リアルタイムMRI動画日本語調音運動データベースの設計

著者	前川 喜久雄, 西川 賢哉, 浅井 拓也, 能田 由紀子, 正木 信夫, 島田 育廣, 竹本 浩典, 北村 達也, 斎藤 純男, 籠宮 隆之, 石本 祐一, 菊池 英明, 藤本 雅子, 八木 豊
雑誌名	言語資源活用ワークショップ発表論文集
巻	5
ページ	209-230
発行年	2020
URL	http://doi.org/10.15084/00003161

リアルタイム MRI 動画日本語調音運動データベースの設計

前川喜久雄（国語研）[†]，西川賢哉（国語研），浅井拓也（早稲田大），能田由紀子（国語研），正木信夫（ATR-Promotions），島田育廣（ATR-Promotions），竹本浩典（千葉工大），北村達也（甲南大），斎藤純男（拓殖大），籠宮隆之（国語研），石本祐一（国語研），菊池英明（早稲田大），藤本雅子（早稲田大），八木豊（ピコラボ）

Design of Real-Time MRI Articulatory Movement Database

Kikuo Maekawa (NINJAL), Ken'ya Nishikawa (NINJAL), Takuya Asai (Waseda University), Yukiko Nota (NINJAL), Shinobu Masaki (ATR), Yasuhiro Shimada (ATR-Promotions), Norihiro Takemoto (Chiba Institute of Technology), Tatsuya Kitamura (Konan University), Yoshio Saito (Takushoku University), Takayuki Kagomiya (NINJAL), Yuichi Ishimoto (NINJAL), Hideaki Kikuchi (Waseda University), Masako Fujimoto (Waseda University), Yutaka Yagi (Picolab)

要旨

われわれは、日本語に関する調音音声学的研究の新しいインフラ提供をめざして、調音運動データベースの構築を進めてきている。声道全体の形状変化を毎秒 14 ないし 25 フレームのリアルタイム MRI 動画として記録したデータが、現時点で東京方言 16 名、近畿方言 5 名分収録済である。1 名あたりの発話量は 25〜30 分である。データには個々の発話の開始時刻と終了時刻のタグが付与されており、他に発話内容と話者に関するメタデータを検索に利用できる。

１．はじめに

われわれは 2017 年以来、科学研究費の補助を受けて、日本語音声の調音運動に関するデータベースの構築を進めてきた。本稿では、この研究プロジェクトの背景と目標を紹介した後、成果の現状を報告する。本稿の主目的はデータベースに関する情報提供であるが、最後にこのデータベースを用いた研究上の成果にも簡単に触れる。

１．１　調音音声学とデータ

音声学は言語研究全般の基礎科学として位置づけられることが多いが、わけても調音音声学(articulatory phonetics)の知識は、音韻論・形態論に始まる言語の記述的研究や音韻論における弁別素性理論の基礎知識として重要な役割を果たしている。一般に調音音声学を含む音声学全般は、言語研究の諸領域のなかでもっとも自然科学に近い方法に基づいているものと理解されており、言語学者のなかには音声学を言語研究に必要な自然科学の一領域として理解しているむきもある。

[†] kikuo@ninjal.ac.jp

しかし調音音声学の基礎は、実は自然科学と呼ぶにはかなり脆弱である。例として、音声学の教科書には必ず掲載されている正中矢状断面図による調音位置の説明を考えてみよう。あの断面図は、ほとんどの場合、客観的な観測データに基づいて描かれたものではない。教科書の著者が、調音音声学の知識に基づいてかくあらんと想像した音声器官の状態を表現したものであり、心理学で言う認知地図としての性格を帯びている。

認知地図は人間が脳のなかで感じとる認知事象を理解や解釈・推論のために図式化したものであるから、当然、主体となる人間による異同が生じうる。何らかの点で主体 A と B の認知地図間に齟齬が見出された場合、現在の調音音声学では、両者の主張の優劣をデータに基づいて客観的に判断することが多くの場合に不可能であり、水掛け論に終始する可能性が高い。

このような調音音声学上の問題を解決するためには、ふたつの条件が満たされる必要がある。まず、調音音声学の基礎をなす調音運動を客観的に観測し、定量的に測定する手法が開発され普及すること、そして、測定されたデータが公開され共有化されることのふたつである。

