

音楽情報処理 (第1回)

音楽情報処理領域概観

片寄晴弘

関西学院大学工学部情報工学課程

講義の特徴

- 音楽情報処理関連の研究をはじめるとにあたっての基本的知識の獲得
- 反転授業スタイル（講義資料事前提示
他の科目と比べて予習作業量が多い）
- 履修条件：
（音楽に関する耳 or 基礎的な数理能力）
and（音楽がめちゃくちゃ好き or 学習意欲）
条件満たされると天国、でないと大変なことも！
- 座学講義として実施
- 成績：レポート（50%）＋テスト（50%）

キーワード

- デザイン
- アート
- 自己表現
- コンテンツ
- 認知・心理
- プログラミング
- 人工知能
- 信号処理
- インタラクション

後半は数理情報処理、「数学」必須！

本日ミニレポート

1. 好きな音楽？音楽経験（あれば）？
2. 音楽がわかるとはどういうこと？

本日のメニュー

- アイスブレイク
- 音楽の楽しみの歴史的概観
- 音楽情報処理の歴史的概観
- これからの講義の予定

「超」イントロクイズ (2024年度版)

1  

2  

3  

4  

5  

6  

7  

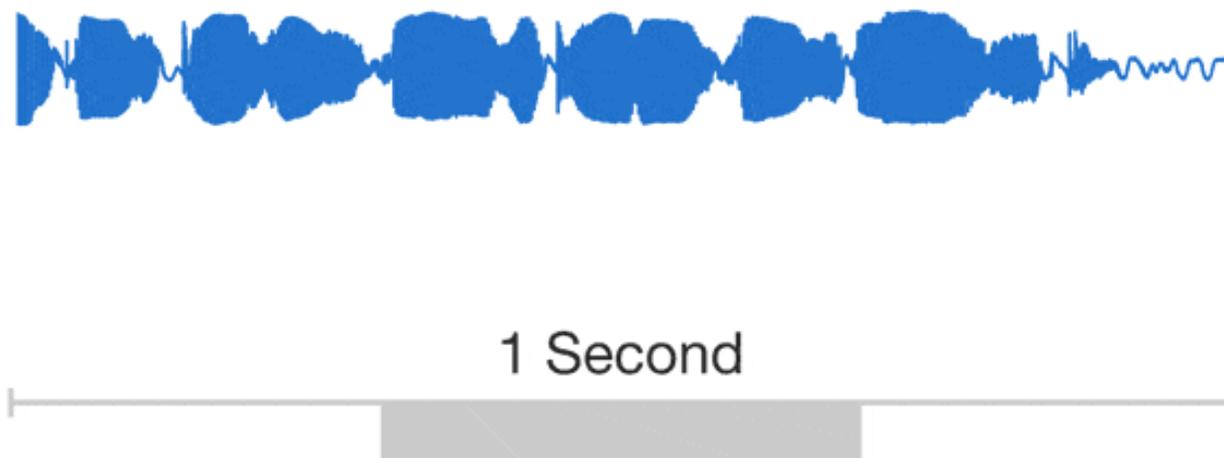
8  

9  

10  

本日のメニュー

- アイスブレイク 音の時間分解能

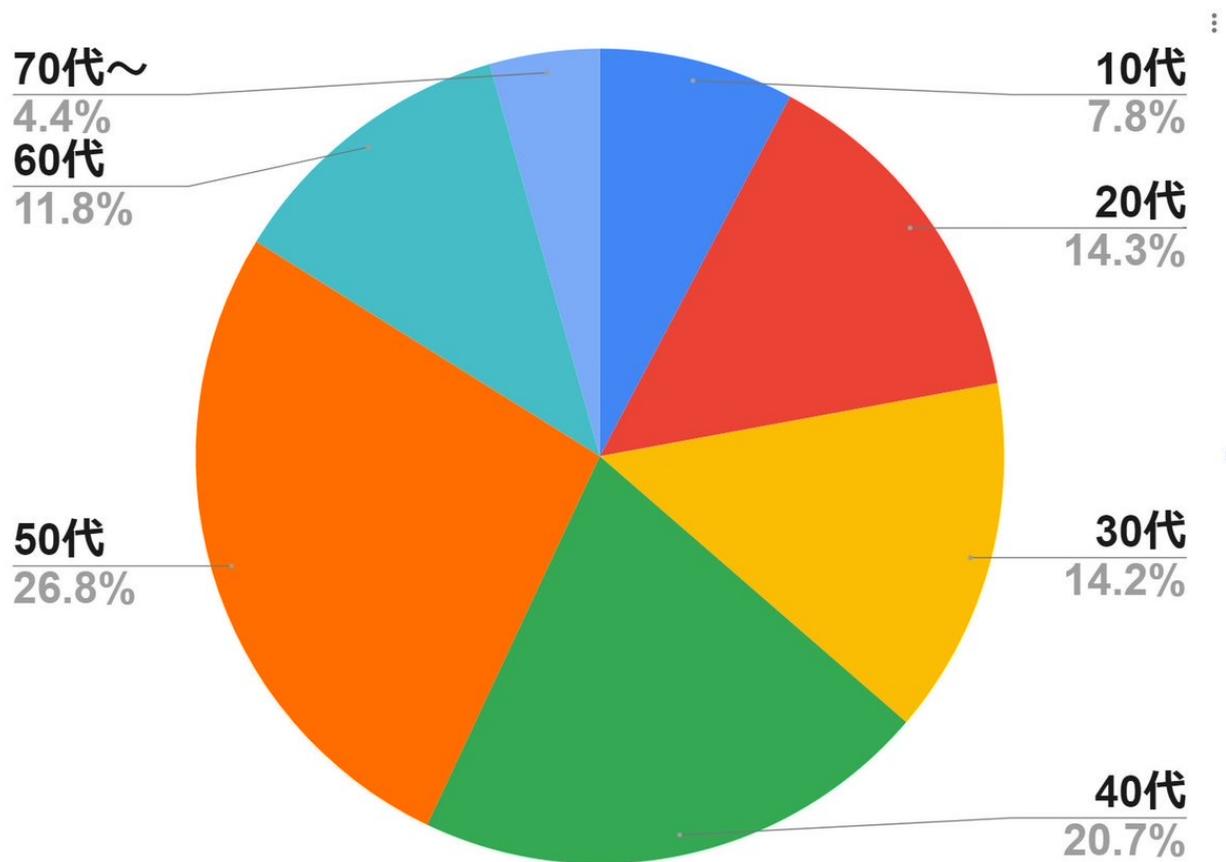


<https://medium.com/@naotokui/deep-learningを用いた音楽生成手法のまとめ-サーベイ-1298d29f8101> より

本日のメニュー

- アイスブレイク
- 音楽の楽しみの歴史的概観
- 音楽情報処理の歴史的概観
- これからの講義の予定

DTM人口 (2023年)



