



Lead-vocal level in recordings of popular music 1946–2020

Karsten Gerdes^{a)} and Kai Siedenburg

Department of Medical Physics and Acoustics, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany karsten.gerdes@uol.de, kai.siedenburg@uol.de

Abstract: Lead vocals constitute the central element of popular music. Here, the lead-vocal-to-accompaniment level ratio (LAR) was estimated from representative recordings of popular music. Measuring the LAR from 1946 to 2020, two distinct phases were observed: the average LAR decreased from around 5 dB to 1 dB until around 1975 but remained static from thereon. Comparing the LAR across musical genres, positive values were observed for Country, Rap, and Pop, values around zero for Rock, and negative values for Metal. Solo artists featured consistently higher LAR values compared to bands. These results establish a baseline for a central aspect of music mixing. © 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/license/by/4.0/).

[Editor: Vasileios Chatziioannou]

https://doi.org/10.1121/10.0017773

Received: 11 January 2023 Accepted: 16 March 2023 Published Online: 25 April 2023

1. Introduction

Vocals arguably constitute the central element of popular music. Singers tend to be perceived as the most important and identifiable members of music groups and great care is taken in music production to sculpt and convey specific musical identities via the singing voice. According to Wicke (2011), it is a "basic rule of music production" that lead-vocals of a piece are mixed in such a way as to be perceived as part of the musical "foreground." In multi-track music mixing, the relative dominance of a track or instrument in the overall musical mix can be adjusted via its amplitude level and the careful adjustment of level relationships is among the most important aspects of music mixing (Owsinski, 2013). Here, we consider the *lead-to-accompaniment ratio* (LAR), denoting the level ratio of the lead vocals and the accompaniment, and use the LAR to explore the dominance of lead vocals in music recordings from 1946 to 2020.

The discourse around level in popular music has generally been much dominated by questions around the socalled "loudness war," denoting the increase in the overall level of music recordings from the 1980s onward (Serrà *et al.*, 2012; Vickers, 2011). Hove *et al.* (2019) provided an analysis of bass level in the top two songs per year from the Billboard Hot 100 year-end list across several decades and observed an increase in spectral flux in the lowest frequency bands (up to 100 Hz). Studies on level relationships among sound sources in a mix have shown that lead-vocals are mixed at higher levels compared to all other instruments (Ward *et al.*, 2017). The latter study used a set of songs that was compiled for the purpose of evaluating musical source separation (Vincent *et al.*, 2012) and is not necessarily representative of Western popular music and its development. To the best of our knowledge, there yet does not exist any evaluation of vocal level for a representative set of popular songs spanning several decades.

The level of vocals in musical mixes may also be viewed from the perspective of the intelligibility of lyrics. Naturally, level is the critical factor determining the audibility of elements in musical mixes. Condit-Schultz and Huron (2015) observed an increasing intelligibility of the lyrics with increasing subjectively perceived loudness of the vocals, paired with strong differences across musical genres. It was further suggested that listeners with cochlear implants prefer a significantly higher LAR in music mixes (Buyens *et al.*, 2014). In speech perception, the drastic way in which speech intelligibility of normal-hearing listeners degrades for negative signal-to-noise ratios (SNR) has been carefully documented. For instance, using closed-set sentence tests, performance was near 100% at $-4 \, \text{dB}$ SNR, but degraded to 0% at $-14 \, \text{dB}$ SNR; 50% thresholds were around $-7 \, \text{dB}$ SNR (Wagener *et al.*, 1999). However, Bürgel *et al.* (2021) found that vocals attracted auditory attention in a mix likely owing to specific acoustic cues of human voices—even at very low LARs (up to $-15 \, \text{dB}$). Therefore, there may in fact be some leeway in music production in dealing with vocal level that could be used to convey specific artistic goals. For instance, it seems plausible that the LAR could reflect the degree to which a mix intends to direct the musical focus on the singing voice alone or whether the focus is distributed across various musical instruments or elements. The specific range of LAR values that are common across representative samples of popular music are yet unknown.

