Wav2Karaoke: A Pipeline to Align Notes and Lyrics from a Song

Hong Hai Ngo[1] Yunmeng Li[1] Steven Coyne[1,2] Kamila Szelag-Levine[1]
Keisuke Sakaguchi[1,2]
[1]Tohoku University [2]RIKEN
ngo.hong.hai.s5@dc.tohoku.ac.jp

Abstract

The synchronization of lyrics with timing and musical notes is desirable for applications such as karaoke systems and lyrics-based music retrieval. Although related problems have been studied in Automatic Speech Recognition (ASR) and Music Information Retrieval (MIR), most existing approaches operate at the word or phone level, while singing voice fundamentally requires syllable-aware alignment. We propose **Wav2Karaoke**, a pipeline that transforms English-language song recordings directly into karaoke-style notation. The system is evaluated using standard lyric alignment and pitch estimation metrics. Our results demonstrate that end-to-end performance is substantially lower than may be expected from isolated benchmarks, highlighting the challenges of integrated deployment.

1 Introduction

Song is a unique and rich form of expression, where semantic content and melodic structure operate in parallel to convey meaning across diverse cultural contexts. To analyze, synthesize, or translate songs computationally, it is essential to align these components. Linking each syllable to its corresponding note enables machines to represent songs in a structured karaoke-style format (Figures 1 and 2), capturing both lyrics and melody as an organized, interpretable sequence.

The singing voice is an ongoing topic in Music Information Retrieval (MIR) and Automatic Speech Recognition (ASR) communities. However, existing approaches mainly focus on word- and phone-level alignment and pitch estimation. This is insufficient for singing, which requires syllable-level alignment, since an individual word may span multiple musical notes. To date, few open-source datasets offer note-level synchronization between lyrics and music, with DALI [1, 2] being one of these examples. To the best of our knowledge, DALI is the only publicly available dataset providing note-level lyric synchronization on syllables.

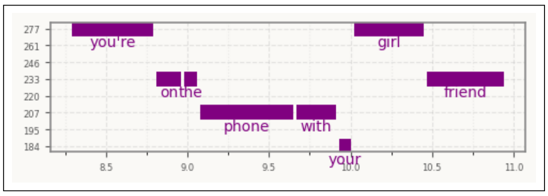

Figure 1 Example of Karaoke-style visualization

start	end	freq	syl
8.28	8.80	277.961	you're
8.80	8.97	234.073	on
8.97	9.07	233.469	the
9.07	9.66	206.048	phone

Figure 2 Example of Karaoke-style notation

Despite the apparent research and industrial demand, accessible tools for note-level synchronization remain scarce. To our knowledge, such features are currently restricted to a limited number of commercial platforms only, most notably AceStudio.[1] To address this gap, we propose **Wav2Karaoke**,[2] a pipeline designed to transform English-language song tracks directly into karaoke-style notation, featuring modular components that enable extension to other languages.

2 Related Work

2.1 Pitch Estimation

Pitch estimation is the process of identifying the fundamental frequency of a musical or vocal signal over time.

1) https://docs.acestudio.ai/ai-tools/vocal-to-midi
2) https://github.com/NyokoKei/Wav2Karaoke

Historically, the field was dominated by autocorrelation-based [3] methods [4], followed by Hidden Markov models [5] and Viterbi decoding [6], which enables a frequency estimator [7] to improve stability in musical contexts. In recent years, the state-of-the-art has shifted toward Neural Network (NN) based architectures. Unlike classical methods, these models learn robust representations directly from raw audio or spectrograms. Most available NN tools [8, 9, 10, 11] operate strictly at the frame level, providing a frequency estimate for every short window (10-20 ms). These methods have been applied not only to monophonic singing [12] but also to instrumental tracks [13] and synthetic audio [14].

2.2 Speech Recognition and Forced Alignment

Forced Alignment (FA) synchronizes an audio text transcription with its recording. By analyzing the speech signal, it identifies the exact start and end times for specific linguistic units. Modern ASR systems, which directly transcribe audio files into the temporal boundaries of words, such as Wav2Vec 2.0 [15] and Whisper [16], appear to compromise [17] on phonetic alignment compared to traditional HMM-based systems, such as Montreal FA [18], limiting their application in timing-sensitive tasks such as lyrics alignment.