• <https://twitter.com/kenfujimoto/status/1620791366233251840> より

音楽の楽しみの歴史的概観

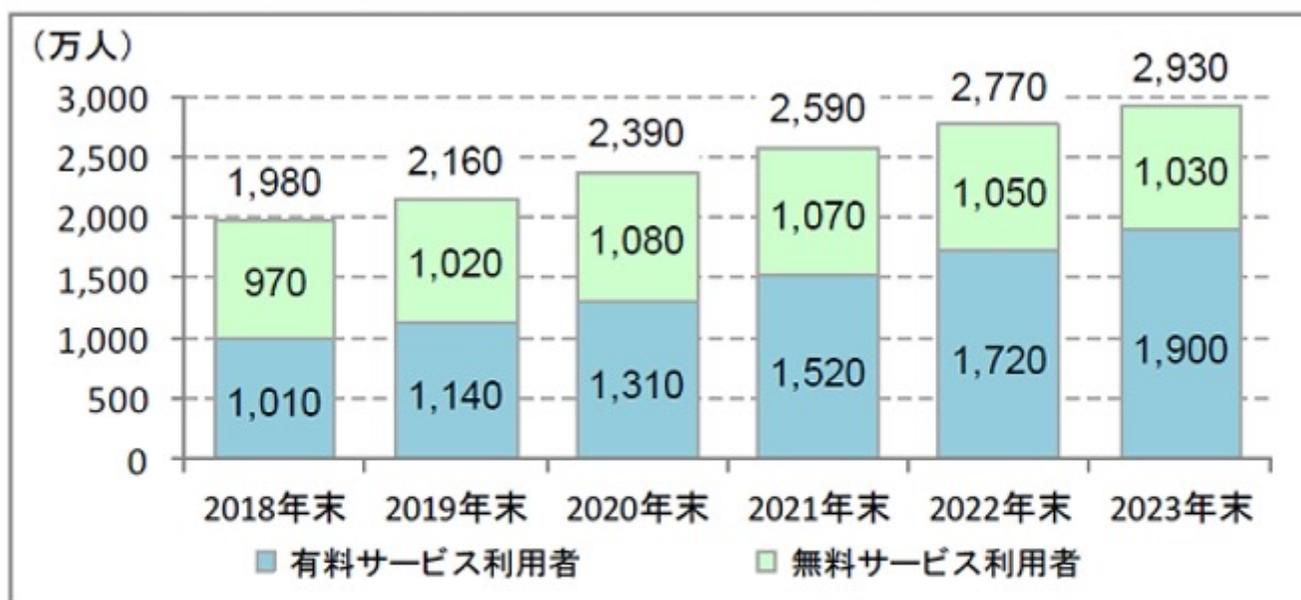


<https://prtimes.jp/main/html/rd/p/000000165.000019470.html>
<https://01earth.jp/creative/other/analog-record/>
https://belluna.jp/interior/01/010401/d/OCZB/00910/goods_detail/
<https://im.belluna.jp/interior/ph/O/9126/1007279126/DLARGE.JPG>
<http://analogsound.blog.fc2.com/blog-date-201209.html>
<https://joshinweb.jp/av/1651/4548736105409.html>
<https://japan.cnet.com/article/20355791/> より