^{a)}Author to whom correspondence should be addressed.

From a general perspective, two hypotheses on the role of vocal level in music production may be differentiated: First, vocal level may have been fixed throughout history of popular music to guarantee the intelligibility of the lyrics and audibility of the main melody on the one hand and the audibility of the accompaniment on the other. Second, one could hypothesize that vocal level is used more flexibly as a result of the evolution of music technology and as a means to convey specific artistic choices and intentions during music production. The goal of the present study was to empirically test these hypotheses on a large dataset of more than 700 songs. We thus quantified the LAR in a representative sample of well-known songs of recordings of popular music spanning several decades. Because only stereo mixtures were available from these songs, we used music-source separation software to separate the vocals from the accompaniment. The precision of the source-separation-based estimate of the LAR was validated in experiment 1. In experiment 2, we characterized the evolution of the LAR for the top four songs from the Billboard Hot 100 list since 1946. In experiment 3, we characterized effects of musical genre and whether the LAR varied as a function of whether songs were performed by solo artists or groups. We argue that the LAR is a valuable acoustical descriptor that directly mirrors aesthetic choices during music production.

2. Methods

2.1 Music source separation

Because multi-track formats are unavailable for most commercial music, we used the *Rebalance* module of the software IZOTOPE RX 8 to separate the vocals from the mixture. This software is built to separate music recordings into four tracks: vocals, bass, percussion/drums, and all remaining sounds. Specifically, the gain of these tracks can be adjusted. Here, the lead-track was created by setting the vocals to 0 dB (meaning no change in gain) and setting the remaining groups to –Inf. To obtain the accompaniment-track, the gain of the vocals was set to –Inf and all of the other tracks to 0 dB. A separation of 100% and the quality setting "best" was chosen in the software.

2.2 LAR calculation

The LAR was computed as the level ratio between the extracted lead-vocal signal x_L and the accompaniment signal x_A (i.e., the remaining audio in the mixture),

$$LAR = 20\log_{10}(\overline{x_L}/\overline{x_A}). \tag{1}$$

Here, $\overline{x_{L/A}}$ denotes the root-mean-squared (rms) amplitude across one-second segments after A-filtering of $x_{L/A}$. A-filtering was used to emphasize frequency regions for which human hearing is most sensitive and reduce the influence of bass-frequencies shown to increase from 1955 to 2015 (Hove *et al.*, 2019). For integrating the LAR across the full song, segments were discarded if the levels of the lead vocals or the accompaniment fell below -20 dB relative to their respective maximum value in a given song. The threshold of -20 dB was chosen empirically as a cutoff to differentiate between the actual signal (usually above -20 dB) and lower-level residual artifacts of the source separation algorithm (below -20 dB). The LAR of song was defined as the mean over the remaining segments and the left and right channels of the audio signals.

2.3 Song databases

Due to its reliance on source separation algorithms, the definition of the LAR was an estimate based on a potentially imperfect music-source separation of the lead vocals. Therefore, in experiment 1, the LAR estimate was validated and corrected, for which separated tracks were required. To test the estimation accuracy, we used the same database of popular music recordings as in Bürgel *et al.* (2021), comprising 65 professional multi-track sound-alikes (i.e., imitations) of popular music recordings from different genres and decades (1950–2020). Because the source-separation software was unable to differentiate between lead- and backing-vocals, 41 songs with and 14 songs without backing vocals were considered separately. To obtain similar sample sizes, the backing vocals of the songs were removed, yielding a total of 55 songs without backing vocals.

In experiment 2, the Year-End-List of the *Billboard Hot 100* was used (Billboard, 2023). This list integrates weekly counts of physical and digital sales in addition to airplay. Due to their diversity, the *Billboard Charts* are considered independent and representative (Brockhaus, 2017). To represent the charts of a year, the present study used the four highest ranked songs containing lead vocals of the Year-End List from 1946 to 2020. For a few songs, remixes or remasters from other years had to be used because the original mix was not available. In total, this results in a database of 300 songs for the years 1946–2020.

