

Confidential

NTT DATA
Trusted Global Innovator

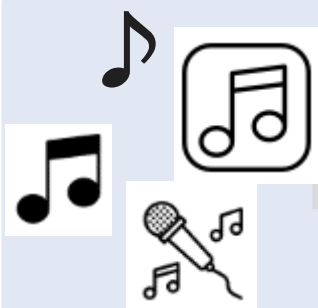
**Billboard Japan × NTTデータ
NeuroAIによる楽曲分析プロジェクト**

2020年9月3日
株式会社NTTデータ
株式会社NTTデータ経営研究所

プロジェクト全体像と研究開発範囲

リリースされている楽曲とその売上やストリーミング数などの各種データを活用し、楽曲特徴からユーザーの脳を含めた反応を予測することにより、レコード会社等に対し、新たな有用指標の提供ができるかの事前検証を双方持ち出しの共同研究を行う。

Virtual Listeners Brain



Virtual Listen!



楽曲パワーを予測

- ✓ ダウンロード数
- ✓ CD売上枚数
- ✓ ストリーミング数
- ✓ ... 等

※billboard様で発表しているデータ項目



【保有】
・楽曲を聴いた時の
脳活動予測技術
・データ解析AI技術

×

JAPAN
billboard

【保有】
・過去数年分の楽曲ランクデータ
・楽曲

本共同研究開発プロジェクト範囲

Customer & Case

指標予測

マーケティング

レコード会社

★プロモーション戦略/売上予測/マーケティング

アーティスト事務所

★アーティストコンサル・マーケティング

CDショップ

★在庫管理 / デジタルマーケティング活用

プロモーション

映画制作配信会社

★タイアップ選曲

CM広告主

★CM選曲

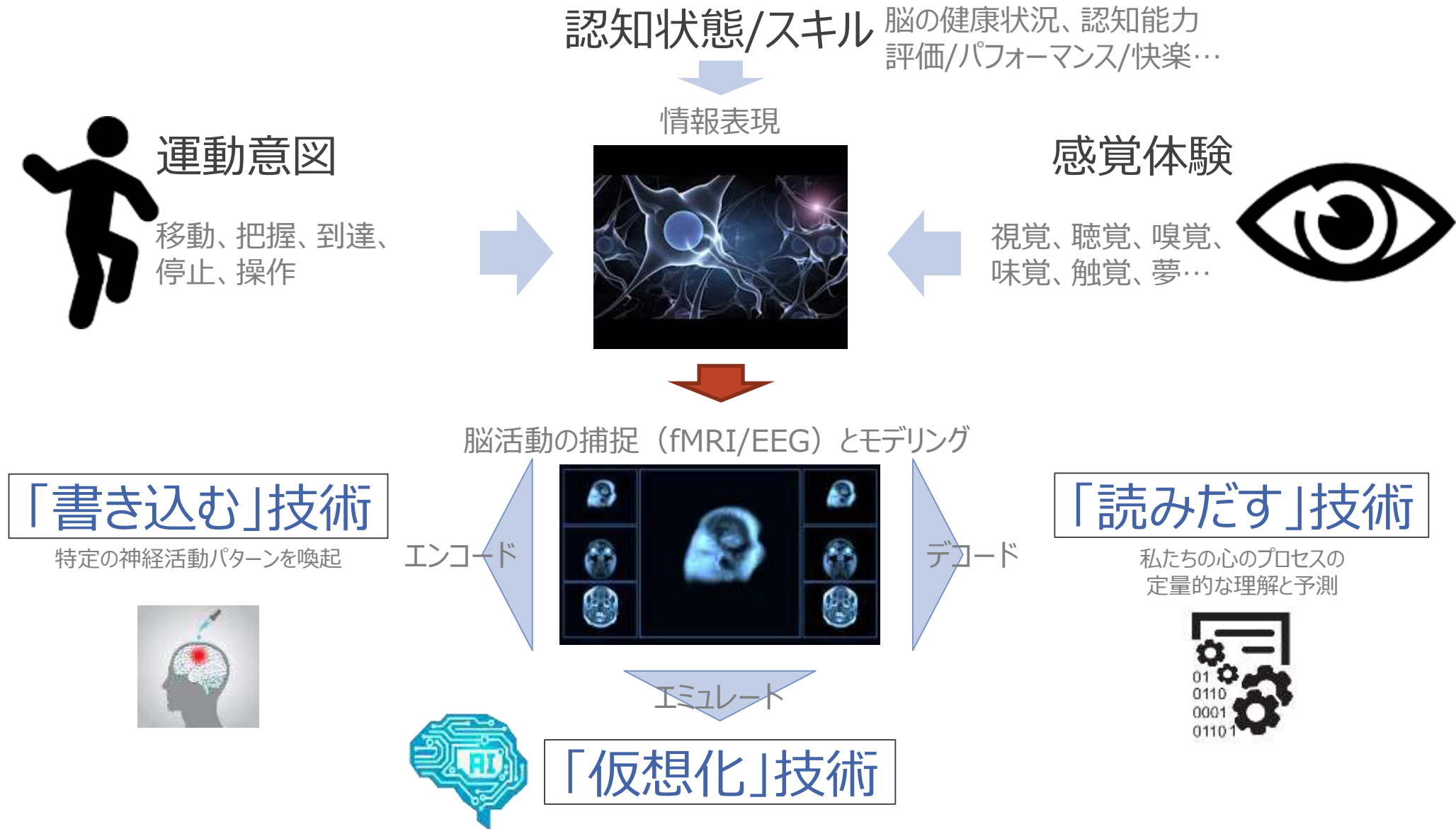
楽曲トレンドのコンテンツ

TV・ラジオ・Youtube等のコンテンツに活用

★トレンド予測・評価

ニューロテクノロジー & NeuroAI

「ニューロテクノロジー」



音楽に「気持ちよさ」を感じているとき脳の線条体でドーパミンが放出されている

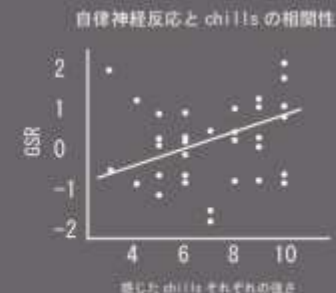


音楽を聴いて感動のあまり戦慄する体験 (= Chills) をしている時、
脳の中の報酬系と呼ばれる領域では、ドーパミンが放出されていることがわかっている。(PET 実験)

また、感動を「予期している」時と「体験」している時では活動する場所が違っても分かった。(fMRI 実験)

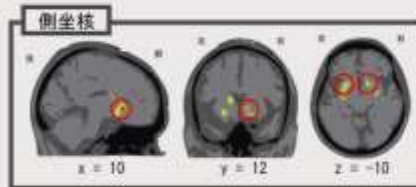
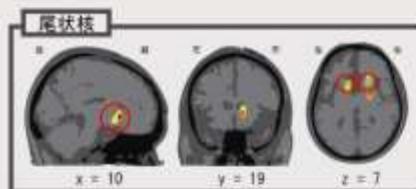
♪ 音楽鑑賞時の情動的反応の測定 ♪

被験者が感じた chills それぞれの強さは、全ての生理計測項目と有意な相関を示し、chills の強さについての主観報告が音楽による感情の高ぶりの指標となることが示された。



PET 実験

被験者に、chills を感じる好ましい音楽と中立的な音楽の 2 つを試聴させ、PET 計測を行った。
chills を感じる音楽を試聴すると、**尾状核・被核・側坐核**でドーパミンが放出されていることがわかった。

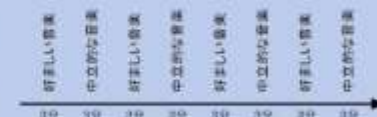


fMRI 実験

ドーパミン放出が起こった脳領域の活動量の時間的変化を fMRI 計測を利用して調べた。

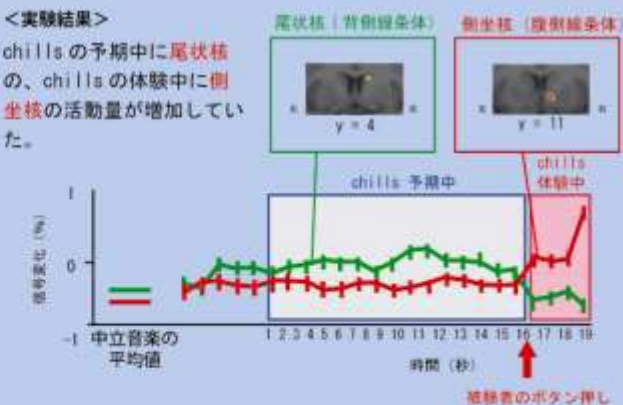
<実験構成>

好ましい音楽と中立的な音楽を 3 分ずつ交互に聴き、chills を感じたタイミングをボタンを押して報告する。



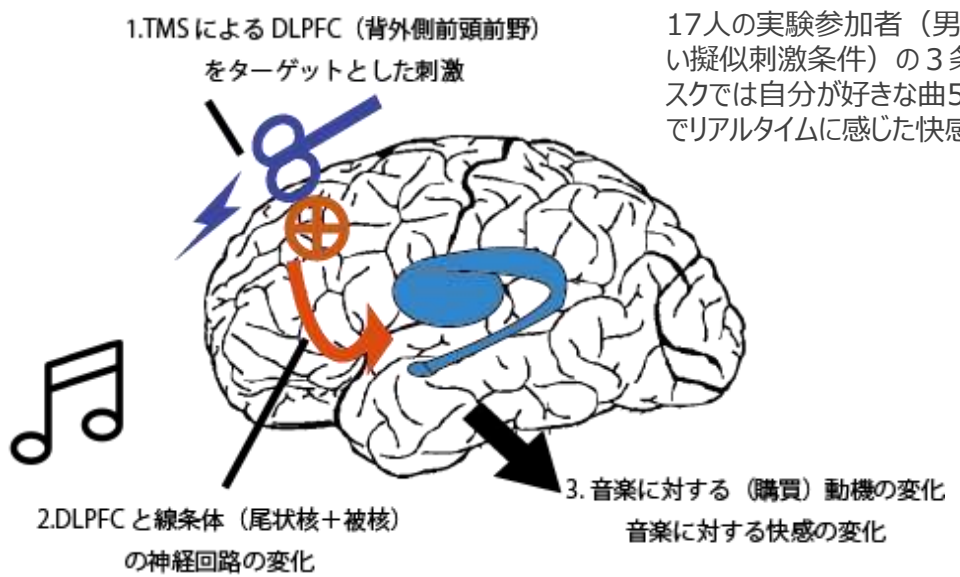
<実験結果>

chills の予期中に**尾状核**の、chills の体験中に**側坐核**の活動量が増加していた。



Valorie N Salimpoor 氏らの研究論文 (2011) を参考に作成

外部からの脳情報操作により、「より気持ちよく」音楽を聴くことが可能



17人の実験参加者（男性9名、平均年齢25.3歳）に対して、3日間に分けて、iTBSとcTBS、そしてシャム（音だけして実際には刺激をしない擬似刺激条件）の3条件をそれぞれの日に行った。参加者は全条件で計600回の刺激を受けた後、音楽を聴くタスクを行った。音楽を聴くタスクでは自分が好きな曲5曲と実験者が選んだ10曲を聴き、「1：ニュートラル、2：小さな快感、3：大きな快感、4：ゾクゾクする」の4段階でリアルタイムに感じた快感を報告