音楽の楽しみの歴史的概観（～2023）

サブスク（定額制サービス）

■表1. 定額制音楽配信サービス利用者数 需要予測



* ICT総研による利用者数推計。

* 無料サービス利用者には、お試し無料サービス期間中の利用者などが含まれる。

* 有料サービスと無料サービスの両方を利用する場合は有料サービス利用者としてカウント。

ICT総研調べ

ICT総研 2020年度調査 <https://ictr.co.jp/report/20201113.html/> より

音楽の楽しみの歴史的概観（～2018）

2018年の日本の音楽コンテンツ市場

- CD等：前年比 9% 減少
- 配信市場：前年比13%増

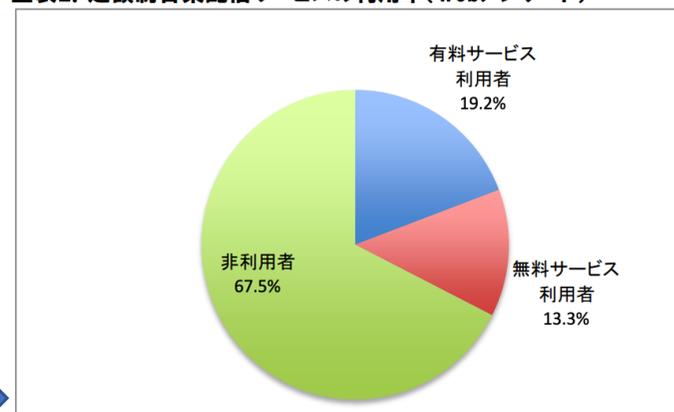
（日本レコード協会の統計）

- 配信の中でも
 - Prime Music（Amazon）Apple Music, LINE, Spotify

サブスク（定額制サービス）

ICT 総研が 2020 年 10 月に実施した
Web アンケート調査の結果より

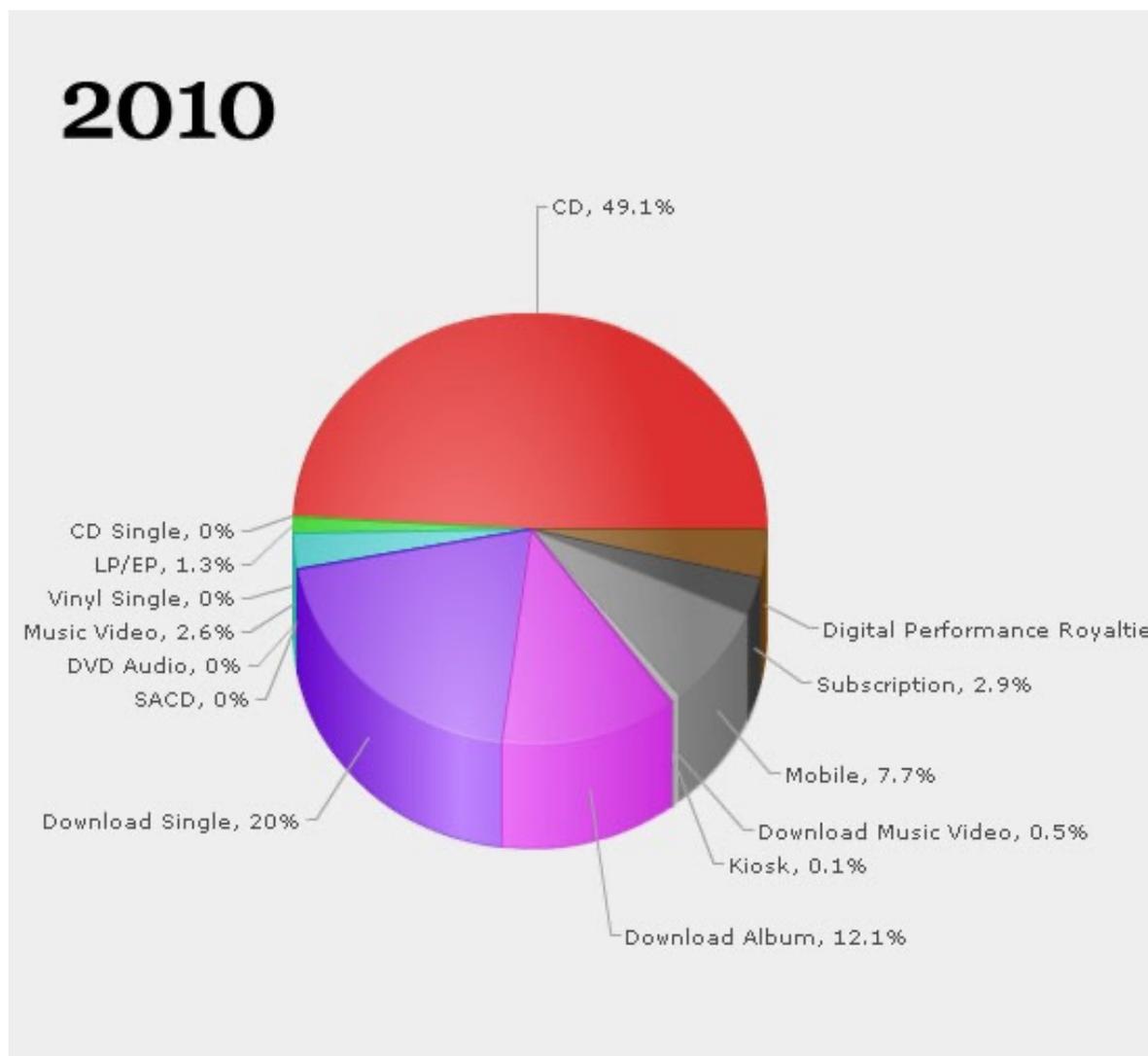
■表2. 定額制音楽配信サービスの利用率(Webアンケート)



* 4,409人に対するWebアンケート調査結果。複数回答。

* 有料サービスと無料サービスの両方を利用する場合は有料サービス利用者としてカウント。

音楽の楽しみの歴史的概観（振り返り）



<http://www.digitalmusicnews.com/stories/081611thirty> より US-based data

音楽の楽しみの歴史的概観（振り返り）

もともとは．．感情・意図の伝達
権威者の権威付け
その場に居合わせた人のみ．． 吟遊詩人スタイル？

1455 **グーテンベルグ 聖書の印刷**
作曲・編曲，演奏（分業スタイルの確立）
西洋的合理主義？

1877 **エジソン レコードの発明**
1892 エジソン、円筒レコード複製に成功、量産化へ拍車
1896 マルコーニが、無線通信に関する特許
1920 ピッツバーグに初のラジオ局
音楽鑑賞エンタテインメントの普及
1926 高柳健次郎、「イ」の字の映出に成功
1935 テレビの試験放送の開始（ドイツ，フランス）

1979~ **Walkman**

1982 **MIDI の制定，パソコンの普及**
DTMの普及
作曲・編曲，演奏の一体，インタラクティブアート
誰もが音楽家の時代へ

2001~ **iPod（データダウンロード型）**

2015 **サブスクの時代へ**

宮廷音楽 1800年頃
ベートーベン



音楽を持ち歩く

音楽の原点への回帰？



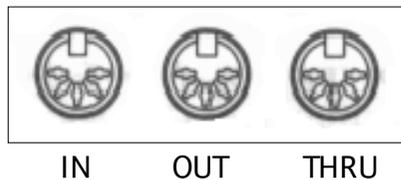
無尽蔵の音楽を楽しむ時代へ

音楽の楽しみの歴史的概観

DTM (Desk Top Music)

- 1988年頃ら、パソコンの普及とともに普及
- MIDI, MIDI音源モジュールを主体とした音楽編集、再生システム (ソフトウェア)
- 2000年頃に衰退？

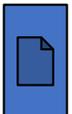
トピック MIDI



MIDIインタフェースの形状
(5ピンのDINコネクタ)

- 1982年に発表されたデジタル楽器間のデータ転送及びハードウェアに関する設計規約
- 31.25Kbpsの電流ループによるシリアル通信で、鍵盤楽器をモデルとした各種の演奏操作イベント情報を転送する（1音の発音データに約1msかかる）
- MIDIメッセージ：1バイト（8ビット）のデータの並び。
 - 1~16のMIDIチャンネルの区別を持ったチャンネルメッセージとチャンネルの違いを越えてシステム全体に共有されるシステムメッセージ
 - チャンネルメッセージ：ノートオン（音を鳴らす）、ノートオフ（音を切る）、モジュレーションホイールなど
- デジタル楽器にとどまらず、スタジオ機器や舞台照明の制御まで、用途は広がっている。

[2017 梯さん死去](#)

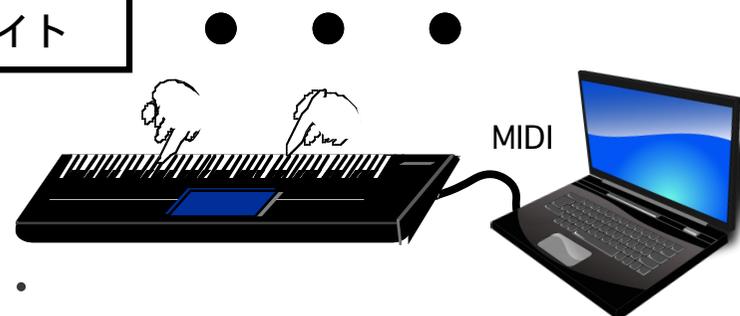


トピック MIDI

ステータスバイト

データバイト

データバイト



- キーボードの中央ドの音を目一杯押すと
 - 144, 60, 127という値がMIDIを通じて送信される.
 - 144 : 1 chで示される音源のノートオン
 - ステータスバイトと呼ばれる
 - 連続するこのメッセージが何であるかという情報を示している.
 - 128~255までの値が使われる
 - 60 : ノートナンバーと呼ばれ, 中央ドの音を示している.
 - 127 : 音量
 - (光センサを用いて, キーボードの押される速さを検出し, それを音量に割り当てたため, ベロシティと呼ばれるようになった.)
 - 指をはなすと, 128, 60, XX (144, 60, 0) という数値が送信されている.
- 計算機側からの同様のデータを送ると, キーボードに直接接触らずとも同様の発音がある.