このうち第二の要請はいわゆるオープンデータ・オープンサイエンスの問題である。国立国語研究所は 1960 年代に本データベースと問題意識を共有する貴重な X 線映画を撮影しており、その分析に基づく大部の報告書を 2 冊刊行しているが（国立国語研究所 1978, 1990）、X 線映画資料そのものは 2010 年代に入るまでながらく非公開であった。このことは、上述の報告書の評価にも大きく影響したと考えられる。この反省にたって、われわれは 2016 年秋に科学研究費を申請するに際して、当初から研究終了後の適当な時期にデータを一般公開することを前提とした計画を立案した。本論文の末尾に述べるように、以下に紹介するデータベースの一部は近く公開する予定である。

次に、上記の第一の要請については、19 世紀末の実験音声学の誕生以来、様々な手法が開発されてきた。いま対象を、調音音声学にとって最も重要な音声器官であり、同時に客観的な観察が困難な器官でもある舌の形状に関する情報を得るための手法に限っても、静的 X 線写真、X 線映画、X 線マイクロビーム装置、超音波断層撮影、EMA と総称される Electromagnetic Articulography 装置・WAVE 装置などの手法が開発され、利用されてきている。

ただ、これらの手法には様々な制約が存在した。X 線を利用した手法には被曝の危険性があるため、同一の被験者から多量のデータを集めることができない。X 線映画の場合、だいたい 10 数分が上限である。X 線マイクロビーム装置は髪の毛よりも細く絞られた X 線ビームをコンピュータ制御して、舌などの音声器官に接着されたペレット（金属球）の位置を予測し観測するシステムであり、被曝量を実際上無視できる程度まで軽減させることに成功した。ただし、この装置によって得られる情報は、ペレットの点位置情報であり、同時に追跡可能なペレットは最大で 10 個ほどであるため、声道の形状を全体として把握することは困難であった。また装置が極めて大規模であったため設置・運用できる組織に強い制限があった。

EMA 装置は、静電誘導の原理を利用して、磁界中を移動するセンサーコイルの位置を記録する装置である。被曝の可能性は原理的に存在せず、同一個人からの多量データ収集が可能であるが、X 線マイクロビームと同じく、得られるのは点的なセンサーの位置情報に限られる。

超音波断層撮影も原理的に被曝の危険性はない。また組織の断面の情報を得ることができる点で EMA よりも情報量が多い。ただし、組織の表面で生じる超音波の反射波（エコー）を捉えて画像を構成するという原理上、空間的に隔てられたふたつの器官を観察すること

ができない。例えば下顎にプローブをあてて舌の形状を観察した場合、口蓋の形状を観察することはできない。

１．２　リアルタイム MRI 動画データの特徴

以上の諸手法に比べると、われわれがデータ収集に利用しているリアルタイム MRI 動画法（以下 rtMRI と書くことがある）は、多くの面ですぐれた特性が認められる。rtMRI は、医療用 MRI 装置を次節で説明する特殊な設定で稼働させることによって、毎秒数十フレームの時間解像度の動画を撮像する技術である。リアルタイムという形容詞が用いられるのは、多数繰り返された発話からひとつの動画を構成する時間同期法(Masaki et al. 1999)ではないことを示した名称である。

rtMRI の特長は、声道を含む頭部矢状断面の全体像が高い空間解像度で把握できる点にある。図 1 に示すように、声道を構成する唇・舌・口蓋・喉頭蓋・咽頭壁・喉頭などが明瞭に画像化されている。図 1 の場合、図全体で 256×256 ピクセルの解像度があり、1 ピクセルは 1 ㎜に該当するので、子音調音における狭窄の発生位置や母音調音における舌と口蓋の距離などの情報を正確に把握することができる。さらに、前鼻棘・後鼻棘や頸椎などのように、声道形状の正規化のために利用される器官の位置情報も入手できる。また時間解像度が十分とはいえないものの、音声器官の運動が可視化されることによって、調音音声学の教材としては卓越した価値がある。

