

キャラクターをより魅力的に！ ゲーム向けリップシンクミドルウェア

音声・映像を専門としたミドルウェア開発会社



1990年 人工知能・CD・音声・映像技術の研究開発

1995年 セガサターン用ADX

1998年 CRI Clipper SDK Ver.1.00 をリリース

2001年 マルチプラットフォーム展開を開始

2011年 iOS/Android版 CRIWARE

2012年 Unity 対応

2013年 Unreal Engine 4対応

2016年 ブラウザ版ADX2 リリース

2019年 STADIA 対応

採用タイトル数 5,000 を突破

CRIのミドルウェア群の総称



サウンド演出

CRI ADX2



ムービー演出

CRI Sofdec2



ファイルパッキング

ファイルマジックPRO



リップシンク

ADX LipSync

ADX[®] LipSync

NEW!

キャラクターの魅力をアップするための表現

ロパク表現の実装はキャラクターの表情付けの一環。

キャラの存在感・実在感の演出に欠かせない。



作るには考えることが多くて大変

アルゴリズム
選択

CPU負荷

手作業

物量

キャラモデルの
仕様

音とアニメの
タイミング合わせ

自動化

ボイスチャット

品質

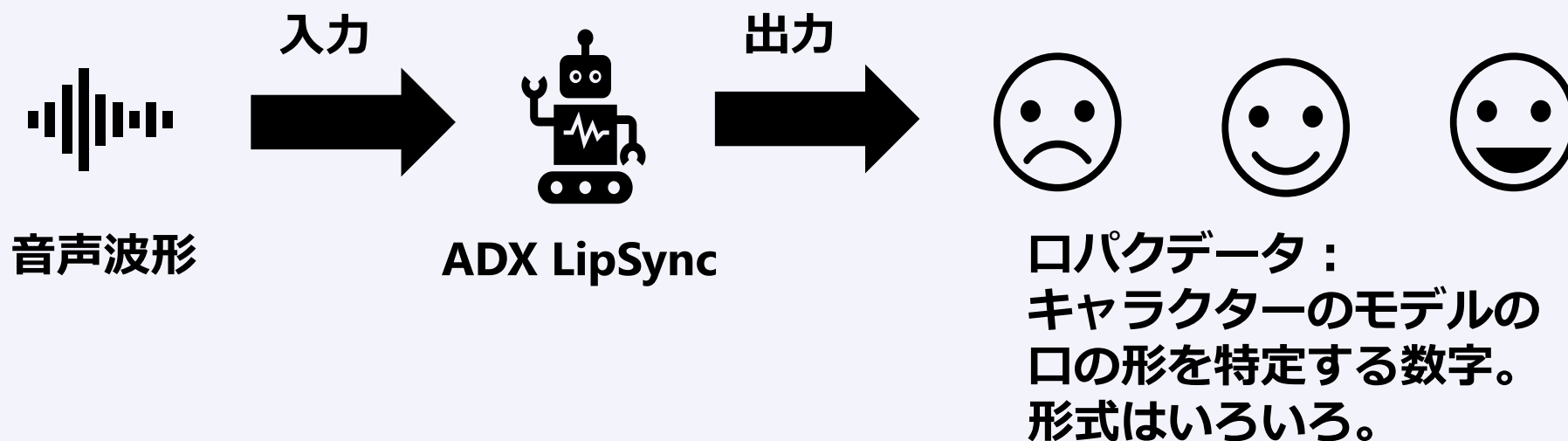
多言語対応

いい感じにやりたい...

ADX LipSync で口パク実装作業を効率的 & 柔軟に

- 音声波形を解析して口パクデータを自動生成！
- ディープラーニングを駆使

ADX[®] LipSync



なんかすごそう！でもまだピンと来ない……

- 「口パクって普通はどう実装するの？」
- 「そもそも何が大変なの？」
- 「ADX LipSync を使うと何がうれしいの？」



ロパク実装の基本



音声を聴いて口の形を手作業で仕込む

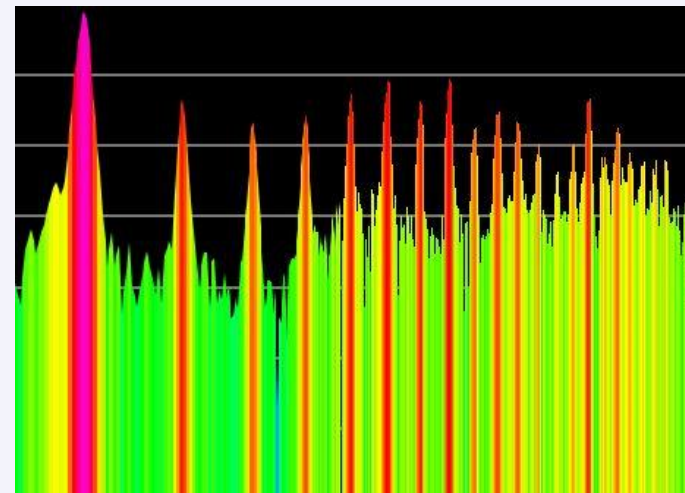
- 音声を聞いてモデルの口アニメーションを手付け
- 高品質だがタイトル規模が大きいと大変！
- 可能な限り自動化したい……

ロパクデータを自動的に生成する

1. フェイシャルキャプチャ：モーションキャプチャで顔の動きを記録
2. 音声解析：音声波形を元にロパクデータを生成

今回は**音声解析**に注目！

……なんか難しそう？



基本的な手法

1. 音量ベース
2. フォルマントベース

再生音声の音量変化を見るだけ！

- 音量変化にあわせて開口度を変える
- シンプルな見た目だが作りやすい
- 3D より 2D タイトルでよく見かける
- 言語を選ばず使える手法
- 機械的な見た目になるのが難点

