

【研究ノート】

音声合成システム Google Cloud Text to Speech を利用したスラッシュ ・リーディング／リスニングのための英語音声教材作成方法について

丸 山 真 純
鈴 木 齊

Abstract

This technical note describes how to create an audio text with pauses for “slash” reading/listening (aka “phrase” or “chunk” reading/listening) by using a voice synthesis system. In general, English textbooks do not have audio CDs or files for slash reading. This is also true for the variety of English texts and passages freely available on the Internet. In these cases, the instructors still might want to create audio files for pedagogical purposes. Therefore, this paper explains a step-by-step procedure to create audio files with pauses for slash reading by utilizing Google Cloud Text to Speech, along with “break,” an SSML tag. Although the created audios are not natural voice, they should still help students learn English more effectively.

Keywords: slash reading/listening, voice synthesis system, Text to Speech (TTS)

1. はじめに

本稿では、音声合成システム (voice synthesis system) である Google Cloud Text to Speech¹, および、それがサポートしている音声合成マークアップ言語² (Speech Synthesis Markup Language; SSML) の無音声部分を意味する “break” タグ³を用いて、スラッシュごとに無音で区切られた英語音声教材を作成する技術的方法について説明する。

英語を学習する際に、意味のまとまり(チャンク)ごとにスラッシュを入れ、そのスラッシュごとに読んだり、聞いたりすることは効果的な英語学習方法として推奨されている。そのよう

¹ <https://cloud.google.com/text-to-speech/> 2020年3月13日時点

なお、Google Cloud Text to Speech 以外の音声合成システムでも、本稿で説明するような読み上げが可能である。

² <https://www.w3.org/TR/speech-synthesis/> 2020年3月13日時点

³ Google Cloud Text to Speech で利用可能なタグは <https://cloud.google.com/text-to-speech/docs/ssml> に説明がある。

な方法は、スラッシュ・リーディング、フレーズ・リーディング、チャンク・リーディングなどと呼ばれる。また、通訳訓練の一つとして用いられるサイト・トランスレーション法なども、意味のまとまりごとに聞いて、そのまとまりごとに訳をするというという意味では、スラッシュ・リーディングと同列に考えることができよう。

(意味の切れ目に)スラッシュを入れて、英語を読んだり、聞いたりすることで、返り読みをせず語順通りに内容を把握することができ、直読直解が可能となり、それゆえに、速読も容易となる。したがって、リーディングの際に、単に区切って読むだけでなく、実際に、スラッシュごとに区切って読まれる音声を区切り聞きのために併用することも、リーディング学習に有効であると考えられる。

しかし、通常、英語学習用の教科書(とりわけ、リーディング用教科書)には、本文を読み上げた音声オーディオ CD あるいは mp3 ファイル形式で提供されてはいるものの、スラッシュごとに無音声部分(ポーズ)を入れたものは、スラッシュ・リーディング(やリスニング)を意図した教材を除けば、非常に限られているのが現状である。また、近年では、ネット上には膨大な英語で書かれた文章が存在するが、そのようなものを教材として利用しようとしても、その音声は、英語学習などを目的としたものを除けば、利用することができない。まして、スラッシュごとに区切られた音声はほぼ皆無である。

このような場合に、(スラッシュ)音声を使用したいと考えるならば、英語話者にそのような音声を逐一吹き込んでもらうことになるだろう。しかし、例えば、1 学期分の教材の吹き込みを依頼することは、個人的な好意に頼るには膨大過ぎる作業量となるであろう。

そこで、次の選択肢として考えられるのは、音声合成読み上げシステムを利用して、音声教材を作成するという方法である。そのようなシステムへの懸念としては、人工音声であるがゆえに、その音声が現実の音声と比べて、英語学習にふさわしいかということであろう⁴。しかし、近年、このような読み上げ機能は大規模なデータと機械学習の恩恵から、発音等が格段に向上し、かなり自然音声に近いものとなってきたため、自然音声に代替する教材音声として、英語学習に十分耐えうるものになってきている。

そこで、本稿では、Google が提供する音声合成読み上げシステムである Google Cloud Text to Speech を利用した音声教材の作成方法、とりわけ、スラッシュごとに無音声部分を入れた音声教材の作成とそのファイル化の技術的方法について解説する。