図 1 : rtMRI 画像の例（休止状態の声道）

一方、rtMRI 法にもいくつかの問題ないし限界がある。①まず現在の rtMRI がとらえているのは正中矢状断面の情報である。現在の撮像速度を維持する場合、冠状面の情報を同時に取得することはできない。②次に MRI 装置は稼働時にかなり大きな騒音を発するので、実験時に収録した音声にはその騒音が重畳されてしまう。③また被験者は仰臥位（うわむきに横たわった体位）で発話を行うので、日常発話の大部分とは重力のかかる方向に違いがある。そして、④われわれの撮像条件における時間分解能は毎秒 14 ないし 27 fps (frame per second) であり、音声生成研究で要求される分解能として必ずしも十分でない。

このうち④は、われわれの撮像条件に限った問題であり、技術上の限界ではない。実際、海外では毎秒 100 フレーム以上での撮像も行われている（6 節参照）。その場合、例えば毎秒 90 フレームの撮像が可能であれば、毎秒 30 フレームで矢状断面を 3 箇所撮像することが可能になるので、①と④は実際上同一の問題（撮像速度）に帰着する面がある。

②については、デジタル信号処理によって、耳に聞こえるノイズを相当程度まで低減させることが可能である。われわれのデータにもその処理を施しているが、強力なノイズ軽減処理は、音声の周波数特徴にも影響を及ぼすので、その匙加減が問題になる。

③については、体位が音声に及ぼす影響についていくつかの研究が行われているが、報告された影響は研究毎にかなり異なっている（Kitamura et al. 2005; Shiller, Ostry & Gribble

1999; 能田ほか 2019）。大方の結論として、体位の影響は否定できないが、音声研究において rtMRI データが有する価値を毀損するほどのものではない（したがって rtMRI データベースは公開する価値がある）というのがわれわれの判断である。

２．データベースの内容

本節ではわれわれのデータベースにどのような発話が収録されており、どのようなメタデータが提供されており、どのようなアノテーションが施されているかを説明する。本データベースを構成する発話項目を表 1 に示した。このうち「発話項目」については 2.1 節で、それ以外については 2.2 節で説明する。

２．１　発話項目

MU(mora unigram)は現代日本語で用いられるほぼすべてのモーラを周辺的なものまで含めて収録している。一部に母音間のラ行子音（アラ・アリ・アル…）や長母音（アー・イー・ウー…）、連母音（アイ・アウ・アエ…）などの 2 モーラからなる項目が含まれている。

MB(mora bigram)は日本語の調音結合を定量的に観察するために考案された項目である。26 個のモーラ「カキクケコキャキュキョハヒフヘホサシスセソシャシュショマミムメモ」のすべての組み合わせ「カハ・カヒ・カフ…」等 676 個を「これが＿＿型」というキャリアセンテンスに入れて収録した。26 個のモーラは子音音素として/k, kj, h, s, sj, m/のいずれかを含んでいる。これらの子音は、調音位置（両唇音・歯茎音・硬口蓋音・軟口蓋音）と調音様式（破裂音・摩擦音・鼻音）と音韻的な口蓋化（拗音化）に配慮して選択した。また上述のキャリアセンテンスで用いられている接尾辞「〜型」は前節する語のアクセントを消す作用をもっているので、MB 項目はすべて無核語として発音されている。

MP(mora phoneme)は、日本語の音韻を特徴づけるモーラ音素を含む語のリストである。大部分は「新案・真円・心音・心因…」のような有意味語であるが、一部に「ケヘ・バダ・スッ…」のような無意味語も含まれている。

TT(tongue twister)には以下の 3 個の早口言葉が含まれる。「菊栗、菊栗、三菊栗、あわせて菊栗、六菊栗」「この竹垣に竹立てかけたのは竹立てかけたかったから竹立てかけたのだ」「虎を捕るなら虎を捕るより鳥を捕り鳥を囮に虎を捕れ」。

PL(para-language)には 2 種類のパラ言語情報に関する課題が含まれる。ひとつは意図を指定して同一のテキストを読み分ける課題である。テキストは「山田さんが」であり、「反問」「感心」「落胆」の意図、およびパラ言語的な含意のない「中立」の 4 種類を発話する。もうひとつの課題は対比の強調(contrastive focus)である。「家賃の高いマンションに入った」というテキスト中の「家賃の」「高い」「マンションに」の部分を対比強調する発話と対比強調を含まない発話の 4 種類で構成される。