音楽の楽しみの歴史的概観

DTM (Desk Top Music)

- 1988年頃ら、パソコンの普及とともに普及
- MIDI, MIDI音源モジュールを主体とした音楽編集、再生システム (ソフトウェア)
- 2000年頃に衰退?

DAW

- オーディオ信号を直接編集する音楽システム
- 2000年頃にハードからソフトウェアシステムへ
 - ProTools, GarageBand, Cubase VST 等

Pickup! 藤本健のDTMステーション

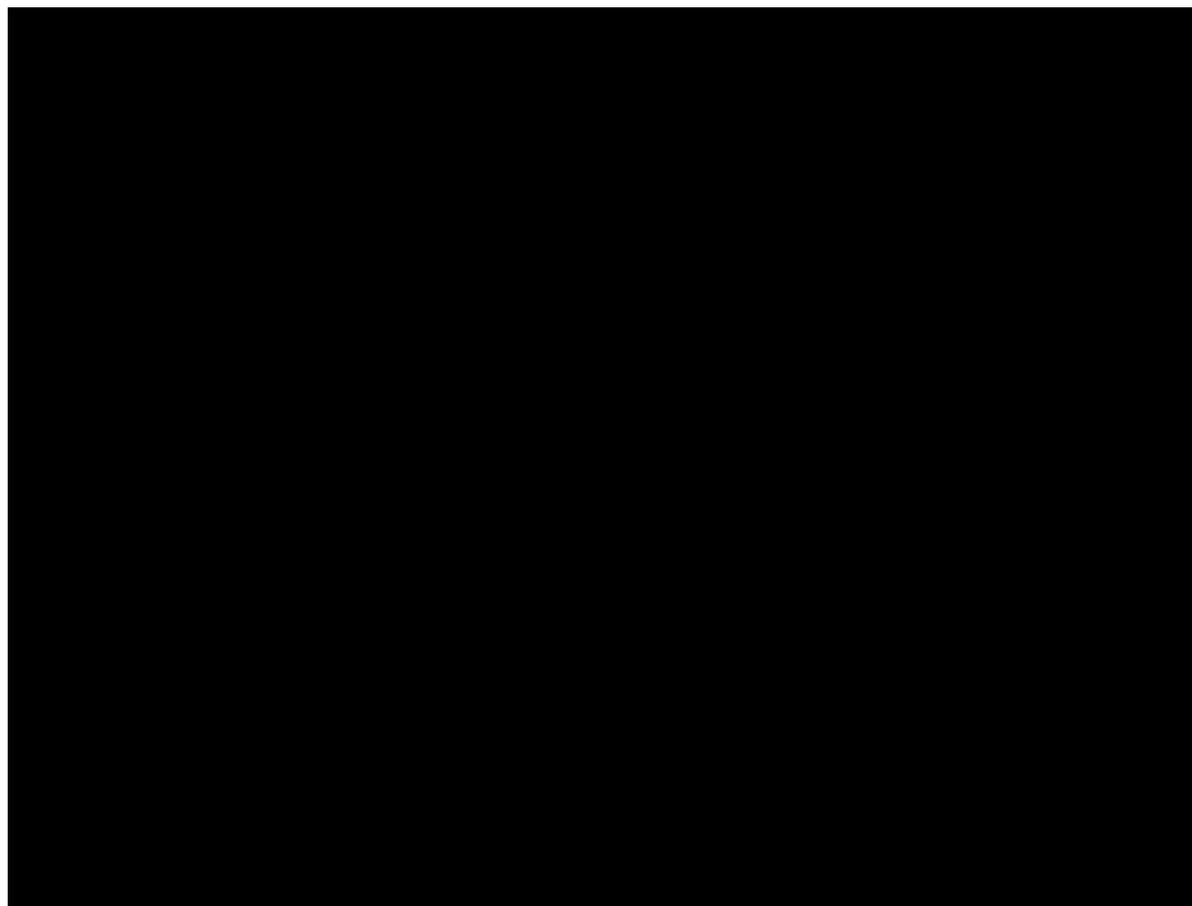
<https://www.dtmstation.com/>

本日のメニュー

- アイスブレイク
- 音楽の楽しみの歴史的概観
- 音楽情報処理の歴史的概観
- これからの講義の予定

音楽情報処理の歴史的概観

KGBroadBandStation005 (2008)



<https://crestmuse.jp/klab/lecture/mi/KGBroadBandStation.mp4>

音楽情報処理の歴史的概観

- 講義資料
- <https://crestmuse.jp/klab/lecture/mi/>
- 電子情報通信学会知識ベース2群9編
(音楽情報処理)
- http://www.ieice-hbkb.org/portal/doc_557.html

音楽情報処理の歴史的概観 (~2010)

1960 1970 1980 1990 2000



2010年以降は？ 深層学習応用の時代

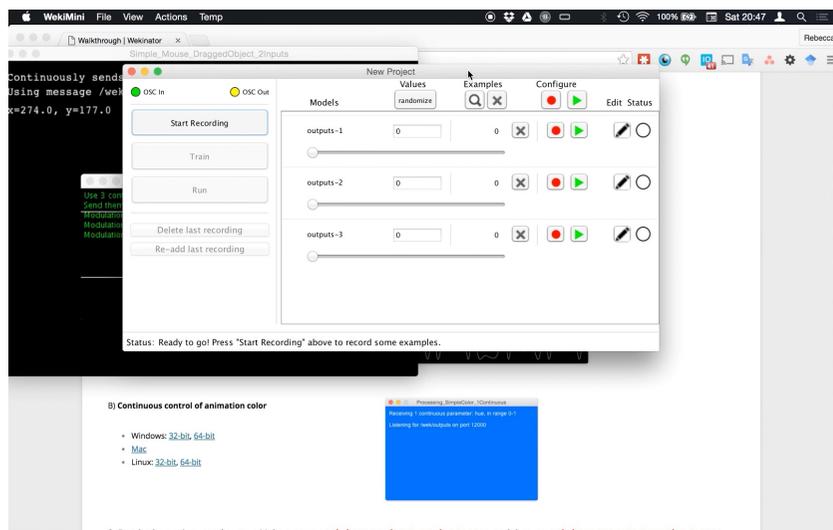
音楽領域における機械学習応用

- Wekinator (R. Feibrink 2009~) [新世代楽器系]
 - DeepBach (G. Hadjeres 2016) [編曲系]
 - Mann のシステム (Y. Man 2019) [新世代楽器系]
 - Dear Glenn (Maezawa 2019) [演奏表情付け系]
 - Music Transformer (Google 2018) [自動作曲系]
 - Jukebox (P. Dhariwal 2020) [自動作曲系]
 - DDSP (Tensorflow Magenta 2020) [新世代楽器系]
 - Mubert (Mubert Inc. 2022) [プロンプト系]
 - MusicLM (Google 2023) [プロンプト系]
 - 川口+片寄 (関学 2023) [プロンプト系]
 - udio (<https://www.udio.com/> 2024) [プロンプト系]
- * サウンド生成系は VAE → Diffusion model へ (後ほど)