以下では、まず、英語テキストを Google Cloud Text to Speech で読み上げさせる方法について説明する。次いで、スラッシュが入った英文を、SSML のタグ“break”を利用して、スラッシュごとにポーズを入れ、それを Google Cloud Text to Speech で読み上げさせる方法について解説する。最後に、これらの読み上げられた音声を録音し、音声ファイル(mp3)として保存する方法について説明する。なお、最後の録音と保存については、Macintosh コンピュータを使用した場合についての説明である⁵。

⁴ このような懸念を解消するために、近い将来、受講生の評価も含めた客観的検証を行う予定である。

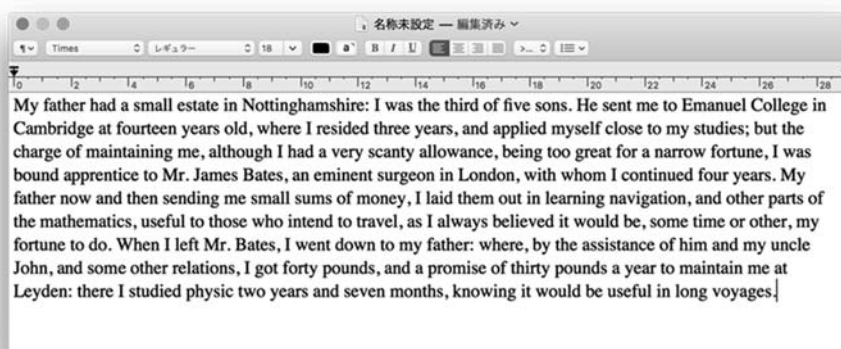
⁵ 例は、Macintosh コンピュータで示すが、Windows PC においても、録音・保存はもちろん可能である。

2. 音声合成システムによる英文の読み上げ

本節では、任意の英語テキストを音声合成システムで読み上げさせる方法を解説する。

まず、読み上げさせたい英語テキストを手入力するか、または、ネット上の著作権的に問題の無いものを、テキストエディタなどに切り貼り（Copy & Paste）し、ファイルとして保存する（図1の文章は、Jonathan Swift の Gulliver's Travelーガリバー旅行記ーの冒頭部分⁶である）。

図1 英語テキストの文書ファイル化



次に、Google の音声合成テキスト読み上げシステム Google Cloud Text to Speech のページ (<https://cloud.google.com/text-to-speech/>) を開く。このサイトは、Google が有償で提供するサービスのデモサイトの位置づけといえる。従来のテキスト読み上げ機能を提供するサービスは、変換結果の一部を提示するのみであったが、このサイトでは利用者が希望するテキストでの性能評価を実質的に回数制限も無く試すことが可能となっている。

先のテキストをページ内のテキストボックス（Text to speak）に入力、あるいはペーストする（図2参照）。特段、設定等を変更する必要はないが、Language にて米語・英語などの言語の指定ができる（図2では、デフォルトのまま English (United States) を指定している⁷）。また、Voice name にて、スピーチ音声を変更することも可能である（男性音声・女性音声など）。そして、Speed にて、読み上げられる速さを調節⁸することもできる。

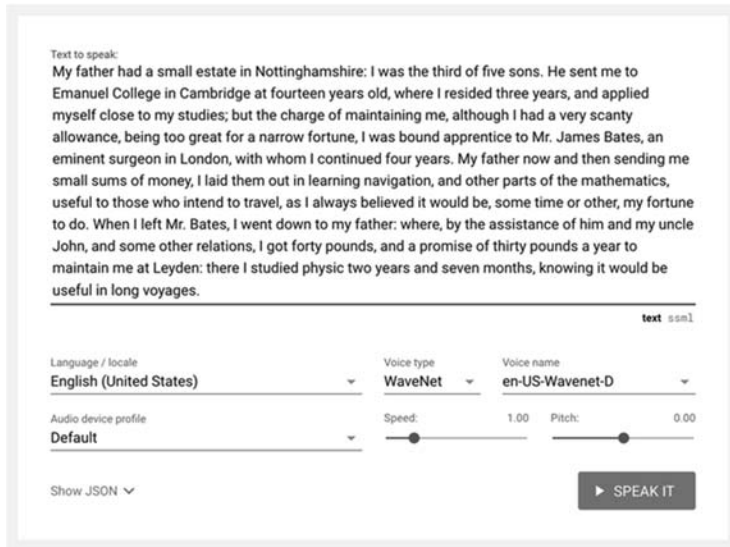
以上の設定後、「SPEAK IT」をクリックすると、合成された音声を読み上げられる。

