



Electronics for the Future

ローム 音声合成LSI
LSI選定から
音声コードデータ書き込み・出荷まで

2024年7月
ローム株式会社
LSI事業本部MCU商品開発部
MCUマーケティング2課

LAPIS TECHNOLOGY™は、ローム株式会社の商標または登録商標です。

powered by **LAPIS**
TECHNOLOGY

本資料は、ロームの音声合成LSIを初めてお使いいただくお客様に、以下のことをご理解いただくために用意したものです。
本資料は、3部から構成されます。

- ・第1章： FJXLSPEECHLSI_CHAP1-**.pdf（本書）
LSI選定から音声コードデータ書き込み・出荷までの工程概要
- ・第2章： FJXLSPEECHLSI_CHAP2-**.pdf
音声合成LSIで音を鳴らすためのデータの作成&音の試聴方法の概要
- ・第3章： FJXLSPEECHLSI_CHAP3-**.pdf
音声合成LSIに必要なツールおよびツール操作方法の概要

ラピステクノロジーの音声合成LSI（以下「音声LSI」と呼びます）は、原音があれば “手軽” に音を鳴らすことができます。

注意点は、音を鳴らすには必ず原音(wavデータ)の準備が欠かせないということです。

全体の流れは以下の通りです。STEP1~STEP5について以降のページで説明していきます。

【全体の流れ】

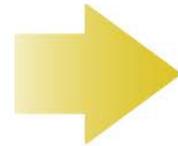
お客様作業

STEP 1 用途、ROM容量でLSIを選定

- 用途、再生時間、必要機能等、ご要求内容に応じてLSIを選択。

STEP 2 フレーズリスト作成

アドレス	音声
1	おはよう
2	こんにちは
3	ぶつ



当社 and/or お客様 作業

STEP 3 原音作成

- スタジオ録音、メロディ・効果音作成、音声ライブラリから選択

STEP 4 音声分析、編集

- 切り分け、音量調節、つなぎ合わせ、イコライジング

STEP 5 サウンドデータ完成、ROM書き込み、出荷

音声合成LSIセレクションガイドにより、お客様の製品の用途（車載/民生・産機）や再生時間、ROM容量（音声メモリサイズ：音声LSIの中に入れられる音の容量）等に応じて、最適な音声合成LSIをご提案いたします。
お客様は、表示された音声合成LSIのリストの中からお選びください。

◆【音声合成LSIセレクションガイド】：

<https://www.rohm.co.jp/calculators/speech-synthesis-lsi#TabPage1>

① 用途や再生時間、メモリタイプなど音声合成LSIの基本設定を入力

② 音声データの合成方式やフレーズ数、サンプリング周波数などを入力して音声メモリサイズを算出

③ ①、②の入力内容に基づいて、最適な音声合成LSIをリスト表示します。この中から音声合成LSIをお選びください。

お客様のシステムに必要な音声メモリサイズを計算します！

お客様の製品の音声メモリサイズを算出して、最適なラピステクノロジー音声合成LSIをご提案いたします。

1. 音声合成LSIの基本設定を入力

2. 音声データのスペックを入力してメモリサイズ算出

合成方式	フレーズ数	音の長さの平均	サンプリング周波数	メモリサイズ
HQ-ADPCM	10 個	3 秒	6.4kHz	0.6 Mbit
HQ-ADPCM	0 個	0 秒	6.4kHz	0 Mbit
HQ-ADPCM	0 個	0 秒	6.4kHz	0 Mbit
HQ-ADPCM	0 個	0 秒	6.4kHz	0 Mbit

音声データを格納するメモリサイズ合計は、**0.6 Mbit**

3. お客様のシステムに最適なラピステクノロジー音声合成LSIを下記リストにてご提案いたします。

項目	データシート	動作電圧 (V)	動作消費電力 (mW)	発振子	動作温度 (°C)	ROM容量 (bit)	フレーズ数	最大再生時間	CPU I/F	SPアンプ出力(W)/クラス	ミキシング数 (内蔵)	その他	PKG	車載対応	ネット
ML22Q553-NNNMB(Tray)		2.7 to 5.5	4.096	External	-40 to 105	Flash 4M	1024	202sec	Clock synchronous serial	1.0/AB-class	4	Speaker terminal Short circuit detection function	SSOP30		
ML22Q573-NNNMB(Tray)		2.7 to 5.5	4.096	External	-40 to 105	Flash 4M	1024	202sec	Clock synchronous serial	1.0/AB-class	4	Fail safe	SSOP30		
ML22Q573-NNNMB(Taping)		2.7 to 5.5	4.096	External	-40 to 105	Flash 4M	1024	202sec	Clock synchronous serial	1.0/AB-class	4	Fail safe	SSOP30		