2.3 Syllabification

To segment words into syllables, computational approaches rely on phonetic representations, such as International Phonetic Alphabet (IPA),[3] and established linguistic rules, such as the Sonority Sequencing Principle (SSP) [19] or the Maximal Onset Principle (MOP) [20]. Modern toolkits, including NLTK[4] or Pyphen[5], primarily utilize rule-based or dictionary-driven resources, such as the CMU Pronouncing Dictionary[6] or CELEX2 database [21], to segment words into smaller units across covered languages.

3 Proposed Wav2Karaoke Pipeline

Wav2Karaoke can be conceptually divided into four blocks, each corresponding to a distinct stage in the processing pipeline. In this section, we provide a systematic description of each block, outlining the underlying technology and its functional role within the overall system.

3.1 Singing Voice Separation

For audio files containing both instrumental and vocal components, the singing voice is initially separated from the accompaniment for further processing. We apply HT-Demucs [22, 23], a hybrid music source separation (MSS) model, to retrieve the vocal track from a song. Although there are higher-performance alternatives,[7] such as SCNet-large [24] and BS-RoFormer [25], they require more training or computational resources. In comparison, HT-Demucs offers a balance of quality, efficiency, and usability sufficient for this task.

3.2 Melody Extraction

Unlike ordinary speech, singing involves controlling one's voice in a specific manner to shape melody and expression, requiring the capture of musical information along with linguistic content.

Therefore, we employ CREPE [8], one of the state-of-the-art pitch estimators,[8], which demonstrates high performance on both instrumental and vocal audio. The output, consisting of frame-level frequency (measuring the sound signal every 10 ms) with confidence values, will subsequently be aligned with the corresponding lyrics.

3.3 Lyrics Transcription & Syllabification

We use MFA [18] to obtain word- and phoneme-level alignments from the audio recordings. The module matches a lyrics transcription with its vocal track, annotating linguistic units: words and phones, with their start and end times. We employ a pre-trained `english_us_arpa` acoustic model [26] and its corresponding ARPAbet pronunciation dictionary [27] to conduct the alignment. We use a beam search width of 200 and a retry beam width of 800 to achieve better alignment.

Initially, we attempted to transcribe the recordings automatically; however, we noticed that for certain songs, particularly in the rock genre, even Whisper's [16] `Whisper-large-V3` model produces flawed output. To ensure proper alignment, we used the original

3) https://www.internationalphoneticassociation.org/
4) https://www.nltk.org/
5) https://pyphen.org/
6) http://www.speech.cs.cmu.edu/cgi-bin/cmudict
7) https://beta.hyper.ai/en/sota/tasks/music-source-separation/benchmark/music-source-separation-on-musdb18-hq
8) https://github.com/lars76/pitch-benchmark

Table 1 Gender distribution of performers in the evaluation dataset

Category	Male	Female	Mixed	Total
Solo Singers	12	11	–	23
Groups & Bands				
1 Vocalist	31	2	–	33
Multiple Vocalists	9	3	4	16
TOTAL	52	16	4	72

lyrics provided by the dataset (Section 4.1) instead.

Assuming that each syllable corresponds to one note, we perform syllabification based on the word- and phoneme-level alignments obtained with MFA. First, the algorithm maps each word to its associated phonemes based on their time boundaries. Phoneme-level syllables are then constructed by grouping them around vowel nuclei, indicated with numbers to mark stress, following the Maximal Onset Principle (MOS) [20], which assigns as many consonants as possible to the onset of the following syllable while maintaining a natural pronunciation pattern. This principle also ensures the accurate temporal structure of the song. Secondly, we apply the hyphenation module Pyphen[9] to create orthographic syllables, while adjusting mismatches, such as elisions. As a result of this algorithm, we obtain phone- and orthographic-level syllables of the lyrics aligned with timestamps.