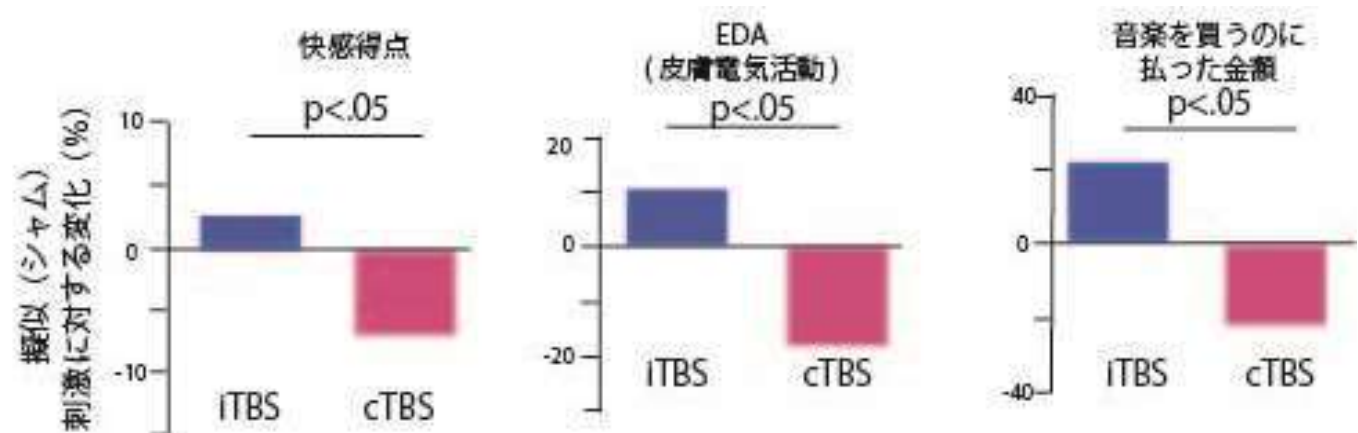


Figure 2 実験のスキーム



【出所】ミユキ技研社HP
<http://www.miyuki-net.co.jp/jp/product/researchEquipment.shtml>

Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2017). Modulating musical reward sensitivity up and down with transcranial magnetic stimulation. Nature Human Behaviour.

概要

- 音楽が、脳活動(横側頭回と上側頭回の位相同期)・生理的な反応(ガルバニック皮膚反応*・心拍)にどのような影響を与えるのかを検証。
*皮膚表面の発汗度合いを測定し高揚状態を測る
- VARモデル(多変量自己回帰モデル)は人の無意識な反応の予測に効果的であることが分かった。

分析方法

- 使用した音楽：ストリーミングサイト・SNS上でhappyとsadのタグが付いた曲の中から3曲に絞った。
「Fyrstathe」悲しい気持ちになる短い歌(256秒)、
「Discovery of the Camps」悲しい気持ちになる長めの歌(515秒)、「Race Against The Sunset」ハッピーな気持ちになる歌(168秒)
 - 被験者40人(平均24歳)：3曲を聴き fMRI装置で脳を撮影。
 - 被験者60人(平均19歳)：3曲を聞く間ガルバニック皮膚反応と心拍を測定。また聴いた時の感情変化を10段階で評価させた。
- ①位相同期の測定→脳の横側頭回とSTGの時間的ダイナミクスを評価：fMRI位相同期ツールで計測
 - ②ガルバニック皮膚反応：SCRソフトウェアで測定。カーネル密度推定(KDE)関数を用いて最適な帯域幅(=8秒)を求め、被験者間のSCRプロットを同一のものに近づけた。
 - ③心拍を計測し2つのディストリビューション間の最小のKullback-Liebler情報量をもたらす帯域幅を持つKDEを作成し帯域幅0.5秒を求めた。また心拍間隔の標準偏差を平均して心拍変動を計算した。

分析結果

- VARモデルは、横側頭回(両側)と上側頭回(両側)の位相同期を予測するのに最適であった。
 - VARモデルはガルバニック皮膚反応のSCRと心拍を予測するのに最適であった。
- VARモデルは音楽に対する無意識な反応の予測に有効。

横側頭回

	Left			Right		
	SS	SL	H	SS	SL	H
AVG	.335	.253	.277	-.269	.253	.199
LASSO-T	.307	.230*	.262	-.269	.248	.194
Ridge-T	.330	.230*	.271	.270	.250	.196
LASSO-VAR	.289	.258	.223*	-.287	.212*	-.173*
Ridge-VAR	.281	.305	.307	-.286	.243	-.176
Features	40,41	40,43	4	N/A	40,43	37,40

上側頭回

	Left			Right		
	SS	SL	H	SS	SL	H
AVG	.311	.296	.383	.339	.311	.416
LASSO-T	.310	.288	.383	.324	.282	.408
Ridge-T	.310	.287	.382	.327	.324	.416
LASSO-VAR	.311	.268*	.359*	.291*	-.273*	-.379*
Ridge-VAR	.304	.276	.304	.300	.279	-.379
Features	N/A	40,43	4	43,51	40,43	4

ガルバニック皮膚反応のSCR

	Sad Short	Sad Long	Happy
	AVG	0.164	0.159
LASSO-T	0.153	0.150	0.107
Ridge-T	0.142	0.137	0.105
LASSO-VAR	0.126*	0.138	0.098*
Ridge-VAR	0.128	0.102*	0.136
TPA	1733	1440	1197
Features	43,63,49	40	61

心拍

	KDE			HRV		
	SS	SL	H	SS	SL	H
AVG	.253	.186	.541	.924	.888	.918
LASSO-T	.230	.175	.536	.925	.888	.919
Ridge-T	.240	.175	.545	.923	.888	.914
LASSO-VAR	.240	.186	.510*	.927	.890	.915
Ridge-VAR	.218*	.173*	.531	.928	.891	.914
TPA	20.8	20.7	19.6	42.6	55.8	12.9
Features	41,42	54,61	4	N/A	N/A	N/A

T Greer, M Sachs, et al.(2019) A Multimodal View into Music's Effect on Human Neural, Physiological, and Emotional Experience. MM19 Proceedings of the 27th ACM International Conference on Multimedia Pages 167-17

Table 2: Feature Index
予測に使う音楽の特徴量

Number	Feature Name	Description
1-13	1st-13th MFCC	Timbre
14-26	1st-13th Δ MFCC	Change in timbre
27-39	1st-13th $\Delta\Delta$ MFCC	Change in timbre
40	Pulse Clarity	Strength of beats
41	Brightness	% of high-end frequencies
42	Key Strength	How likely a key is
43	RMS	Loudness
47	Kurtosis	Change in spectrum
48-59	1st-12th Chroma	Strength of C, C#, D..., B
61	Compressibility	Complexity
63	Spectral Flux	Harmonic change

Table 3: STG - Test RMSE
上側頭回の活動予測の精度と選択された特徴量

	Left			Right		
	SS	SL	H	SS	SL	H
AVG	.311	.296	.383	.339	.311	.416
LASSO-T	.310	.288	.383	.324	.282	.408
Ridge-T	.310	.287	.382	.327	.324	.416
LASSO-VAR	.311	.268*	.359*	.291*	.273*	.379*
Ridge-VAR	.364	.270	.364	.380	.279	.379*
Features	N/A	40,43	4	43,51	40,43	4

Table 5: SCR - Test RMSE

S C Rの予測の精度と選択された特徴量

	Sad Short	Sad Long	Happy
AVG	0.164	0.159	0.108
LASSO-T	0.153	0.150	0.107
Ridge-T	0.142	0.137	0.105
LASSO-VAR	0.126*	0.138	0.098*
Ridge-VAR	0.128	0.102*	0.136
TPA	1733	1440	1197
Features	43,63,49	40	61

Table 6: Heart Activity - Test RMSE ($Entries \times 10^{-1}$)
心拍の予測の精度と選択された特徴量

	KDE			HRV		
	SS	SL	H	SS	SL	H
AVG	.253	.186	.541	.924	.888	.918
LASSO-T	.230	.175	.536	.925	.888	.919
Ridge-T	.240	.175	.545	.923	.888	.914
LASSO-VAR	.240	.186	.510*	.927	.890	.915
Ridge-VAR	.218*	.173*	.531	.928	.891	.914
TPA	20.8	20.7	19.6	42.6	55.8	12.9
Features	41,42	54,61	4	N/A	N/A	N/A

コード進行の不確実性と意外性は快感をもたらす & 脳の扁桃体・海馬・聴覚野の活動を予測することができる

実験概要

機械学習モデルを用いて、ヒットしたポップス745曲に使われている8万コードを対象としてリスナーの「予測しにくさ（不確実性）」と「意外性」スコアを算出し、音楽を聴いているときの主観的快感が、コード進行に対するリスナーの予測（期待）と関連しているか、またそれが人間の脳活動にどのように反映されているかを検証。

実験方法

実験1：米国ビルボードにランキングするポップス745曲から選んだ30種のコード進行のいずれかに当てはまる1,039コードを、成人39人に聞かせ、各コードが好みか好みでないか（快感度）を評価させた。

