Appendix

A1.1. Study 1: Participants and Sampling

Based on population statistics from Eurostat (2016), a German commercial online provider and its subsidiaries in UK and Spain realised 'crossed' target quotas for *generation* (gen Y: 18-34, gen X: 35-51, gen B: 52-68), *education level* (ISCED 0-2, ISCED 3-4, ISCED 5-8) and *gender* (female/male), arriving at an overall gross sample of 10,180 subjects with approximate equal participants from each of the three countries. Due to internal logics of the project producing the data, survey experiments were conducted in two waves, December 2016 (n = 3,485) and May 2017 (n = 6,659).

A1.2. Study 1: Stimulus Material

Employed musical excerpts stem from a pool of 549 popular music tracks that had been selected from the digital music library of the collaborating audio branding agency *HearDis (see Acknowledgements)*. Frequencies of selected tracks across genres were deliberately unbalanced, in order to represent typical frequencies of genres in a music library of the audio branding industry (see Figure 4). The 30s excerpts were cut from the original recordings (if possible) to always contain the end of first verse and the beginning of first chorus. Subsequently, resulting audio files were technically and perceptually normalised in terms of loudness. Randomised excerpt selection and playback for individual participants was programmed so that each track had the same selection probability within each combinatory cell of sociodemographic variables and country of residence.

A1.3. Study 1: Data Pre-processing

The obtained gross sample of both online survey experiment waves underwent extensive data cleansing. Subjects were filtered out as irregular respondents if less than half of the items had been answered or if there was insufficient variance across all items (< 0.6 for wave 2, < 0.4 for wave 1), resulting in a dataset of n = 9,197 subjects with a structure approximately resembling population statistics in terms of socio-demographics (see Figure 5). In the overall net sample, gender is nearly equally distributed with 50.2% women. Also, the overall shares of generations (gen Y: 32.8%, gen X: 33.2%, gen B: 34.0%) and countries (UK: 34%, Germany: 35%, Spain: 30%) are very similar. The median age is 43 with an interquartile range of 23.

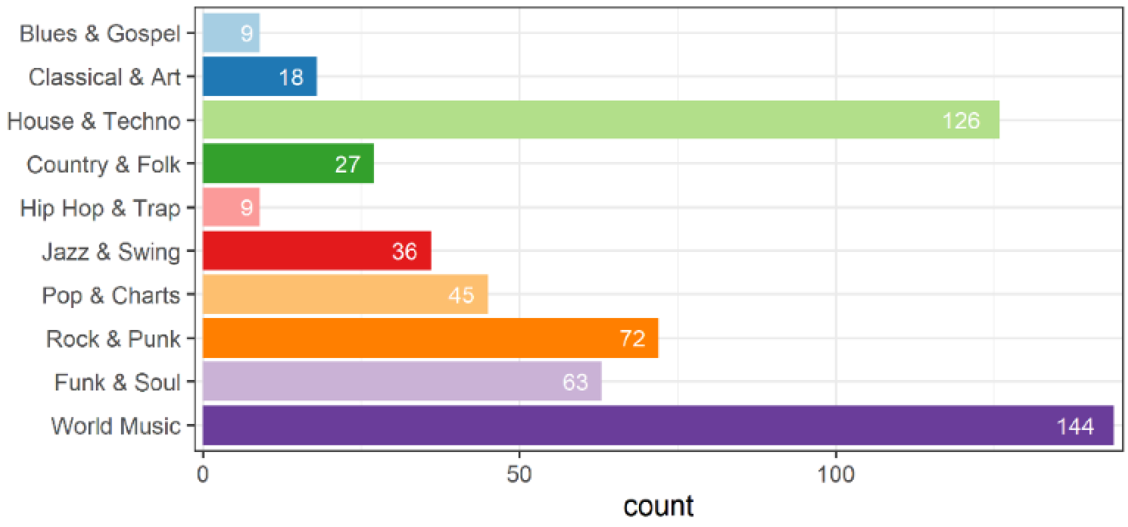

Figure 4. Composition of musical excerpt stimulus pool by genre (track excerpt counts), overall: 549 tracks