In experiment 3, we sought to analyze genre dependency of the LAR for the genres Country, Rap, Pop, Rock, and Metal. Here, songs were assigned to a genre based on nomination for the *Grammy Awards*, differentiating between different musical genres. Three songs per genre and per year between 1990 and 2020 were selected for the database. Because the categories of the Grammy Awards are not fixed across time, songs from separate categories were grouped together: For Country, a total of 93 songs were selected from the categories *Best Female Country Vocal Performance* (1990–2011), *Best Male Country Vocal Perf.* (1990–2011), *Best Country Solo Perf.* (2012–2020), *Best Country Perf. by a*

asa.scitation.org/journal/jel

Duo or Group with vocals (1990–2011), and Best Country Perf. by a Duo or Group (2011–2020). For Rap, 83 songs were selected from the categories Best Rap Perf. (1990–2001, 2003–2020), Best Rap Perf. by a Duo or Group (1991–2011), Best Female Rap Solo Perf. (2001–2002) or Best Male Rap Solo Perf. (2001–2002). For Pop, a total of 93 song were selected from the categories Best Female Pop Vocal Perf. (1990–2011), Best Male Pop Vocal Perf. (1990–2011), Best Pop Solo Perf. (2012–2020), and Best Pop Duo/Group Perf. (1990–2020). For Rock, a total of 83 songs were selected from the categories Best Rock Vocal Perf., Male (1990–2003), Best Rock Vocal Perf, Female (1990–2003), Best Rock Vocal Perf. By A Duo or Group with Vocal (1990–2011), Best Solo Rock Vocal Perf. (2004–2011), and Best Rock Perf. (2012–2020). For Metal, 62 songs were selected from the categories Best Metal Perf. (1990z–2012, 2014–2020) and Best Hard Rock/Metal Perf. (2013). This resulted in 414 songs in total.¹

In experiments 2 and 3, stereo recordings were obtained from YouTube (2023) and the presence of backing vocals was checked manually by K.G.

3. Results

3.1 Experiment 1: Validation

With the availability of clear vocal tracks in multi-track recordings, the accuracy of the source-separation-based estimation of the LAR could be validated. This was conducted separately for songs with and without backing vocals. Figure 1 shows the comparison of the "true" LAR and the estimated LAR as outlined above in Sec. 2.2. The fit for the songs without backing vocals was $R^2 = 0.91$, and the fit for songs containing backing vocals was $R^2 = 0.64$. For tracks with backing vocals, it is visible that there was a systematic offset because the music-source separation software separated the backing vocals together with the lead vocals. This offset resulted from the backing vocals increasing the level of the separated lead vocals, leading to an increase in the estimated LAR compared to the "true" LAR.

To correct the estimate, we used the correction terms derived from a linear regression fit, with LAR* denoting the corrected estimate. The corresponding correction for songs without backing vocals was

$$LAR^* = 0.83 \cdot LAR + 0.33 \tag{2}$$

and for songs with backing vocals the correction was, respectively,

$$LAR^* = 0.75 \cdot LAR - 0.87. \tag{3}$$

Compared to the true LAR, the corrected estimate LAR* exhibited mean absolute errors of 1.6 and 0.71 dB for songs with and without backing vocals, respectively. Based on these values and given the 15 dB range of original LAR values in the dataset, we conclude that the described estimate LAR* yields a sufficiently precise measure of relative vocal levels for songs with and without backing vocals, even without the availability of multi-track recordings. In the following, we will use the corrected LAR* from hereon but skip the asterisk notation for the sake of simplicity.

