

Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners

Journal Article**Author(s):**

Kirchberger, Martin; Russo, Frank A.

Publication date:

2016-01-01

Permanent link:

<https://doi.org/10.3929/ethz-b-000117818>

Rights / license:

[Creative Commons Attribution-NonCommercial 3.0 Unported](#)

Originally published in:

Trends in Hearing 20, <https://doi.org/10.1177/2331216516630549>

Dynamic Range Across Music Genres and the Perception of Dynamic Compression in Hearing-Impaired Listeners

Trends in Hearing
2016, Vol. 20: 1–16
© The Author(s) 2016
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/2331216516630549
tia.sagepub.com



Martin Kirchberger¹ and Frank A. Russo²

Abstract

Dynamic range compression serves different purposes in the music and hearing-aid industries. In the music industry, it is used to make music louder and more attractive to normal-hearing listeners. In the hearing-aid industry, it is used to map the variable dynamic range of acoustic signals to the reduced dynamic range of hearing-impaired listeners. Hence, hearing-aided listeners will typically receive a dual dose of compression when listening to recorded music. The present study involved an acoustic analysis of dynamic range across a cross section of recorded music as well as a perceptual study comparing the efficacy of different compression schemes. The acoustic analysis revealed that the dynamic range of samples from popular genres, such as rock or rap, was generally smaller than the dynamic range of samples from classical genres, such as opera and orchestra. By comparison, the dynamic range of speech, based on recordings of monologues in quiet, was larger than the dynamic range of all music genres tested. The perceptual study compared the effect of the prescription rule NAL-NL2 with a semicompressive and a linear scheme. Music subjected to linear processing had the highest ratings for dynamics and quality, followed by the semicompressive and the NAL-NL2 setting. These findings advise against NAL-NL2 as a prescription rule for recorded music and recommend linear settings.

Keywords

dynamic range, music genre, compression, hearing loss, hearing aids

Date received: 3 September 2015; revised: 21 December 2015; accepted: 13 January 2016

Introduction

Compression in Music Production

Dynamic range refers to the level difference between the highest and lowest-level passages of an audio signal. Dynamic range compression (or dynamic compression) is a method to reduce the dynamic range by amplifying passages that are low in intensity more than passages that are high in intensity. Dynamic compression can serve different purposes in the music production process. It may have an aesthetic purpose in the mastering process to make the mix more coherent and to minimize excessive loudness changes within a song (Katz & Katz, 2003). It may also have a pragmatic purpose if it is employed to adapt the dynamic range of music to the technical limitations of recording or playback devices. Dynamic compression, however, can and has infamously been used to increase the loudness of a song. It is widely believed in the music industry that loudness levels and record sales are correlated (Vickers, 2011). A very strong compressor called a limiter is employed to reduce peak levels. A so-called *makeup gain* then amplifies the whole signal until the peaks reach full

scale again. This method increases the overall energy of the signal but often introduces distortion (Kates, 2010) and compromises signal quality. Even when distortion is not perceptible, highly compressed music can become physically or mentally tiring over time (Vickers, 2011).

Compression in Hearing Aids

Hearing aids incorporate dynamic range compression to compensate for higher absolute hearing thresholds and the effects of loudness recruitment, which are commonly experienced by people with sensorineural hearing loss. Loudness recruitment is an abnormally rapid growth in loudness that accompanies increases in suprathreshold stimulus intensity (Villchur, 1974). A hearing aid must

¹ETH Zürich, Switzerland

²Ryerson University, Toronto, ON, Canada

Corresponding author:

Frank A. Russo, Ryerson University, 350 Victoria Street, M5B 2K3 Toronto, Canada.

Email: russo@ryerson.ca



amplify soft passages more than loud passages so as to increase audibility while maintaining a comfortable listening experience. Fortunately, hearing aids provide some flexibility in the extent to which compression is applied. Important compression parameters are attack time, release time, compression ratio (CR), compression threshold, and number of channels (Giannoulis, Massberg, & Reiss, 2012). The parameterizations vary across hearing-aid manufacturers (Moore, Füllgrabe, & Stone, 2011) and may also depend on the detected signal class, such as speech or music.

There are established prescription rules that define gain targets for speech as a function of frequency, sound level, and hearing loss. CAMEQ (Moore, 2005; Moore, Glasberg, & Stone, 1999) and its successor CAM2 (Moore, Glasberg, & Stone, 2010) are fitting recommendations from the University of Cambridge. The later version CAM2 uses a revised loudness model for the gain calculations (Moore & Glasberg, 2004). Furthermore, the gain recommendations were extended from frequencies up to 6 kHz in CAMEQ to frequencies up to 10 kHz in CAM2. In general, the gains between 1 and 4 kHz are between 1 and 3 dB lower in CAM2 than in CAMEQ (Moore & Sek, 2013).

Another established fitting method is DSL—Desired Sensation Level—from the National Centre for Audiology at Western University, Canada (Scollie et al., 2005). The DSL prescriptions were originally developed to address the specific amplification needs of children (Seewald, Ross, & Spiro, 1985). A later version of DSL, DSL v5 Adult, supports hearing instrument fitting for adults (Jenstad et al., 2007; Scollie et al., 2005).

The National Acoustic Laboratories in Australia provide the prescription rule NAL-NL1 (Dillon, 1999) and its successor NAL-NL2 (Dillon, Keidser, Ching, Flax, & Brewer, 2011; Keidser, Dillon, Flax, Ching, & Brewer, 2011). NAL-NL1 is based on the assumption that speech is fully understood when all speech components are audible. NAL-NL2 accounts for the fact that as the hearing loss gets more severe, less information is extracted even when it is audible above threshold (Keidser et al., 2011). NAL-NL2 recommends gains for frequencies up to 8 kHz, whereas NAL-NL1 is limited to 6 kHz (Moore & Sek, 2013). NAL-NL2 prescribes more low- and high-frequency gain and less mid-frequency gain than NAL-NL1 (Johnson & Dillon, 2011). In addition, the gender and hearing-aid experience of the patient can be taken into account for the gain precalculation with NAL-NL2.

Johnson and Dillon (2011) compared the latest prescription rules CAM2, NAL-NL2, and DSL v5 Adult with regard to insertion gain, loudness, and CR. For speech at a level of 65 dB sound pressure level (SPL), DSL v5 Adult provides most gain in the high frequencies. Regarding overall loudness, CAM2 is louder than DSL and NAL-NL2. With regard to the CRs, CAM2 and

NAL-NL2 provide generally more compression than DSL v5 Adult. For sloping hearing loss, however, the CR of DSL v5 Adult is higher than the CR of NAL-NL2.

These prescription rules were primarily designed for speech and have not been adapted for music. If the dynamic range of music is different than the dynamic range of speech, then the established prescription rules may be inappropriate for music. Further, different genres of music might be best handled using their own prescription rules.

Music Perception With Hearing Aids

A recent Internet-based survey by Madsen and Moore (2014) showed that many hearing-aid users experience problems with their hearing aids when listening to music. Many of these problems may be attributed to distortions introduced by the hearing aid. Hockley, Bahlmann, and Fulton (2012) argued that live music will often generate distortions in hearing aids due to the presence of high sound levels and a large dynamic range.

With regard to recorded music, there are several studies that have explored optimal hearing-aid compression schemes for music perception (see Table 1 for examples of these). In general, the linear or least compressive settings received the best preference or quality ratings. This outcome can be interpreted in the following manner. Hearing-impaired listeners may tolerate not hearing soft passages in favor of rejecting distortions or a reduction in dynamic range caused by dynamic compression. This might be partly due to different priorities when we listen to music or speech. It has been argued that the primary focus in music listening is enjoyment rather than intelligibility (Chasin & Russo, 2004). If a passage in music is rendered inaudible, it likely affects the enjoyment of that passage alone. By contrast, not hearing parts of speech affects the inaudible passages as well as the ability to follow the entire conversation. It therefore seems probable that the optimal trade-off in music between audibility and quality is shifted toward quality so that less compression is more appropriate for music than for speech.

Nevertheless, the size of the audio corpora used in the studies listed in Table 1 was consistently small, and it is thus difficult to generalize the results, given the broad diversity of music that exists in the real world. For these reasons, we chose to undertake a formal study of dynamic range across a broad range of recorded music.¹

Dynamic Range of Recorded Music

Dynamic properties of recorded music may vary with genre due to differences in various practices including instrumentation and mastering. Previous surveys of dynamic range have focused on differences in dynamic range across eras rather than across genres

Table 1. Peer-Reviewed Research Exploring the Effect of Different Compression Settings on Music Perception in Hearing-Impaired Listeners.