3.4 Aligning Notes with Lyrics

Finally, the results from sections 3.2 and 3.3 are used to align the lyrics with musical pitch, creating karaoke-style notation of the song. For each syllable, the pipeline groups all frequency values within its timestamp and, following the one-to-one assumption, assigns the median of high-confidence pitches ($>$ 0.99 percentile) as the target pitch. For interpretability and further musical applications, we also include the corresponding musical note names (e.g., C4, A#3), and MIDI note numbers representation.

4 Experiments

4.1 Dataset

The evaluation dataset is derived from DALI version 1.0 [1, 2], a multilingual collection of 5,358 songs with synchronized full-duration audio, lyrics, and notes. From this pool, 105 songs are marked as ground-truth English-language recordings, of which 99 are stated to have functional YouTube links corresponding to the dataset's alignments. In practice, only 80 audio files are available for use.[10] These songs are performed by a total of 72 performers (table 1), as some of them appear several times in the dataset. Obtained musical works were released between the years 1958 and 2015, and range from pop and rock to country and electro genres, as described in Appendix A. The dataset also exhibits a mix of solo artists and bands. There are 23 solo singers (12 males and 11 female), while the remaining performers are predominantly male bands with one vocalist, as summarized in table 1.

4.2 Evaluation Metrics

We evaluate the performance of our pipeline across three categories: text, pitch, and alignment (timing).

Text-based metrics measure the accuracy of lyrics' syllabification. For this category, we implement the following ASR metrics [28] by treating syllables as independent words:

- *Word Error Rate (WER)*: the percentage of syllables that were incorrectly predicted.
- *Match Error Rate (MER)*: the percentage of syllables that were incorrectly predicted and inserted.
- *Character Error Rate (CER)*: the percentage of characters that were incorrectly predicted.

Pitch-based metrics evaluate the accuracy of the extracted melody and frequency prediction, based on standard practices of MIR [29]:

- *Raw Pitch Accuracy (RPA)*: the ratio of accurately labeled melody frames within half a semitone.
- *Raw Chroma Accuracy (RCA)*: similar to RCA, but ignores octave errors.
- *Overall Accuracy (OA)*: the ratio of both melody and non-melody frames labeled correctly.

Alignment Score estimates the temporal synchronization between the predicted alignment and references. For syllable-level notation, we adopt the alignment score, defined in MFA evaluation framework,[11] which calculates the timed overlap cost between alignments. For each segment, the cost increases with the difference between the

9) https://pyphen.org/

10) As of January 8, 2026

11) https://montreal-forced-aligner.readthedocs.io/en/v3.3.8/user_guide/implementations/alignment_evaluation.html

Table 2 Average performance of the pipeline

Metric	
WER ↓	20.00% ± 10.77%
MER ↓	19.81% ± 10.56%
CER ↓	6.79% ± 5.49%
RPA ↑	16.72% ± 23.18%
RCA ↑	45.59% ± 15.97%
OA ↑	13.18% ± 18.43%
Alignment ↑	-5.19

predicted start and end times and the corresponding reference interval. An additional penalty is added if corresponding syllables do not match. These per-syllable overlap values are then averaged across all syllables and produce a single overlap score, interpreted as a measure of temporal synchronization between the predicted and reference alignments at the syllable level.

4.3 Results

The evaluation results of the pipeline are summarized in Table 2. Text-based metrics show relatively poor accuracy, particularly when compared to the performance of the underlying forced aligner. Low CER versus high WER and MER indicates that a substantial portion of errors arise from misaligned syllable segmentation, highlighting the difference between phonetic and linguistic syllabification. For instance, the word *"typical"*, syllabified as "*ty-pi-cal*" in the ground truth but predicted as "*typ-i-cal*" by the pipeline, results in harsher punishment in word-level metrics, explaining their disparity.

Pitch estimation also shows unexpected results. Metrics' values are notably below benchmark performances reported in previous studies[12)] on other evaluation datasets [30, 14, 31, 13, 12, 32]. RCA is higher than RPA, which implies the system often correctly identifies the pitch class, but on different octaves or tonality. These discrepancies may stem from error propagation or limitations within the pitch estimation pipeline, which require further analysis. Additionally, it may be necessary to reexamine the evaluation subset, as previous studies [33], including the dataset creators [2] themselves, have noted potential mismatches between audio recordings and the provided alignments due to raw annotations storing the notes as interval differences with respect to an unknown reference. The alignment score indicates minor systematic offsets in syllable timing, highlighting an ongoing challenge for ASR systems and forced aligners.