⁶ Lit2Go (<https://etc.usf.edu/lit2go/>)。 2020年3月13日時点

⁷ 英語のみならず他言語の指定も可能である。

⁸ 話速は日本語を母語とする学習者には、0.85程度の少しゆっくりとした速度に調整するとちょうどよいかもしれない。録音時は1.0に設定しておき、再生時にプレーヤーの話速変換機能で再生スピードを落としてもよい。

図 2 Google Cloud Text to Speech 上の入力欄



3. スラッシュ音声の読み上げ方法

本節では、スラッシュの入った音声を、スラッシュごとにポーズ（無音声部分）を入れて読み上げさせる方法について説明する。

3.1. スラッシュを入れた英文の用意

テキストエディタなどに読み上げさせたい英文を用意するのは先と同様である。先と異なるのは、英文をスラッシュ毎にポーズをおいて、読み上げさせるための手続きである。

手順は以下の通りである。まず、先の英文テキストに、意味のまとまり毎に、スラッシュを入れる。スラッシュを入れる手間を省くには、田中省作（立命館大学）、行野顕正・富浦洋一（九州大学）らによる共同研究の成果がスラッシュ英文作成サイト⁹として提供されているので、それを利用することができる。

このサイトでは、図3にある「Model」からひとつを選択¹⁰することで、そのモデル（方法）

⁹ <http://www.cl.ritsumei.ac.jp/CALL/SR/> 2020年3月13日時点

¹⁰ 「らくらく英文解釈」「All in One」「Kitao」「Kimura」「Kamiya」の5モデルが用意されている。それぞれのサンプルを表示させることもできる。また、スラッシュの入れ方には、絶対的原則があるわけではないので、このサイトでスラッシュを入れる作業の省力化のために使用するというのが現実的であろう。

なお、これらのモデルの技術的側面については、以下の論文などを参照：

行野顕正 田中省作 富浦洋一 柴田雅博. (2007). 統計的アプローチによる英語スラッシュ・リーディング教材の自動生成. *情報処理学会論文誌*, 48, 1, 365-374. <https://ci.nii.ac.jp/naid/110006152210/> 2020年3月24日時点

にしたがって、入力された英文にスラッシュを自動的に入れてくれる。

図3 スラッシュ英文作成サイト

Slash Reading

使い方 RES O
システムについて

☆ 本システムで作成した教材集

本システムは、スラッシュ・リーディング用の教材作成を支援するものです
システムは、任意の英文書にスラッシュを付与することができます

① My father had a small estate in Nottinghamshire: I was the third of five sons. He sent me to Emanuel College in Cambridge at fourteen years old, where I resided three years, and applied myself close to my studies; but the charge of maintaining me, although I had a very scanty allowance, being too great for a narrow fortune, I was bound apprentice to Mr. James Bates, an eminent surgeon in London, with whom I continued four years. My father now and then sending me small sums of money, I laid them out in learning navigation, and other parts of the mathematics, useful to those who intend to travel, as I always believed it would be, some time or other, my fortune to do. When I left Mr. Bates, I went down to my father: where, by the assistance of him and my uncle John, and some other relations, I got forty pounds, and a promise of thirty pounds a year to maintain me at Leyden: there I studied physic two years and seven months, knowing it would be useful in long voyages.