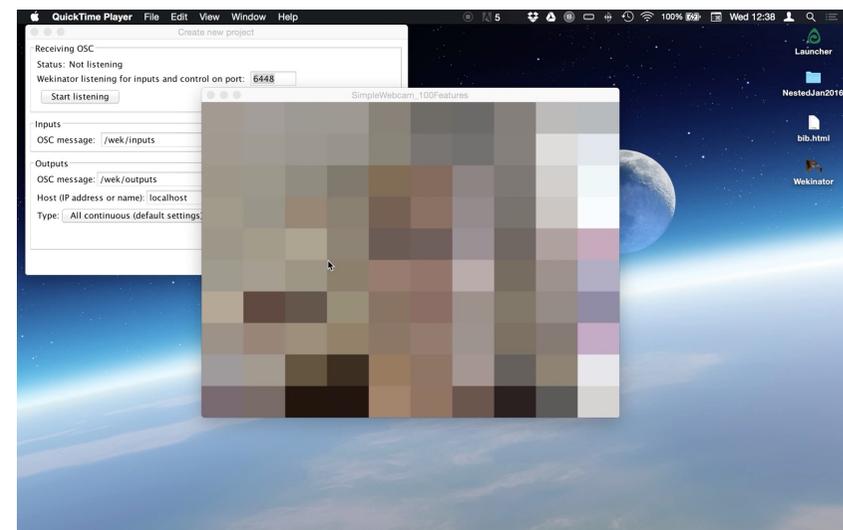
Wekinator [新世代楽器系] (R. Fiebrink 2009)

<http://www.wekinator.org/>

- ジェスチャとコントロールパラメータとのマッピングの支援環境
- 応用例も多数（オープンソース）



1:12

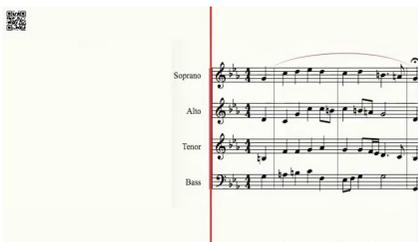
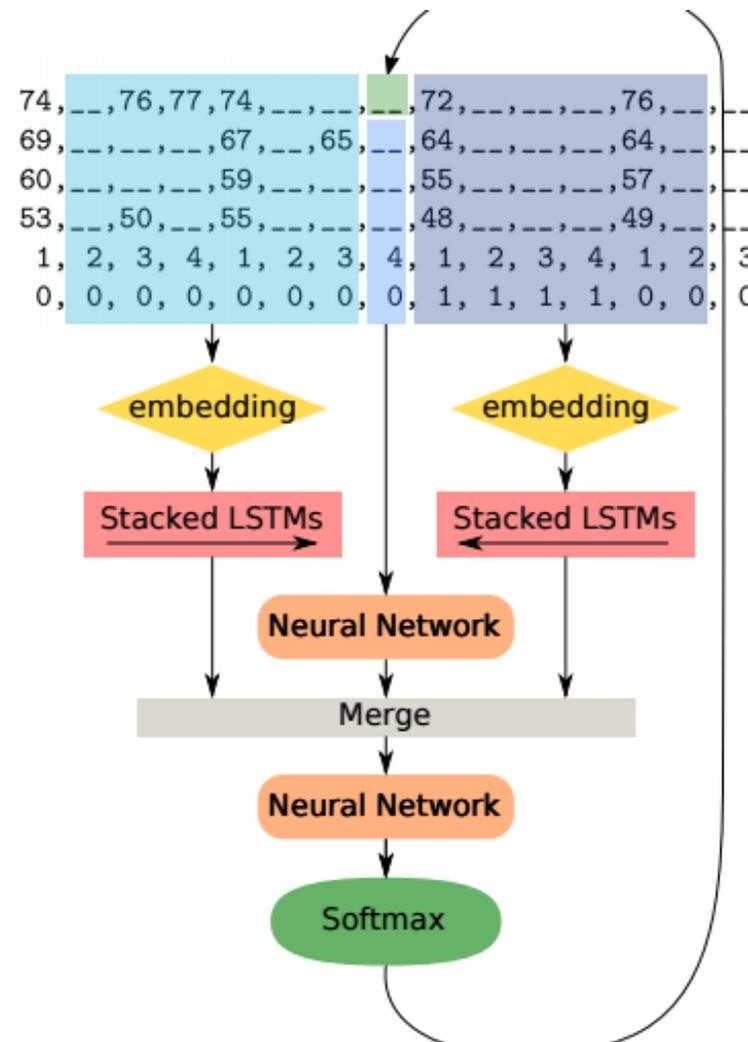


2:40

DeepBach [編曲系]

(Gaëtan Hadjeres, François Pachet 2016)

- 四声体和声課題（メロディーに対してバス、アルト、テノールパートを生成する課題）を実施
- バッハの賛美歌曲 352曲とそれらを転調させたものの2503曲を元に学習
- 過去と未来の系列にLSTM (Deep-RNN) を適用するのがポイント
- 400名のプロのミュージシャンや音大学生を含む1600人テストで、半数がオリジナルだと間違える精度
- オープンソース