3.2 LAR from 1946 to 2020

Figure 2 shows the estimated LAR values for four songs per year in the period 1946–2020. Notably, only 41 out of the 300 songs did not contain backing vocals. In addition to the individual values, a broken stick regression with two continuously connected segments was used as a model of the general trend across songs and years. The estimated breakpoint between regression segments one and two was 1975 (95% CI: [19 661 983]). The estimated slope was $\beta_1 = -0.16$ (CI: [-0.23,



Fig. 1. Relation between the true LAR and the LAR based on source separation using IZOTOPE RX8. (A) Songs without backing vocals. (B) Songs with backing vocals.

asa.scitation.org/journal/jel



Fig. 2. Evolution of the LAR from 1946 to 2020 for the top four songs on the Billboard Charts. The line corresponds to a broken-stick regression (i.e., continuous and piece-wise linear fit) for all songs. Songs with and without backing vocals (BV) as indicated in the legend.

-0.09]) in segment one and $\beta_2 = 0.01$ (CI: [-0.02, 0.04]) in segment two. That is, among the top four songs of the Billboard Top 100, there was a significant downward trend of the LAR up until the mid 1970s from which on the LAR appeared to stagnate on average between values of 1 and 2 dB.

Several developments in music technology may underpin this observation. The increasing electrical amplification of guitar and bass, that was started in the 1930s, meant that vocal levels had to be increased significantly in order to remain intelligible, simply by positioning the vocals closer to the microphone compared to the band. The development of multitrack recording technology that developed in the 1950s provided much more control of vocal level. Importantly, stereophonic recording technology provided spatial release from masking for vocals, potentially allowing for higher levels of accompaniment level with similar audibility of lead vocals. Another potential interpretation of this reversal point at the beginning of the 1970s concerns the stylistic evolution within popular music. Over time, different genres and style of music have evolved, all of which are "designed" differently. In experiment 3, we thus set out to test the extent to which genre-specific factors influenced vocal level.

3.3 LAR in five musical genres 1990-2020

Here, we tested whether vocals would be mixed differently in different musical genres. Figure 3 shows extracted LAR values for songs from solo and band artists from the five examined genre categories together confidence intervals for the mean value. There were two marked outliers in the Country and Pop categories. In Country, the song *Lovesick Blues* (2001) by Ryan Adams with a LAR of close to 20 dB. In Pop, the song *Don't Know Why* (2002) by Norah Jones reached a LAR of around 15 dB. Otherwise, the distribution was relatively homogeneous without any marked multi-modal tendencies. The genre Country yielded the highest average LAR value, M = 3.9 dB with 95% CI [3.4, 4.5], followed by Rap, M = 3.2 dB [2.7, 3.6], and Pop, M = 2.7 dB [2.1, 3.4]. An average LAR close to zero dB was observed for Rock, M = 0.2 dB [-0.3, 0.7]. A clearly negative average LAR was observed for Metal, M = -3.1 dB [-3.6, -2.6]. This confirms genrespecific LARs in popular music with highest vocal levels in Country and lowest vocal levels in Metal.

As depicted in Fig. 3, we also considered whether songs stemmed from solo artists (singers) or bands. For Country, the LAR was about 1.5 dB higher for solo artists compared to bands, and this difference was also present for Pop (3.5 dB) and Rock (2.5 dB). We abstain from drawing conclusions for Rap and Metal due to their limited number of bands and solo artists, respectively. In addition to the genre specificity of the LAR, these observations indicate that vocals of solo artists are generally mixed at higher levels compared to the lead vocals in bands.

In the Rock (Band) and Metal categories, we observed an overwhelming majority of negative LAR values. Guitar riffs are a distinctive feature of Rock and Metal, with guitars taking a position comparable to the lead vocals (Elflein, 2010; Fast, 2014; Walser, 2001). Quantitatively, this was confirmed by computing the guitar-to-accompaniment-Ratio (GAR) for Metal songs, which yielded a positive value of 2.4 dB. Note that this was only possible in the genre category Metal that did not contain other instruments separated into the "Other" category of the IZOTOPE REBALANCE software. That is, in the Metal genre, guitars are mixed to possess a similarly prominent position as vocals in other genres of popular music.

asa.scitation.org/journal/jel



Fig. 3. Estimated LAR of each song for five different genres 1990–2020. Horizontal lines next to individual data points show the mean, vertical bars the corresponding 95% confidence interval. Solo artists are represented by violet circles and bands by green squares.