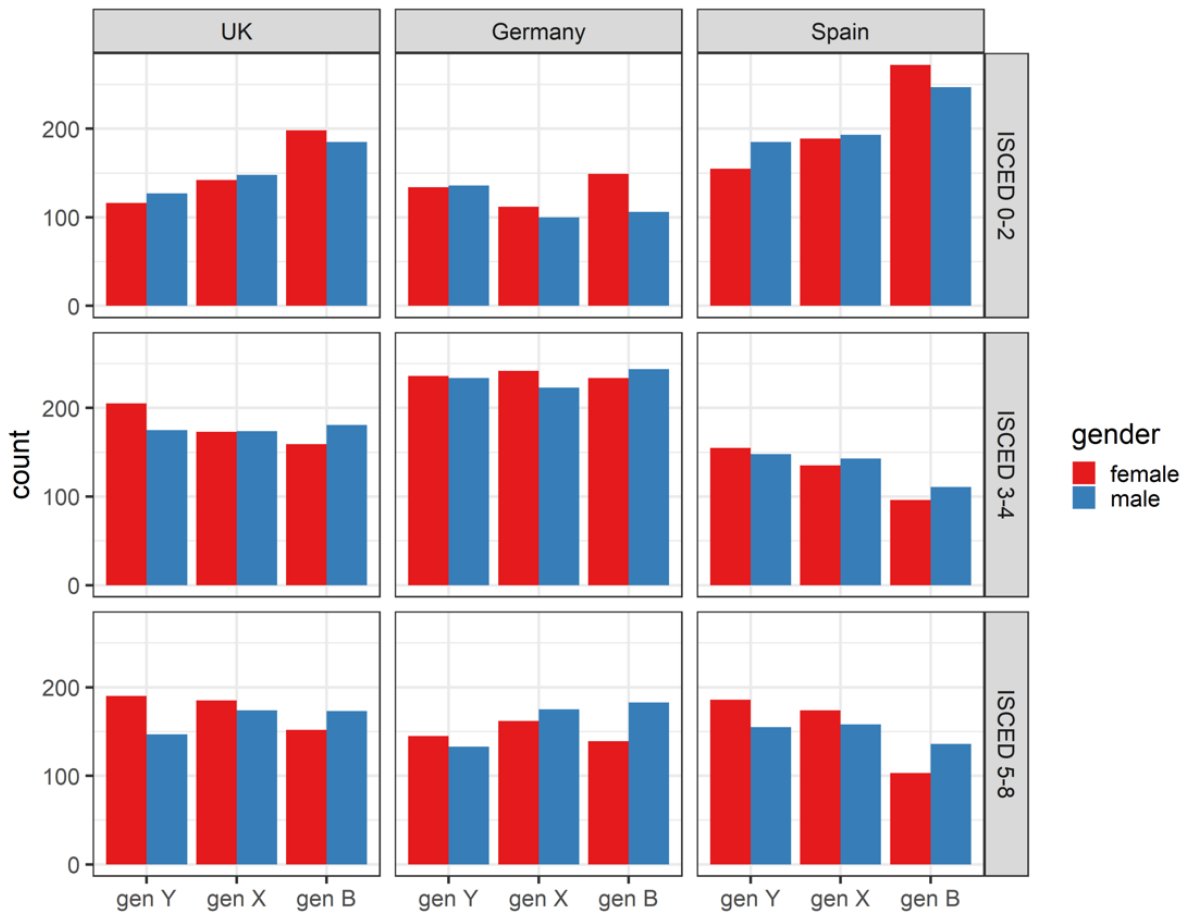

Figure 5. Net sample structure in terms of socio-demographics (case counts)

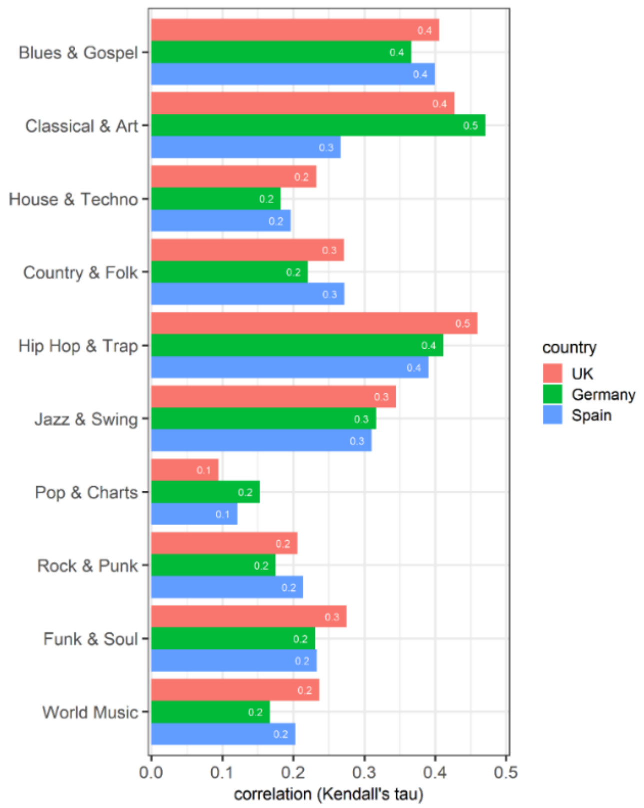

Figure 6. Correlations of stated genre affinities with actual liking of musical genres