実験2：実験1と同じコードを実験1とは異なる参加者40人にfMRIスキャナーの中で聞かせ、音楽が誘起する感情に関与するとされている脳領域の活動を計測した。

結果

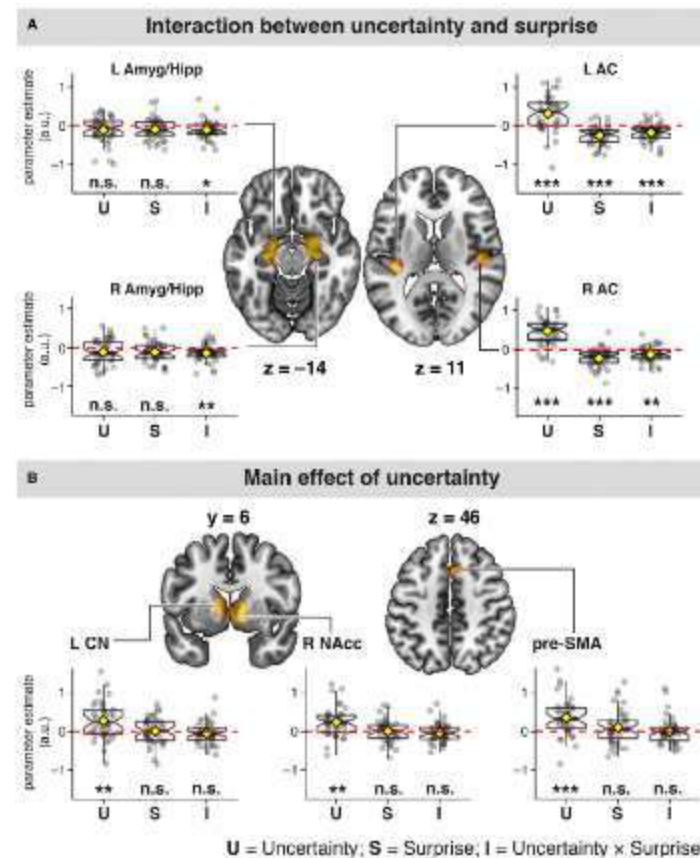
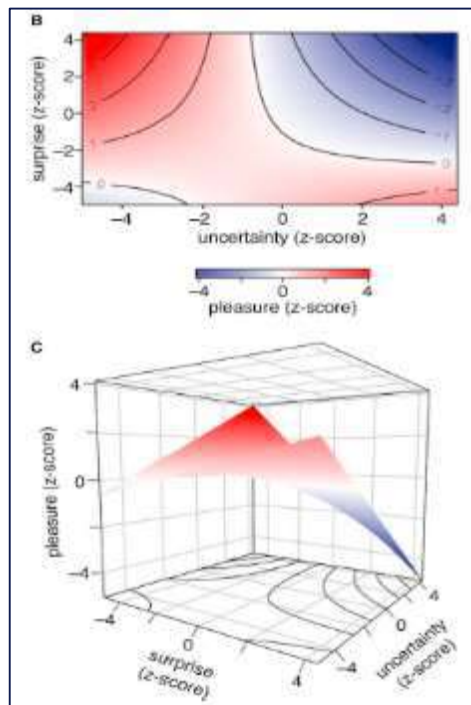
実験1：コードの快感度評価は、事前に統計学習モデルを用いて算出したコード進行の「不確実性」と「意外度」から予測することが可能（＝快感度はリスナーのコード進行予測精度と音色調和に基づく確率の相乗影響を受ける）ことが明らかとなった

- 曲の終盤など、コード進行の予測が容易（不確実性が低い）なとき、「意外」なコード進行が続くと「快感度」評価が高くなる

実験2：

両側扁桃体および両側海馬のBOLD信号と不確実性と意外性間に相関が見られた
不確実性が高いとき、右側坐核（RNACC）、左尾状核（LCN）、前補足運動野（Pre-SMA）の活動が賦活した

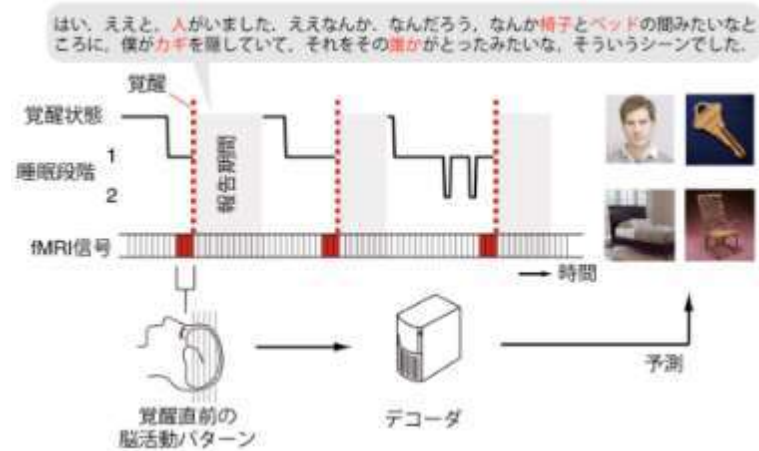
コード進行の「予測しにくさ（不確実性）」と「意外性」が主観的な音楽視聴の快感および扁桃体、海馬、聴覚野の活動に影響を与えていることが明らかとなった



Cheung et al. (2019) Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. Current Biology

感覚体験の解読と再構築

夢



https://youtu.be/inaH_i_TjV4

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, 340(6132), 639-642.

自然な視覚



<https://youtu.be/6Fsh7RK1S2E>

Shinji Nishimoto, An T. Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu & Jack L. Gallant (Current Biology 2011)

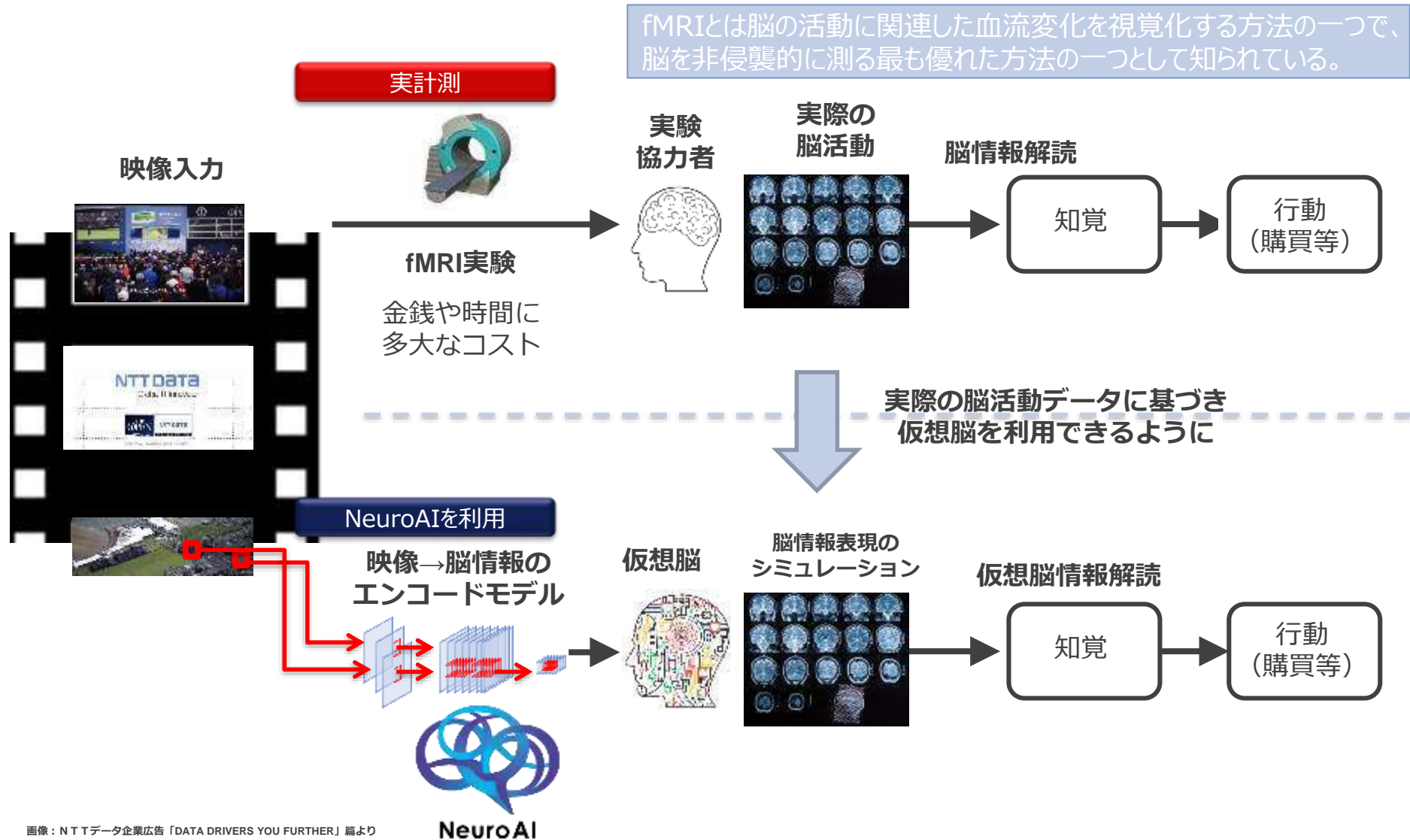
自然言語モデルを用いた脳活動の解読例

The screenshot displays a software interface for NeuroAI. At the top left is the NeuroAI logo, and at the top right is the NTT DATA logo. The interface is divided into two main sections. The left section, titled '視聴中コンテンツ' (Content being watched), shows a video player with the text 'nttdata commercial: The 146th Open 「DATA DRIVES YOU FURTHER」'. The right section, titled '脳活動解読・知覚推定結果' (Brain activity interpretation and perception estimation results), is divided into three columns: '名詞' (Nouns) in red, '動詞' (Verbs) in green, and '形容詞' (Adjectives) in yellow. The content area is currently blank. At the bottom right of the interface is a white silhouette of a human head with a brain icon inside.