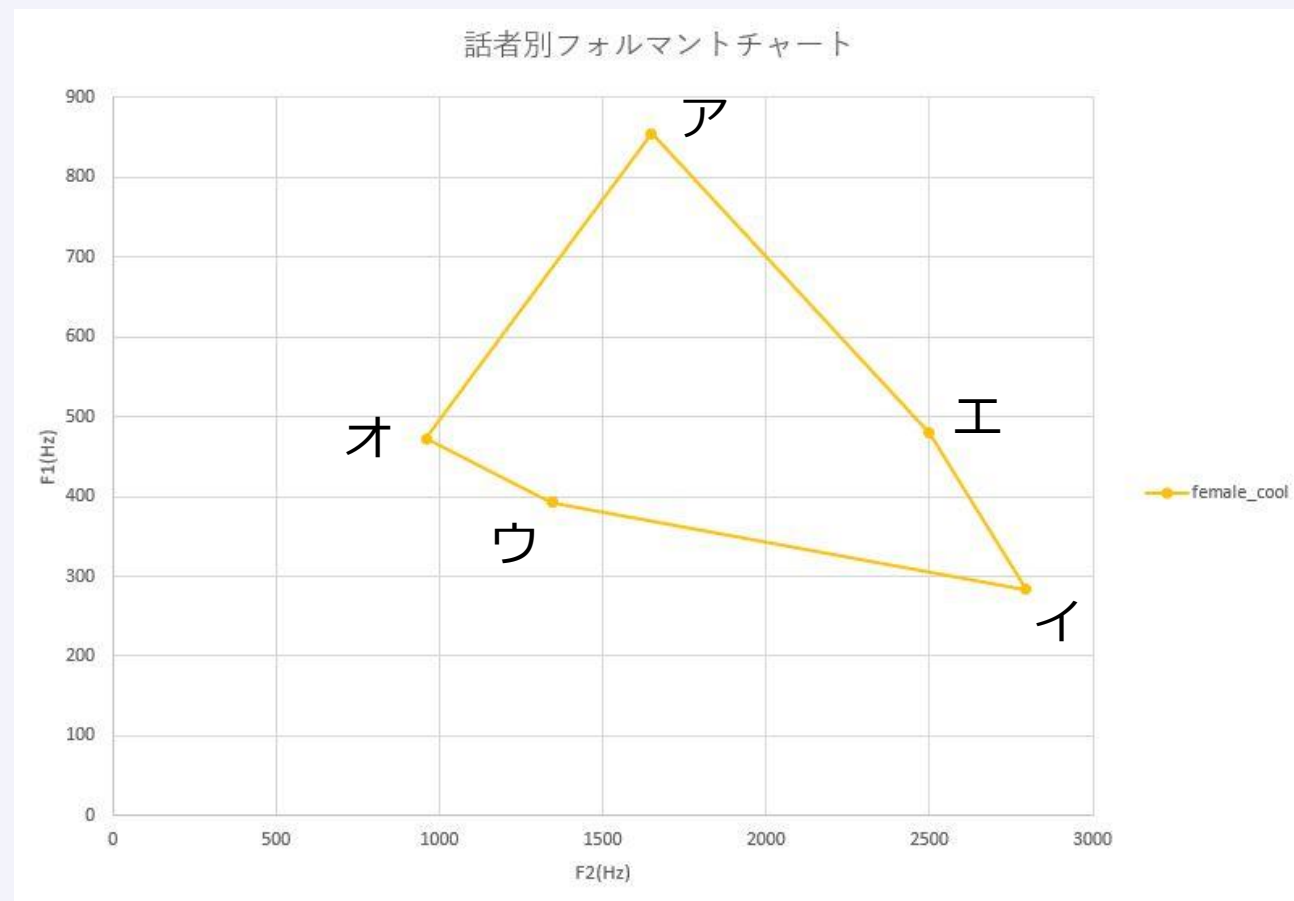


「フォルマント」を調べれば母音を特定できる

フォルマント：音声の周波数のピーク

2種類のフォルマントの位置関係で
母音を だいたい 特定可能！

音量ベースよりも自然な口パクになる



注意点もいろいろ

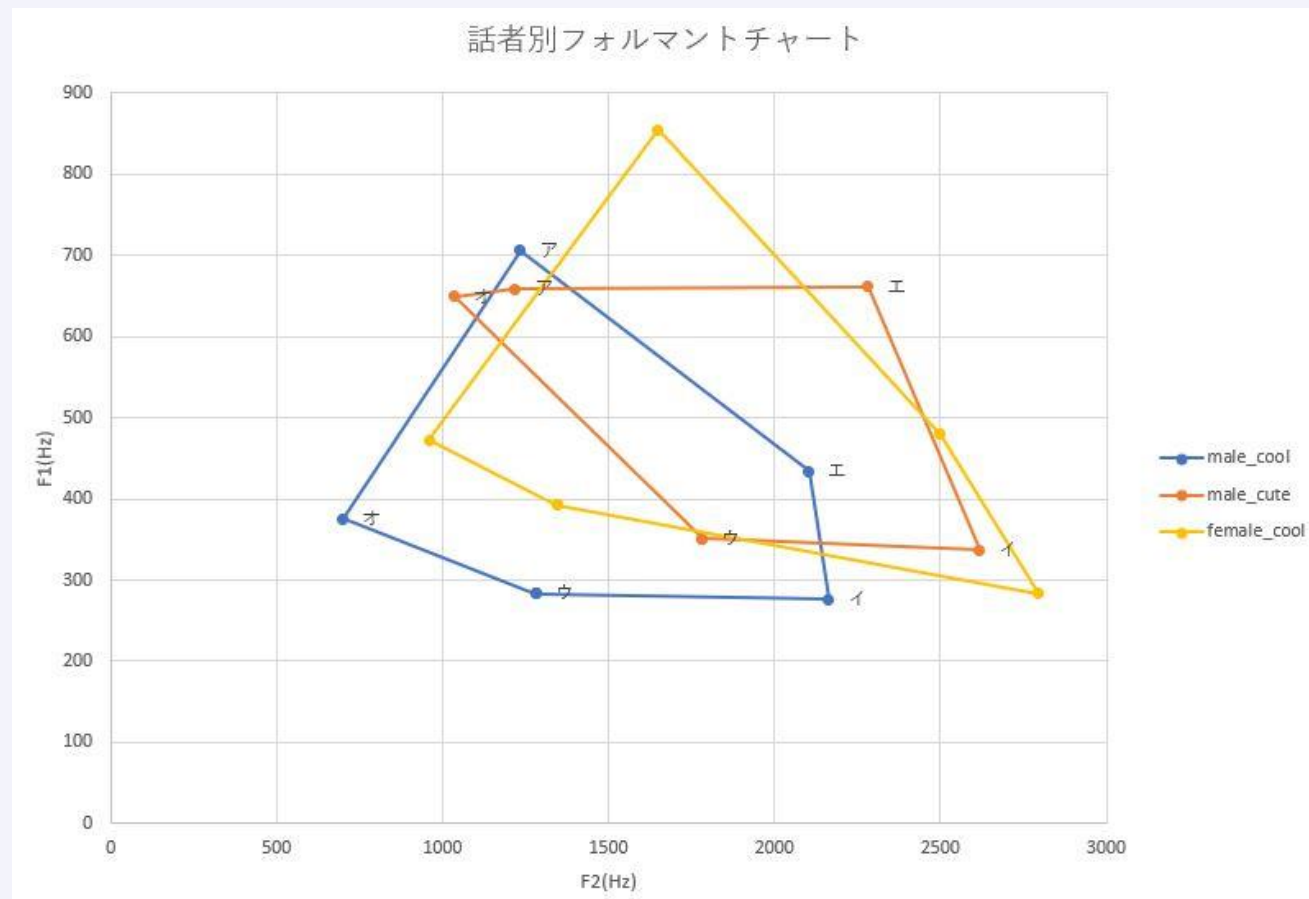
話者ごとに傾向が違う

フォルマント抽出の精度も問題

子音含む全ての音素を判別は困難

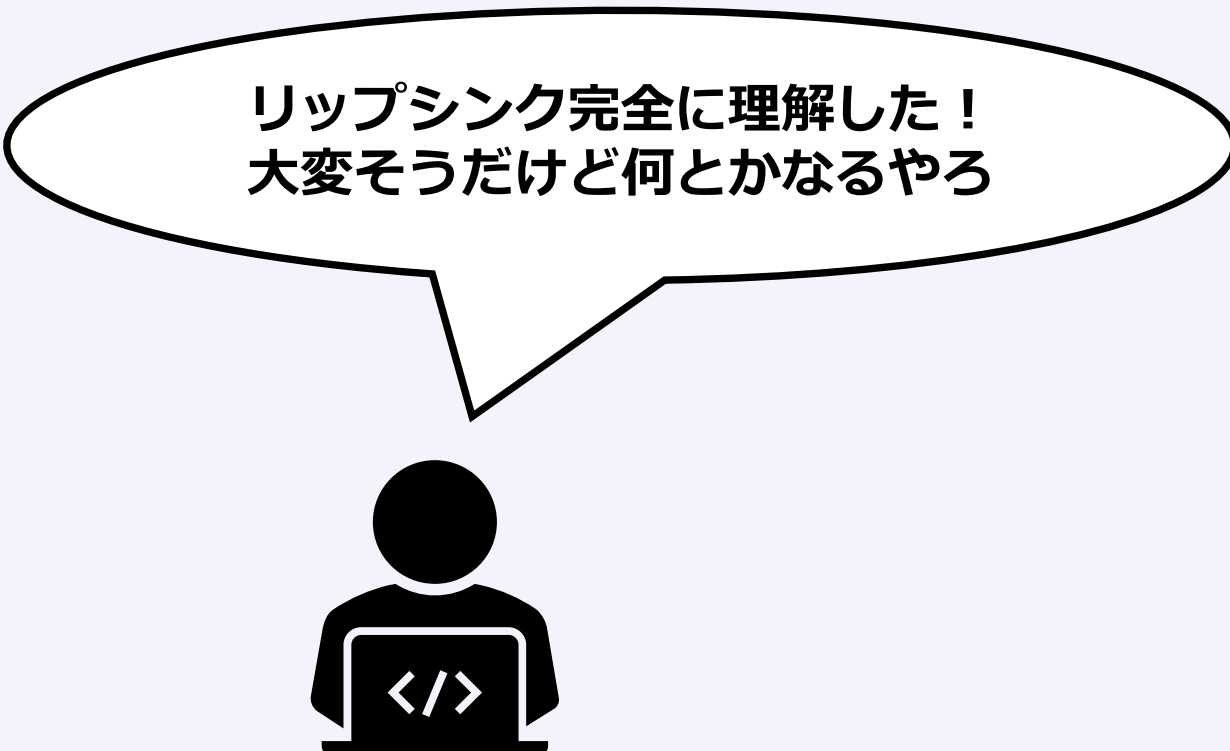
母音別にモデルの表情パターンが必要

自前で実装するのは面倒



基本的な手法

1. 音量ベース
2. フォルマントベース



リップシンク完全に理解した！
大変そうだけど何とかなるやろ

……本当？

ロパク実装の悩みどころ



①モデルの口構造に合うロパクデータが必要

例えば以下の構造タイプがある (注：名称は適当です……)

- **音素別表情パターン型**：音素別に用意した口形を切り替えorブレンド。言語選ぶ
- **1軸型**：縦方向のみ口形を変化。3Dなら下アゴ用ボーン等。言語選ばず
- **2軸型**：縦、横方向で口形を変化。言語選ばず

モデル構造にロパクデータ形式を合わせる。またはその逆

② 「マ」「バ」など一度口を閉じる発音への対応

いわゆる「唇音 (しんおん)」に対応できていないと……

- 「バーノッパ」は「アーアアア」で口が開きっぱなし
- 「右耳(ミギミミ)」は「イイイイ」で口が開きっぱなし

③ 母音の数は言語で異なる

- 日本語は5母音
- 英語は9母音 (諸説あり)

フォルマントベースのロパクは言語間の差異を吸収しない。

モデル側にも表情追加が必要。

やっぱり考えることが多くて大変

アルゴリズム
選択

CPU負荷

手作業

物量

キャラモデルの
仕様

音とアニメの
タイミング合わせ

自動化

ボイスチャット

品質

多言語対応

いい感じにやりたい... (2回目)