We further evaluate the pipeline with stratification by gender type and genre, shown in Appendix B. In summary, female performers, whether as soloists or in multiple-vocalist groups, tend to achieve better results on text-based and pitch estimation metrics. Additionally, multiple singers do not really affect the score, unless the genders of the singers are mixed. Genre-level evaluation reveals that song types do affect text alignment, transcription, and pitch analysis. Slower-paced songs tend to have better performance than rapid ones (Appendix B).

5 Conclusion

In this paper, we introduced Wav2Karaoke, a pipeline that cascades forced aligner and pitch tracking tools to transform English-language song recordings directly into syllable-to-note alignment. Objective evaluations reveal moderate text alignment performance, primarily due to syllabification mismatch, while pitch estimation demonstrates better performance on female voices and slower-paced songs.

6 Future Work

While Wav2Karaoke demonstrates a functional pipeline for syllable-to-note alignment, the current performance quality is limited. We plan to conduct an ablation study to identify which components contribute most to performance degradation and explore strategies to improve both syllable alignment and pitch estimation accuracy.

Due to its modular design, Wav2Karaoke can be extended to multilingual datasets. The forced aligner component, based on MFA [18], already supports multiple languages through pretrained acoustic and pronunciation models.[13)] Pyphen also provides language-specific syllabifications based on LibraOffice's[14)] Hunspell hyphenation dictionaries, covering more than 30 languages. Accordingly, the pipeline has the potential to align lyrics in languages beyond English. Evaluating alignment quality across these diverse linguistic scenarios can provide valuable insights and help guide future adaptations of the system to improve its generalizability and robustness.

12) https://github.com/lars76/pitch-benchmark

13) https://mfa-models.readthedocs.io/en/latest/acoustic/index.html

14) https://git.libreoffice.org/dictionaries

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP25K03175.

References

[1] Gabriel Meseguer-Brocal, Alice Cohen-Hadria, and Geoffroy Peeters. DALI: a large Dataset of synchronized Audio, LyrIcs and notes, automatically created using teacher-student machine learning paradigm. In ISMIR, editor, **19th International Society for Music Information Retrieval Conference**, September 2018.

[2] Gabriel Meseguer-Brocal, Alice Cohen-Hadria, and Geoffroy Peeters. Creating DALI, a Large Dataset of Synchronized Audio, Lyrics, and Notes. **Transactions of the International Society for Music Information Retrieval**, Jun 2020.

[3] J. C. R. Licklider. A duplex theory of pitch perception. **Experientia**, Vol. 7, No. 4, pp. 128–134, Apr 1951.

[4] Alain de Cheveigné and Hideki Kawahara. YIN, a fundamental frequency estimator for speech and music. **The Journal of the Acoustical Society of America**, Vol. 111 4, pp. 1917–30, 2002.

[5] Leonard E. Baum and Ted Petrie. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. **Annals of Mathematical Statistics**, Vol. 37, pp. 1554–1563, 1966.

[6] Andrew Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. **IEEE Transactions on Information Theory**, Vol. 13, No. 2, pp. 260–269, 1967.

[7] Matthias Mauch and Simon Dixon. PYIN: A fundamental frequency estimator using probabilistic threshold distributions. **2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, pp. 659–663, 2014.

[8] Jong Wook Kim, Justin Salamon, Peter Li, and Juan Pablo Bello. Crepe: A convolutional representation for pitch estimation. In **2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)**, pp. 161–165. IEEE, 2018.

[9] Lars Nieradzik. SwiftF0: Fast and Accurate Monophonic Pitch Detection, 2025.

[10] Haojie Wei, Xueke Cao, Tangpeng Dan, and Yueguo Chen. RMVPE: A Robust Model for Vocal Pitch Estimation in Polyphonic Music. In **Interspeech**, 2023.