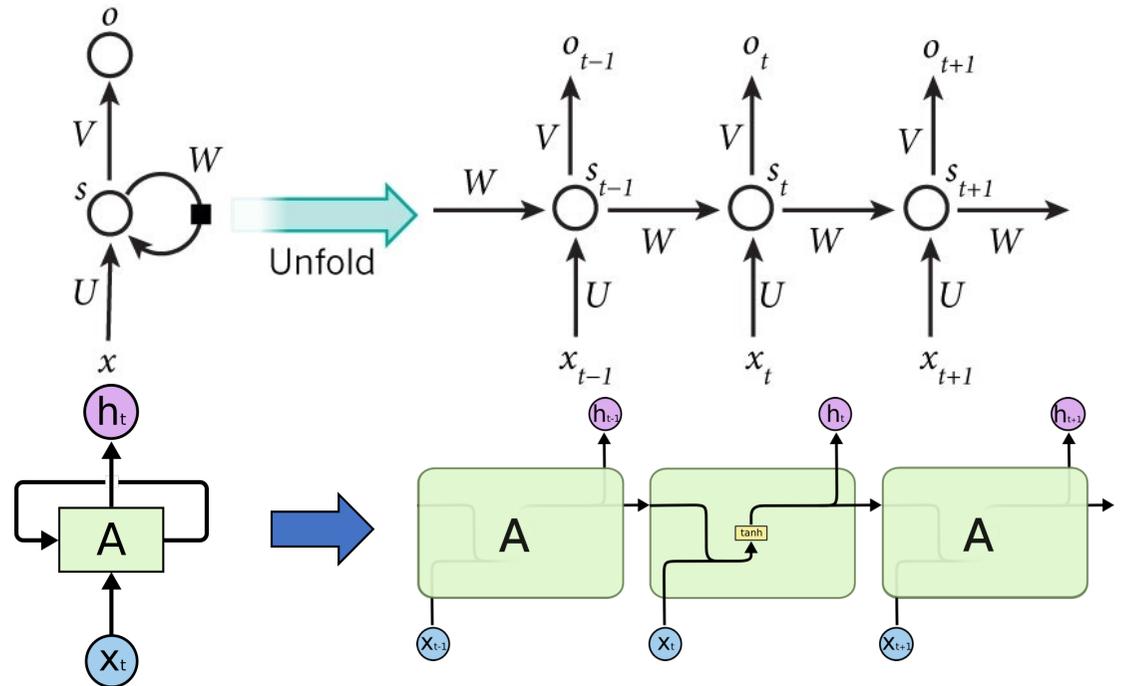
0:59

DeepBach [編曲系]

(Gaëtan Hadjeres, François Pachet 2016)

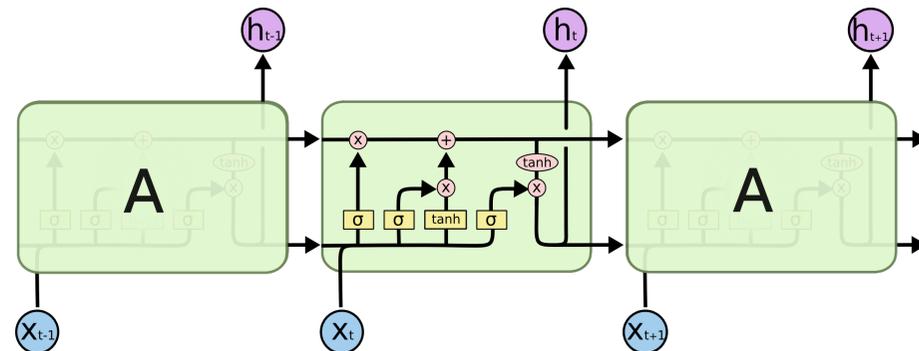
RNN(1986~) :

- ノードにフィードバックループを設けることで、時系列情報（前後関係）を扱うことを可能としたネットワークモデル
- 自然言語や音声情報の処理が応用範囲



LSTM(1997~) :

- 基本的にはRNNと同じだが、内部に直前ノードの状態を参照できるチャンネルを有する
- この仕組みにより現在から離れた情報のコーディングができる



Mann のシステム [新世代楽器系]

(Y. Mann, Magenta Studio Google TensorFlow 2019)

口笛音をオーケストラ（和声付き）風に変換

<https://twitter.com/yotammann/status/1197264489684623362>

Real-time audio
timbre transfer using
a GAN trained on
Mozart symphonies.
Whistling ends up like
a flute with some
accompaniment.
anyone want a guitar
pedal?



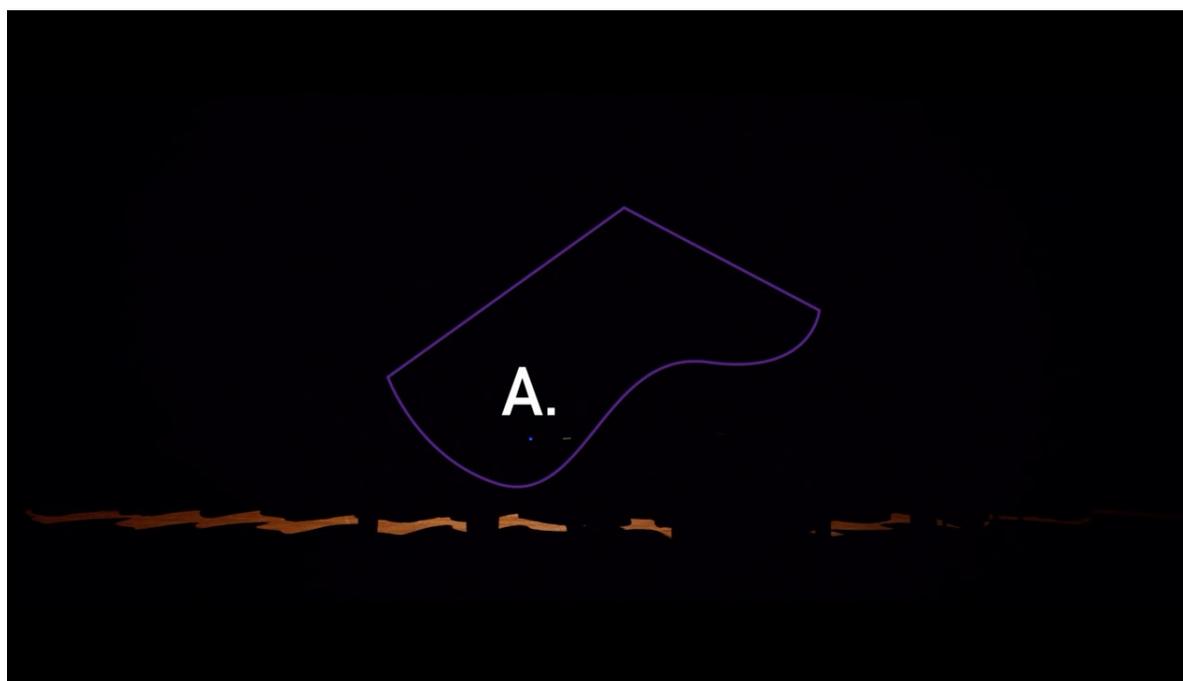
0:30

Dear Glenn [演奏表情付け系]

(Maezawa, YAMAHA 2019)

https://www.yamaha.com/ja/about/ai/dear_glenn/

https://www.yamaha.com/ja/news_release/2019/19102301/

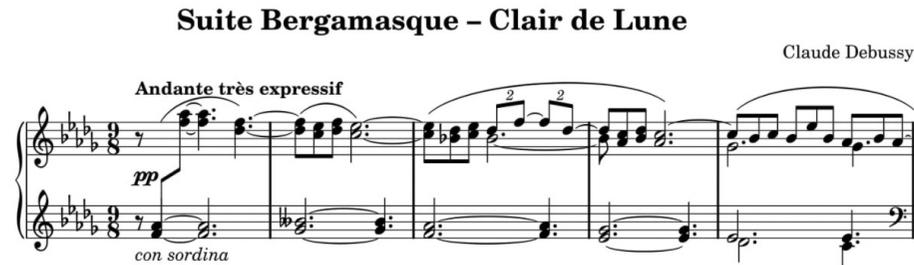


<http://youtu.be/vyUgY3R9-GQ>

Music Transformer [自動作曲系]