ADX LipSync はどう変わる？

①モデルの口構造に合うロパクデータが必要

ADX LipSync は 2 種類のロパクデータ形式に対応

A) Width-Height 形式：入力音声に対する口形の幅、高さを解析。言語選ばず

B) 音素別ブレンド量形式：入力音声に含まれる各音素の割合を解析。言語選ぶ

モデルの口構造に依存しない！

※ 上記 2 形式とは別に、入力音声に対する「**舌の高さ**」も解析可能

② 「マ」「バ」 など一度口を閉じる発音への対応

ADX LipSync は入力音声「唇音」かどうかを判定可能！

唇音なら、一瞬の「閉じ口状態」に相当する口パクデータを返す

③言語ごとに異なる母音の数への対応

ADX LipSync の「Width-Height 形式」の口パクデータは言語を選ばない！

※ キャラクターモデルの口構造は言語を「2軸型」がおすすめ



ADX LipSync をどう使う？

ADX LipSync は CRI ADX2 SDK と連携する

ADX2 既存の仕組みを活用。使用感やインターフェースは ADX2 的。

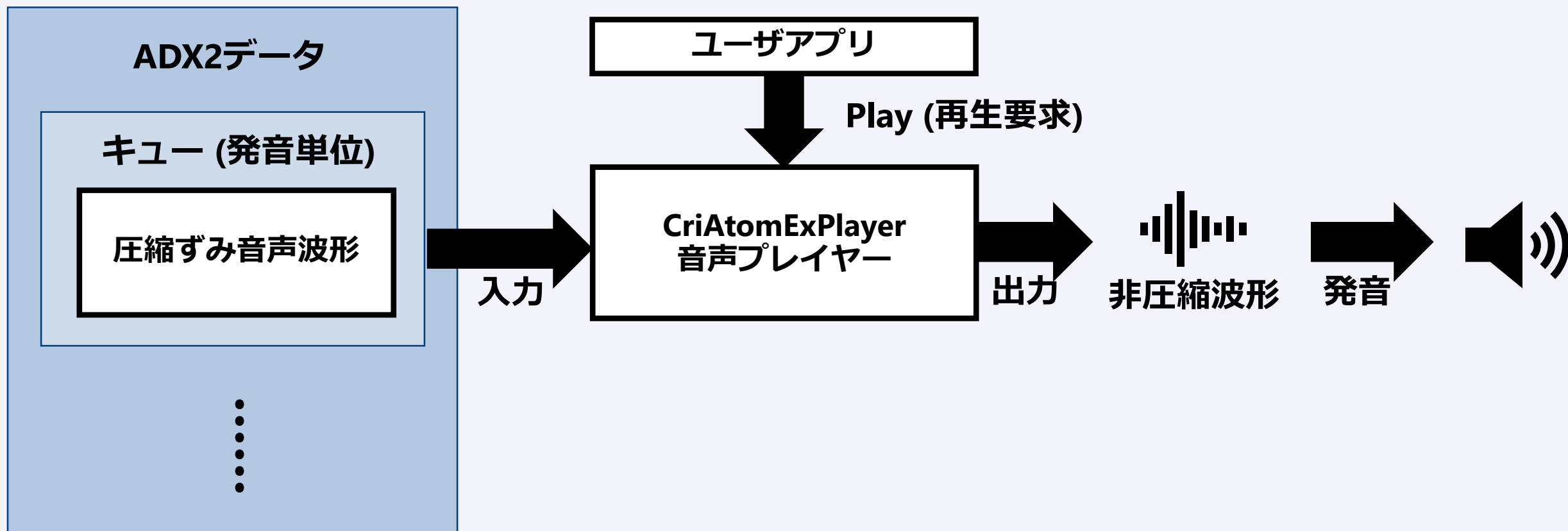


×

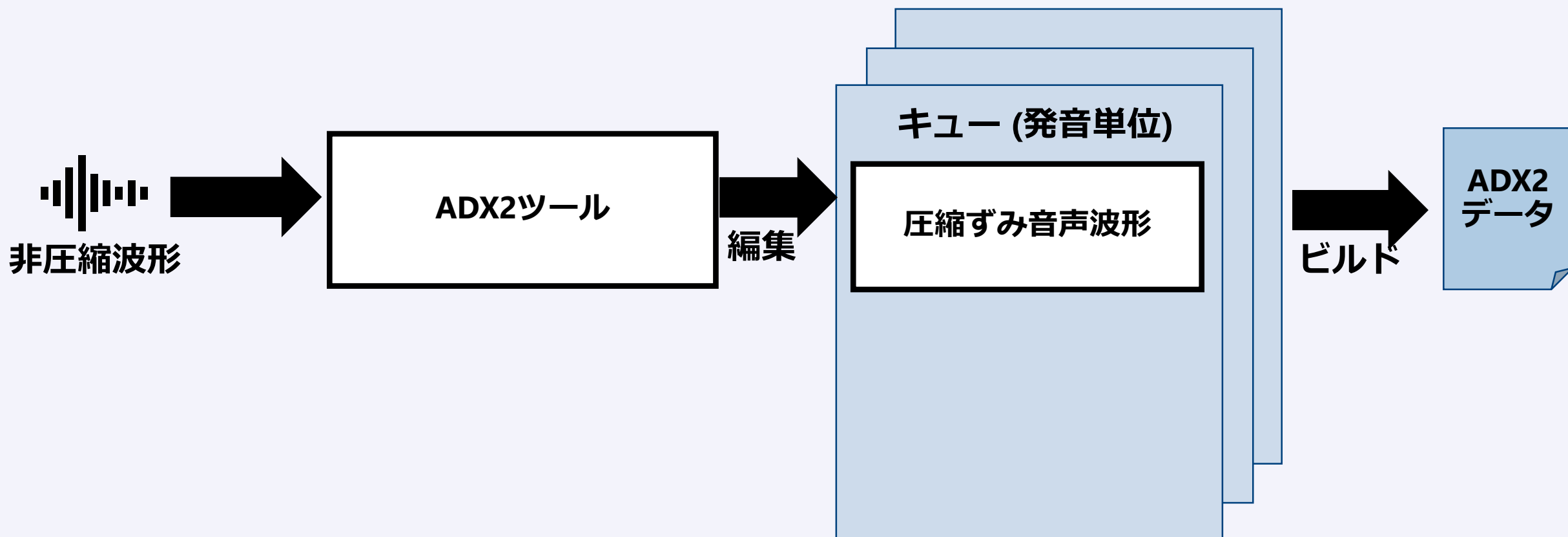
ADX[®] LipSync

.....つまり具体的にはどんなプログラムを書くことになる？

前提：ADX2 での音声再生はどんな感じ？



前提：ADX2データってどう作るの？



ケース 1：音声再生と同時に口パクデータを生成

リアルタイム式

- メリット：便利！楽ちん！
- デメリット：解析精度の関係で品質には限度あり。解析負荷も。

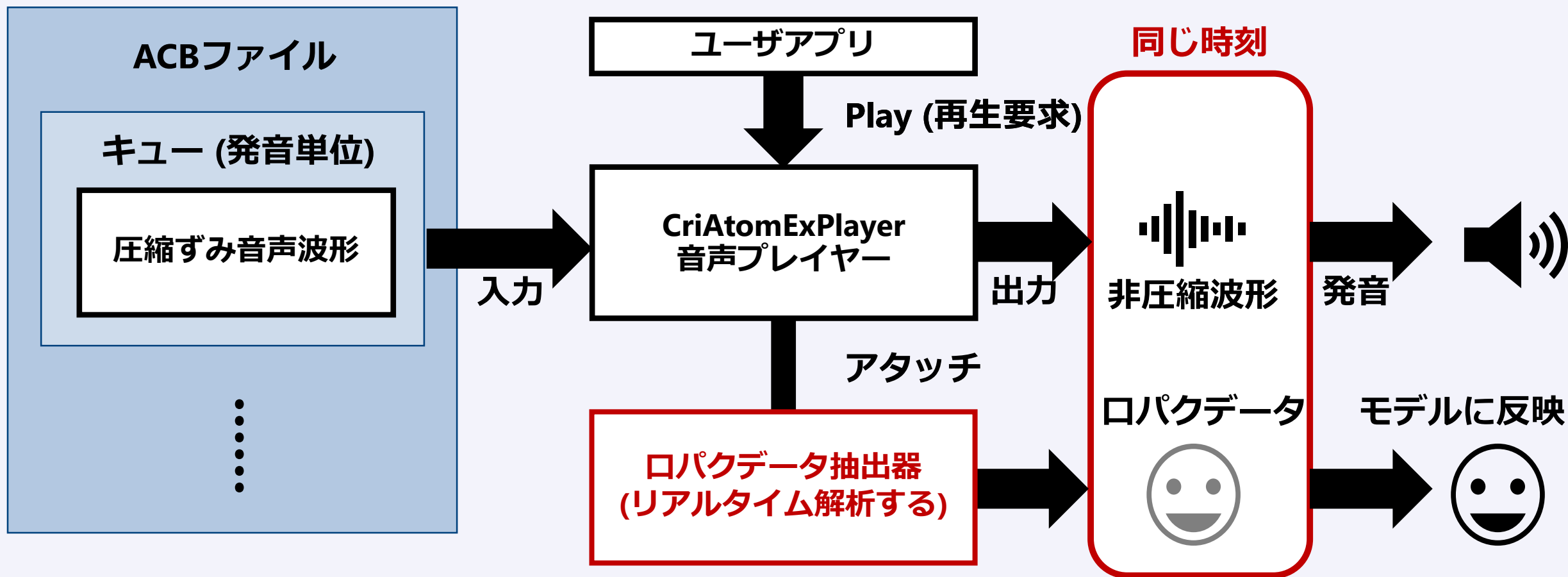


×

ADX[®] LipSync

ADX2 と連携したプログラムを書く

ADX2 と ADX LipSync を連携するときは.....