Reference	No. of stimuli/genres	Methods/conditions	Results
van Buuren, Festen, & Houtgast (1999)	4/piano, orchestra, pop, lied	Playback to one ear only. Conditions include all combinations of four different CRs (0.25, 0.5, 2, 4) and 3 different number of bands from (1, 4, 16) plus an uncompressed condition as benchmark.	Highest pleasantness ratings for linear amplification.
Hansen (2002)	4/orchestra, chamber music, pop (not specified which genre contributed two stimuli)	Experiment I: Conditions included 4 AT/RT combinations: 1/40 10/400 1/4000 100/4000. Gain shape was defined by NAL-R rule. CR was fixed at 2.1:1 for all channels and subjects. Experiment II: Conditions were 4 combinations of AT/RT/CT/CR: 1/40/20/2.1, 1/40/50/3, 1/40/50/2.1, 1/4000/20/2.1	Highest preference for longer RT, smaller CR, and lower CT.
Davies-Venn, Souza, & Fabry (2007)	2/classic instrumental, lied	Comparison of peak clipping, compression limiting, and WDRC for 80 dB SPL (loud) signals.	WDRC preferred over peak clipping and limiting.
Arehart, Kates, & Anderson (2011)	3/orchestra, jazz instrumental, single voice	WDRC with 18 bands in 4 combinations of CR/RT: 10/10 ms, 10/70 ms, 2/70 ms, 2/200 ms	Lower CR and longer RT were preferred. All compressive settings were rated worse than the uncompressed setting (linear gain shape NAL-R).
Higgins, Searchfield, & Coad (2012)	3/classic, jazz, rock	Comparison of ADRO and WDRC. ADRO was less compressive than WDRC.	ADRO preferred to WDRC. However, the experimental setup does not allow to attribute preference only to the linearity but also to further differences such as gain shape.
Moore et al. (2011)	2/jazz, classic	Comparison of slow (AT: 50 ms, RT: 3 s) and fast compression speeds (AT: 10 ms, RT: 100 ms) at 3 input levels (dB SPL): 50, 65, and 80.	Longer time constants were preferred for classic at 65 dB SPL and 80 dB SPL and for jazz at 80 dB SPL.
Croghan et al. (2014)	2/classic, rock	Three levels of music industry compression (no; mild, CT: -8 dBFS; heavy: -20 dBFS) combined with 3 levels of HA compression (linear; fast, RT: 50 ms; slow RT: 1000 ms) with either 3 or 18 channels.	WDRC generally least preferred. For classic, linear processing and slow WDRC was equally preferred. For rock, linear was preferred over slow WDRC. The effect of HA WDRC was more important than music-industry compression limiting for music preference.

Note. CR = compression ratio; AT = attack time; RT = release time; CT = compression thresholds; WDRC = wide dynamic range compression (multiband dynamic compression); ADRO = adaptive dynamic range compression; SPL = sound pressure level; HA = hearing aid.

(Deruty & Tardieu, 2014; Ortner, 2012; Vickers, 2010, 2011). Insights regarding variability in dynamic range across genres may inform the optimization of compression schemes for music.

Experiment I: Dynamic Range of Music Across Genres

Stimuli

The music corpus used for analysis contained 100 songs in each of 10 genres: chamber music, choir, opera, orchestra, piano music, jazz, pop, rap, rock, and *schlager*.² Song selection for the corpus was constrained to release dates between 2000 and 2014 to minimize the impact of the historic rise in compression standards that occurred prior to this era (Deruty & Tardieu, 2014; Ortner, 2012). All songs were commercially available and were retrieved in CD quality with 44.1-kHz sampling rate and 16-bit resolution. The songs were chosen from a wide range of composers and labels to avoid a bias from a single composer or mastering engineer. For the dynamic-range analysis, a 45-s segment was excerpted from the center of each song, converted to mono and normalized in root mean square (RMS) level.

The taxonomy of genre is not standardized, but music retailers are an influential source of genre classification (Pachet & Cazaly, 2000). The biggest Internet retailer of music is iTunes with a database of more than 30 million songs (Neumayr & Joyce, 2015). Although iTunes provides a convenient classification, it classifies albums rather than individual songs (Palmason, 2011). We therefore chose to use iTunes as a first classifier and added further classification criteria based on the properties described in Table 2.

For comparative purposes, speech stimuli of one female and one male native speaker of 14 different languages including Chinese, English, Spanish, French, Japanese, German, Italian, Portuguese, Hungarian, Bulgarian, Polish, Dutch, Slovenian, and Danish were added to the audio corpus. All monologues are translations of the same original German text. The monologues were recorded in professional studios in noise-free environments with a 1-m distance between the speaker and microphone. The recordings were provided by Phonak and are available in the Phonak iPFG fitting software.

Analysis

There are several definitions for measuring the dynamic range of music (Boley, Danner, & Lester, 2010). The EBU-Tech 3342 (2011) recommendation from the European Broadcasting Union defines loudness range as the difference between the 95th and 10th percentiles measured with overlapping windows of 3-s duration. Another

Table 2. Additional Criteria for Each Genre That the Songs and Stimuli (45-s Segments) Had to Meet Beyond the iTunes Genre Classification.

Genre	Additional criteria
Chamber	Instrumentation: string quartets.
Choir	Stimulus contains only vocal elements and no accompanying instruments.
Opera	Stimulus contains vocal and instrumental elements.
Orchestra	Only nonvocal excerpts accepted.
Piano	Only solo piano accepted.
Jazz	Stimulus contains drums, bass, piano, and a lead brass instrument.
Pop	Stimulus contains vocal and instrumental elements.
Rap	Stimulus contains rapped vocal elements and instrumental elements.
Rock	Stimulus contains vocals and a distorted guitar.
Schlager	Stimulus contains vocals.

common measure, the crest factor, is defined as the sound level difference between some estimation of peak level and some estimation of central tendency (e.g., average). The exact determination of the peak or time-averaged level, however, varies among studies (Croghan, Arehart, & Kates, 2014; Deruty & Tardieu, 2014; Ortner, 2012). The IEC 60118-15 (2008) recommendation was developed to characterize signal processing in hearing aids. It uses a standardized test signal that is composed of short speech segments from female Arabic, English, French, German, Mandarin, and Spanish speakers as hearing-aid input signal to analyze the signal processing. The processed signal is partitioned into third-octave bands, and the dynamic range analysis is conducted for each frequency band individually. The analyzing window is set at 125 ms with 50% overlap. The signal duration for analysis specified in the standard is 45 s. The dynamic range is usually reported for the 30th, 65th, and 99th percentiles.

The IEC standard was used as the method of analysis in this experiment. It is the most appropriate method for dynamic range analysis in the context of hearing-aid signal processing, as it uses window lengths that correspond with the time resolution of loudness perception (Chalupper, 2002; Moore, 2014) and provides a frequency-dependent analysis. It is therefore possible to analyze the dynamic range of any kind of acoustic signal including music. All stimuli were RMS equalized prior to analysis to allow for averaging across segments within genres.

Results

The results of the dynamic range analysis for each genre (100 samples) and speech (28 samples) are depicted in Figure 1. The percentiles are reported in dB SPL and shown across frequency in kHz.