4. Discussion

In this study, we measured the lead-vocal-to-accompaniment ratio (LAR) for vocals in representative recordings of popular music. Because recordings from these popular songs were only available in stereo format, we used a commercially available music-source separation software (IZOTOPE RX8). We quantified the extent to which the presence of backing vocals affected the precision of the LAR estimation and optimized our approach by using the data from experiment 1.

In experiment 2, we considered the top four songs from the Billboard Hot 100 list from 1946 to 2020. Using a broken stick regression model, we identified a period up until around 1975, where LAR values generally decreased on average from around 5–1 dB. From the mid 1970s onwards, average LAR values remained relatively constant in the range 1–2 dB. This finding complements large-scale analyses of harmonic and timbral properties of popular music from the Billboard lists (Mauch *et al.*, 2015). The authors had identified structural turning points in the history of music. Here, we consider an analysis of mixing properties and show that there are two phases in the adjustment of vocal level with a turning point in the mid 1970s. Several factors in the history of music and music technology may be associated with our observation. Among these are the emergence of multi-track recording, the stereo format, but also trends in the evolution of popular music that is reflected in the Billboard list. Accordingly, it has be argued that with multi-track recording technology and improved sound engineering, the lead vocals were inserted less exaggeratedly into the mix and still remained intelligible (Schiffner, 1991).

In experiment 3, we considered popular songs representative to five different musical genres as categorized by Grammy awards. We found that Country, Rap, and Pop generally had positive LAR values; Rock had LAR values around zero dB; and Metal had negative LAR values. These results suggest genre-specific mixing of vocals. We acknowledge the limitations of the used notions of genre used to categorize songs in separate groups. Here, we relied on a categorization through an expert jury for the Grammy Awards that supposedly honor outstanding achievements for specific genres. Even though our data demonstrate, on average, distinct acoustical properties of mixes in these genres, there may be debate around whether genre boundaries can be drawn clearly or are subject to fluent change (Mauch *et al.*, 2015).

The clearly negative LAR for Metal and band-based Rock together with the high levels of guitar sounds in these recordings corresponds to the notion that guitars are of equal (if not greater) importance in these genres compared to the vocals. This may further explain, among other things such as particular vocal technique, the incomprehensibility of song lyrics in metal described by Weindl (2005). Moreover, these observations appear to contradict a basic rule of music production as outlined by Wicke (2011), namely, that lead-vocals are always positioned in the very foreground of a mix. In the metal and band-based rock genres, it is common for groups to operate as collectives rather than featuring individual standout members. This idea was corroborated by contrasting recordings from solo artists with bands. Our data indicated that the former had higher LAR values compared to the latter. This result may reflect the circumstance that there is a direct correspondence between the way in which popular music is mixed and how musicians are promoted. Often, solo artists are accompanied by interchangeable instrumentalists, who may not be critical for the typical sound of an artist and thus also play a subordinate role in the mix. Bands, on the other hand, view themselves as collectives in which all musicians play a more equal role and hence not a single element, such as the singer, is mixed to stand out in a dominant way. All this aligns with the notion that the final mix of a song serves as a representation of the hierarchy of importance within a music ensemble.

17 February 2024 05:48:20



ARTICLE

5. Conclusion

In summary, the level of lead vocals in representative recordings of popular music have decreased until the mid 1970s and remained relatively constant from thereafter. Further, vocal level possesses characteristic levels in different musical genres and tends to be higher for solo artists compared to bands. Vocal level may be interpreted as an index of prominence of vocals in the mix, which was shown to vary over time and genre, rather than being a static factor of music production.

Acknowledgments

The authors wish to thank Mario Dunkel and Michael Hove for valuable comments. This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project ID 352015383–SFB 1330 A6. This study was also supported by a Freigeist Fellowship from the Volkswagen Foundation to K.S.