ケース2：先にロパクデータを生成しておく

オフライン式

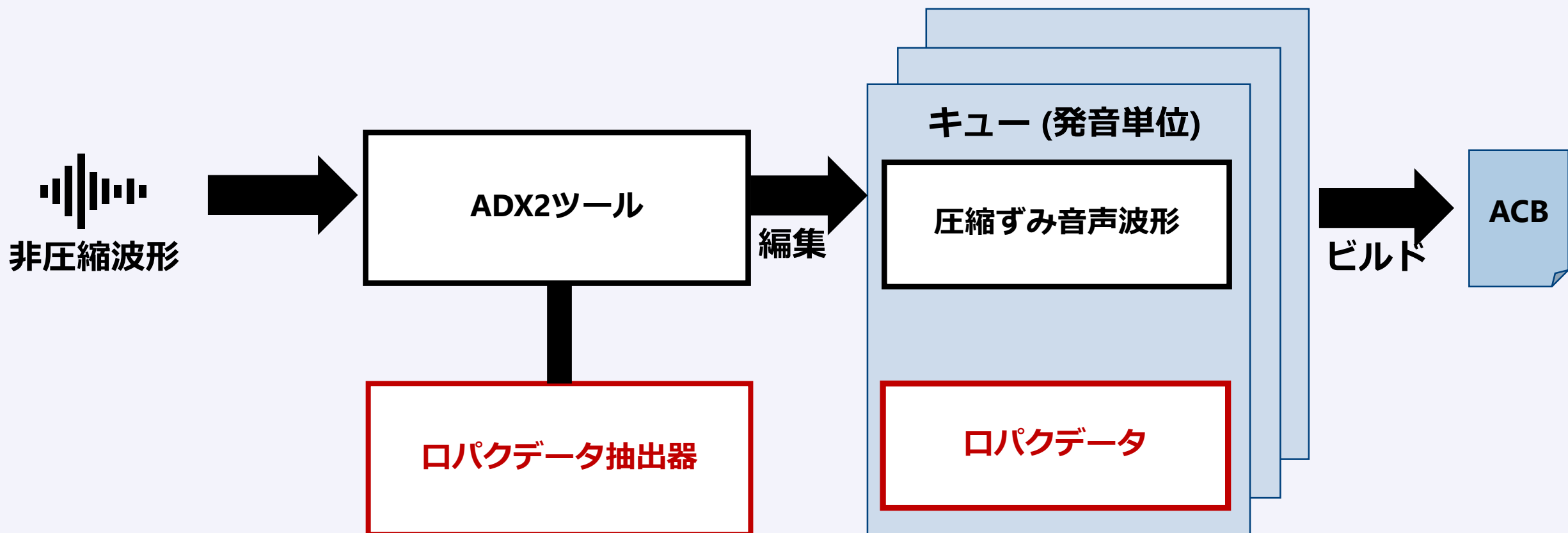
- メリット：予め解析で再生時は負荷なし。ロパク編集で品質調整可能
- デメリット：ACBファイルサイズが5%くらい大きくなる