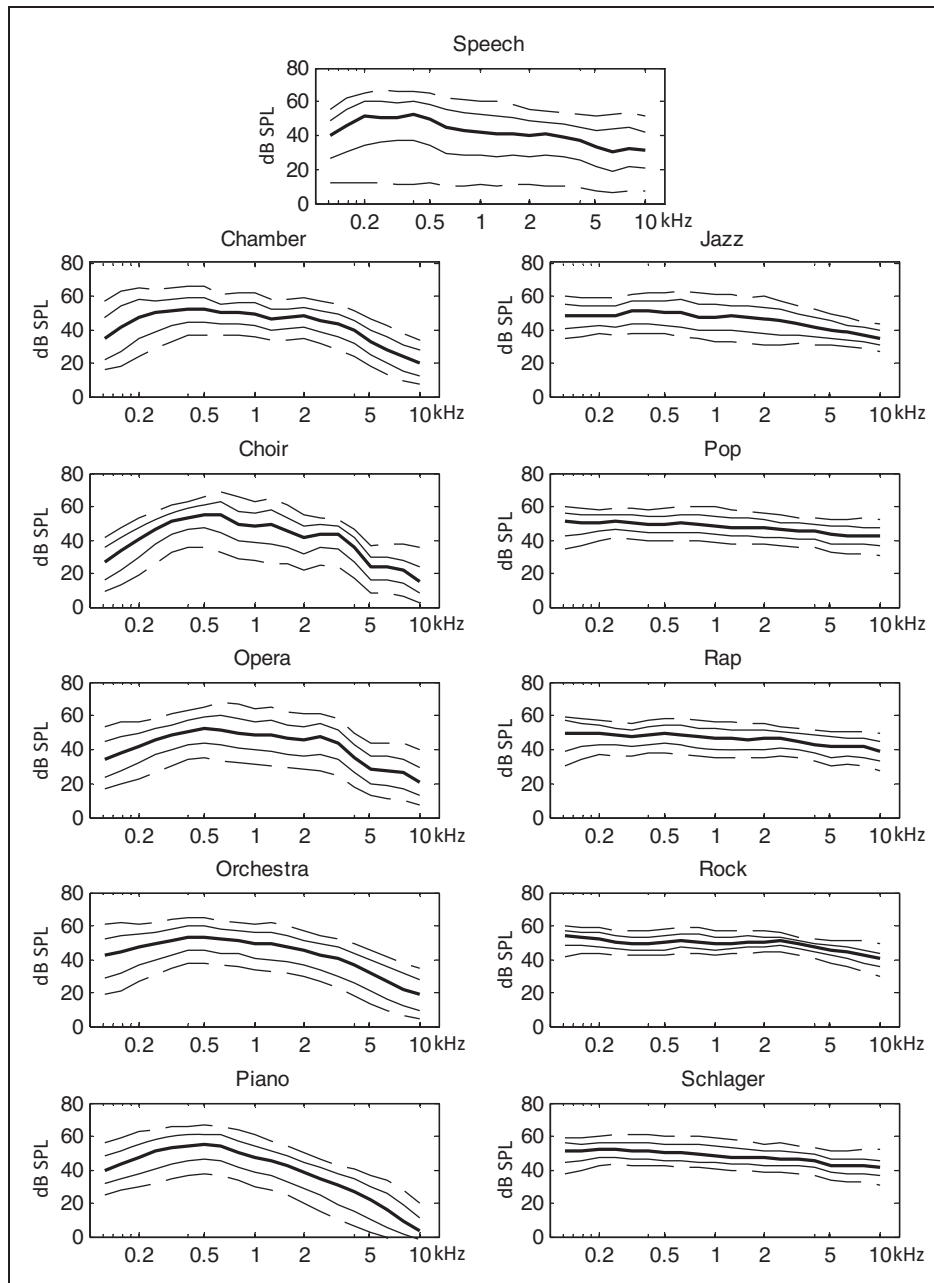


Figure 1. Dynamic range of speech and 10 different music genres. The lines represent percentiles in dB (SPL) across frequency in kHz (99th: upper dashed line, 90th: upper solid line, 65th: thick line, 30th: lower solid line, and 10th: lower dashed line). SPL = sound pressure level.

The percentiles of the modern genres (pop, rap, rock, and schlager) cluster together more than the percentiles of the classical genres (chamber, choir, opera, orchestra, piano), with the extent of clustering in jazz falling somewhere in between. Within the classical genres, opera and choir show higher differences between the highest and lowest percentiles than chamber, orchestra, and piano, especially in the frequency region between 0.5 and 2 kHz.

Figure 2 shows a dynamic range comparison of all genres calculated as the difference between the 99th and

30th percentile. According to this analysis, speech is generally largest in dynamic range in all frequency bands followed by the classical genres, jazz, and the modern genres. Speech is only locally surpassed by chamber music in the lowest two frequency bands ([110 Hz to 140 Hz]; [140 Hz to 177 Hz]) and by orchestra and opera in the lowest band.

The findings indicate that the dynamic range of music is generally smaller than the dynamic range of speech in quiet. The differences in dynamic range across genres can be attributed to acoustic properties such as

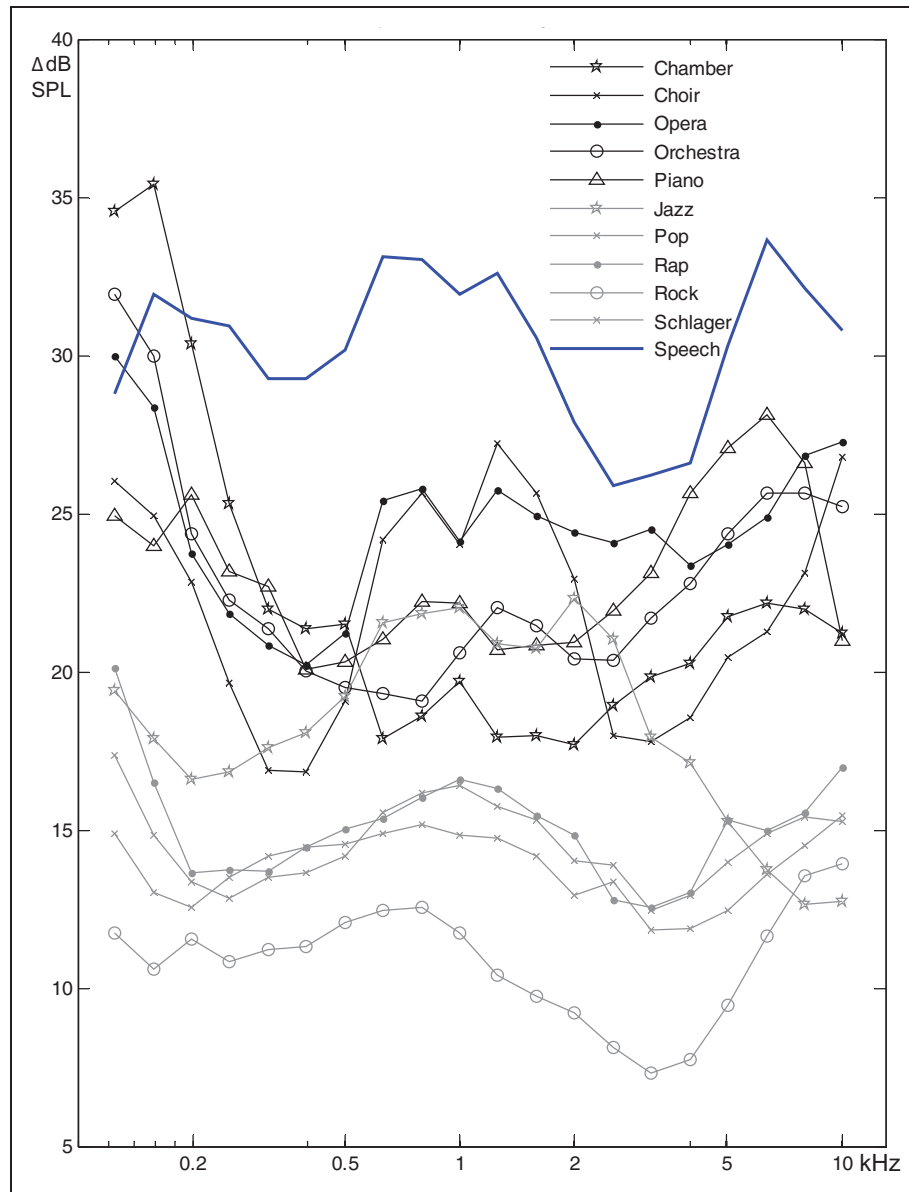


Figure 2. Dynamic range comparison between different music genres and speech across frequency. The dynamics are calculated as the difference between the 99th and 30th percentiles according to the IEC 60118-15 standard.

SPL = sound pressure level.

instrumentation and to genre-dependent compression preferences. A further investigation to reveal the extent to which acoustic properties or compression preferences contribute to overall differences in dynamic range, however, is beyond the scope of this article.

Conclusion

The dynamic range of recorded music across genres based on an audio corpus of 1,000 songs was found to be smaller than the dynamic range of monologue speech in quiet. Samples from modern genres such as pop, rap, rock, and schlager generally had the smallest dynamic range,

followed by samples from jazz and classical genres such as chamber, choir, orchestra, piano, and opera. Only in the lower frequencies was the dynamic range of speech surpassed by the dynamic range of music, and then only in the case of chamber music, opera, and orchestra.

Experiment 2: Effect of CR on Music Perception

Rationale

Dynamic compression reduces the dynamic range of a stimulus to provide audibility for low-level passages

without reaching uncomfortable loudness levels for high-level signals. Signals with a small dynamic range need less compression than signals with a large dynamic range to ensure audibility and comfortable listening levels. Based on the analysis, the dynamic range of music and particularly the dynamic range in samples from modern genres are smaller than those found in monologue speech in quiet. We therefore hypothesize that less compression is preferable for music relative to speech, especially for music from modern genres.³

To test our hypotheses, we assessed the sound quality of music stimuli from a set of genres in three different conditions: no compression (linear), full compression (NAL-NL2), and semicompression (half the CR of NAL-NL2). Apart from the CR, all other compression parameters remained equivalent across conditions. To keep the session time for the participants below 2 hr, we tested only half of the genres from Experiment I: choir, opera, orchestra, pop, and schlager. We predicted that the linear condition would provide the best sound quality for the stimuli from the modern genres (pop and schlager) and that the semicompressive condition would provide the best quality for the stimuli from the classical genres (choir, opera, and orchestra). In addition to sound quality, participants were asked to provide direct judgments of dynamics. Dynamics was defined as the perceptual correlate to dynamic range. The dynamics ratings were used to verify whether the differences in dynamic range across the three dynamic compression schemes were perceptible. Potential differences between conditions in spectral shape and loudness were controlled, and participants were additionally asked to provide direct judgments of loudness.

Participants

Thirty-one hearing-impaired listeners (ages 48–80, mean age 69) were recruited from the internal database of Sonova AG headquarters, Stäfa, Switzerland. The audiometric data were assessed within 3 months of the first test date. All participants were Swiss and native German speakers.

The participants' music experience was assessed with a questionnaire according to Kirchberger and Russo (2015). The questionnaire asks participants a series of questions about music training and activity to arrive at an overall measure of music experience.

All participants were fitted with Phonak Audéo V50 receiver-in-canal hearing aids according to the NAL-NL2 prescription rule. If the participants did not accept the first fit, changes were made until full acceptance was achieved. The changes affected the overall gain or gain shape but not the compression strength of the setting. For the coupling, standard domes (open, closed, and power) and receivers (standard, power) were used.