フレーズリストとは、音声LSIから鳴らしたい音のリストであり、料理で言えばレシピにあたります。
このフレーズリストを基にしながら【STEP3・4】を進めていくので、とても大切な工程になります。

✿Let's✿手順に沿ってフレーズリストを作成してみましょう。

手順④再生方式とサンプリング周波数の選定

STEP1で選定した音声LSIの機種・容量に合わせて設定しますが、STEP4までは変更が可能です。
「再生方式」、「サンプリング周波数」については、弊社HPの「音声データの基礎知識」を参照してください。

<https://www.rohm.co.jp/calculators/speech-synthesis-lsi#TabPage2>

① フレーズ番号	② アドレス	③ 発声語	④ 再生方式	サンプリング周波数(Hz)
0	0	今日の天気は晴れです	HQ-ADPCM	16000
:	:	:		
10	A	「ピーツ(1000ms)」 7(しち)時です	<div style="border: 1px solid red; padding: 2px;"> HQ-ADPCM(1/5) ✓ HQ-ADPCM(1/5) 8bit Non-linear PCM 8bit Straight PCM 16bit Straight PCM 4bit ADPCM2 </div>	<div style="border: 1px solid red; padding: 2px;"> 48000 ✓ 6400 8000 10666 11025 12000 12800 16000 21333 22050 24000 25600 32000 44100 48000 </div>

手順①フレーズ番号の割り当て

1フレーズ1発声語でフレーズ番号を割り当てます。
No.0からはじまり 10進数で記載します。この時点では、0から順番に割り当てます。

手順②アドレスと発声語の割り付け 1

アドレスは音声LSIメモリの場所を示すもので、発声語の部屋番号のような役割をします。

No.0からはじまり 16進数で記載します。この時点では、0から順番に割り当てます。

手順③アドレスと発声語の割り付け 2

アドレスに音声LSIから鳴らしたい発声語を割り付けます。

POINT 1 :

「効果音」+ ガイダンスの場合は「効果音」の表記もします。
(音が複雑な場合は「メロディ①」、「効果音①」など)

POINT 2 :

数字が含まれるなど読みに注意がある場合はフリガナを記入します。(例: 7(なな)時間、7(しち)時)

原音については、wavデータで用意をします。
既に原音（wavデータ）の用意がある場合はそちらをご使用いただけます。
✿Let's✿フレーズリストを元に、原音データを用意しましょう。