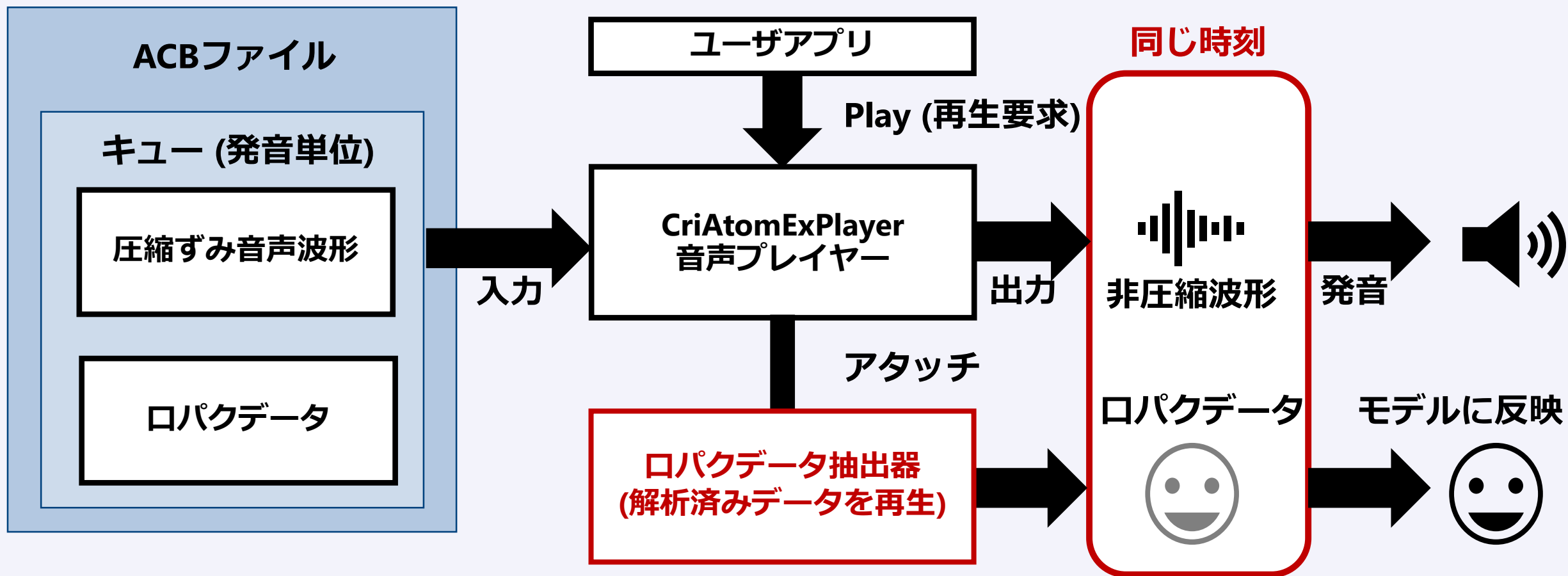
×

ADX[®] LipSync

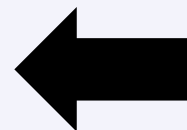
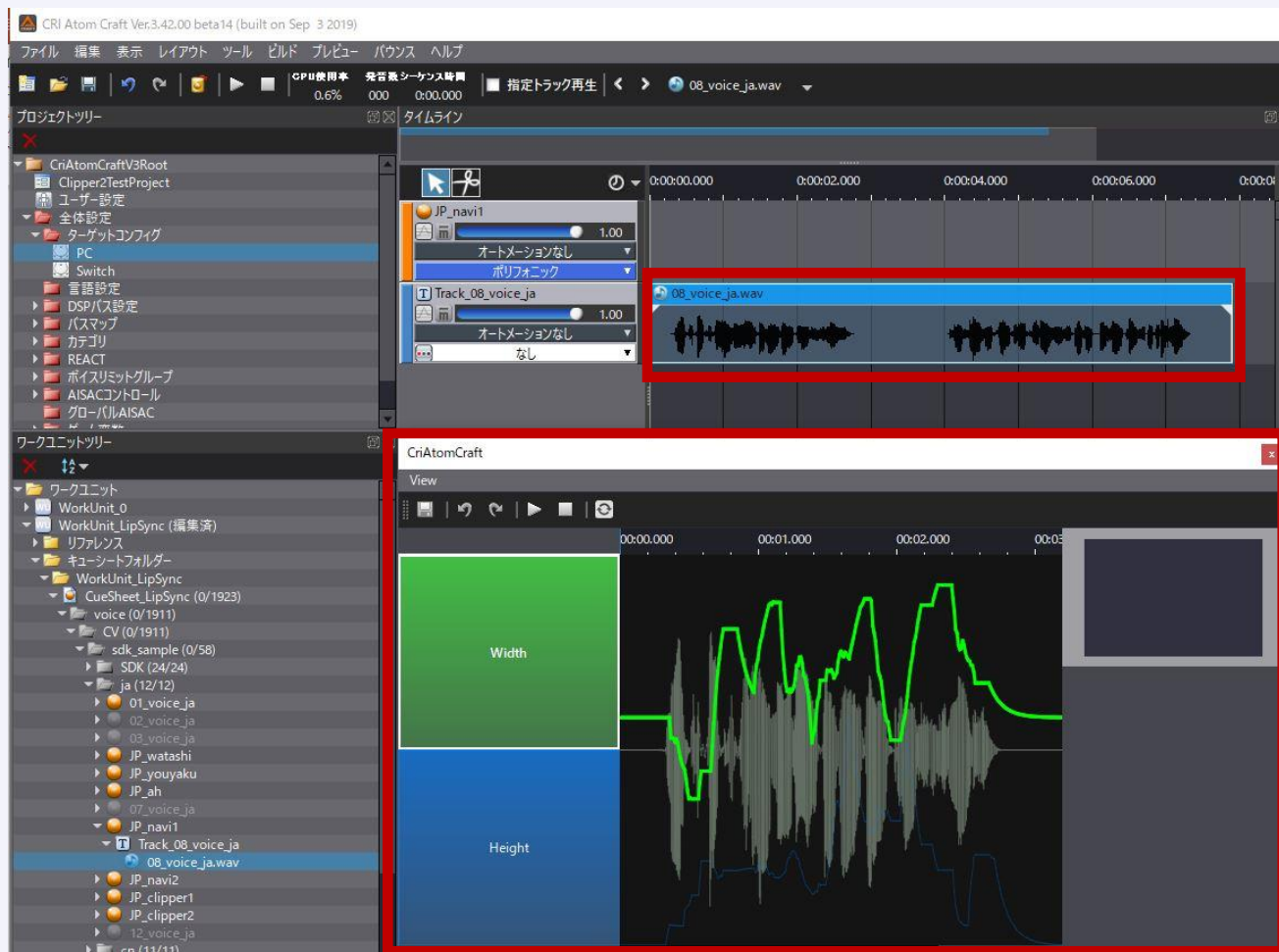
ロパクデータも生成して、まとめてパック