The choice of the individual dome was based on the recommendation from the Phonak Target™ fitting software but was subject to change according to the participant's preference. Participants had worn the hearing aids for at least 5 weeks and 2 months on average before the test sessions began. All information regarding the participants is provided in Table 3.

Test Conditions

In the experiment, participants compared the effect of linear, semicompressive, and compressive parameterizations of hearing aids on the perception of 20 music segments. In the compressive condition, the compression parameters of the participant's individual NAL-NL2 fitting were used. In the linear and the semicompressive condition, the CR of the compressive condition was modified to $CR_{linear} = 0$ and to $CR_{semi} = 1/2 \cdot CR_{comp}$, respectively. All other compression parameters remained the same. The hearing aids had 20 compression channels. The time constants were band-dependent and ranged from 10 ms for attack and 50 ms for release in the lower frequencies to 6 ms for attack and 37 ms for release in the higher frequencies.

Control of spectral shape and level. Dynamic compression can change the overall level of a signal. Moreover, in multiband dynamic compression systems, such as that which can be found in state-of-the-art hearing aids, the gain calculation and application differs across bands. As a consequence, multiband dynamic compression can change the level of each band independently and therefore also modify the spectral shape of a signal (Chasin & Russo, 2004). The experiment, however, aims at investigating the perception of different dynamic ranges. Changes in spectral shape across conditions would impose a bias. To limit this potential bias, controls were put in place so that within each band, the same gain was applied (on average) in the linear, semicompressive, and compressive setting. Specifically, the gain curves of the three different conditions within each band were aligned to intersect at the RMS level of the input in the corresponding band (Figure 3). As the intersection was already defined by the compressive gain curve (NAL-NL2 fitting) and the RMS levels of the stimuli, the linear and semicompressive gain curves were adjusted so that they intersected at the same point. The RMS levels of the stimuli across bands were measured with the same setup as described later in section 'Test Stimuli'. The measurements were retrieved from the hearing aid so that any potential inaccuracies of the transfer functions for the loudspeakers, the head, and the microphones were accounted for. To further increase the precision, the RMS measurements were conducted for both ears so

that two different RMS sets were available for the right and left hearing aid correspondingly.

The gain curve alignment was carried out manually for both conditions, each band (20), segment (20), participant (31), and both ears resulting in $2 \times 20 \times 20 \times 31 \times 2 = 49,600$ manually adjusted curves.

Control of loudness. We further controlled the loudness of all stimuli with the dynamic loudness model (DLM) by Chalupper and Fastl (2002). In contrast to other loudness models such as the well-established Cambridge

loudness model developed by Moore (2014), DLM supports the loudness calculation of nonstationary signals in hearing-impaired listeners. To simulate each participant's individual hearing loss, the air-conduction thresholds, bone-conduction thresholds, and uncomfortable loudness levels at 0.5, 1, 2, and 4 kHz were entered into the model. As the lengths of the stimuli were between 9 and 16 s, it was necessary to average the loudness values of the model across time. The long-term loudness of the whole stimulus was determined according to Croghan, Arehart, and Kates (2012) as the mean of all long-term

Table 3. Characteristics of the Test Participants: Audiometric Data (dB HL), Age (Years), Hearing Aid Experience/HAE (Years), Music Experience/ME (Range From -3: Low to 4: High), and Coupling (Dome: o. = Open, cl. = Closed, po. = Power; Rec./receiver: s = Standard, p = Power).

	Left ear, frequency (kHz)								Right ear, frequency (kHz)								Cpl					
	0.1	0.25	0.5	1	2	4	6	8	0.1	0.25	0.5	1	2	4	6	8	Sex	Age	HAE	ME	Dome	Rec.
P1	30	30	35	60	60	70	115	105	30	35	35	45	60	70	90	80	m	80	11	2.5	cl.	s
P2	20	10	5	20	65	75	85	100	25	20	10	5	40	70	90	90	f	64	>10	1	o.	s
P3	10	10	5	25	45	60	85	100	15	10	15	10	55	90	100	100	m	72	7	-1.5	o.	s
P4	5	10	20	55	65	70	80	75	10	10	15	30	75	70	90	80	m	74	12	-2	cl.	s
P5	40	45	45	65	70	80	90	80	35	35	40	50	60	80	80	75	m	67	>10	1.5	po.	s
P6	25	40	40	75	75	100	105	105	15	20	30	35	45	65	75	85	m	73	8.5	-1	cl.	s
P7	10	20	55	70	75	85	95	90	10	10	20	50	65	95	100	100	m	72	6	-2.5	cl.	p
P8	15	15	25	50	55	75	100	95	15	15	20	25	70	85	100	105	m	70	0.2	-2	o.	s
P9	20	15	5	30	45	90	105	95	10	10	10	45	40	100	100	100	m	70	10	-3	o.	p
P10	25	30	55	70	75	75	80	85	25	25	35	50	70	70	75	75	m	65	8	0	cl.	s
P11	15	30	55	75	75	80	85	90	40	45	50	65	75	85	100	85	m	73	16	-2.5	cl.	p
P12	25	40	65	90	75	75	105	121	30	40	45	60	90	75	90	121	m	68	50	-3	po.	p
P13	20	30	35	35	50	60	70	65	30	30	35	40	40	65	85	80	m	69	9	-2.5	po.	s
P14	15	15	10	55	60	65	90	95	15	10	10	15	55	65	80	95	m	76	12	0	cl.	s
P15	40	50	55	80	85	95	100	NT	20	25	40	40	55	80	NT	85	m	71	5	0.5	cl.	p
P16	30	30	65	80	80	85	90	85	25	25	35	40	55	80	90	90	m	75	21	1.5	po.	p
P17	10	5	10	45	65	80	105	85	5	10	10	15	65	80	90	75	f	55	7	-2	cl.	s
P18	35	55	60	65	65	65	80	65	45	50	55	55	60	55	65	65	f	66	25	0	cl.	s
P19	55	60	55	55	55	55	65	70	35	55	55	55	60	70	70	65	m	73	10	-2	po.	s
P20	35	60	55	65	60	75	105	106	15	40	50	50	60	60	60	100	m	68	30	-2	po.	p
P21	60	70	80	85	75	75	80	80	55	60	60	65	85	75	80	85	m	66	25	-2	po.	p
P22	20	40	45	60	55	60	65	60	20	25	35	50	55	60	65	65	m	67	13	2	cl.	p
P23	35	55	55	65	70	75	80	80	40	40	55	50	60	75	75	85	m	77	30	0.5	po.	s
P24	50	60	50	65	60	65	80	75	35	40	35	40	55	50	60	70	f	78	3	-2	po.	s
P25	35	55	55	65	75	80	90	105	55	60	60	60	70	85	100	105	m	76	>10	-1.5	po.	p
P26	30	60	70	55	65	70	75	85	30	35	60	70	65	80	85	105	m	67	37	1	po.	p
P27	25	45	45	50	70	60	60	55	20	30	40	50	50	65	55	60	m	48	7	-2	cl.	s
P28	15	35	45	70	60	65	70	95	20	20	40	60	75	70	65	85	m	70	16	1.5	po.	p
P29	30	55	60	55	45	45	50	70	30	40	45	55	50	40	50	65	f	69	10	-3	po.	s
P30	25	45	65	70	60	70	75	85	30	30	45	65	70	70	75	70	m	51	>10	4	cShell	s
P31	40	45	55	50	55	60	65	90	45	45	45	55	60	55	80	80	f	72	28	0.5	po.	p

Note. NT indicates a hearing threshold that was not tested.

loudness levels that were above two phons (corresponding to absolute threshold). If loudness differences between the three conditions of a music segment were greater than one phon, the linear or semicompressed versions were amplified so that the deviations were within one phon of the compressed version. Figure 4 illustrates the loudness curves for one segment and participant combination (Segment 3, Participant 29).

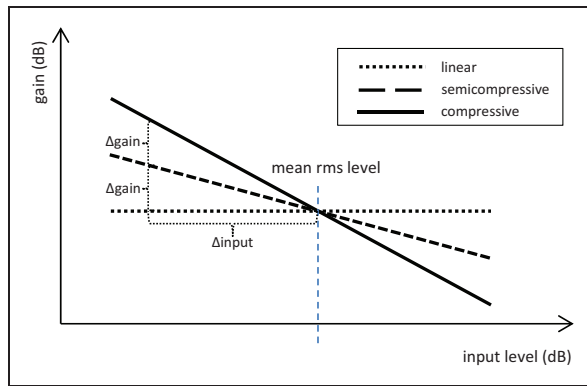


Figure 3. Schematic gain curves within a compression band for the linear, semicompressive, and compressive condition. The compression ratio of the semicompressive condition ($CR^{semi} = \Delta gain / \Delta input$) is half the compression ratio of the compressive condition ($CR^{comp} = 2 * \Delta gain / \Delta input$).