② Model: All In One

サンプル表示 [らっくらく英文解釈 / All In One / Kitao / Kimura / Kamiya / すべてまとめて表示]

③ Submit Clear

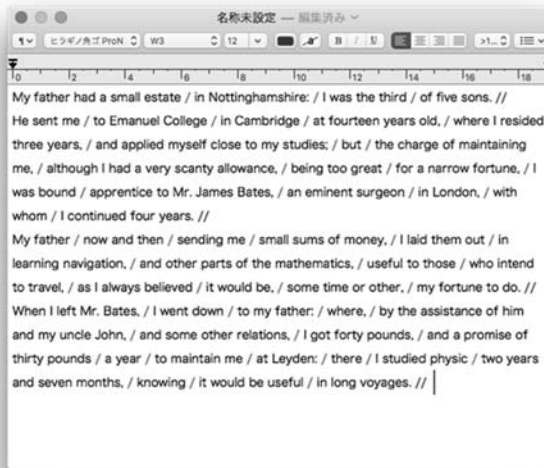
「Submit」をクリックすると、選択したモデルにしたがって、スラッシュ英文が出力される(図4参照)。その出力結果をテキストエディタにペーストし、学習目的、学習者のレベル等を勘案して、適宜、スラッシュの位置を修正、変更する(図5参照)。なお、このスラッシュの入った文書を受講者用配布資料とするために、例えば、Microsoft Wordなどに貼りつけて保存しておくとい。

図4 スラッシュ英文作成サイトでの出力例

My father had a small estate / in Nottinghamshire: / I was the third / of five sons. //
He sent me / to Emanuel College / in Cambridge / at fourteen years / old, / where I resided three years, / and
applied myself close / to my studies; but / the charge / of maintaining me, / although I had a very scanty
allowance, / being too great / for a narrow fortune, / I was bound / apprentice / to Mr. James Bates, / an
eminent surgeon / in London, / with whom / I continued four years. //
My father now and then sending me small sums / of money, / I laid them out / in learning navigation, and other
parts / of the mathematics, / useful / to those / who intend to travel, as / I always believed / it would be, some
time or other, / my fortune to do. //
When I left Mr. Bates, / I went / down / to my father: / where, / by the assistance / of him and my uncle / John,
and some other relations, / I got forty pounds, and a promise / of thirty pounds / a year / to maintain me / at
Leyden: / there / I studied physic two years and seven months, / knowing / it would be useful / in long voyages.
//

Return

図5 スラッシュ位置の調整後



3.2. スラッシュの break タグでの置き換え

次に、スラッシュ部分をポーズ（無音声部分）とするために、スラッシュを SSML で無音声部分を意味する “break” タグへと置き換える。“break” タグは次のように用いる：

```
<break time="XX ms" />
```

XX の部分には、無音時間をミリ秒で入力する。例えば、1 秒の無音声部分であれば、`<break time="1000ms" />` と入れる。`<break time="1 s" />` でも同様に 1 秒の無音声部分が作られる。

スラッシュ箇所の無音時間としては、1 秒とり、文の終わり、つまり、ピリオドの後には、1.5 秒とることが適当と思われる（必要に応じて、調整可能である）。これに従えば、「/」部分は、「`<break time="1000ms" />`」で置き換え、スラッシュ・リーディングで通常、文の終わりに打たれる「//」は、「`<break time="1500ms" />`」へと置き換える。

つまり、テキストエディタの検索・置き換えにて、「/」を「`<break time="1000ms" />`」に、ピリオドの後に打たれる「//」を「`<break time="1500ms" />`」へと置き換える。

なお、一般的なテキストエディタに備わっている置換機能は、テキスト前方からの単純な置き換え処理を行うものであるため、最初に「/」を「`<break time="1000ms" />`」で置き換える時点（図 6 参照）で、ピリオドの後の「//」も置き換え対象と判定される。その結果として、ピリオドの後の「//」は「`<break time="1000ms" /><break time="1000ms" />`」へと置き換えられる。

そこで、先の置き換え後に「`<break time="1000ms" /><break time="1000ms" />`」を検索し、「`<break time="1500ms" />`」へと置換を行うことで、ピリオドの後の「//」は無事、「`<break time="1500ms" />`」へと置き換えられる（図 7 参照）。