[11] Rachel M. Bittner, Juan José Bosch, David Rubinstein, Gabriel Meseguer-Brocal, and Sebastian Ewert. A Lightweight Instrument-Agnostic Model for Polyphonic Note Transcription and Multipitch Estimation. In **Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)**, Singapore, 2022.

[12] Rachel M Bittner, Katherine Pasalo, Juan José Bosch, Gabriel Meseguer-Brocal, and David Rubinstein. vocadito: A dataset of solo vocals with f_0, note, and lyric annotations. **arXiv preprint arXiv:2110.05580**, 2021.

[13] Chao-Ling Hsu and Jyh-Shing Roger Jang. On the improvement of singing voice separation for monaural recordings using the MIR-1K dataset. **IEEE transactions on audio, speech, and language processing**, Vol. 18, No. 2, pp. 310–319, 2009.

[14] Jesse Engel, Cinjon Resnick, Adam Roberts, Sander Dieleman, Mohammad Norouzi, Douglas Eck, and Karen Simonyan. Neural audio synthesis of musical notes with wavenet autoencoders. In **International conference on machine learning**, pp. 1068–1077. PMLR, 2017.

[15] Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: A framework for self-supervised learning of speech representations. **Advances in neural information processing systems**, Vol. 33, pp. 12449–12460, 2020.

[16] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In **International conference on machine learning**, pp. 28492–28518. PMLR, 2023.

[17] Rotem Rousso, Eyal Cohen, Joseph Keshet, and Eleanor Chodroff. Tradition or Innovation: A Comparison of Modern ASR Methods for Forced Alignment. pp. 1525–1529, 2024.

[18] Michael McAuliffe, Michaela Socolof, Sarah Mihuc, Michael Wagner, and Morgan Sonderegger. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. In **Proc. Interspeech 2017**, pp. 498–502, 2017.

[19] Elisabeth Selkirk. On the major class features and syllable theory. In Mark Aronoff and Richard T. Oehrle, editors, **Language Sound Structure: Studies in Phonology**, pp. 107–136. MIT Press, Cambridge, MA, 1984.

[20] Ton Broeders and Carlos Gussenhoven, editors. **An Introduction to American English Phonetics. A theoretical guide to American English pronunciation**. 2025.

[21] R. Harald Baayen, Richard Piepenbrock, and Leon Gulikers. CELEX2 LDC96L14. Web Download, 1995. Available from Linguistic Data Consortium.

[22] Simon Rouard, Francisco Massa, and Alexandre Défossez. Hybrid Transformers for Music Source Separation. In **ICASSP 23**, 2023.

[23] Alexandre Défossez. Hybrid Spectrogram and Waveform Source Separation. In **Proceedings of the ISMIR 2021 Workshop on Music Source Separation**, 2021.

[24] Weinan Tong, Jiaxu Zhu, Jun Chen, Shiyin Kang, Tao Jiang, Yang Li, Zhiyong Wu, and Helen Meng. SCNet: sparse compression network for music source separation. In **ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, pp. 1276–1280. IEEE, 2024.

[25] Wei-Tsung Lu, Ju-Chiang Wang, Qiuqiang Kong, and Yun-Ning Hung. Music source separation with band-split rope transformer. In **ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, pp. 481–485. IEEE, 2024.

[26] Michael McAuliffe and Morgan Sonderegger. English (US) ARPA acoustic model v3.0.0. Technical report, https://mfa-models.readthedocs.io/acoustic/English/English(US)ARPAacousticmodelv3_0_0.html, Feb 2024.

[27] Kyle Gorman, Jonathan Howell, and Michael Wagner. Prosodylab-aligner: A tool for forced alignment of laboratory speech. **Canadian Acoustics**, Vol. 39, No. 3, pp. 192–193, 2011.

[28] Andrew Morris, Viktoria Maier, and Phil Green. From wer and ril to mer and wil: improved evaluation measures for connected speech recognition. In **Proc. Interspeech 2004**, 10 2004.

[29] Colin Raffel, Brian McFee, Eric J. Humphrey, Justin Salamon, Oriol Nieto, Dawen Liang, and Daniel P. W. Ellis. MIR_EVAL: A Transparent Implementation of Common MIR Metrics. In ISMIR, editor, **International Society for Music Information Retrieval Conference**, 2014.