Test Stimuli

Twenty songs (four from each genre choir, opera, orchestra, pop, and schlager) were used from the audio corpus described in the first experiment. Segments were selected at points consistent with musical phrasing. The average dynamic range of the segments of each genre was similar to the dynamic range of the larger sample used in Experiment I (Figure 5).⁴ Further details about the segments are provided in Table 4.

The test stimuli for each participant were generated by recording the music segments with a KEMAR manikin (model 45BB by G.R.A.S.) that had hearing aids attached to the ears. Music segments were played back in stereo via two loudspeaker pairs in 1.2 m distance at an angle of 30° and -30° , as common practice in audio engineering (Dickreiter, Dittel, Hoeg, & Wöhr, 2014). Each loudspeaker pair consisted of a mid- to high-range speaker (Meyer Sound MM-4-XP) and an aligned subwoofer (Meyer Sound MM-10-XP). The output level of each music segment was normalized to 65 dB SPL to ensure realistic and comparable listening levels (Croghan et al., 2012, 2014). The stimuli were prepared for each participant individually. Exact copies of the participant's hearing aids were fitted to the KEMAR, including the coupling (open, closed, or power dome); receiver (stand-alone or power); and individual fitting. For each of the 60 test stimuli (20 segments \times 3 conditions), the

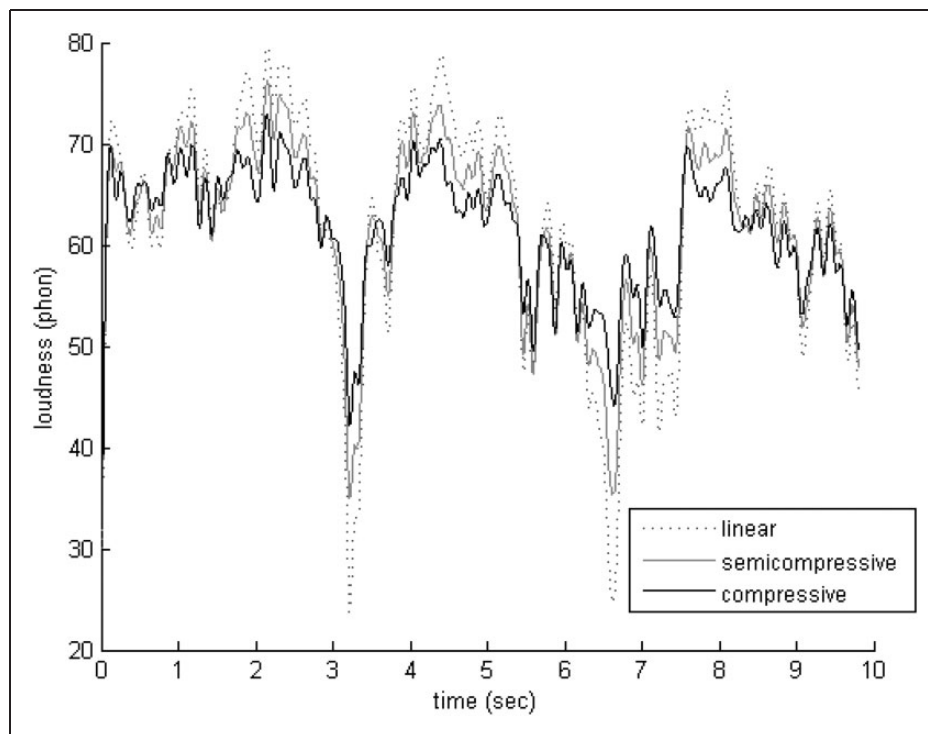


Figure 4. Example loudness curves of the linear, semicompressive, and compressive version (here for participant 29 and stimulus 3).

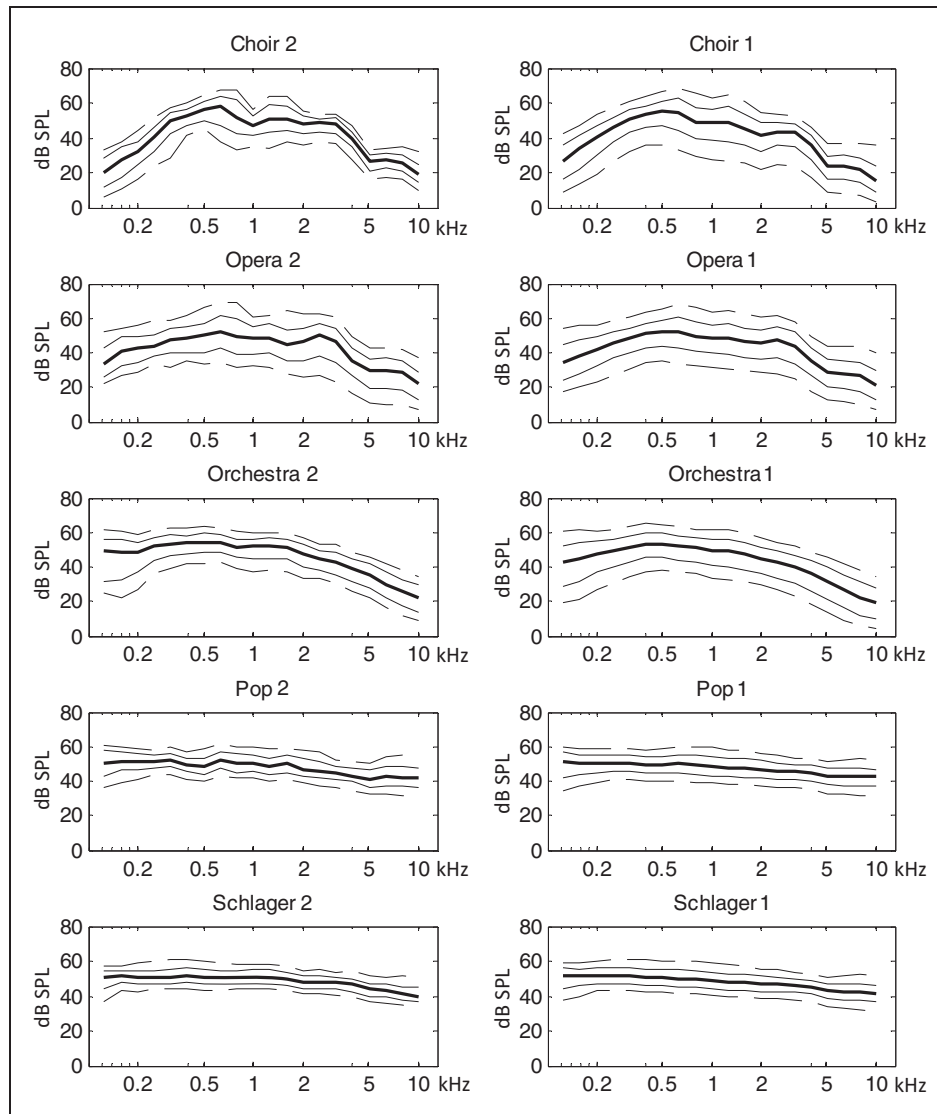


Figure 5. Dynamic range of the test sample in Experiment 2 (left column) and the audio corpus in Experiment 1 (right column) for the genres choir, opera, orchestra, pop, and schlager. The lines represent percentiles in dB (SPL) across frequency in kHz (99th: upper dashed line, 90th: upper solid line, 65th: thick line, 30th: lower solid line, and 10th: lower dashed line).

corresponding hearing-aid parameterizations were uploaded to the hearing aids prior to recording the individual stimulus. Recordings were made in stereo with microphones located in both KEMAR ears at the position of the eardrums. The recordings were equalized before further processing to compensate for the ear resonance of the KEMAR and the frequency response of the Sennheiser HD 600 headphones that were used for playback in the test sessions.

Listening Test

The participants conducted listening tests on two separate occasions to compare the linear, semicompressive, and compressive parameterizations. The setup of the

music test was a double-blind multistimulus test method similar to a Multiple Stimuli with Hidden Reference and Anchor (MUSHRA) setup as described in the recommendation ITU-R BS 1534-1 (2001). In each trial, participants had to rate 3 stimuli on a scale from 0 to 100. The three stimuli differed in dynamic range and were processed by a linear, a semicompressive, or a compressive (NAL-NL2) hearing-aid setting.

Participants were asked to make judgments along two dimensions: sound quality and dynamics. Dynamics was explained as the difference between loud and soft passages. The scales used for the ratings ranged from *poor* to *good* for sound quality and from *low* to *high* for dynamics. Participants were instructed to focus on the relative differences between the conditions within one trial rather

Table 4. Details About the 20 Music Segments of Experiment 2 (Music Taste: Participant's Average Music Taste Ratings [−1: Do Not Like; 1: Like]).