NS(narrative story)は、まとまった意味をもつ文章 2 種類の朗読である。ひとつはグリム童話の「北風と太陽」、もうひとつは自然科学に関する新書から抜粋した「膨張する宇宙」である。後者は『日本語話し言葉コーパス』の朗読課題のひとつとして用いたものを再利用した（籠宮 2004）。

SR と TEST は一部の話者の収録時に試験的に実施した項目である。SR(speaking rate)は、MU, MB, MP 項目に含まれる一部の発話を、通常よりも低い発話速度発話する課題である。TEST は本データベースの構築開始後に 27 fps での撮像が可能になったのを受けて、14 fps データとの比較のために、MU, MB, MP 項目の一部を撮像したものである。話者数は SR が 8 名、TEST が 11 名である。

表 1：データベースに収録された発話項目

項目クラス	内容	発話数	スライド数	項目追加
MU	単独モーラ	109-142	5	有
MB	26 種のモーラの組み合わせ	676	34	無
MP	撥音・促音・二重母音・長母音を含む語	100-149	6	有
TT	早口言葉	0, 3	1	無
PL	パラ言語情報を指定した発話	0, 8	2	無
NS	文章の朗読	0, 1, 2	2	有
SR	発話速度を落とした発話	0, 31	1	有
TEST	27 fps での収録	0, 45, 52	2	有

２．２　スライドとセッション

被験者は MRI 装置内部のスクリーン（正確にはコンピュータスクリーンを映した鏡）に投射されるスライドを読み上げる形で発話をおこなう。表 1 の「スライド数」は、1 回のデータ収録作業（以下「実験」と呼ぶ）で被験者に示されるスライドの枚数を示している。スライド 1 枚には複数の発話項目が掲載されており、例えばＭＵ項目のスライドであれば、27〜32 項目、MP 項目であれば 25 項目、MB 項目であれば 20 項目が印刷されている。各スライドに印刷された項目数の合計が表 1 の「発話数」である。

実験では、スライド 1 枚ごとに MRI 装置を稼働させて、36 秒（14 fps の場合）ないし 19 秒（27 fps の場合）の連続した調音運動を記録する。この 1 回の連続した記録をセッションと呼んでいる。

理想的に実施された実験では、表 1 のスライド数は MRI のセッション数と一致する。しかし現実には、さまざまな理由で、1 枚のスライドが複数のセッションに対応することがある。例えば、発話速度が遅すぎたり、セッションの途中で読み直しが数回生じたりすると、37 秒内にスライドの最後まで発話できないことがある。そのような場合は、１枚のスライドを複数（通常は 2 回）のセッションにわけて記録することになる。その際、2 回目のセッションでは、1 回目で発話しきれなかった項目だけを発話するのではなく、スライドの途中から初めて最後まで進み、その後スライド冒頭にもどって発話を継続することにしている。また TT や PL 項目ではスライド 1 枚分の発話が十数秒で終了することが多いので、同じスライドを繰り返し発話するよう話者に指示をあたえている。そのような理由によって、本データベースには、同一話者の同一実験において、同一の発話項目が 2 回以上記録されていることがある。こうした情報はメタデータの一部としてデータベースに記録される。

２．３　非発話データ

以上の発話項目に加え、一部の話者では歯列の情報および静止状態にある声道の立体情報を得るための撮像も実施した。MRI の撮像原理上、水分・脂肪分を含まない骨は撮像することができない。しかし、前歯の形態は調音音声学上重要であるし、声道の全体形状を知るためには歯列全体の情報も必要である。

われわれの実験では、話者に前舌面を上顎歯列に強く密着させつつ舌尖を上下の前歯で軽く噛んだ状態を維持することを要求し、その状態で矢状面を左右に 1 ㎜ずつずらした撮像を繰り返すことで、歯列の情報を得た。その際、発話は行っていない。

また 5 母音および一部の子音について、静止状態にある声道の立体形状を同じ方法で撮像した。矢状断面(sagittal plane)を多数撮像するので、パラサジタル(para-sagittal)なデータと呼んでいる。これらのデータは、本データベースには格納されないが、別途、利用者が利用可能にしたいと考えている。