- 客観的/主観的な体験意味内容を脳活動から約1万単語について（一定精度で）定量
- NTTデータの動画解析・評価サービスの基盤技術として利用

Nishimoto et al., 2015, 2016a/b *patented*; Nishida and Nishimoto 2017 *NeuroImage*

NeuroAI AIを利用した脳の仮想化



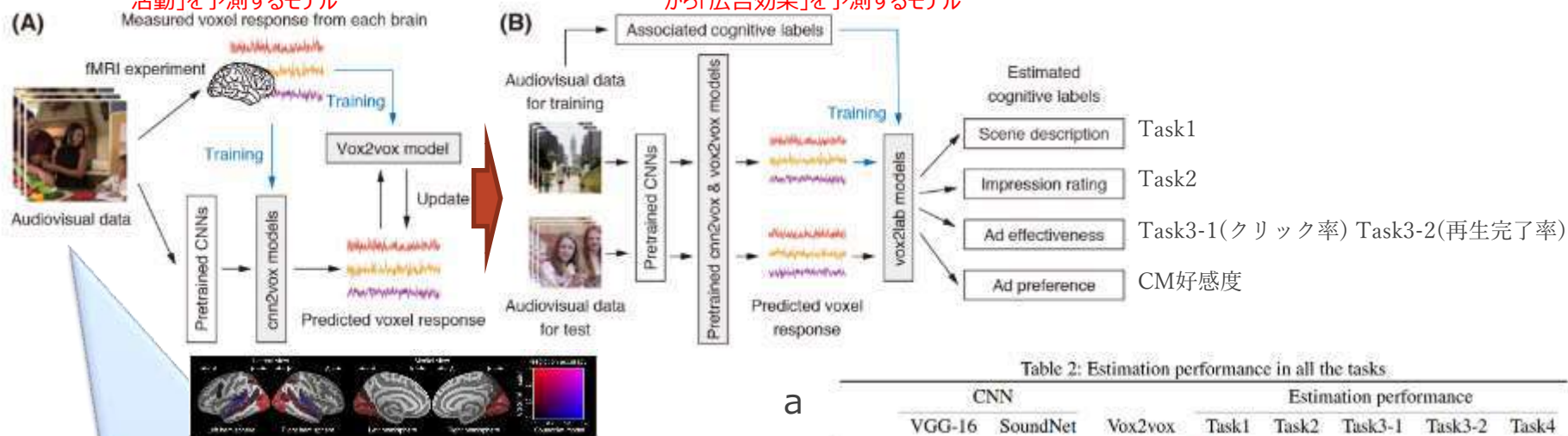
画像：NTTデータ企業広告「DATA DRIVERS YOU FURTHER」篇より

NTTデータは神経科学者と共に人間の脳を仮想化することに成功 人工知能の最高峰 A A A I (2020) で採択された

- ① fMRI という脳スキャナ内で動画広告を大量に視聴させ脳活動を記録
 - ② 深層学習を利用し映像と音声の情報から、脳活動を予測（仮想脳の構築）
 - ③ 仮想脳に動画を見せ、仮想的な脳活動情報を取得し、それを利用して動画広告の効果を予測
- 広告に対する好意は仮想脳を利用することで最新の深層学習等を凌駕する制度で予測することが可能になった

STEP1 「先行する脳活動」と「映像・音声」から「脳活動」を予測するモデル

STEP2 「予測脳活動」から「広告効果」を予測するモデル



WEBCM368本(2015-2018) TVCM2452本(2011-2017)

Table 1: The number of movie clips belonging to individual product/service categories

Categories	Web ad movies	TV ad movies
Electronic & Precision	4	50
Audiovisual	5	6
Appliance	16	23
Car	31	145
Food & Confectionery	7	369
Beverage & Alcoholic drink	20	236
Medical & Health	35	156
Cosmetics	49	85
Sundries & Home equipment	10	254
Garment/apparel	9	43
Entertainment	42	237
Media & Education	41	82
Distribution & Retailer	12	112
Communication & Service	35	328
House & Construction	9	90
Finance	9	145
Enterprise, Public service, & Others	34	91

Table 2: Estimation performance in all the tasks

a	CNN		Estimation performance					
	VGG-16	SoundNet	Vox2vox	Task1	Task2	Task3-1	Task3-2	Task4
BTL	✓			0.546	0.498	0.505	0.440	0.354
	✓		✓	0.546	0.500	0.502	0.444	0.360
		✓		0.146	0.234	0.094	0.223	0.244
	✓	✓	✓	0.146	0.235	0.093	0.227	0.252
TL	✓			0.536	0.516	0.446	0.376	0.381
	✓		✓	0.536	0.517	0.447	0.375	
		✓						
	✓	✓	✓					
BD								
	✓		Single	0.504	0.450	0.478	0.217	0.320
	✓		Multiple	0.537	0.482	0.510	0.288	0.334
		✓	Single	0.127	0.208	0.067	0.134	0.176
BD			Multiple	0.143	0.232	0.065	0.166	0.213
	✓	✓	Single	0.504	0.471	0.475	0.245	0.336
	✓	Multiple	0.535	0.501	0.502	0.303	0.365	
				0.396	0.349	0.287	0.468	-

仮想脳利用



機械学習のみ

実際の脳情報解読

Nishida S, Nakano Y, Blanc, A, Maeda N, Kado M, Nishimoto S. Brain-mediated Transfer Learning of Convolutional Neural Networks. Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, 2020.

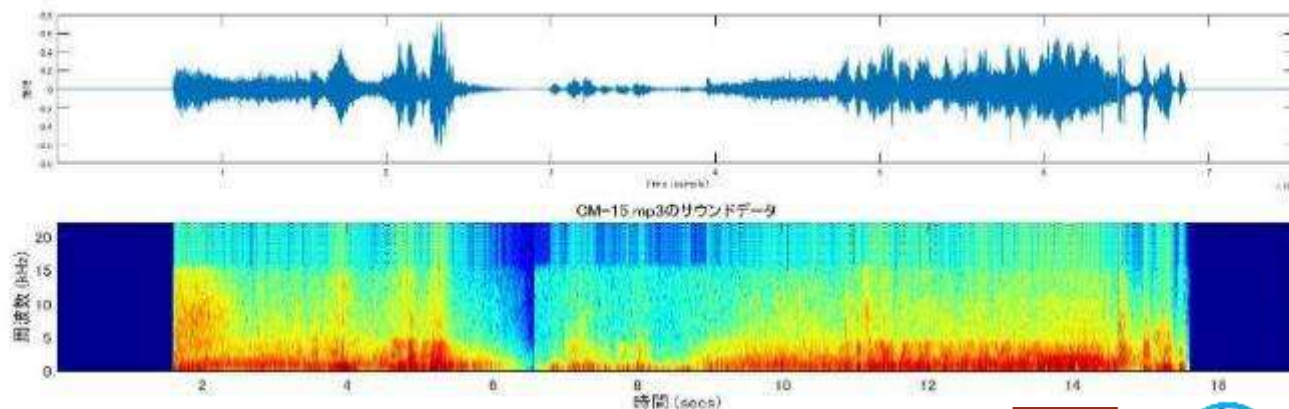
楽曲の脳情報化

対象楽曲

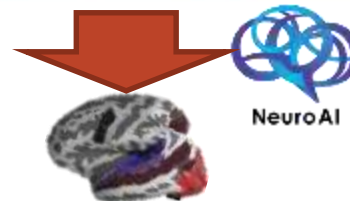
(2017/1-2020/5 で B B J 年間チャート300位以内にいずれかの指標が入った曲)

- 対象楽曲データ : **2185曲**
(2017上半期-2020上半期(2020/05まで)の各半期いずれかのチャートの100位以内にランクインしたもの)
- チャートデータ (週次) : **2047曲**
- ※ 2017上半期(2016/12/5週)一直近(2020/7/26週までを対象)

★楽曲特徴の抽出



- ① NeuroAI で特徴抽出 (音楽→脳活動)
- ② MFCC で特徴抽出 (音色の特徴)



総合順位	総合ポイント	Radio	CD	DL	LU	TW	MV	STR	Karaoke
総合ポイントの順位	各指標の総合値	ラジオでの再生数	CDの販売数	ダウンロード数	ルックアップ数 (PC取り込み数)	Twitterでの投稿数	MVの再生数 (Youtube)	ストリーミング再生数	カラオケでの歌唱数

「音」の特徴を定量的に取り出す手法

confidential

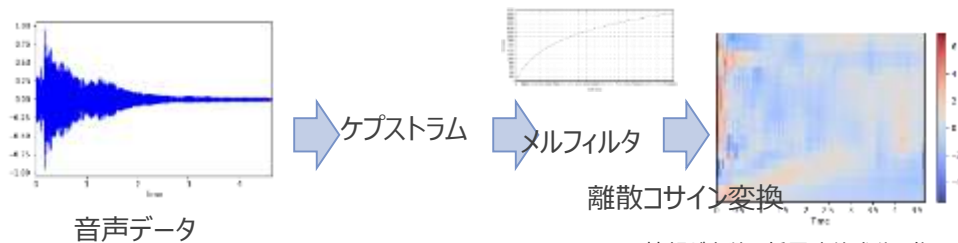
機械学習に関しては以下の2種類で実装

MFCC (メル周波数ケプストラム係数) による特徴抽出

<https://qiita.com/martin-d28jp-love/items/34161f2facb80edd999f>

メル尺度：音高の知覚的尺度。メル尺度の差が同じであれば、人間が感じる音高の差が同じになることを意図。可聴域の下限に近い音は高め、上限に近い音は低めに聞こえることを考慮。

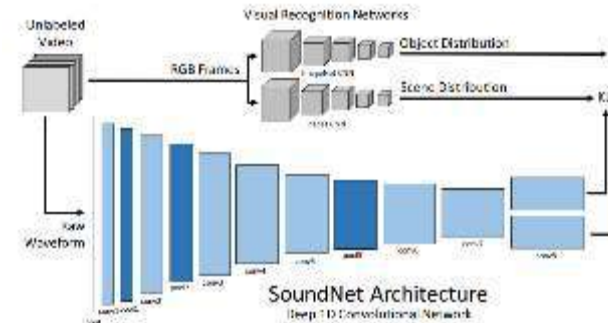
ケプストラム：音声信号をフーリエ変換して周波数スペクトルに変換した後、その対数を取ったものを逆フーリエ変換して時間空間に戻す。対数スペクトル領域で見ることによってスペクトル微細構造（細かい動き）と包絡構造（なだらかな変化）が分離できるメリットがある



情報が少数の低周波数成分に集中するため、低次項から適当な数で打ち切ってしまうことで情報をなるべく残した状態で次元圧縮

SoundNetによる特徴抽出 (今回はなし)

MITが開発。自然音のデータを利用して特徴認識(クラス分類)を実現。学習に際しては、ラベルを学習済みの物体/シーン認識のモデルから取得し、それを音声の教師ラベルにするという手法。これで教師有りより高い性能を出せた。



Hearing the Hidden Representation

Although the network is trained without ground truth labels, it learns rich sound features. We visualize these features by finding images that maximize the activation of each unit. Check the images below to learn what sounds activate that unit. Turn on your speakers. You will hear the top 5 sounds that activate that unit.