References and links

¹See supplementary material at https://www.scitation.org/doi/suppl/10.1121/10.0017773 for the complete list of songs

Billboard (2023). https://www.billboard.com/charts/hot-100/ (Last viewed November 18, 2021).

- Brockhaus, I. (2017). Kultsounds: Die Prägendsten Klänge der Popmusik 1960–2014. Musik und Klangkultur (Kultsounds: The Most Iconic Sounds of Pop Music 1960–2014. Music and Sound Culture) (Transcript-Verlag, Bielefeld), Bd. 23.
- Bürgel, M., Picinali, L., and Siedenburg, K. (2021). "Listening in the mix: Lead vocals robustly attract auditory attention in popular music," Front. Psychol. 12, 769663.
- Buyens, W., van Dijk, B., Moonen, M., and Wouters, J. (2014). "Music mixing preferences of cochlear implant recipients: A pilot study," Int. J. Audiology 53(5), 294–301.
- Condit-Schultz, N., and Huron, D. (2015). "Catching the lyrics," Music Perception 32(5), 470-483.
- Elflein, D. (2010). Schwermetallanalysen (Heavy Metal Analysis) (Transcript Verlag, Bielefeld), Bd. 6.
- Fast, S. (2014). "Rock," https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561 592630-e-1002257208 (Last viewed September 19, 2022).
- Hove, M. J., Vuust, P., and Stupacher, J. (2019). "Increased levels of bass in popular music recordings 1955–2016 and their relation to loudness," J. Acoust. Soc. Am. 145(4), 2247–2253.
- Mauch, M., MacCallum, R. M., Levy, M., and Leroi, A. M. (2015). "The evolution of popular music: USA 1960–2010," R. Soc. Open Sci. 2(5), 150081.

Owsinski, B. (2013). The Mixing Engineer's Handbook (Nelson Education, Toronto).

- Schiffner, W. (1991). Einflüsse Der Technik Auf Die Entwicklung Von Rock/Pop-Musik (Influences of technology on the development of rock/pop music).
- Serrà, J., Corral, A., Boguñá, M., Haro, M., and Arcos, J. L. (2012). "Measuring the evolution of contemporary western popular music," Sci. Rep. 2, 521.

Vickers, E. (2011). "The loudness war: Do louder, hypercompressed recordings sell better?," J. Audio Eng. Soc. 59(5), 346-351.

- Vincent, E., Araki, S., Theis, F., Nolte, G., Bofill, P., Sawada, H., Ozerov, A., Gowreesunker, V., Lutter, D., and Duong, N. Q. K. (2012). "The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges," Sign. Process. 92(8), 1928–1936.
- Wagener, K. C., Kühnel, V., and Kollmeier, B. (**1999**). "Entwicklung und evaluation eines satztests für die Deutsche sprache I: Design des Oldenburger satztests," ("Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence tests"), Z. Audiol./Audiol. Acoust. **38**(1), 5–14.

Walser, R. (2001). "Heavy metal," Grove Music Online, https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/ 9781561592630.001.0001/omo-9781561592630-e-0000049140 (Last viewed September 19, 2022).

- Ward, D., Wierstorf, H., Mason, R. D., Plumbley, M. D., and Hummersone, C. (2017). "Estimating the loudness balance of musical mixtures using audio source separation," in *Proceedings of the 3rd Workshop on Intelligent Music Production (WIMP 2017), 3rd Workshop* on Intelligent Music Production, Salford, UK.
- Weindl, D. (2005). Musik & Aggression: Untersucht Anhand Des Musikgeneres Heavy Metal. Mensch Und Gesellschaft (Music & Aggression: Examines the Music Genre Heavy Metal. People & Society) (Peter Lang, New York), Bd. 12.

Wicke, P. (2011). Rock Und Pop: Von Elvis Presley Bis Lady Gaga (Rock and Pop: From Elvis Presley to Lady Gaga) (C.H. Beck, Munich). YouTube (2023). www.youtube.com (Last viewed September 19, 2022).