(Huang et al. 2018)

Attention (Transformer)系 後続予測型作曲システム



冒頭2小節
入力で後続作曲

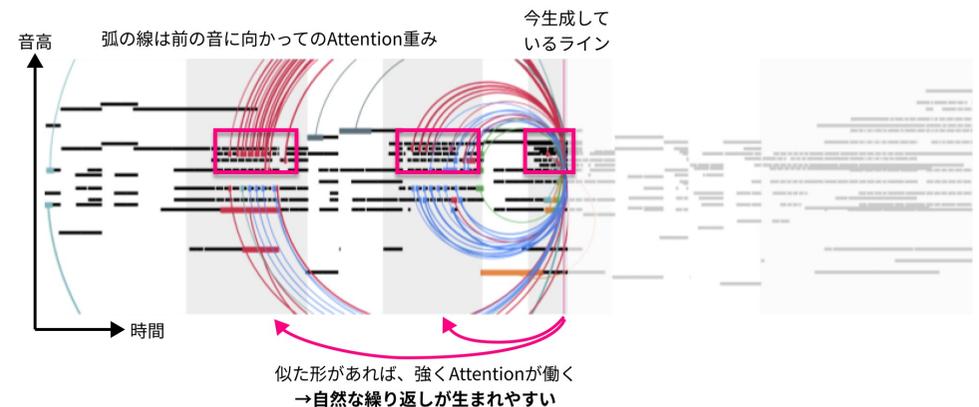


Attention機構

- 時系列（順番が意味を持つ）データの現状最強学習モデル
- 参照「系列」の Attention重みによる（長期・短期）の探索と後続予測 → 繰り返し構造の出現

<参考>

- [Music Transformer 日本語解説](#)
- [Attention機構 日本語解説](#)



Jukebox [自動作曲系]

(P. Dhariwal 2020)

VAE系作曲システム

(歌手名・ジャンル・歌詞入力で audio出力)

<https://openai.com/blog/jukebox/>

←サイトから音も聞けます。

(日本語 [Jukebox紹介記事](#) by 徳井直生先生 (慶應大))



3:30

<http://naotokui.net/2020/05/openai-jukebox/>

この実装では「歌詞」が変数ですが、システムが「歌詞が何か」わかっているわけではありません。→音楽の表象の一つということで、ほぼ同じアーキテクチャでメロディやコード、テンション等への適用（モデル化）が見込まれます。また、[スタイル変換深層学習](#) や[CycleGAN](#) との組み合わせで、さらにすごいことになりそうです。

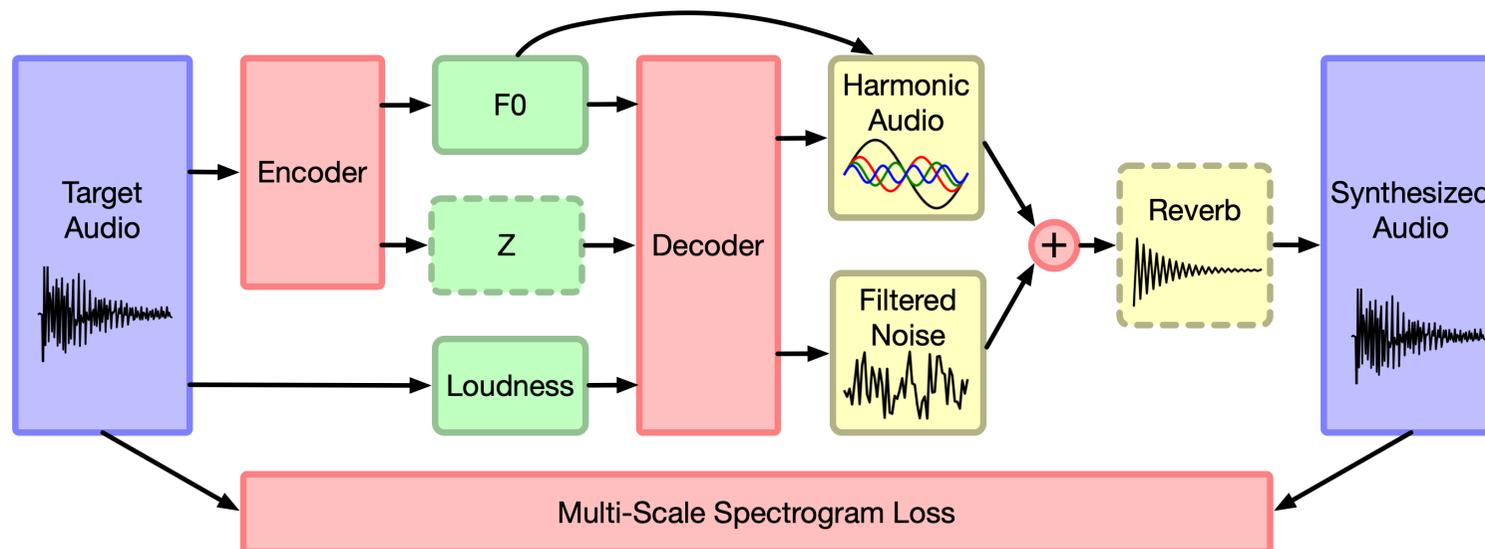
DDSP [新世代楽器系]

(TensorFlow Magenta Group 2020)

深層学習音声変換器

<https://magenta.tensorflow.org/ddsp>

前ページと似ているけれど別バージョン？



Mubert [プロンプト系自動作曲]

(Mubert Inc. 2022)

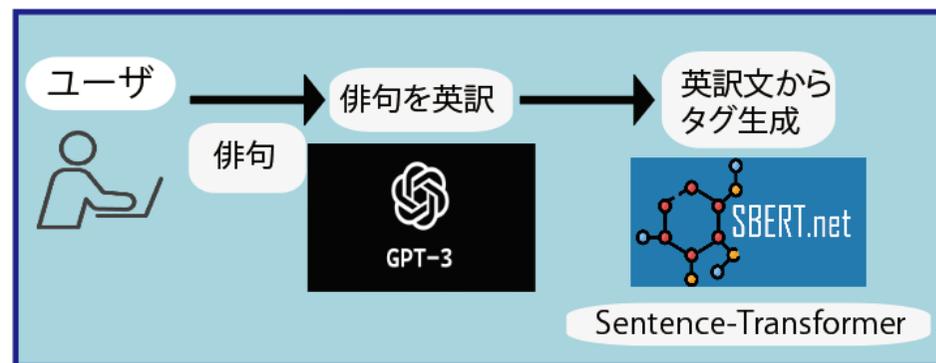
2022 ~流行りのプロンプト系 (Text to Music) [セットアップ](#)

俳句のBGMサンプル (池永 22卒論)



柿食えば
鐘が鳴るなり
法隆寺

から作成
した音源



- MubertAIと一般的な翻訳エンジンで生成



生成タグ:

['lullaby', 'bells', 'romantic', 'techno']

翻訳結果:

「Eating a persimmon, you hear the bell ring」

生成された楽曲
タグが一部微妙...

