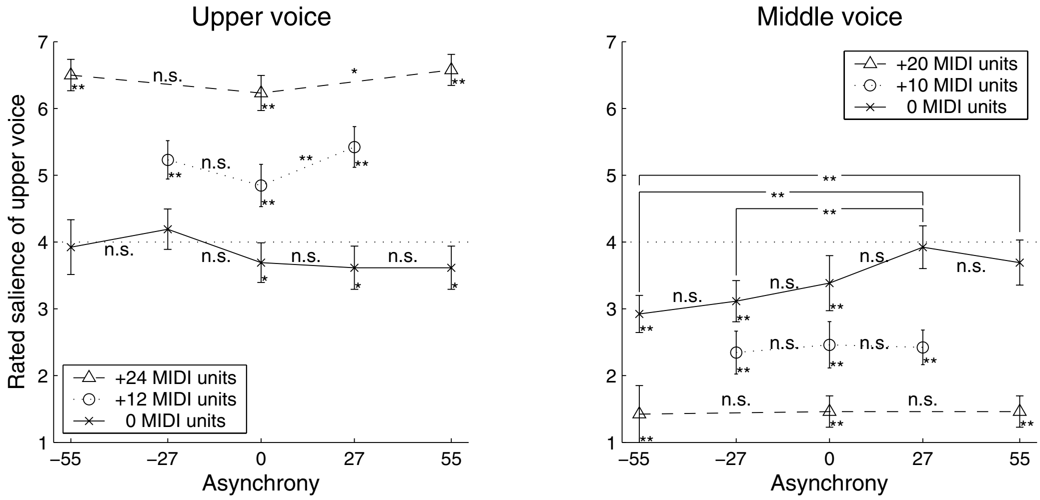


Figure 4: Mean ratings over 26 participants, separately for different voices (panels), intensity combinations (different markers), and asynchronies (x axes). Error bars denote 95% confidence intervals. The asterisks *against* the error bars indicate significant differences to a rating of 4 ("The two melodic interpretations sound equally important to me") according to *t*-tests for single means (* $\alpha < 0.05$, ** $\alpha < 0.01$). Results of Bonferroni post-hoc tests are marked *between* adjacent asynchrony conditions either by asterisks or 'n.s.' (non significant). Significant non-adjacent asynchrony conditions are marked, if they were significant (only in the right panel).

The results of this experiment suggested that intensity was the primary cue guiding the listener's attention to individual voices in a musical excerpt. Temporal effects could be observed, but they were rather small and unpredictable.

5. CONCLUSIONS

We investigated the influence of onset asynchrony and intensity on the perceived dominance of a particular tone or voice in three different experimental conditions. We presented three-tone chords, sequences of chords, and a real music excerpt to the participants. In all experiments, we found intensity to be the dominating cue. Effects of asynchrony were present in the data, but they were comparatively small. There was an overall trend for an anticipated voice to be perceived louder than a delayed voice. However, in the musical excerpt (Experiment III) we also found the opposite trend in the upper melody voice when it was played louder than the middle voice; in this case, the upper voice was perceived even louder when it was delayed.

When the rated tone or voice was very soft, anticipation evidently helped it to avoid masking by louder tones. We found surprisingly little evidence of streaming enhancing the perceived salience of a voice. Perhaps the effect of streaming lies more in clarifying the transparent simultaneous perception of more than one voice rather than in changing one voice's salience.

6. ACKNOWLEDGEMENTS

This research is part of the project Y99-INF, sponsored by the Austrian Federal Ministry of Education, Science and Culture in the form of a START research Prize to Gerhard Widmer. The Austrian Research Institute for Artificial Intelligence acknowledges basic financial support from the Austrian Federal Ministry for Education, Science, and Culture. The authors thank the Acoustics Research Institute of the Austrian Academy of Sciences for generously providing technical equipment and the Bösendorfer company for providing the SE290 grand piano for experimental use. Thanks also to Oliver Vitouch, Elias Pampalk, and Simon Dixon for helpful comments.

7. REFERENCES

Bregman, A. S. (1990). Auditory scene analysis. The perceptual organization of sound. Cambridge, Massachusetts: The MIT Press.

Bregman, A. S., & Pinker, S. (1978). Auditory streaming and the building of timbre. Canadian Journal of Psychology, 32, 19–31.

Goebl, W. (2001). Melody lead in piano performance: Expressive device or artifact? Journal of the Acoustical Society of America, 110(1), 563–572.

Goebl, W., & Parncutt, R. (2002). The influence of relative intensity on the perception of onset asynchronies. In C. Stevens & D. Burnham & G. McPherson & E. Schubert & J. Renwick (Eds.), Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney 2002 (pp. 613–616). Adelaide: Causal Productions.

Palmer, C. (1989). Mapping musical thought to musical performance. Journal of Experimental Psychology: Human Perception and Performance, 15(12), 331–346.

Palmer, C. (1996). On the assignment of structure in music performance. Music Perception, 14(1), 23–56.

Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. Perception and Psychophysics, 56(3), 301–312.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19(2), 457–470.

Rasch, R. A. (1978). The perception of simultaneous notes such as in polyphonic music. Acustica, 40, 21–33.

Repp, B. H. (1996a). The art of inaccuracy: Why pianists' errors are difficult to hear. Music Perception, 14(2), 161–184.

Repp, B. H. (1996b). Patterns of note onset asynchronies in expressive piano performance. Journal of the Acoustical Society of America, 100(6), 3917–3932.