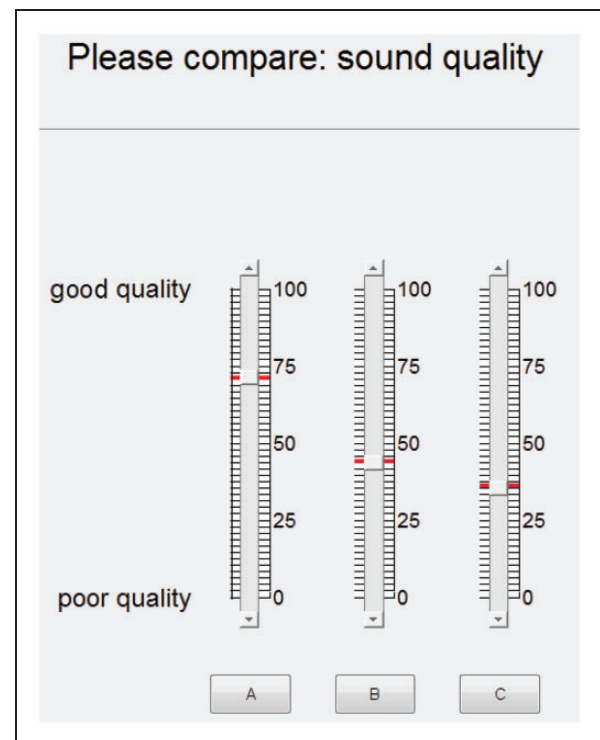
Genre	Title	Artist	Length (s)	Release (year)	Music taste
Choir	Op. 113 No. 5—Frauenchöre Kanons—Wille, wille, will	Brahms	9.9	2003	0.31
	Op. 29—Zwei Motetten—Es ist das Heil uns kommen her	Brahms	12.2	2003	0.17
	Jube Domine, for soloists & double chorus in C major	Mendelssohn	10.0	2002	0.17
	Op. 69 No. 1—I Herr nun lassest du deinen Diener in Frieden fahren	Mendelssohn	15.1	2002	0.28
Opera	Carmen—Act 1—Attention! Chut! Attention! Taisons-Nous!	Bizet	11.2	2003	0.21
	Don Giovanni—Act 1—Udisti? Qualche bella	Mozart	10.1	2007	0.41
	Boris Godunov—Act 2—Uf! Tyazheló! Day Dukh Perevedú	Mussorgsky	16.0	2002	−0.21
	Rosenkavalier—Act 3—Haben euer Gnaden noch weitere Befehle	Strauss	16.3	2001	0.14
Orchestra	Symphonie No. 9 in C-major KV 73—Molto allegro	Mozart	10.7	2002	0.52
	Symphony No. 1 in E-flat major—Molto Allegro	Mozart	8.8	2002	0.79
	Symphonie No. 8—Tempo di Menuetto	Beethoven	10.0	2002	0.69
	Symphonic Poem Op. 16—Ma Vlast Hakon Jarl	Smetana	11.0	2007	0.31
Pop	Downtown	Gareth Gates	9.2	2002	0.17
	Tears and rain	James Blunt	11.3	2004	−0.07
	Shape of my heart	Backstreet Boys	12.5	2000	0.31
	Rock with you	Michael Jackson	8.4	2003	0.21
Schlager	Kleine Schönheitsfehler	Sylvia Martens	12.6	2011	0.38
	Sternenfänger	Leonard	10.9	2011	0.21
	Tränen der Liebe	Peter Rubin	10.3	2011	0.00
	Der Mann ist das Problem	Udo Jürgens	14.7	2014	0.17

than trying to make absolute ratings across trials. Participants were assigned to one of two groups and carried out 20 trials per dimension, one trial for each music segment. Group A started with judgments about the sound quality dimension, while Group B started with judgments about the dynamics dimension. Allocation of participants to the two groups was controlled in a manner that minimized differences in age, hearing loss, or music experience.

The tests were implemented in MATLAB and displayed as a graphical user interface on a touch screen in front of the participants. The stimuli were randomly assigned to one of the three channels: A, B, or C (cf. Figure 6). The stimuli were looped endlessly, and the transitions from the end to the beginning of each loop were not noticeable, as the music segments were selected to preserve musical phrasing.

Participants were freely able to switch between channels (stimuli) at any given time. A 5-ms cross-fade was applied while channel switching to avoid switching artifacts such as pops.

Although loudness was well controlled in the experiment, we additionally asked participants to compare the loudness of the test stimuli across conditions. In 20 trials, participants had to compare and rate the loudness of the linear, semicompressive, and compressive version of a

**Figure 6.** Example screen of the main test.

music segment on a scale from 0 to 100. Scale ends were labeled from *soft* to *loud*. Participants were instructed not to focus on singular events but on the stimuli as a whole to provide an overall impression of loudness.

Finally, participants indicated their music taste by rating how much they liked the music segments. They listened to the semicompressed versions of the music segments and rated them on an absolute three-step scale (-1: do not like, 0: neutral, 1: like).

Results

Quality. For the statistical analysis, the IBM SPSS Statistic 22 software program was used. The quality ratings were subjected to a repeated-measures analysis of variance (ANOVA) with session (test, retest); condition (linear, semicompressive, compressive); and genre (choir, opera, orchestra, pop, schlager) as within-subjects factors. In cases where sphericity was violated, Greenhouse-Geisser corrections were used if the epsilon test statistic was lower than .75; otherwise, the Huynh-Feldt corrections were applied as proposed by Girden (1992). There was no effect of session, $F(1, 30) = 0.115$, $p = .737$, but a significant effect of condition, $F(1.2, 38.5) = 21.09$, $p < .001$, and genre, $F(4, 120) = 2.705$, $p = .034$. There were no interactions between session and condition, $F(1.39, 41.6) = 0.130$, $p = .800$; session and genre, $F(2.49, 74.5) = 0.728$, $p = .514$; or condition and genre, $F(5.5, 164.4) = 1.746$, $p = .120$. The linear condition was rated highest in quality (60.53) followed by the semicompressive (54.30) and the compressive condition (47.43; Figure 7). A Bonferroni-corrected post-hoc comparison of the conditions revealed significant differences between all pairwise comparisons: linear versus semicompressive, $F(1, 30) = 13.08$, $p = .001$; linear versus compressive, $F(1, 30) = 24.25$, $p = .001$; and semicompressive versus compressive, $F(1, 30) = 21.74$, $p = .001$.

Dynamics. For the data on dynamics, the same statistical analysis was applied as for the data on quality. There was no effect of session, $F(1, 30) = 0.067$, $p = .798$, but a significant effect of condition, $F(1.1, 32) = 49.17$, $p < .001$, and genre, $F(2.52, 75.5) = 4.030$, $p = .015$. There was no interaction between session and condition, $F(1.45, 43.5) = 0.060$, $p = .890$, or session and genre, $F(2.84, 85.3) = 0.981$, $p = .402$, but a significant interaction between condition and genre, $F(3.5, 105.9) = 3.47$, $p = .014$. The linear condition had the highest ratings for dynamics (63.63), followed by the semicompressive (52.32) and the compressive condition (43.88; Figure 7). The mean difference scores are displayed in Figure 7. A Bonferroni-corrected post-hoc comparison of the conditions revealed that differences were significant for all pairwise comparisons: linear versus semicompressive, $F(1, 30) = 51.73$, $p < .001$;

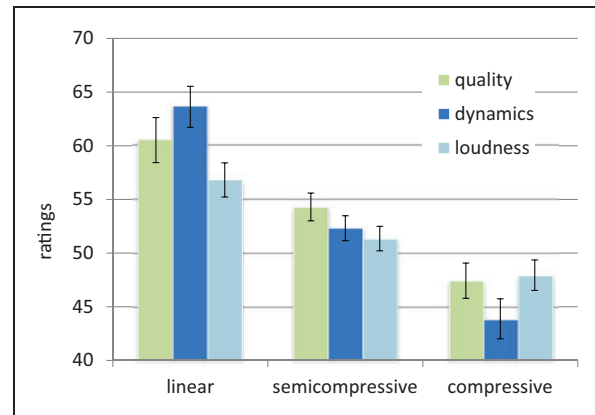


Figure 7. Ratings and standard error for the data on dynamics, quality, and loudness in the linear, semicompressive, and compressive condition.

linear versus compressive, $F(1, 30) = 50.61$, $p < .001$; and semicompressive versus compressive, $F(1, 30) = 39.52$, $p < .001$. With regard to the interaction between condition and genre, the differences in dynamics between the linear and the compressive condition were largest for opera ($\Delta = 23.12$) followed by orchestra ($\Delta = 21.02$), choir ($\Delta = 20.01$), pop ($\Delta = 18.55$), and schlager ($\Delta = 16.05$).

Loudness. An ANOVA was carried out with condition (linear, semicompressive, compressive) and genre (choir, opera, orchestra, pop, schlager) as within-subject factors.

There was a significant effect of condition, $F(1.1, 33.5) = 26.99$, $p < .001$, and genre, $F(2.16, 64.7) = 4.030$, $p = .020$, but no interaction between condition and genre, $F(5.1, 153.7) = 2.157$, $p = .06$. The linear condition was rated loudest (56.81), followed by the semicompressive (51.35) and the compressive condition (47.94). A Bonferroni-corrected post-hoc comparison of the conditions revealed significant differences between all pairwise comparisons: linear versus semicompressive, $F(1, 30) = 29.28$, $p < .001$; linear versus compressive, $F(1, 30) = 14.68$, $p = .001$; and semicompressive versus compressive, $F(1, 30) = 28.11$, $p < .001$.