- 提案システムで生成したBGM



生成タグ: ['nature', 'drill', 'meditative', 'bells']

追加タグ: ['zen', 'nostalgic']

新たにタグを追加して、より俳句の意味合いに！

MusicLM: Generating Music From Text 2023

<https://google-research.github.io/seanet/musiclm/examples/>

We introduce MusicLM, a model generating high-fidelity music from text descriptions such as *"a calming violin melody backed by a distorted guitar riff"*. MusicLM casts the process of conditional music generation as a hierarchical sequence-to-sequence modeling task, and it generates music at 24 kHz that remains consistent over several minutes. Our experiments show that MusicLM outperforms previous systems both in audio quality and adherence to the text description. Moreover, we demonstrate that MusicLM can be conditioned on both text and a melody in that it can transform whistled and hummed melodies according to the style described in a text caption. To support future research, we publicly release MusicCaps, a dataset composed of 5.5k music-text pairs, with rich text descriptions provided by human experts.

GPT-4による対話的音楽生成のための プロンプトエンジニアリング 2023

<https://youtu.be/Zk99iQE-TxE>

① 作りたい曲のイメージをテキスト入力

カフェっぽいゆったりとした曲を作って欲しい

前の曲をもっと伸ばして

ユーザー

② プロンプトをGPT-4に送信



システム

ABC記譜法

```
M: 4/4, K: G, Q: 1/4=60
| "Gmaj" G3/2 |
| D/2 E3/2 C/2 |
| "Cmaj" C2 E2 |
| "Bb" _B2 D2 |
```

生成曲の例
(few-shot)

作る手順
制約



GPT-4

明るさを0.3ぐらい上げて欲しい

暗 ————— 明

④ 音響データに変換

⑤ 聴いてさらに改良

③ 結果をテキスト出力

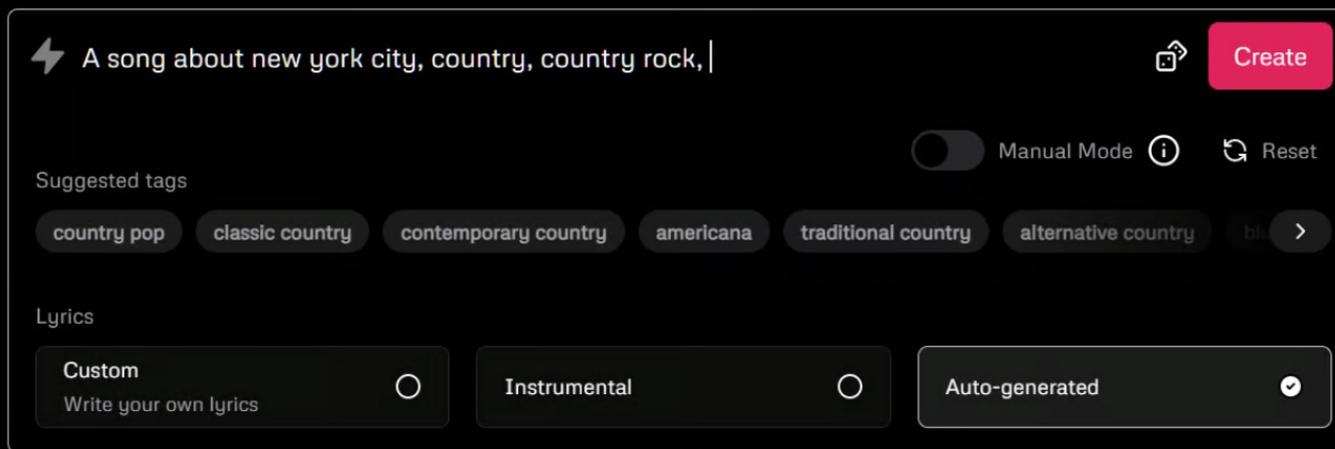
udio: Generating Music From Text 2024

The logo for 'udio beta' is displayed in white text on a black rectangular background. The word 'udio' is in a bold, sans-serif font, and 'beta' is in a smaller, lighter font to its right.

<https://www.udio.com/>

<https://twitter.com/udiomusic>

Create your first song

A screenshot of the udio web interface. At the top, there is a text input field containing the prompt "A song about new york city, country, country rock, |". To the right of the input is a pink "Create" button. Below the input field, there are several controls: a "Manual Mode" toggle switch (currently off), an information icon, and a "Reset" button. Underneath, there is a "Suggested tags" section with a row of buttons: "country pop", "classic country", "contemporary country", "americana", "traditional country", and "alternative country". At the bottom, there is a "Lyrics" section with three radio button options: "Custom" (with the subtext "Write your own lyrics"), "Instrumental", and "Auto-generated" (which is selected with a checkmark).

To get started, just type a description of a song you'd like to create. For best results, use descriptive genres, descriptors, and moods.

本日のメニュー

- アイスブレイク
- 音楽の楽しみの歴史的概観
- 音楽情報処理の歴史的概観
- **これからの講義の予定**

スケジュール

1. 本講義
2. [音楽ツール、シーケンサ](#)
3. [インタラクティブ音楽システムと構成手法](#)
4. [楽音合成（シンセサイザ）](#)
5. [流行りの曲を分析してみよう（聞いて学ぶ音楽用語）](#)
6. [流行りの曲を分析してみよう（実践編）](#)
7. [認知的音楽理論](#)
8. [自動・作編曲システムと構成原理](#)
9. [音楽数理情報処理の技術1：ピッチ抽出の諸問題と単音の基本周波数\(\$f_0\$ \)推定](#)
10. [音楽数理情報処理の技術2：複合音の分離・群化](#)
11. [音楽数理情報処理の技術3：音楽パターン認識のための基礎数理（ベイズの定理と最尤推定）](#)
12. [音楽数理情報処理の技術4：音楽パターン認識の実践（メロディ抽出、リズム認識、コードと調性の推定、音源分離）](#)
13. [演奏表現と表情づけ](#)
14. 音楽システム合評会（領域実習Aでの取り組みデモ）