Visualizing conv7

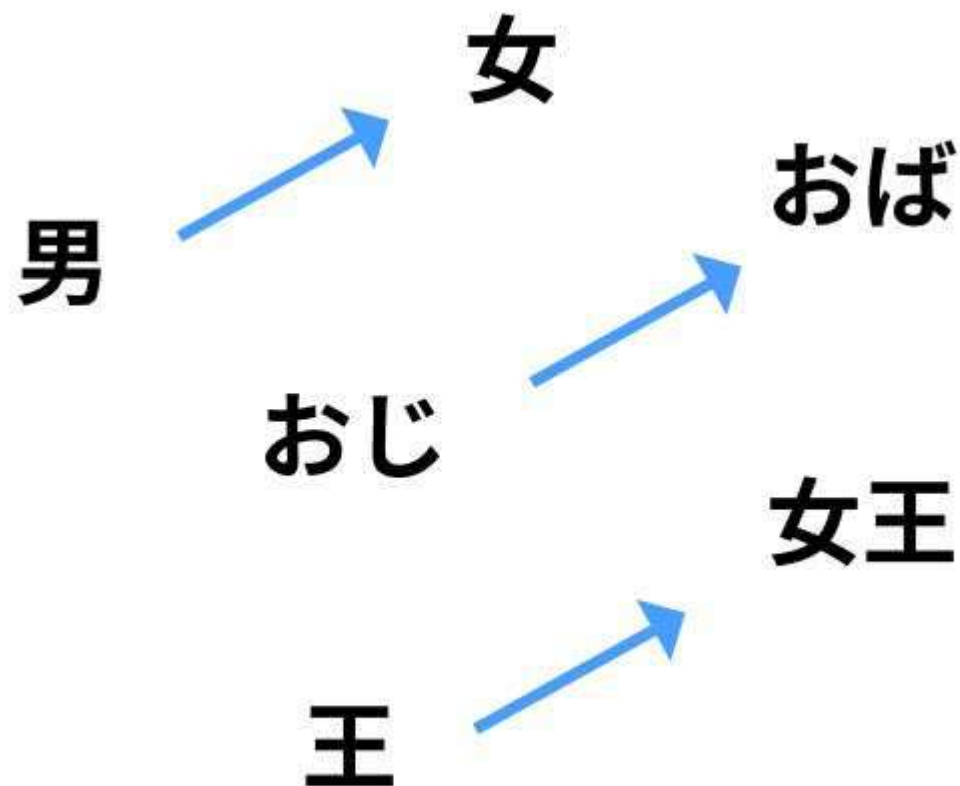
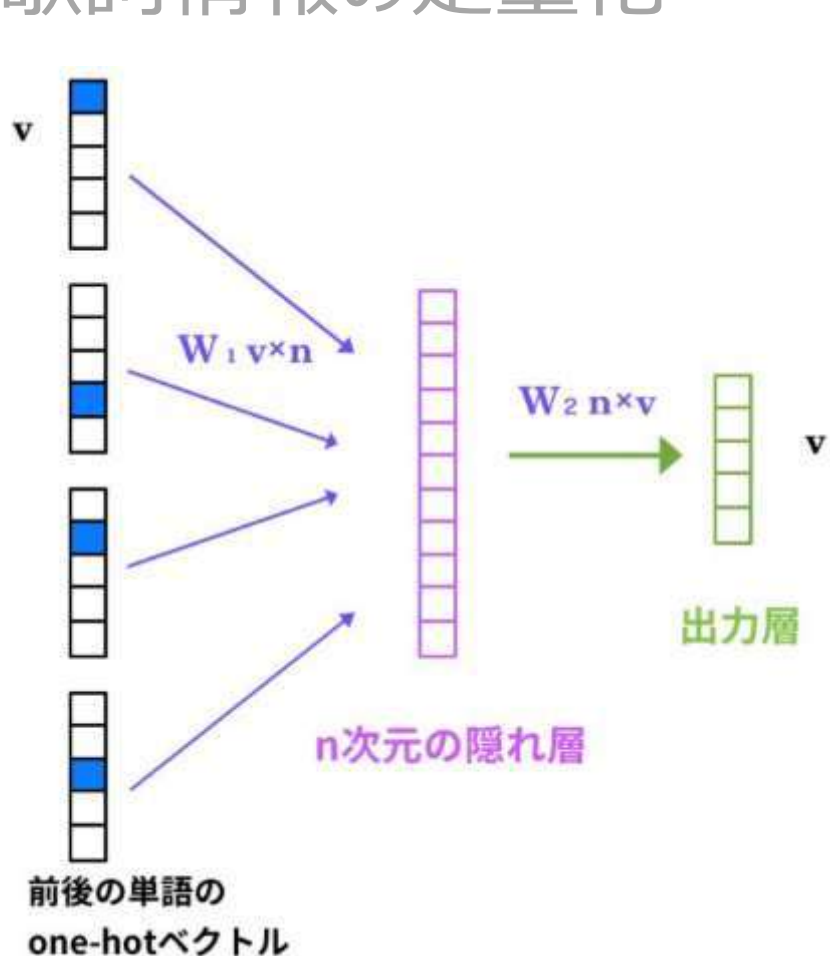
We visualize conv7 as the deep layer at the network. Interestingly, these conv7 features are the network's strongest detectors for high-level concepts that emerge automatically. Note the images are shown only for visual task purposes, and will lose in comparing the sounds.



Method	Accuracy on	
	ESC-50	ESC-10
SVM-MFCC [28]	39.6%	67.5%
Convolutional Autoencoder	39.9%	74.3%
Random Forest [28]	44.3%	72.7%
Piczak ConvNet [27]	64.5%	81.0%
SoundNet	74.2%	92.2%
Human Performance [28]	81.3%	95.7%

<http://soundnet.csail.mit.edu/>

歌詞情報の定量化



今回、歌詞行情報はUta-Netを利用

n=930/2185曲の歌詞の取得に成功。形態素分解と学習済みの分散表現を利用 ⇒曲全体で平均化したものを歌詞特徴ベクトルとした (300D)

P. Bojanowski*, E. Grave*, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information
https://deepage.net/bigdata/machine_learning/2016/08/28/fast_text_facebook.html

コード進行の特徴量の計算

$e_{i-4} \ e_{i-3} \ e_{i-2} \ e_{i-1} \ e_i$
B \flat → B \flat → F → Dm → C7sus4 → F → Am7

コード進行パターンA

Surprise=シャノン情報量

→特定のコードがどれくらい過去にない真新しいものか

$$I(e_i) = -\log_2 p(e_i | e_{i-(n-1)}, \dots, e_{i-1})$$

Surprise

コード進行パターンAのあとコード e_i が来る各確率

Uncertainty=不確実性・情報エントロピー・平均情報量

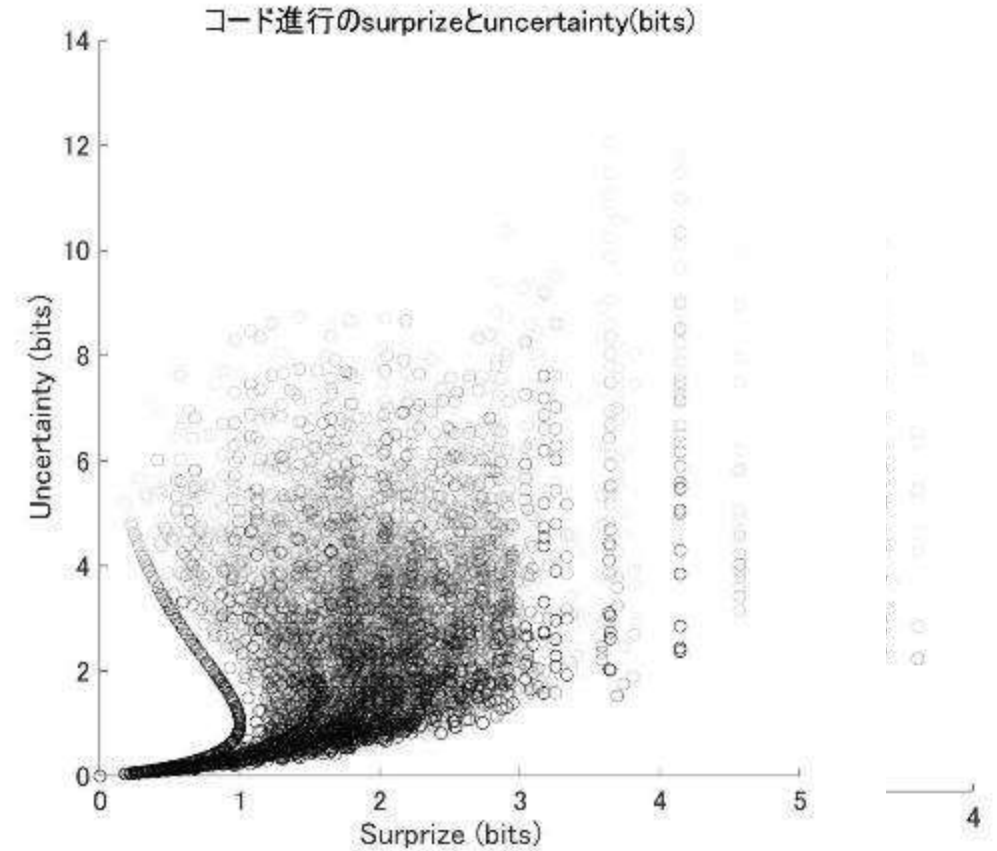
→特定のコードパターンが来た時の次のコード e_i の予測の難しさ

$$H(e_i) = -\sum_{e \in S} p(e_i = e | e_{i-(n-1)}, \dots, e_{i-1}) \log_2 p(e_i = e | e_{i-(n-1)}, \dots, e_{i-1})$$