[30] Gregor Pirker, Michael Wohlmayr, tefan Petrík, and Franz Pernkopf. A Pitch Tracking Corpus with Evaluation on Multipitch Tracking Scenario. In **Interspeech**, 2011.

[31] Justin Salamon, Rachel M. Bittner, Jordi Bonada, Juan J. Bosch, Emilia Gómez, and Juan Pablo Bello. An analysis/synthesis framework for automatic f0 annotation of multitrack datasets. In **Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)**, Suzhou, China, October 2017.

[32] Peter Foster, Siddharth Sigtia, Sacha Krstulovic, Jon Barker, and Mark D. Plumbley. Chime-home: A dataset for sound source recognition in a domestic environment. In **2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)**, pp. 1–5, 2015.

[33] Timon Kick, Florian Grötschla, Luca A Lanzendörfer, and Roger Wattenhofer. Contrastive Lyrics Alignment with a Timestamp-Informed Loss. In **Audio Imagination: NeurIPS 2024 Workshop AI-Driven Speech, Music, and Sound Generation**, 2024.

A Evaluation Dataset Information

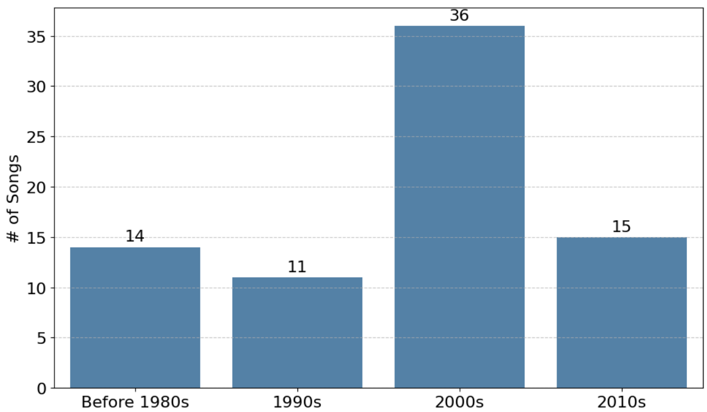

Figure 3 Release year of songs in the evaluation dataset

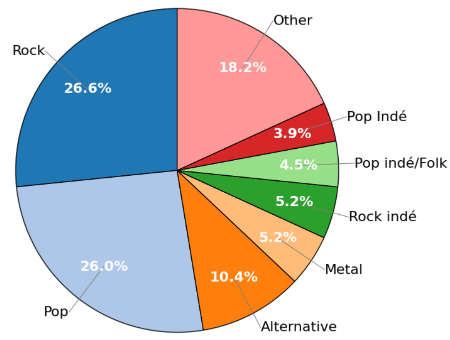

Figure 4 Genre distribution of the evaluation dataset

B Stratified Evaluation Results

Table 3 Average pipeline performance per vocal category.*

Category	WER (%)↓	MER (%)↓	CER (%)↓	RPA (%)↑	RCA (%)↑	OA (%)↑	Alignment ↑
Multi Female	**11.05 ± 3.32**	**10.95 ± 3.14**	4.01 ± 2.21	20.97 ± 25.89	39.88 ± 17.66	15.37 ± 19.24	-6.40
Multi Male	23.59 ± 12.95	23.30 ± 12.56	7.11 ± 5.66	19.56 ± 25.50	44.30 ± 12.22	15.80 ± 20.68	-6.05
Multi Mix	11.46 ± 3.31	11.44 ± 3.33	**3.73 ± 2.40**	2.34 ± 3.31	30.48 ± 36.49	2.29 ± 3.07	-1.54
Solo Female	17.18 ± 7.89	17.07 ± 7.87	5.64 ± 4.55	**30.89 ± 25.77**	**50.75 ± 12.67**	**23.23 ± 20.03**	**-1.24**
Solo Male	21.14 ± 11.03	20.93 ± 10.82	7.49 ± 5.98	11.12 ± 19.92	45.30 ± 16.95	9.14 ± 16.43	-6.40

* **Bold** indicates best performance per column; underline indicates the poorest.