<準備方法>

- ・ 制作会社へ発注（スタジオ録音・音制作）
- ・ 他社TTSソフト（テキスト音声合成）で作成
- ・ 弊社サポートサイトの音声ライブラリから選定

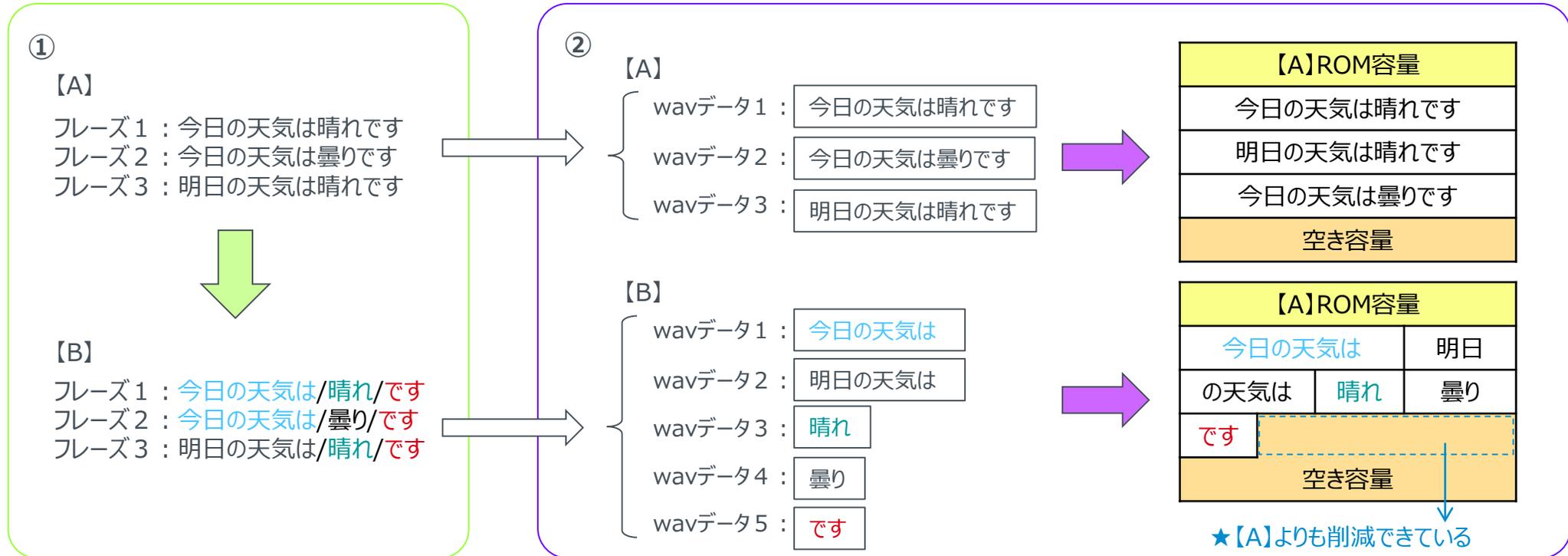
POINT : 鳴らす音のイメージ、クオリティを具体化しましょう

（例）ガイダンスの場合 : 人の声(ナレーター)とTTSを比較
 固有名詞などTTSで作成が困難なフレーズ調査
効果音・メロディの場合 : 音の仕様をきめる（時間、周波数、音の種類など）

制作会社へ発注の場合は、弊社でも紹介可能です。費用のお見積りはご相談ください。
ナレーション・音楽制作：株式会社ビー・エイチ・プロ（<https://bhpro.jp/>）

音声分析とは、原音（wavデータ）の音量調整やノイズ処理の他に、wavデータの切り分けをすることです。切り分け作業は、データ容量を削減することを目的とし、共通の音が多ければ多いほど効果的です。下記の3フレーズを使って分析（切り分け）をイメージしてみましょう。

- 手順①：3フレーズの中に共通した言葉がないか探し、共通部分で切り分けます（【A】切り分けなし【B】切り分けあり）
 手順②：【A】【B】のwavデータ（共用語は1カウント）を、ROM容量に当てはめて並べてみます
 ★結果【A】よりも【B】が、空き容量が増えて削減できていることがわかります



編集とは、分析で切り分けたwavデータを、一連のフレーズに組み立てていく作業のことです。

この編集作業で、最終的に音声LSIから流す音の最終調整をおこないます。

ローム音声合成LSIのページに掲載されている

“Speech_LSI_Utility_Practice-**.zip”のwavデータを使って、フレーズを組み立ててみましょう。



この編集作業は、ロームの開発支援ツールのSpeech LSI Utilityを使用します。

◆ツールの操作方法については、『音声合成LSI 開発支援ツール』の「2. Speech LSI Utilityの使い方」をご覧ください。

「2-6 フレーズの編集」が該当します。

編集により組み立てたフレーズの音声期待通りに聴こえるかどうかを確認するため、音声コードデータを作成し、実機で試聴します。



音声コードデータの作成および音声の実機試聴には、ラピステクノロジーの開発支援ツールのSpeech LSI UtilityおよびSDCBとSDCB Controllerを使用します。

◆ツールの操作方法については、『音声合成LSI 開発支援ツール』の「2. Speech LSI Utilityの使い方」、および「3. SDCB Controllerの使い方」をご覧ください。



音声の試聴にて、期待通りの音声如果能聴こえれば、音声コードデータの完成です。
この後は、音声合成LSIに音声コードデータを書き込んだ後、出荷となります。

<音声合成LSIへの音声コードデータ書込み方法>

- 弊社での書き込み
担当営業にお問い合わせください。
- 貴社での書き込み（サードパーティ製Flashプログラマでの書き込み）
ロームの音声合成LSIに対応しているサードパーティは以下になります。
フラッシュサポートグループ (<http://www.j-fsg.co.jp/>)
アイフォーコム京栄 (<http://www.k-kyoei.jp/>)
Flashプログラマのご購入につきましては、サードパーティに直接お問い合わせください。



Electronics for the Future