Uncertainty

全楽曲に含まれるコード

直前のパターンA → 全コードeの確率 × そのSurprise情報量



今回、コード進行情報はU-FRET (<https://www.ufret.jp/>) からスクレイピングできた

n=842/2185曲を対象とし、曲の冒頭のコードをキーとしてCメジャーに調整。

n=5(曲の冒頭部分はn=1~4)とした

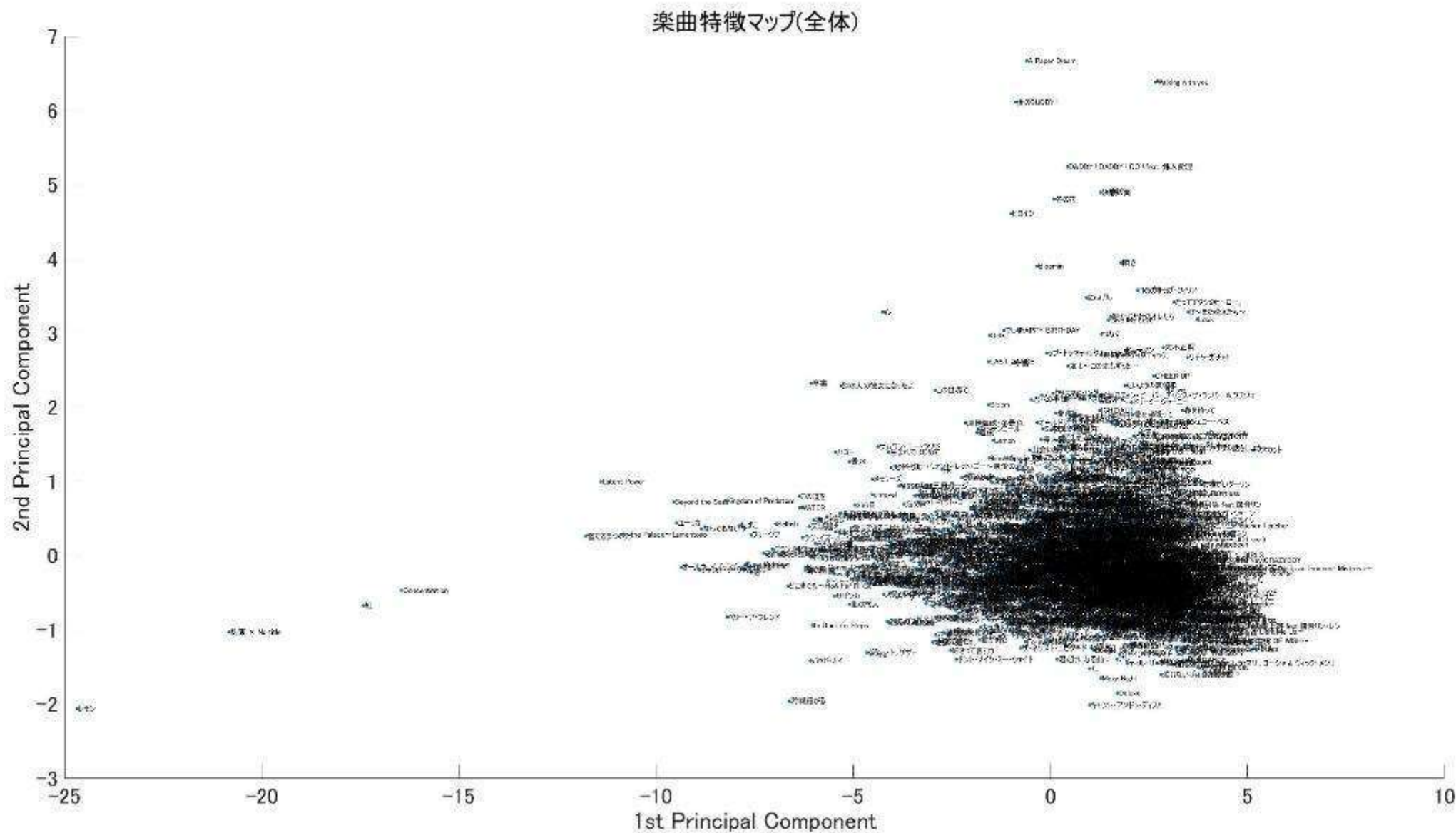
Jeung et al. (2019) Uncertainty and surprise jointly predict musical pleasure and amygdala, hippocampus, and auditory cortex activity. Current Biology