Discussion

Quality. The findings confirmed the hypothesis that less compression benefits the perception of quality. The quality of the linearly processed stimuli was judged to be best, followed by the semicompressive and the compressive setting. Against the background that the participants were acclimatized to hearing aids fitted with the compressive setting, the results appear even stronger. On the basis of the mere-exposure effect (Bradley, 1971; Gordon & Holyoak, 1983; Ishii, 2005; Szpunar, Schellenberg, & Pliner, 2004; Zajonc, 1980), we would expect a bias toward the compressive scheme that

participants had grown accustomed to during the acclimatization period.

The results from this study are consistent with results from previous studies in confirming that less compression for music is beneficial for sound quality. A reason why the least compressive condition consistently yielded the best sound quality might be that the primary focus in music listening is enjoyment rather than intelligibility (Chasin & Russo, 2004). Dynamic compression introduces distortion (Kates, 2010). Listeners might be more sensitive to distortion introduced by compression in music than in speech. Therefore, the trade-off between audibility for soft parts and introducing distortion might shift toward less distortion and therefore to even less compression than the dynamic properties might suggest.

The optimal compression strength, however, varied on an individual level. Seven participants judged the quality of the semicompressive or compressive processing superior to the linear processing. To understand the reason for these perceptual differences, participants were asked after the tests to verbally describe their subjective quality criteria. While instrument separation, liveliness, clarity, bandwidth, and intelligibility of lyrics were assessed as positive factors, inaudible passages or loudness peaks were mentioned as detrimental for sound quality. The latter criterion was shared among all four participants who gave the highest quality ratings for the compressive conditions.

Dynamics. The perception of dynamic differences between the linear, semicompressive, and compressive conditions completely align with the experimental manipulations. Participants rated the linear version highest in dynamics, followed by the semicompressed and the compressed versions. As the pairwise comparisons between versions were significant and there was also no effect of session, it can be assumed that participants reliably perceived differences between the three dynamic conditions.

Furthermore, perceptual differences between the conditions varied across genre. The effect of compression was perceptually bigger for genres with a larger original dynamic range such as opera and orchestra than in less dynamic genres such as pop or schlager.

The fact that participants were asked to judge dynamics as well as quality may have biased participants to consider quality in terms of dynamics. We ran the repeated-measures ANOVA with a between-subjects factor that divided the participants into two subsets: One subset contained the participants who started with evaluating differences in dynamics ($N=15$), and the other subset contained the participants who started with evaluating differences in quality ($N=16$). The analysis revealed that the effect of order was not significant, $F(1.2, 38.5) = 2.562, p = .120$.

Loudness. Although loudness was equalized between conditions prior to the experiment using the DLM loudness model by Chalupper and Fastl (2002), participants rated the stimuli in the linear condition significantly louder than in the semicompressed condition and softest in the compressed condition. Two reasons might have contributed to this deviation: First, our approach to average loudness in phons as proposed by Croghan et al. (2012) might underestimate the long-term loudness perception of dynamic stimuli in hearing-impaired listeners. Second, loudness judgments for stimuli with a length of approximately 10 s are extremely difficult. By design, the linear stimuli have the loudest passages but also the softest passages compared with the semicompressive or compressive stimuli. Participants may overvalue loud passages when trying to determine an average for the overall loudness perception of the stimuli.

To outweigh a potential effect of loudness differences between conditions on quality ratings, an experimental post-hoc repeated-measures ANOVA was conducted in which the loudness scores were subtracted from the quality test scores. As with the original analysis, the differences in the linear condition were highest (3.75), followed by the semicompressive (2.90) and the compressive condition (-0.16). The differences between conditions were significant, $F(1.4, 42.9) = 4.81, p = .022$.

General Discussion

The present study analyzed the dynamic range of an audio corpus of 1,000 recorded songs and 28 monologue speech samples in quiet. A genre-specific analysis revealed that the recorded music samples of all genres generally had smaller dynamic ranges than the speech samples. As a consequence, a further study was conducted in which the compression of the NAL-NL2 prescription rule was compared with linear and semicompressive processing. There was a significant trend that linear amplification yielded the best sound quality, followed by semicompressive and compressive (NAL-NL2) processing.

The current study was based on recorded music. Live music, singing, or practicing an instrument are other forms of music consumption. Further research is required to analyze the acoustic properties in these auditory scenes and to adjust the dynamic compression in hearing aids accordingly. An ongoing challenge for hearing aids is the processing of high-level peaks that are often experienced in live music (e.g., Ahnert, 1984; Cabot, Center, Roy, & Lucke, 1978; Fielder, 1982; Sivian, Dunn, & White, 1931; Wilson et al., 1977; Winkel, 1962).

The genre-based classification as performed in this study is one approach to organize music. Further fragmentation might reveal systematic differences in dynamic

range within genres (e.g., Baroque vs. Romantic orchestra music) that should be addressed by hearing-aid signal processing. Ideally, the compression parameters would adapt to individual songs or even adapt within a song. Streaming services could potentially incorporate information about the dynamic range so that the hearing aids can optimize the compression accordingly. A short delay in playback (look-ahead time) would also allow the possibility of continually adjusting the compression parameters.

There is a secondary finding from the analysis of compression in recorded music that is also worth noting. As may be seen in Figure 1, the spectra of the modern genres, speech, and the classical genres are distinct. The high frequencies are particularly prominent in the modern genres, followed by speech and then classical genres. Because of these differences, one common multi-band dynamic compression setting may apply too little gain for modern genres or too much gain for classical genres in the high-frequency bands. It may be that benefit would be gained by setting the compression curves differently for each of these three signal categories.

Acknowledgments

We particularly want to thank Dr. Markus Hofbauer and Dr. Peter Derleth for their valuable support and our fruitful discussions. We further wish to thank the participants for taking part in this study and for appearing to each session as scheduled. We also thank our colleagues for letting us occupy the laboratory facilities for a long time of fitting and measurements.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research protocol for this study was approved by grant (KEK-ZH. Nr. 2014-0520) from the Ethics Committee Zurich, Switzerland and was registered at <http://clinicaltrials.gov> (ID: NCT02373228). Financial support for this research was provided by ETH Zürich, Phonak AG, and the Natural Sciences and Engineering Research Council of Canada.

Notes

1. We chose to focus our investigation on recorded music exclusively. Live music tends to have a larger dynamic range than the recorded and mastered reproduction; however, nowadays, the consumption of live music is far less common than the consumption of recorded music. A Canadian study with a sample of 1,232 participants (Crofford, 2007) yielded an average of 10.5 live concert attendances per year. Assuming an average of 3 hr per concert, the accumulated amount of live music consumed within a year would be less than 40 hr. Total music consumption,

however, has been estimated at 2.5 hr per day (Bersch-Burauel, 2004), adding up to over 900 hr per year. In this comparison, live music consumption would account for less than 4% of total music consumption.

2. Also known as German entertainer music.
3. Chamber music, opera, and orchestra surpass the dynamic range of speech in the lower frequencies; however, these frequency components are less relevant in the context of dynamic compression with hearing aids. Hearing loss is generally less predominant in the lower frequencies (Bisgaard, Vlaming, & Dahlquist, 2010), and low-frequency gains are constricted by potential feedback from leakage or vents (Cox, 1982) and physical limits of the speaker. As a consequence, the hypothesis refers to frequencies above 200 Hz for which the dynamic range of speech is larger than the dynamic ranges of all analyzed music genres (Figure 2).
4. To conduct the dynamic range analysis, the IEC standard requires a stimulus length of 45 s. As the segments were significantly shorter (Table 4), the segments were looped. To avoid anomalies at the transitions, the segments were linearly cross-faded with a 50-ms window.

References

- Ahnert, W. (1984). *The sound power of different acoustic sources and their influence in sound engineering*. Paper presented at Proceedings of the 75th Convention of the Audio Engineering Society, Preprint 2079. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=11685>.
- Arehart, K. H., Kates, J. M., & Anderson, M. C. (2011). Effects of noise, nonlinear processing, and linear filtering on perceived music quality. *International Journal of Audiology, 50*(3), 177–190.
- Bersch-Burauel, A. (2004). *Entwicklung von Musikpräferenzen im Erwachsenenalter: Eine explorative Untersuchung [Development of musical preferences in adulthood: An exploratory study]* (Doctoral dissertation). University of Paderborn, Germany.
- Bisgaard, N., Vlaming, M. S., & Dahlquist, M. (2010). Standard audiograms for the IEC 60118-15 measurement procedure. *Trends in Amplification, 14*(2), 113–120.
- Boley, J., Danner, C., & Lester, M. (2010, November). *Measuring dynamics: Comparing and contrasting algorithms for the computation of dynamic range*. Paper presented at Proceedings of the 129th Convention of the Audio Engineering Society Convention 129. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=15601>.
- Bradley, I. L. (1971). Repetition as a factor in the development of musical preferences. *Journal of Research in Music Education, 19*(3), 295–298.
- Cabot, R. C., Genter, I. I., Roy, C., & Lucke, T. (1978). Sound levels and spectra of rock music. *Journal of the Audio Engineering Society, 27*(4), 267–284.
- Chalupper, J. (2002). *Perzeptive Folgen von Innenohrschwerhörigkeit: Modellierung, Simulation und Rehabilitation [Perceptive consequences of sensorineural hearing loss: Modeling, simulation, and rehabilitation]*. Herzogenrath, Germany: Shaker.