図6 第1回目の置換操作

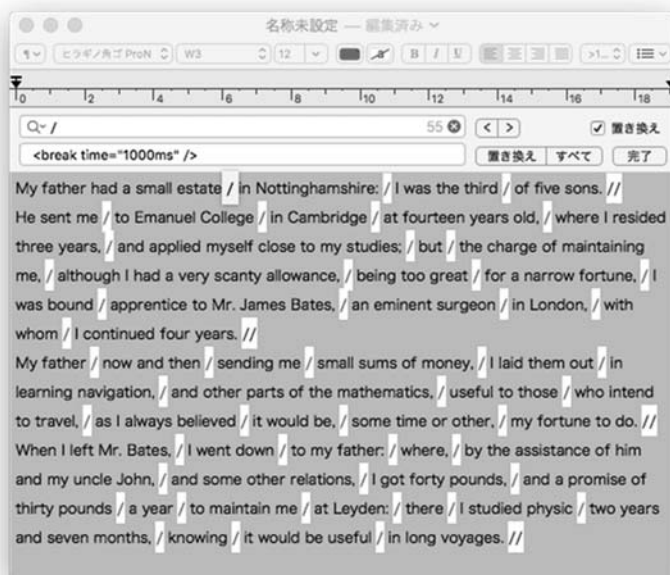
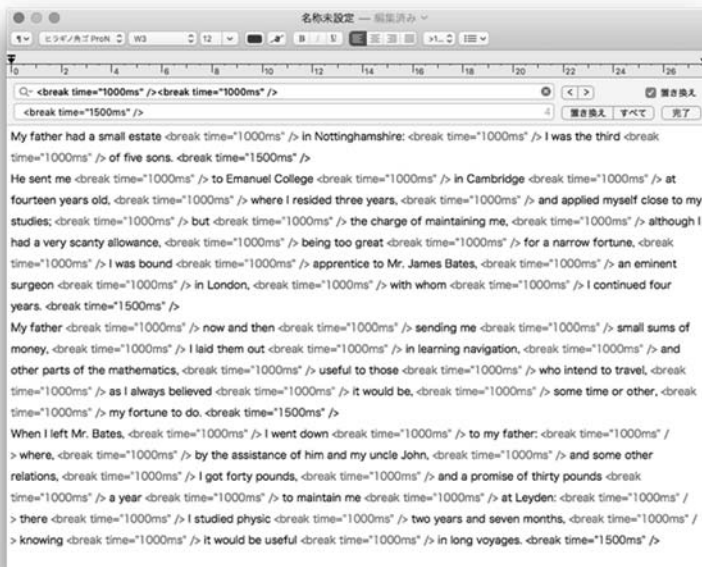


図7 第2回目の置換操作



最後に、置換作業後に、読み上げるテキストの前後に、「< speak>」と「</ speak>」を入力する。

SSML を含め XML 系のマークアップ言語ではタグなどを所謂半角文字で記述する必要がある。しかし、Microsoft Word 等の高機能なエディタには、オートコレクトなどの入力補完機能が存在し、例えば、「"1500ms"」のダブルクォーテーション「"」を「"1500ms"」のように「"」と「"」との組み合わせへと、本来の入力内容を無視して文字コード（番号）の異なるものへと勝手に置き換えるとか、IME（Input Method Editor）の設定によっては半角文字として入力された空白文字が全角文字での空白へと置き換えられるなどの問題がおこりうる。そのようになると、“break” タグであれば、無音声部分が正しく認識されなくなることから注意が必要となる。このため、この編集作業には、極力単純なテキストエディタを使用することが望ましい。

図 8 変換例

```
<speak>
My father had a small estate <break time="1000ms" /> in Nottinghamshire: <break
time="1000ms" /> I was the third <break time="1000ms" /> of five sons. <break
time="1500ms" />
He sent me <break time="1000ms" /> to Emanuel College <break time="1000ms" /> in
Cambridge <break time="1000ms" /> at fourteen years old, <break time="1000ms"
/> where I resided three years, <break time="1000ms" /> and applied myself close to
my studies; <break time="1000ms" /> but <break time="1000ms" /> the charge of
maintaining me, <break time="1000ms" /> although I had a very scanty
allowance, <break time="1000ms" /> being too great <break time="1000ms" /> for a
narrow fortune, <break time="1000ms" /> I was bound <break time="1000ms"
/> apprentice to Mr. James Bates, <break time="1000ms" /> an eminent surgeon <break
time="1000ms" /> in London, <break time="1000ms" /> with whom <break time="1000ms"
/> I continued four years. <break time="1500ms" />
My father <break time="1000ms" /> now and then <break time="1000ms" /> sending me
<break time="1000ms" /> small sums of money, <break time="1000ms" /> I laid them
out <break time="1000ms" /> in learning navigation, <break time="1000ms" /> and
other parts of the mathematics, <break time="1000ms" /> useful to those <break
time="1000ms" /> who intend to travel, <break time="1000ms" /> as I always
believed <break time="1000ms" /> it would be, <break time="1000ms" /> some time or
other, <break time="1000ms" /> my fortune to do. <break time="1500ms" />
When I left Mr. Bates, <break time="1000ms" /> I went down <break time="1000ms" />
```