Table 4 Average pipeline performance per song genre. *

Genre	WER (%)↓	MER (%)↓	CER (%)↓	RPA (%)↑	RCA (%)↑	OA (%)↑	Alignment ↑	BPM$_\mu$
Folk	8.78 ± 3.77	8.78 ± 3.77	2.35 ± 1.26	**64.97 ± 0.13**	**66.28 ± 1.70**	**53.26 ± 4.99**	-0.24	137.82
Jeunesse	12.23 ± 2.93	12.17 ± 2.84	3.78 ± 0.52	44.61 ± 38.53	65.60 ± 5.07	35.35 ± 30.60	-2.58	130.45
Films/Jeux vidéo	15.47	15.30	3.82	0.18	59.86	0.20	-7.35	152.00
Musiques de films	15.47	15.30	3.82	0.18	59.86	0.20	-7.35	152.00
BO TV	15.47	15.30	3.82	0.18	59.86	0.20	-7.35	152.00
Dance	21.66 ± 16.32	21.66 ± 16.32	10.48 ± 10.64	38.37 ± 26.54	56.73 ± 13.07	30.75 ± 20.85	-0.38	115.87
Rock indé	25.56 ± 12.02	24.94 ± 11.24	7.71 ± 3.40	28.17 ± 29.80	55.82 ± 8.85	21.77 ± 22.86	-1.87	124.91
Alternative	22.16 ± 12.33	21.63 ± 11.58	6.56 ± 3.46	21.97 ± 28.90	54.78 ± 11.21	17.48 ± 23.06	-2.35	119.44
Pop internationale	10.27	10.27	2.41	53.18	54.51	39.97	-0.21	172.27
R&B contemporain	10.27	10.27	2.41	53.18	54.51	39.97	-0.21	172.27
Pop indé/Folk	24.52 ± 12.58	24.04 ± 11.82	7.34 ± 3.48	32.14 ± 29.81	53.96 ± 7.70	24.83 ± 22.86	-2.12	118.14
Pop Indé	26.70 ± 12.25	26.14 ± 11.44	8.02 ± 3.26	26.68 ± 28.57	51.70 ± 5.33	20.68 ± 21.96	-2.42	120.60
R&B	17.84 ± 10.71	17.84 ± 10.71	4.43 ± 2.86	28.11 ± 35.45	48.80 ± 8.08	21.10 ± 26.69	-0.14	126.51
Country	28.44 ± 19.87	26.74 ± 17.53	5.86 ± 1.21	24.15 ± 34.15	48.22 ± 0.17	18.64 ± 26.30	-5.16	125.69
Pop	18.21 ± 10.70	18.03 ± 10.46	6.18 ± 4.77	15.37 ± 23.42	48.14 ± 13.98	11.37 ± 17.33	-3.13	126.70
Electro	11.24	11.24	8.73	47.61	47.73	37.30	**-0.02**	129.20
Techno/House	11.24	11.24	8.73	47.61	47.73	37.30	**-0.02**	129.20
Singer & Songwriter	**5.19**	**5.18**	**1.93**	0.24	46.88	0.24	-2.17	129.20
Rock	19.71 ± 11.01	19.46 ± 10.73	6.11 ± 4.13	15.12 ± 23.72	46.13 ± 17.21	12.15 ± 19.05	-8.05	122.38
Soul	25.41	25.41	6.45	3.04	43.09	2.23	-0.07	80.75
Metal	22.72 ± 10.75	22.62 ± 10.78	7.08 ± 3.99	10.17 ± 17.77	41.30 ± 18.54	8.23 ± 14.35	-3.06	125.81
Hard Rock	21.15 ± 14.64	21.11 ± 14.61	7.72 ± 4.89	0.36 ± 0.72	40.20 ± 12.74	0.27 ± 0.47	-4.39	136.44
Rock & Roll/Rockabilly	32.39 ± 14.29	30.44 ± 12.30	5.74 ± 1.39	24.26 ± 33.99	25.54 ± 32.24	18.73 ± 26.18	-5.14	117.69

* **Bold** indicates best performance per column; underline indicates the poorest.
BPM (Beats per minute) is a measurement of tempo in music