- Chalupper, J., & Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica United with Acustica*, 88(3), 378–386.
- Chasin, M., & Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, 8(2), 35–47.
- Cox, R. M. (1982). Combined effects of earmold vents and suboscillatory feedback on hearing aid frequency response. *Ear and Hearing*, 3(1), 12–17.
- Crofford, J. (2007). *Public survey results – Music industry review 2007*. Department of Culture, Youth, and Recreation, Saskatchewan, Canada. Retrieved from www.pcs.gov.sk.ca/Public-Survey.
- Croghan, N. B., Arehart, K. H., & Kates, J. M. (2012). Quality and loudness judgments for music subjected to compression limiting. *The Journal of the Acoustical Society of America*, 132(2), 1177–1188.
- Croghan, N. B., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, 35(5), e170–e184.
- Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology*, 18(8), 688–699.
- Deruty, E., & Tardieu, D. (2014). About dynamic processing in mainstream music. *Journal of the Audio Engineering Society*, 62(1/2), 42–55.
- Dickreiter, M., Dittel, V., Hoeg, W., & Wöhr, M. (2014). *Handbook of recording-studio technology* (vol 1). Berlin, Germany: Walter de Gruyter GmbH & Co KG.
- Dillon, H. (1999). NAL-NL1: A new procedure for fitting nonlinear hearing aids. *The Hearing Journal*, 52(4), 10–12.
- Dillon, H., Keidser, G., Ching, T. Y., Flax, M., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Phonak Focus*, 40, 1–10. Retrieved from <https://www.phonakpro.com/ch/b2b/de/evidence/publications/focus.html>
- EBU-Tech 3342. (2011). *Loudness range: A descriptor to supplement loudness normalization in accordance with EBU R 128*. Retrieved from <https://tech.ebu.ch/docs/tech/tech3342.pdf>
- Fielder, L. D. (1982). Dynamic-range requirement for subjectively noise-free reproduction of music. *Journal of the Audio Engineering Society*, 30(7/8), 504–511.
- Giannoulis, D., Massberg, M., & Reiss, J. D. (2012). Digital dynamic range compressor design—A tutorial and analysis. *Journal of the Audio Engineering Society*, 60(6), 399–408.
- Girden, E. (1992). *ANOVA: Repeated measures*. Newbury Park, CA: Sage.
- Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the “mere exposure” effect. *Journal of Personality and Social Psychology*, 45(3), 492–500.
- Hansen, M. (2002). Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and Hearing*, 23(4), 369–380.
- Higgins, P., Searchfield, G., & Coad, G. (2012). A comparison between the first-fit settings of two multichannel digital signal-processing strategies: Music quality ratings and speech-in-noise scores. *American Journal of Audiology*, 21(1), 13–21.
- Hockley, N. S., Bahlmann, F., & Fulton, B. (2012). Analog-to-digital conversion to accommodate the dynamics of live music in hearing instruments. *Trends in Amplification*, 16(3), 140–145.
- IEC 60118-15. (2008). *Electroacoustics – Hearing aids – Part 15: Methods for characterizing signal processing in hearing aids*. Geneva, Switzerland: International Electrotechnical Commission.
- Ishii, K. (2005). Does mere exposure enhance positive evaluation, independent of stimulus recognition? A replication study in Japan and the USA1. *Japanese Psychological Research*, 47(4), 280–285.
- ITU-R BS 1534-1. (2001). *Method for the subjective assessment of intermediate sound quality (MUSHRA)*. Geneva, Switzerland: International Telecommunications Union.
- Jenstad, L. M., Bagatto, M. P., Seewald, R. C., Scollie, S. D., Cornelisse, L. E., Scicluna, R. (2007). Evaluation of the desired sensation level [input/output] algorithm for adults with hearing loss: The acceptable range for amplified conversational speech. *Ear and Hearing*, 28(6), 793–811.
- Johnson, E. E., & Dillon, H. (2011). A comparison of gain for adults from generic hearing aid prescriptive methods: Impacts on predicted loudness, frequency bandwidth, and speech intelligibility. *Journal of the American Academy of Audiology*, 22(7), 441–459.
- Kates, J. M. (2010). Understanding compression: Modeling the effects of dynamic-range compression in hearing aids. *International Journal of Audiology*, 49(6), 395–409.
- Katz, B., & Katz, R. A. (2003). *Mastering audio: The art and the science*. Oxford, England: Butterworth-Heinemann.
- Keidser, G., Dillon, H. R., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research*, 1(1), e24 Retrieved from <http://audiologyresearch.org/index.php/audio/article/view/audioures.2011.e24/html>.
- Kirchberger, M. J., & Russo, F. A. (2015). Development of the adaptive music perception test. *Ear and Hearing*, 36(2), 217–228.
- Madsen, S. M. K., & Moore, B. C. J. (2014). Music and hearing aids. *Trends in Hearing*, 18. DOI: 10.1177/2331216514558271.
- Moore, B. C. J. (2005). The use of a loudness model to derive initial fittings for hearing aids: Validation studies. In T. Anderson, C. B. Larsen, T. Poulsen, A. N. Rasmussen, & J. B. Simonsen (Eds.), *Hearing aid fitting* (pp. 11–33). Copenhagen, Denmark: Holmens Trykkeri.
- Moore, B. C. J. (2014). Development and current status of the “Cambridge” loudness models. *Trends in Hearing*, 18. DOI: 10.1177/2331216514550620.
- Moore, B. C. J., Füllgrabe, C., & Stone, M. A. (2011). Determination of preferred parameters for multichannel compression using individually fitted simulated hearing aids and paired comparisons. *Ear and Hearing*, 32(5), 556–568.
- Moore, B. C. J., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, 188(1), 70–88.
- Moore, B. C. J., Glasberg, B. R., & Stone, M. A. (1999). Use of a loudness model for hearing aid fitting: III. A general method for deriving initial fittings for hearing aids with

- multi-channel compression. *British Journal of Audiology*, 33(4), 241–258.
- Moore, B. C. J., Glasberg, B. R., & Stone, M. A. (2010). Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF. *International Journal of Audiology*, 49(3), 216–227.
- Moore, B. C. J., & Sek, A. (2013). Comparison of the CAM2 and NAL-NL2 hearing aid fitting methods. *Ear and Hearing*, 34(1), 83–95.
- Neumayr, T., & Joyce, S. (2015). Introducing Apple music – All the ways you love music. All in one place. *Apple Press Info*. Retrieved from <https://www.apple.com/pr/library/2015/06/08Introducing-Apple-Music-All-The-Ways-You-Love-Music-All-in-One-Place-.html>.
- Ortner, R. M. (2012). *Je lauter desto bumm! – The evolution of loud* (Master's thesis). Danube University Krems, Austria.
- Pachet, F., & Cazaly, D. (2000, April). *A taxonomy of musical genres*. Paper presented at the Content-Based Multimedia Information Access Conference (RIAO) (pp. 1238–1245), Paris, France.
- Palmason, H. (2011). *Large-scale music classification using an approximate k-NN classifier* (Master's thesis). Reykjavik University, Iceland.
- Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Lurnagaray, D., ... Pumford, J. (2005). The desired sensation level multistage input/output algorithm. *Trends in Amplification*, 9(4), 159–197.
- Seewald, R. C., Ross, M., & Spiro, M. K. (1985). Selecting amplification characteristics for young hearing-impaired children. *Ear and Hearing*, 6(1), 48–53.
- Sivian, L. J., Dunn, H. K., & White, S. D. (1931). Absolute amplitudes and spectra of certain musical instruments and orchestras. *The Journal of the Acoustical Society of America*, 2(3), 330–371.
- Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 370.
- van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America*, 105(5), 2903–2913.
- Vickers, E. (2010). *The loudness war: Background, speculation, and recommendations*. Paper presented at Proceedings of the 129th Convention of the Audio Engineering Society, San Francisco, CA. Retrieved from <http://www.aes.org/e-lib/browse.cfm?elib=15598>.
- Vickers, E. (2011). The loudness war: Do louder, hypercompressed recordings sell better? *Journal of the Audio Engineering Society*, 59(5), 346–351.
- Villchur, E. (1974). Simulation of the effect of recruitment on loudness relationships in speech. *The Journal of the Acoustical Society of America*, 56(5), 1601–1611.
- Wilson, G. L., Attenborough, K., Galbraith, R. N., Rintelmann, W. F., Bienvenue, G. R., Shampian, J. I. (1977). Am I too loud? (A symposium on rock music and noise-induced hearing loss). *Journal of the Audio Engineering Society*, 25(3), 126–150.
- Winckel, F. W. (1962). Optimum acoustic criteria of concert halls for the performance of classical music. *The Journal of the Acoustical Society of America*, 34(1), 81–86.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151.