できるようになったこと①-1 楽曲特徴の脳活動 / 歌詞 / コード進行か らの可視化



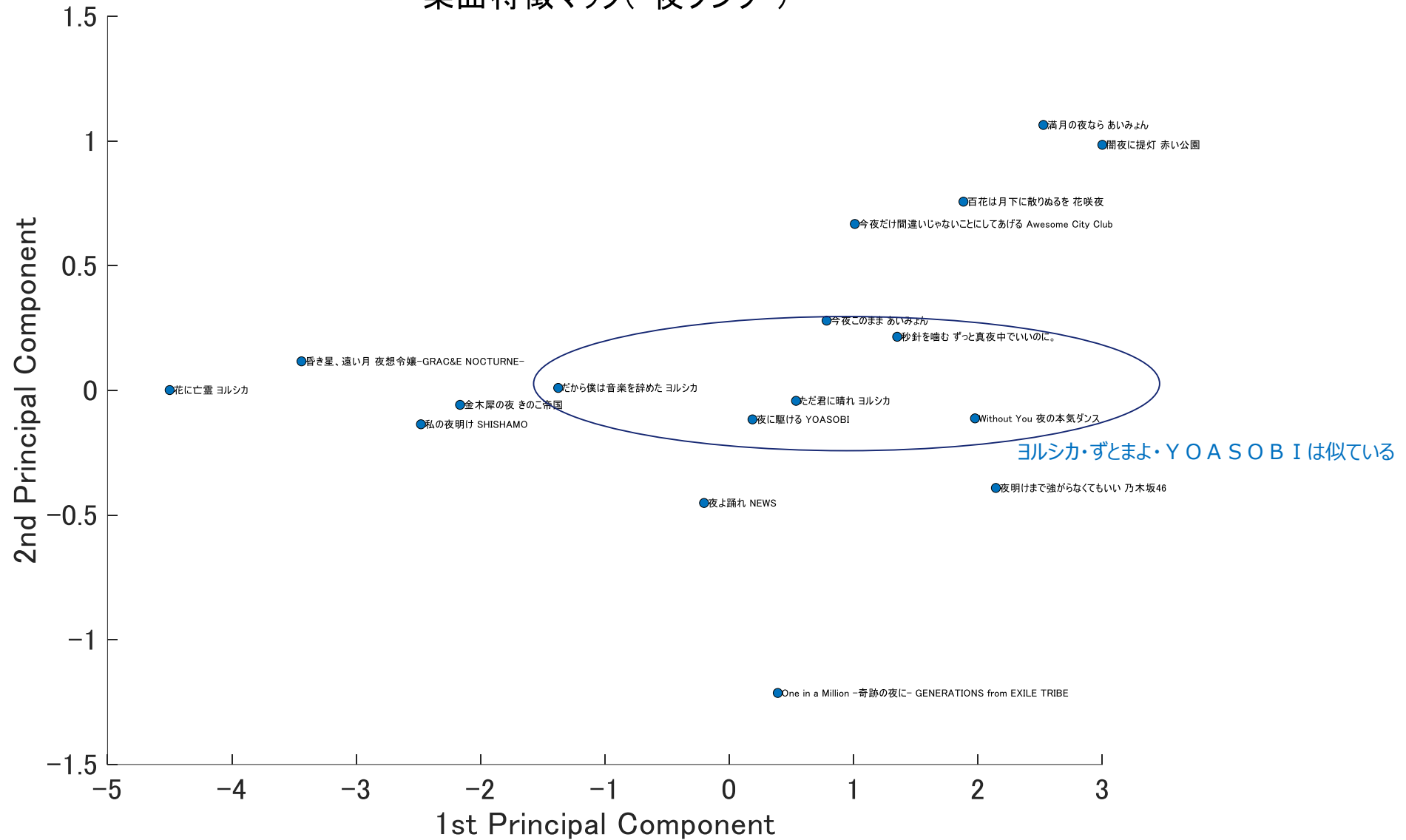
特徴抽出結果 = 曲ごとの特徴ベクトルを2次元に圧縮

→脳活動レベルで似たような曲を評価できるように



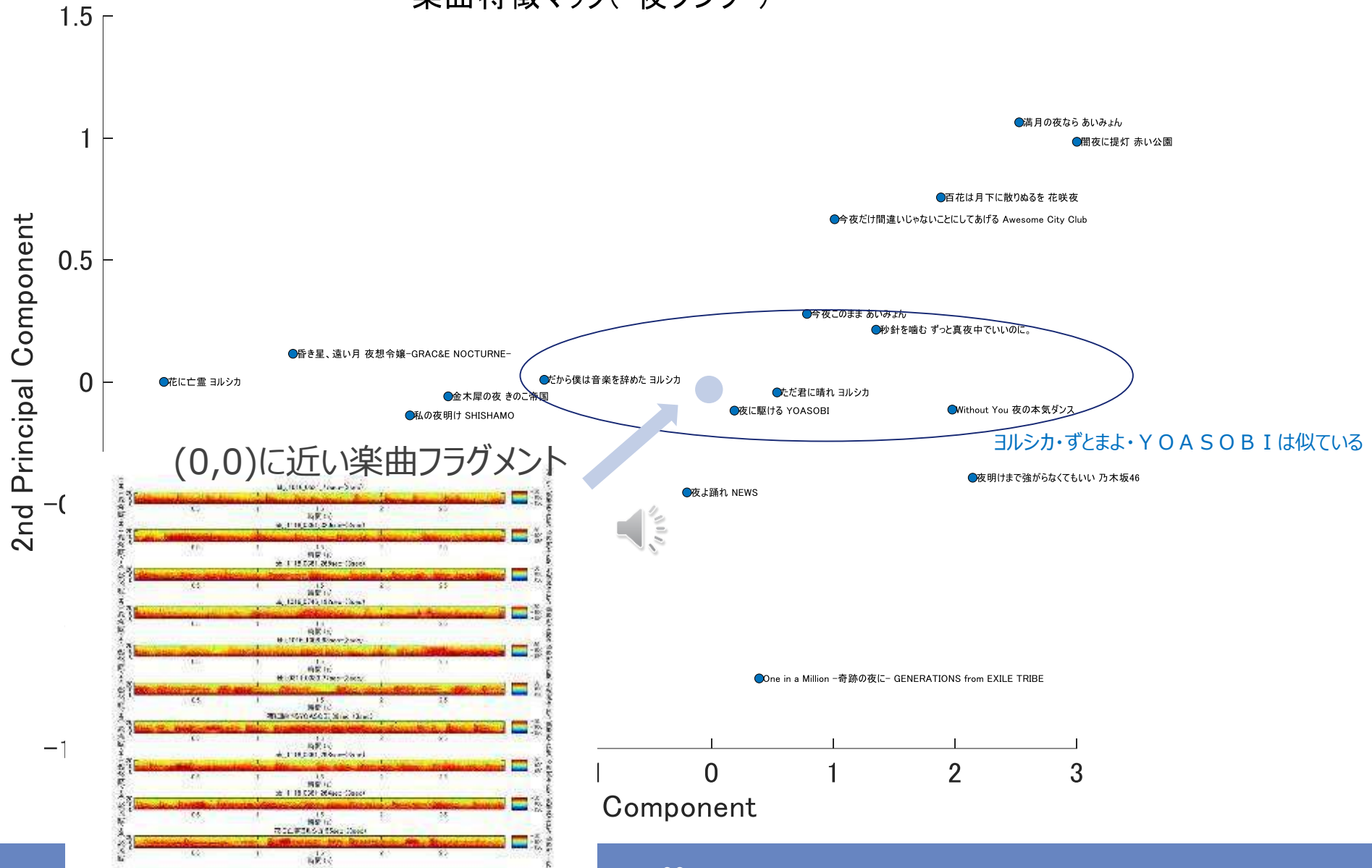
特定の楽曲を絞って可視化することも可能

楽曲特徴マップ(“夜ソング”)

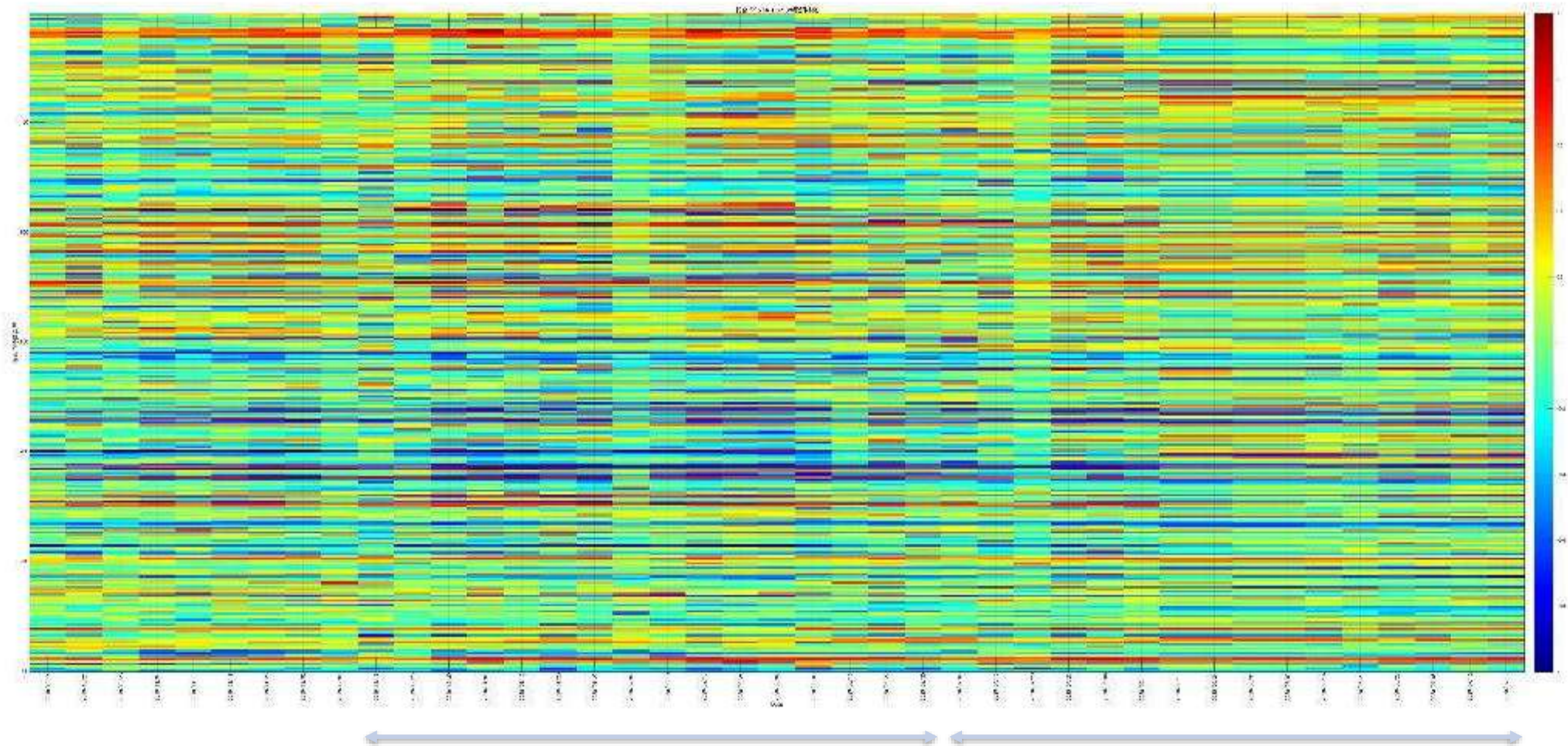


特定の楽曲を絞って可視化することも可能

楽曲特徴マップ(“夜ソング”)



歌詞の特徴の変遷 (2019/10/14-2020/7/19)



コロナ前と後でヒットソングの歌詞にはどんな特徴の差が？

総合ポイントで上位（1～10位）だった曲の頻出歌詞単語
 コロナ前（2019/12-2020/3）とコロナ後（2020/4-2020/7）の比較
 （差が大きかった単語を大きさで示す）

コロナ前（2019/12-2020/3）



コロナ後（2020/4-2020/7）



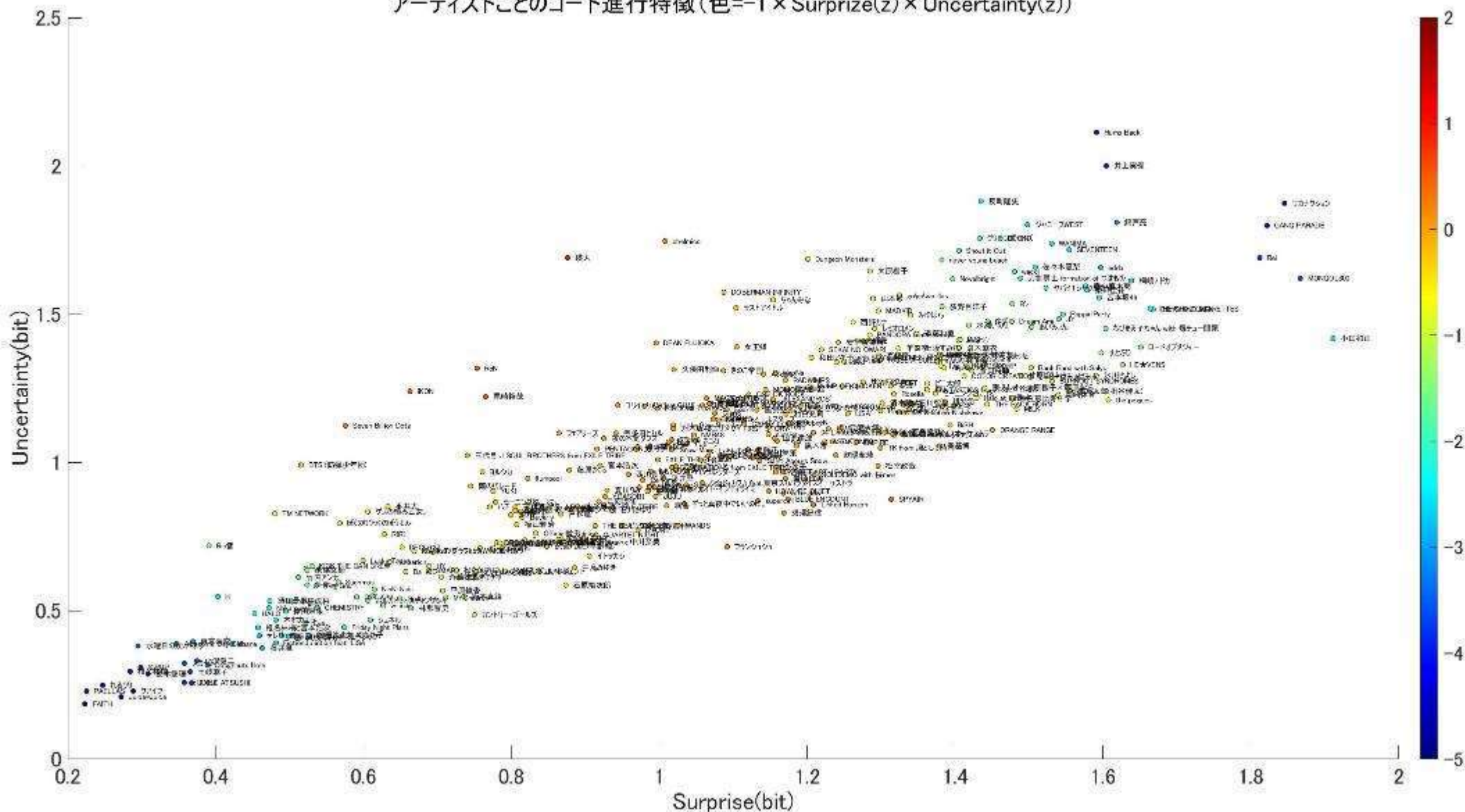
- ・「君」と「僕」の世界に閉じる
- ・「夜」が増える
- ・「夢」「いい」などのポジティブな言葉が減る
- ・「忘れる」「いつか」「きっと」など諦観的な言葉が増える
 →内向的な歌詞の嗜好？