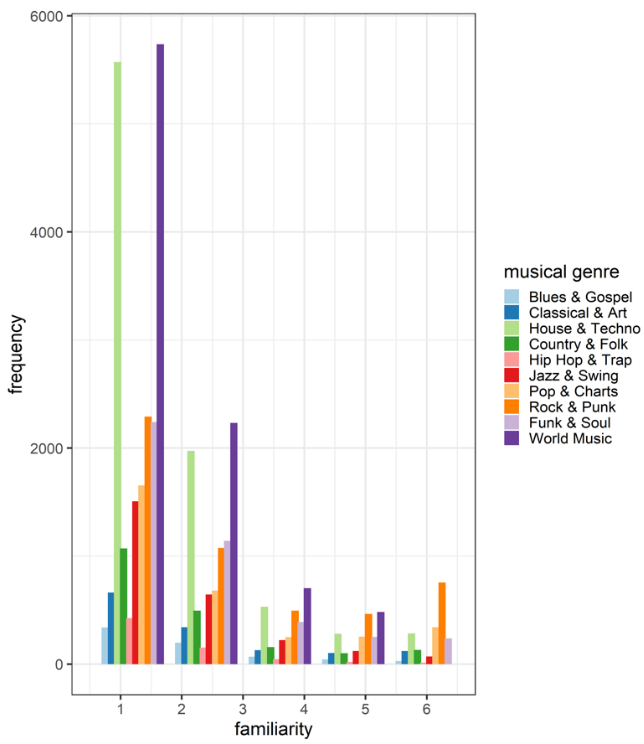

Figure 7. Personal familiarity with musical excerpts by genre

A2.1. Study 2: Development of Audio Descriptors 1 – Machine learning (ML) of expert tags

To make higher-order musical knowledge available for algorithmic classification, we applied supervised learning of music branding experts' annotations (*musical style*, *instrumentation*, *vocals* (yes/no), *vocals gender*, *production timbre*) based on audio signal properties of the musical material contained in the collaborating audio branding agency's digital music archive, training a ML model for each tag family, respectively. In detail, 17,163 representative full music tracks (not used in Study 1) were chosen from the library, in a way to represent at least 100 tracks of each style and all possible combinations of tags. We employed the *IRCAM classification* meta-framework for realising ML, which allows training a classifier given a set of exemplary music tracks belonging to a specific tag (Peeters et al., 2015). The five ML classifiers resulting from this procedure were finally applied to the 549 music tracks used in study 1. Each track was then characterised by its membership probabilities concerning each tag of each tag family. This led to a set of 80 machine learning-based descriptors (sum of all tag classes, Table 4), to be used later as input for the computational prediction models.

A2.2. Study 2: Development of Audio Descriptors – Extraction of further audio descriptors with MIR toolboxes

Additional signal analysis of the 549 music tracks used in Study 1 was conducted based on publicly available MIR software toolboxes to obtain further meaningful audio and music descriptors. The resulting set of content descriptors relates either to musical characteristics (such as *tempo* or *key*) or global sound characteristics (such as the frequency bandwidth of the audio signal). To cover a wide range of audio and music description, we employed *IRCAM beat* (Peeters & Papadopoulos, 2011) in order to represent rhythm and tempo (9 descriptors), *IRCAM keymode* (Peeters, 2006) to represent mode and key (12 descriptors), as well as *IRCAM chord* in an adapted version (Steffens et al., 2017) representing typical chord successions and functional harmonics in popular music (13 descriptors). Moreover, we utilised *IRCAM descriptor* (Peeters, 2004 for representing the overall sound of the music track, for instance in terms of *sinusoidal components*, *roughness*, *mean energy,* or the *Mel Frequency Cepstral Coefficients (MFCCs)* which form a highly-integrated description of the sound energy contained in specific frequency bands (42 descriptors). Finally, we applied the *IBM tone analyser* (IBM Corporation, 2017) to analyse the perceived expression (emotions and character) of the song lyrics. Due to practical reasons, we always analysed the full audio tracks and relied on the toolboxes' default options only. In result, we gathered 76 additional audio and music descriptors.