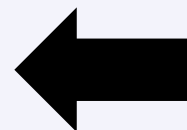
ACB内のロパクデータの再生



ADX2ツール上でロパクデータを編集可

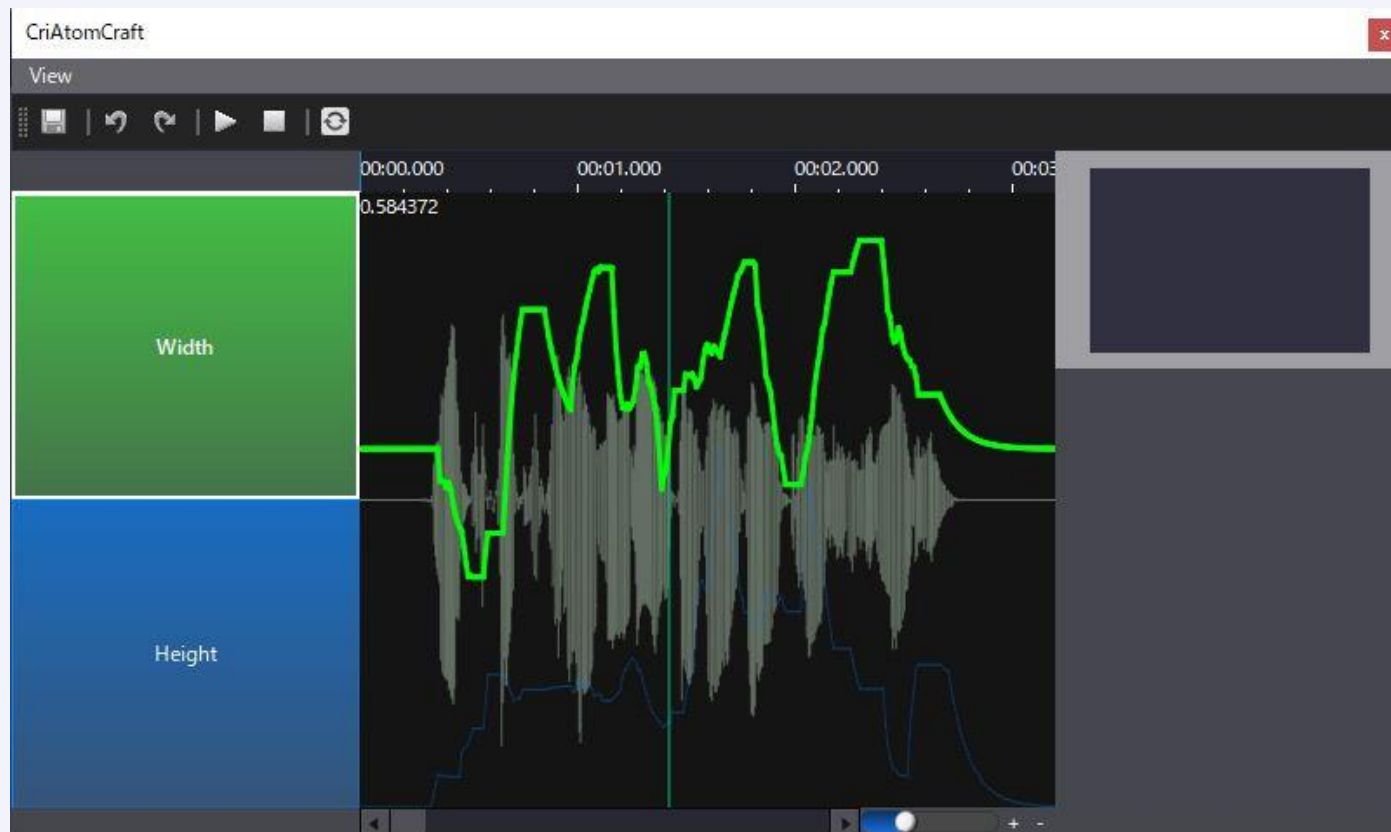


キュー内の音声波形



リップシンクエディター

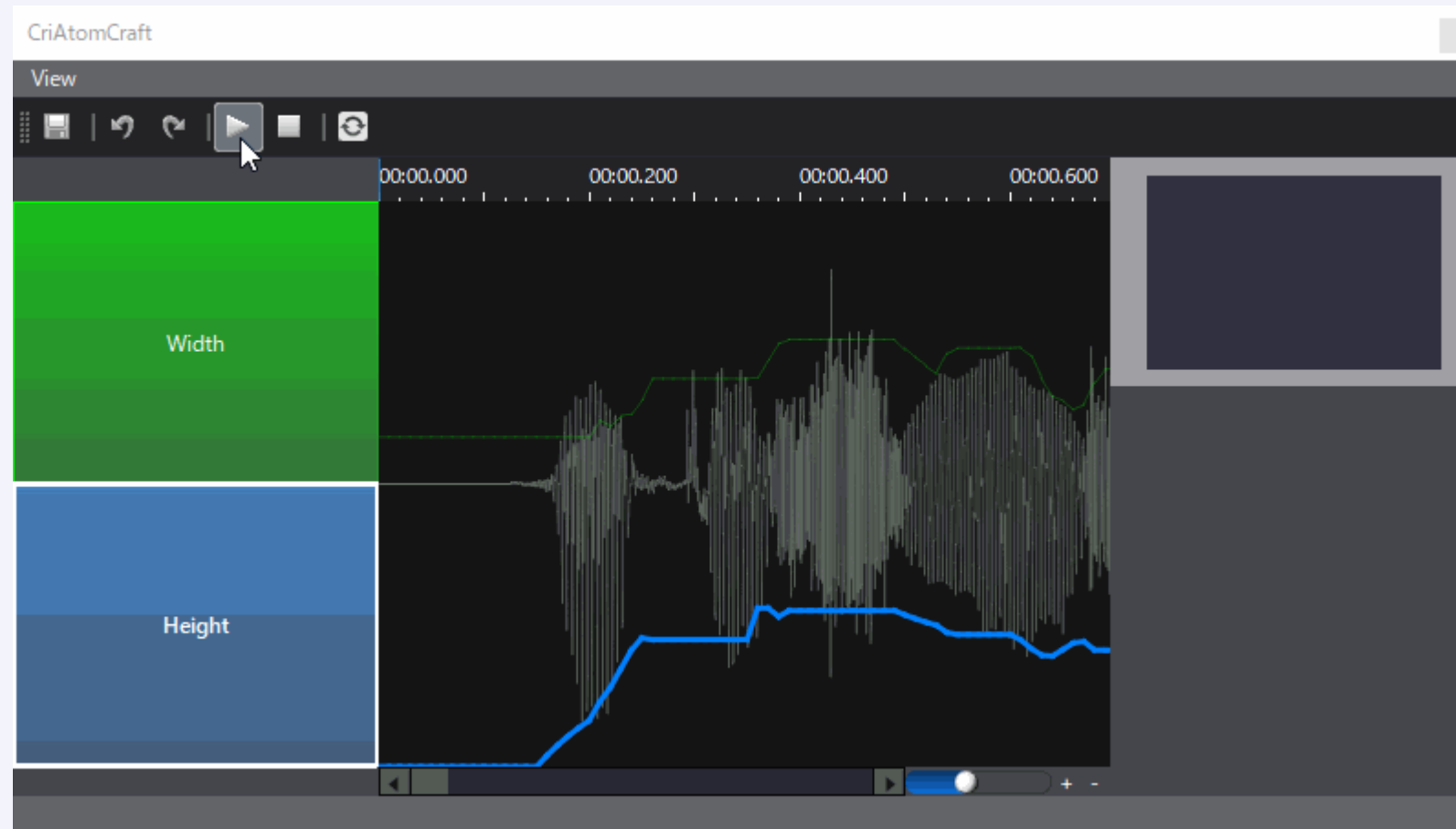
Width-Height 形式データの編集



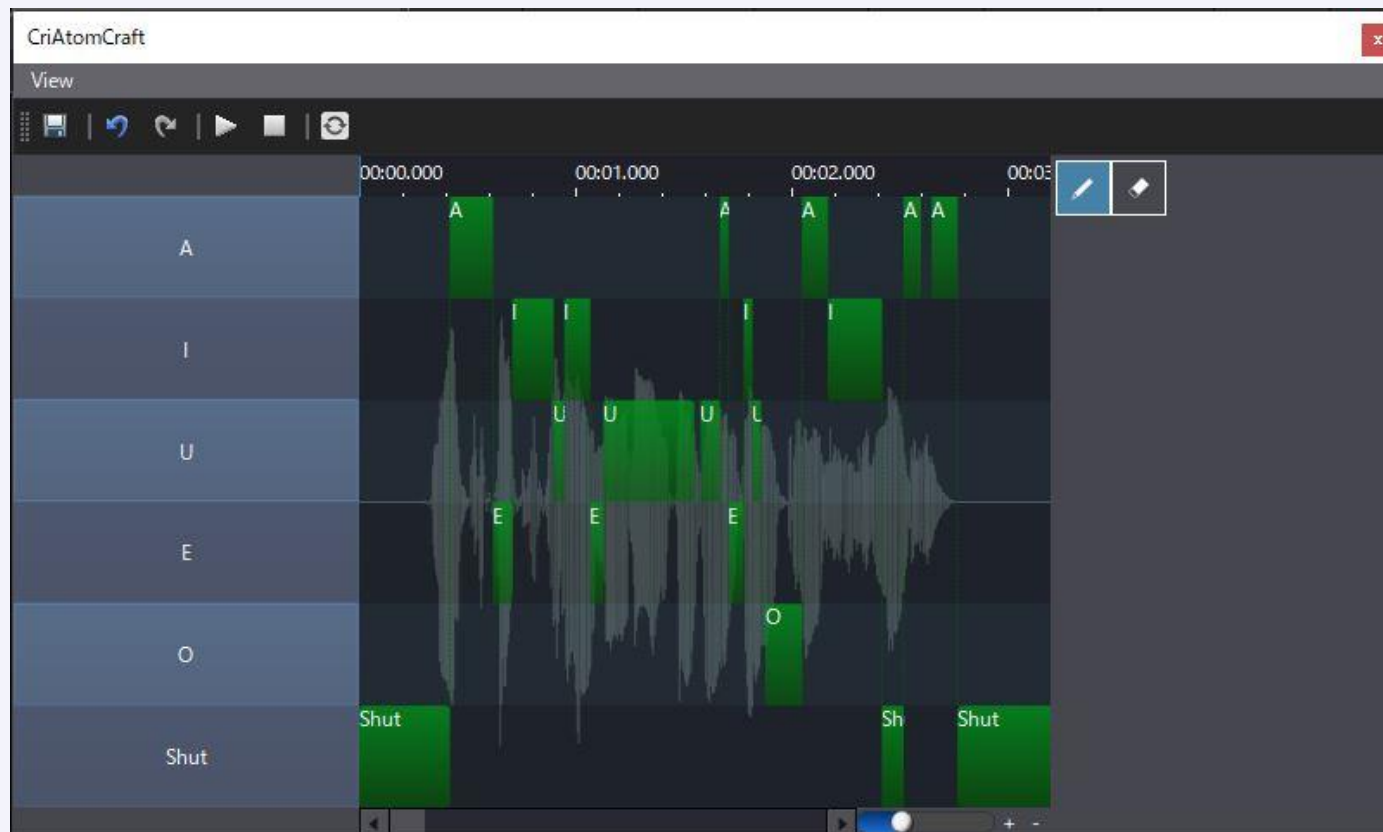
音声波形のタイムライン上に
Width値、Height値が並ぶ

フリーハンドでグラフを修正

Width-Height 形式データの編集の様子



音素別ブレンド量形式データの編集

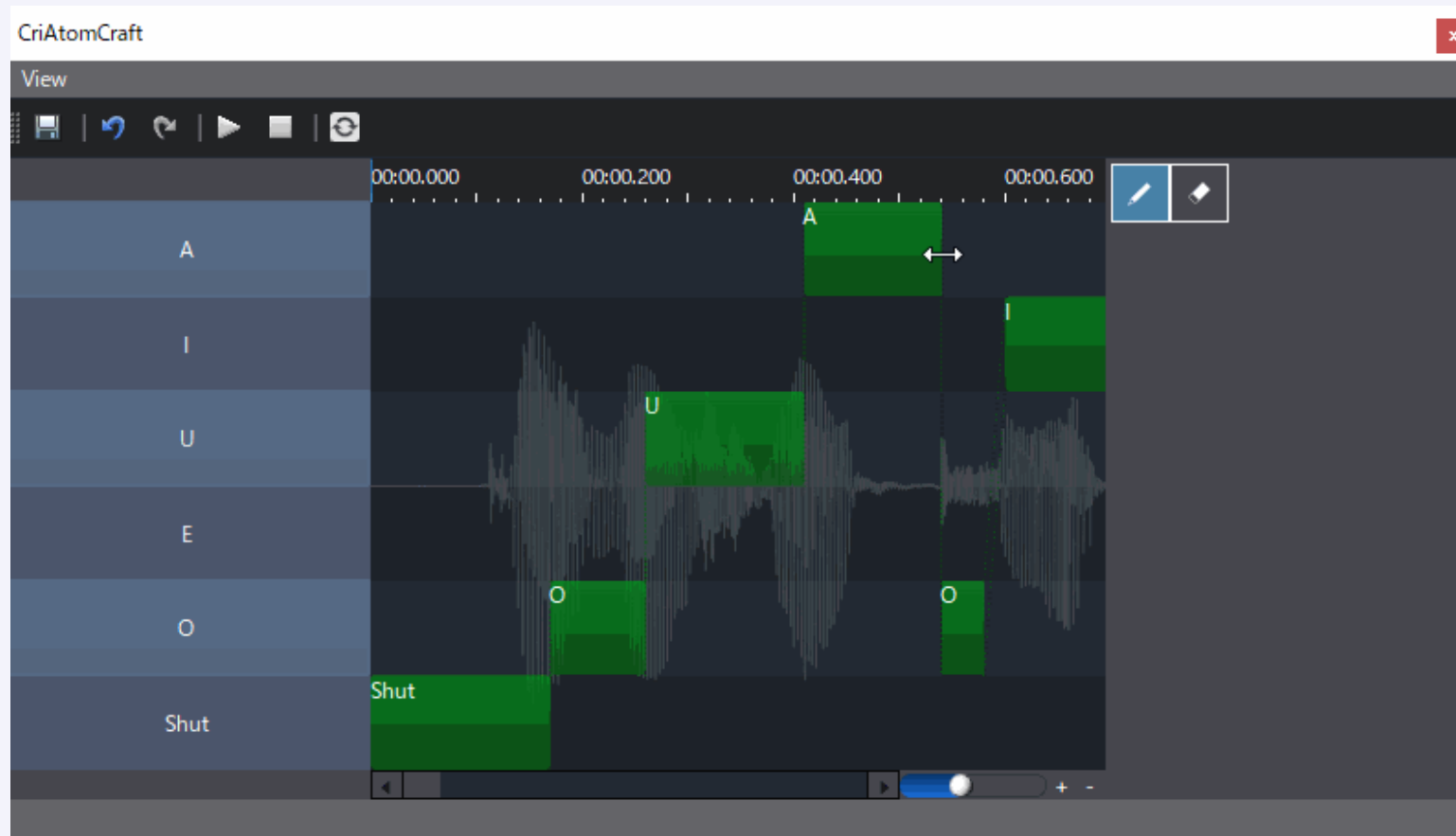


音声波形のタイムライン上に
音素別ブレンド量が並ぶ

MIDIノート風の見たい目

ノートを移動したり伸ばしたり

音素別ブレンド量形式データの編集の様子



いつから使える？

絶賛開発中。今秋リリース予定

別途ご相談ください！

ADX2と一緒に使う方には導入しやすい形を想定

リップシンクミドルウェア「ADX LipSync」

- **柔軟な口パクデータ生成で効率化**
 - キャラクターモデルに依存しない出力形式
 - 多言語対応
- **唇音への対応でよりリアルな表現が可能**
- **ADX2 連携でシンプルなAPI & 編集インタフェース**
 - マイク入力でリアルタイムな口パクも.....

CRIWAREとは

CRIWARE は単なるパッケージではありません。

より素晴らしいゲーム制作をサポートするサービスです。

是非とも、皆様のご意見・ご要望をお聞かせください。



お問い合わせ先

<http://www.cri-mw.co.jp/contact>