アーティストごとのコード進行特徴

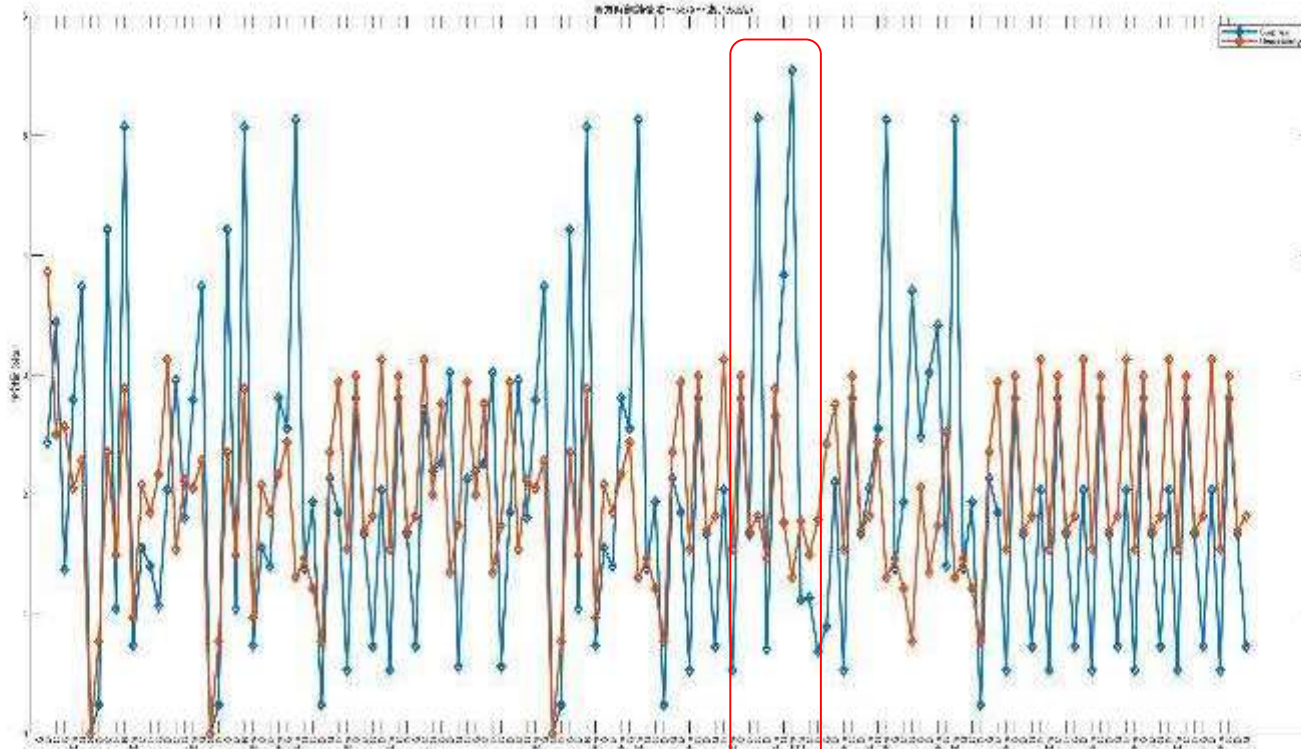
先行研究で Surprise と Uncertainty はどちらかが高く、どちらかが低い曲が心地よいと評価されている

→ 標準得点化し掛け合わせたものをマイナスにすることでその傾向を可視化

アーティストごとのコード進行特徴 (色 = $-1 \times \text{Surprise}(z) \times \text{Uncertainty}(z)$)

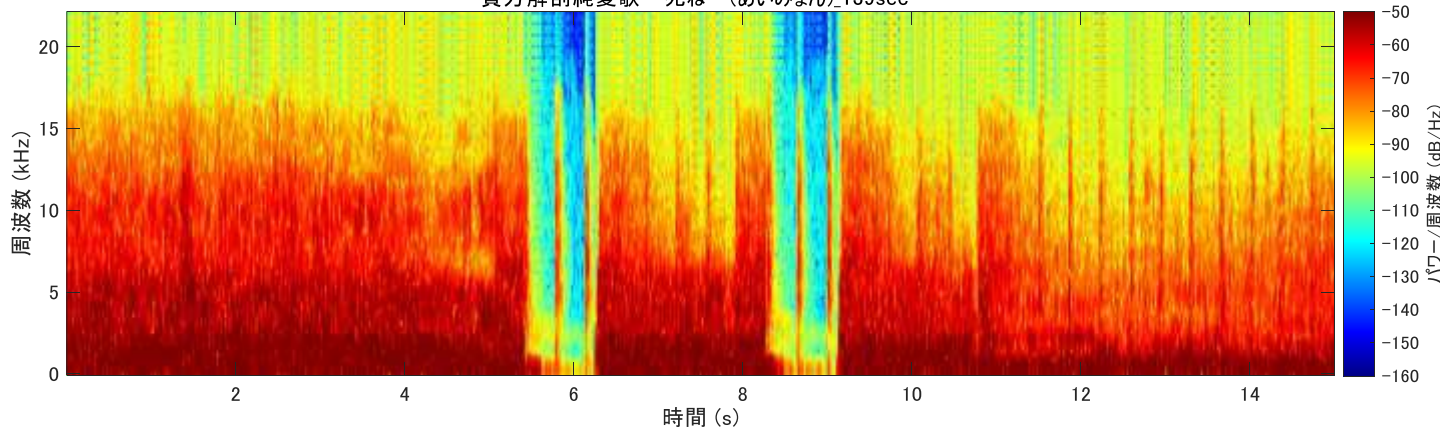


コード進行の
 Surprise
 Uncertainty
 が高い例
 貴方解剖純愛歌～
 死ね～ (あいみょん)



貴方解剖純愛歌～死ね～(あいみょん) 139sec-

間奏 C->D->Em->G->Am の Surprise が大きい

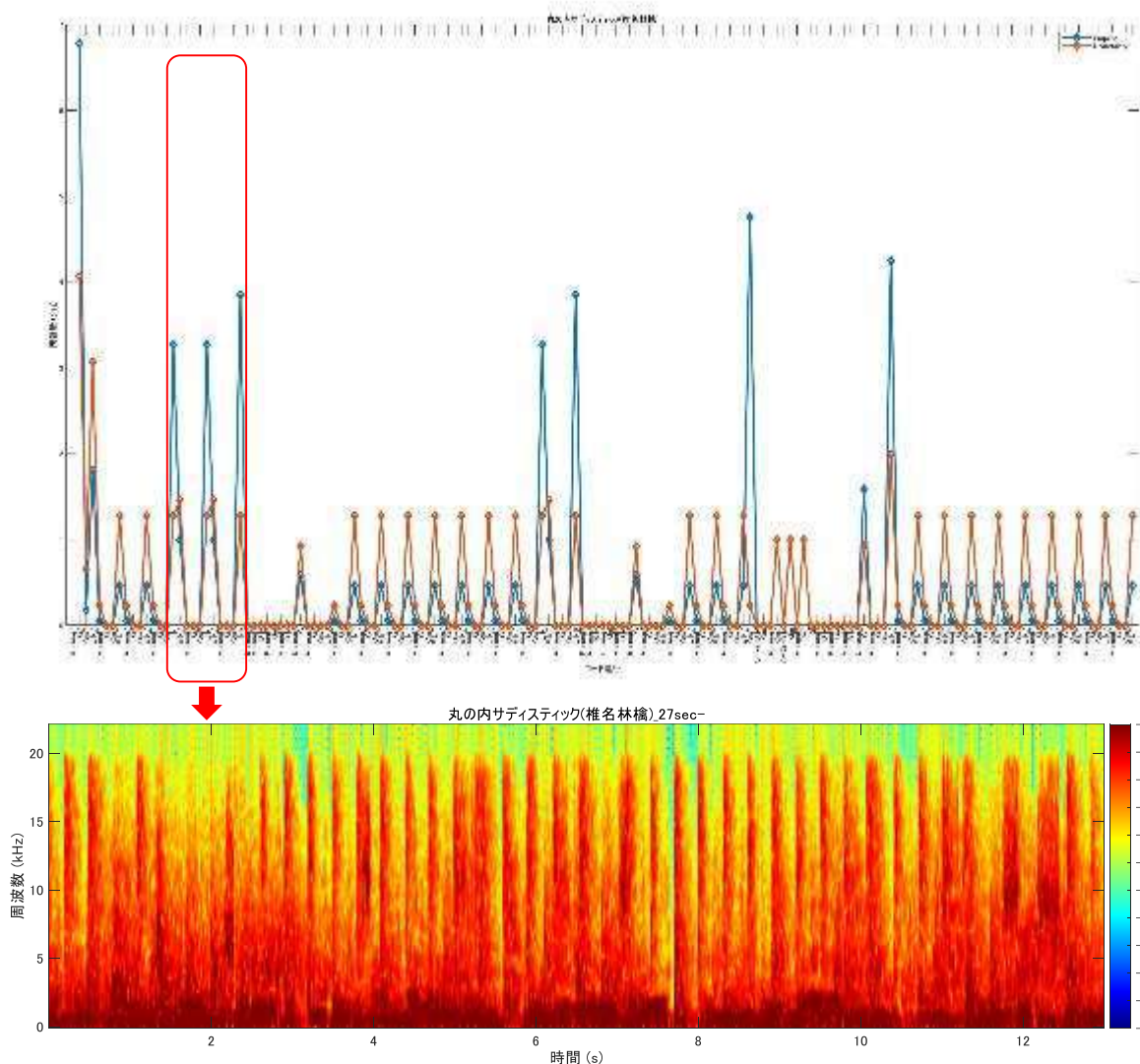


C D C D Em G Am Bm C D
 愛して いるのに

→ 珍しいコード進行



コード進行の
Uncertainty
が低い例
丸の内サディスティック
(椎名林檎)



→全体的にUncertaintyが少ない
→このコードが来たら次はこうと予測できる=こうしたコード進行を使う人が他に少ないオリジナリティが高い曲とも解釈



E \flat E \flat 7 A \flat maj7 G7 Cm7 E \flat E \flat 7 A \flat maj7 G7 Cm7
報酬は入社後平行線で東京は愛せど何も無い

できるようになったこと①-2 脳活動の類似度からプレイリスト