Additional References

Eurostat. (2016). *Eurostat Database. Your key to European* Statistics. Eurostat Database. http://ec.europa.eu/eurostat/data/database

IBM Corporation. (2017). *Tone Analyzer*. https://www.ibm.com/watson/services/tone-analyzer

Peeters, G. (2004). *A large set of audio features for sound description (similarity and classification) in the CUIDADO project [online]*. IRCAM. https://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.html

Peeters, G. (2006). Chroma-based estimation of musical key from audio-signal analysis. *Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006)*.

Peeters, G., Cornu, F., Doukhan, D., Marchetto, E., Mignot, R., Perros, K., & Regnier, L. (2015, July). When audio features reach machine learning. *International Conference on Machine Learning—Workshop on" Machine Learning for Music Discovery."* https://hal.archives-ouvertes.fr/hal-01254057

Peeters, G., & Papadopoulos, H. (2011). Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(6), 1754–1769. https://doi.org/10.1109/TASL.2010.2098869

Steffens, J., Lepa, S., Herzog, M., Schönrock, A., Peeters, G., & Egermann, H. (2017). *High-level chord features extracted from audio can predict perceived musical expression*. Paper presented at 18th International Society for Music Information Retrieval Conference (ISMIR 2017), Suzhou, China.

Table 8. Results of measurement invariance tests across GMBI_15 questionnaire language versions (English, German, Spanish)

Invariance model	X^2	df	RMSEA	CFI	SRMR	Δ CFI
configural	7909.276	240	0.050	0.956	0.036	-
Metric	8610.042	260	0.050	0.952	0.041	0.00
Scalar	9823.706	280	0.051	0.945	0.042	0.01
residual	12793.639	310	0.056	0.928	0.049	0.02

Table 9. Results of variance component estimation for musical expression dimensions

Musical expression dimension	Variance proportion explained by socio-demographics	Variance proportion explained by track identity
Arousal	0.64 %	26.33 %
Valence	0.61 %	12.24 %
Authenticity	0.17 %	11.81 %
Timeliness	0.94 %	24.19 %
Eroticity	2 %	11.97 %

Table 10. Results of machine learning classification of expert tags

Classifier	Class labels (expert tags)	No of classes	Accuracy (in %)	Recall (in %)
Musical style	*Style tags are provided below this table**	61	98	45
Instrumentation	Acapella, Acoustic-Guitar, Brass, Choir, Electric-Guitar, Live Drums, Orchestral, Percussions, Piano, Speech, Strings, Synthetic Drums, Whistle	13	81	42
Vocals	yes, no	2	92	92
Vocals_gender	male, female, mixed	3	76	63
Production timbre	hard, soft, warm, cold, bright, dark	6	82	46

** Class labels of style classifier: Afro, Ambient, AOR, Asian, Balearic, Balkan, Blues, Boogaloo, Boogie, Bossa-Nova, Broken-Beats, Calypso, Chanson, Classical-Jazz, Classic-Rock, Contemporary-Classical, Contemporary-Folk, Country, Dancehall, Deep-House, Disco, Downbeat, Dream-Pop, Drum & Bass, Dubstep, Easy-Listening, EDM, Electro, Electro-Pop, Electro-Rock, Flamenco, Folkloric, Funk, Fusion-Jazz, Hip-Hop, Historical-Classical, House, Indie-Dance, Indie-Pop, Indie-Rock, Krautrock, Latin, Mainstream, Northern-Soul, Nu-Jazz, Oriental, Progressive-Rock, Punk, R&B, Rare-Groove, Reggae, Reggaeton, Rock & Roll, Samba, Schlager, Smooth-Jazz, Soul, Tango, Tech-House, Traditional-Folk, UK-Funky*

Table 11. Relative explanatory potential of predictor blocks in hierarchical stepwise regression models for dependent variables

Predictor block	Valence	Arousal	Authenticity	Timeliness	Eroticity	$\overline{R^2_{adj}}$
IRCAM beat	.077	.141	.214	.297	.028	.151
IRCAM keymode	.032	.028	.020	.000	.015	.019
IRCAM chord (adapted)	.000	.051	.034	.011	.036	.026
ML: instrumentation & vocals	.132	.219	.141	.216	.205	.183
ML: musical style	.177	.239	.169	.213	.155	.191
IRCAM descriptor I: sound	.019	.030	.008	.014	.006	.015
ML: production timbre	.015	.009	.006	.004	.022	.011
IRCAM descriptor II: MFCCs	.000	.026	.019	.012	.017	.015
Lyrics: IBM tone analyser	.011	.002	.000	.003	.008	.005
Σ R^2_{adj}	**.463**	**.745**	**.611**	**.770**	**.493**	**.616**