3.3. スラッシュ英文の読み上げと調整

次に、先の Google の音声合成テキスト読み上げシステム Google Cloud Text to Speech のページ (<https://cloud.google.com/text-to-speech/>) に行く。このサイトでは、入力するテキストの扱い方に関して、text モードと ssm1 モードとの 2 つがある。開いた時点では、text モード(先の単純音声読み上げ)に設定されているため、ssm1 モードに切り替えるために、「ssm1」をクリックする (図 9 参照)。

最後は、英文の読み上げである。先のテキストをテキストエディタからペーストする。あとは先の読み上げと同様である。米語・英語、男性・女性音声などを指定し、「SPEAK IT」のクリックで、音声を読み上げられる。

再生された音声を聞き、無音声部分の長さを調整する。特に、スラッシュの間隔が短い箇所では、1 秒の無音では長すぎると感じられるかもしれない。こうした箇所に調整・修正を加える (i.e. “break” タグの無音声部分の時間を適宜変更する)。なお、固有名詞などでは、一部読み間違えが発生することがある。この場合には、“sub” タグを使用して読み替えを行う方法が

提供されている。例えば、

```
<sub alias="World Wide Web Consortium">W 3 C</sub>
```

として、「W 3 C」を「World Wide Web Consortium」と読み替えさせることができる。

図9 ssm1モードに切り替えた時の画面



4. 音声の録音とファイル化

教室でネットに接続する環境があれば、また、音声速度を変更したり、特定の部分を繰り返したりする必要がなければ、先のサイトをそのまま利用して授業時に使用することができる。しかし、この音声の話速変更や繰り返しなどを行いたい場合、受講生にファイルで渡したい場合(あるいは、単に、ネット環境に教室からアクセスできない場合など)、mp3などの音声ファイルとして保存し、それを活用¹¹したいと考えるであろう。

本節では、先の読み上げられた音声を録音し、ファイルとして保存する方法を説明する。基本的には、先の読み上げられた音声を録音することになるが、2つの方法が考えられる。ひとつ目はケーブル接続により音声を録音する方法であり、もうひとつはケーブル接続と同等の役割を提供する Sound Mixer 系のソフトを使用して音声を録音する方法である。

4.1. ケーブル接続により録音する際の事前準備

この方法は、スピーカー（ヘッドフォン）端子とマイク端子を使用して、スピーカー端子に出力された音声をマイク端子経由でコンピュータに録音するものである。コンピュータがスピーカー端子とマイク端子それぞれを有している場合と、両者が一体となっているコンピュータの場合がある。

¹¹ Google Cloud Text-to-Speech は1ヶ月あたり100万文字の無料枠があるものの有償のサービスである。常用するならば正規の利用手続きを踏む必要がある。

4.1.1. スピーカー端子とマイク端子があるコンピュータの場合

両端がミニプラグ (3.5mm) のステレオ・ミニジャックケーブルを、一方をスピーカー端子、もう一方をマイク端子に挿す。録音される音声をヘッドフォンなどでも聞きたい場合には、分配ケーブルを併せて使用すると良い。

4.1.2. スピーカー端子とマイク端子が一体型のコンピュータの場合

近年では、スマートフォンなどと同様にスピーカー端子とマイク端子が一体となっているコンピュータも多くある。最近の Macintosh コンピュータ (iMac, MacBook Air, MacBook Pro など) はサイト上での表現ではスピーカー端子のみではあるが、スピーカー端子とマイク端子が一体型のものが使われている¹²ようである。一体型端子の場合、スピーカー・マイク分配ケーブルと、上述の両端がミニプラグ (3.5mm) のケーブルを用意する。分配ケーブルのふたつの端子にステレオ・ミニジャックケーブルの両端を差す。再生音声を聞く場合は、もう一本の分配ケーブルを併せて使用することで対応する。

4.2. Sound Mixer 系ソフト<Soundflower>により録音する際の事前準備

ケーブル接続ではなく、Sound Mixer 系のソフトを使用しても同様のことが可能となる。Macintosh コンピュータの場合では、Soundflower というソフトを使用するのが簡単であろう。Soundflower はコンピュータ上で流される音を、特殊な機材を使用すること無く、録音を可能とするソフトである。まず、Soundflower をダウンロードする。Github 上にあるサイト¹³から、対応する OS 用のバージョン、今回の場合では、macOS Mojave (10.14) 以降用の Soundflower-2.0b2.dmg をダウンロードし、インストールする。インストール時には、同サイトにあるセキュリティ上の告知に注意しておく必要がある。

インストールできたら、設定を行う。「システム環境」, 「サウンド」の順にクリックし、入力の設定を「Soundflower (2ch)」, 出力の設定を「Soundflower (2ch)」にする (図10, 図11参照)。なお、出力装置として Soundflower を選択中は内蔵スピーカーからは再生音が流れない。

また、入力音量の調整が不適切な状態では、録音された音声のエコーが強すぎたり、ハウリングを起こしたりすることがある。図10で音量の設定を行うか、後述する録音するソフトウェア上で入力音量の調整を行う。Soundflower を使用する場合は、入力音量を最小とすることでほとんど解決できるはずである。

なお、言うまでも無いことではあるが、一連の録音作業が終了した後は、設定を元に戻しておく必要がある。つまり、「システム環境設定」から「サウンド」, そして、「出力」とクリックし、「内蔵スピーカー」を選ぶ。この作業を行うまでは、音声が内蔵スピーカーから出力されることは無いことに留意する必要がある。また、同様に「入力」から「内蔵マイク」に戻しておく。

¹² 発売元である Apple が公式には認めていないものの、動作結果から一体型であると判断できる。

¹³ <https://github.com/mattingalls/Soundflower/releases/> 2020年3月13日時点

図10 入力装置の選択



図11 出力装置の選択



4.3. 音声の録音と保存

Macintosh コンピュータの場合、デフォルトでインストールされている QuickTime Player を開き、「ファイル」から「新規オーディオ収録」を選ぶ。

真ん中の「丸い赤い録音ボタン」を押して（図12参照）、先の Google Cloud Text-to-Speech サイトで音声を再生させる。入力音量は下部の四角に囲まれた部分を左右に移動させ、調整する。なお、コンピュータ上の音声のみが録音されるため、話し声などに注意する必要は無いが、メール受信の通知音などは同時に録音されてしまう。必要のないソフトを終了しておくとか、通知音の設定を切っておくことが必要となる。

図12 QuickTime Player による録音



録音された音声に問題がなければ、保存する。「ファイル」から「保存」を選び、ファイル名、保存場所を指定し、「保存」をクリックする。指定された場所に、mp3 形式の音声ファイルが保存される。

5. 録音したファイルを再生

いったん、mp3 に保存できれば、プレーヤーを使用して、音声再生が可能である。スマートフォンやタブレットに付属しているプレーヤーを使用してもよいが、語学学習系の授業として考えるならば、話速を変換できたり、特定箇所をリピートできたりする機能を有したプレー

ャー¹⁴を使用することで、より学習効果が上げられるであろう。

6. おわりに

本稿では、英語学習のためのスラッシュ・リーディングに有効なスラッシュごとにポーズを設ける読み上げ／朗読音声を音声合成読み上げシステム（Google Cloud Text-to-Speech）を用いて作成する方法と、Macintosh コンピュータにて、それを録音し、音声ファイルとして保存する方法について解説した。

本稿では、技術的な方法のみの説明であったが、これらの方法の有効性の検証、改善点などの検討、音声合成システムによる音声を活用した効果的な授業方法の考案などを今後行う予定である。

¹⁴ スマートフォンやタブレットであれば、例えば、Audipo (iOS のみ) は0.5～2.0倍までの話速変換や細かく繰り返し箇所が設定可能である：

<https://apps.apple.com/jp/app/id607971056> 2020年3月24日時点