２．４　データ収集の方法

データ収集は㈱ATR-Promotions の脳活動イメージングセンタ(ATR-BAIC, 京都府相楽郡)において、磁界強度 3T の MRI 装置（MAGNETOM Prisma 3T, Siemens)を用いて実施した。空間分解能は 256 × 256 ピクセル（1 ピクセルは 1 ㎜）、時間分解能は 14 fps を原則とし、一部で 27 fps での撮像もおこなった。スライス幅は 10 ㎜である。rtMRI の撮像技術については、付録にまとめて記載したので、興味のある読者は参照されたい。

実験は、1 コマ 90 分を単位として MRI 設備を利用する形で実施した。1 名の実験に要した時間は、事前の準備を含めて 90〜120 分程度であり、そのうち話者が実際に MRI 装置内に入って発話を行った時間は 45 分〜60 分程度であった。これによって 25〜30 分の発話に対応する rtMRI データが収集される。

発話実験中は MRI オペレータが頻繁に声掛けをするなどして話者の体調に配慮し、随時休憩を含めながら実験を進めた。話者は体調の以上に気づいた場合、実験中であっても随時実験の中止を要請できることになっているが、幸い、そのような事態はこれまで一度も生じていない。

MRI 装置によって記録されるのは人体を断層撮像した画像情報だけであり、提供ファイル形式は DCM 形式である。話者が発した音声は口元の光マイクロホンを介して別途 DAT にサンプリング周波数 48 kHz, 分解能 16 bit で録音される。われわれは ATR-BAIC から提供される画像および音声データを合成して、AVI 形式と MP4 形式の 2 種類の動画ファイルを構成して利用している。合成の際、各フレームの右下にフレーム番号を挿入して、アノテーション作業および分析の利便性を高めている。

２．５　データの現状

本稿執筆の時点（2020 年 7 月下旬）で、データ収録を終えた話者数は、東京方言が 16 名（男性 11 名女性 5 名）、近畿方言（大阪・神戸）が 5 名（男性 3 名女性 2 名）である。生年代は 1950 年代 7 名、60 年代 7 名、70 年代 4 名、90 年代 3 名であり、平均年齢は 53 歳である。

表 1 の「発話数」列の数字は「109-142」「0, 45, 52」のように幅を与えられていることがある。これはデータベース構築の過程で、調査票が拡張されたり、改変されたりしたことによる変動を示している。数字「0」は、その項目クラスが実施されなかった実験があることを示している。全被験者が共通して実施した項目クラスは MU, MB, MP の 3 クラスである。そのうち MB だけは全実験を通じて発話数が一定であるが、他のクラスは多少とも変動している。実験期間全体を通しての傾向として、MU と MP が大幅に拡張されている。その結果、実験時間が不足するようになったため、ある時期以降の実験では、当初は必ず実施していた TT と PL を省略することがあった。

このような発話項目の拡張の結果、初期の実験に参加した話者と後期の話者とで発話内容に異同が生じることになった。現在、この問題を解消するために、初期の実験に参加した話者からの追加収録実験を実施している最中である。

３．アノテーション

本節では rtMRI 動画に付与されたアノテーションと、そのためにわれわれが行った技術開発について述べる。

３．１　発話開始・終了時刻

本データベース検索のために最も重要なアノテーションは、個々の発話項目の開始・終了時刻のアノテーションである。検索システムは、この情報を用いて、検索条件に該当する発話を検索して切り出し、再編集して出力する。

通常の音声信号を主体とした音声データベースでは、発話の切り出しは音声信号中のポーズに依拠して実施されることが多い。その場合、発話頭に/k/, /t/, /p/などの閉鎖音が位置する発話では、閉鎖音の破裂調音の直前に発話開始時刻を設定することが多い。

しかし調音運動データベースの場合、事情が異なり、音声信号としては実現されない調音運動もデータに含めなければならない。そのような運動は、閉鎖音の閉鎖区間だけではなく、発話頭に生じるすべての子音・母音に多少とも存在している。本データベースの発話切り出し作業では、対象とする発話のための調音運動が開始される直前の、声道が休止状態にある時刻から、対象発話の調音運動が終了して、声道が休止状態に復帰した時刻までを切り出すことを原則とした。

先に示した図 1 は休止状態にある声道の典型例を示している。その特徴としては、発話の前後では口蓋垂が下降していて鼻腔を介した呼吸が可能な状態にあることと、口腔中に顕著な閉鎖ないし狭窄がないことの二点が挙げられる。そして、(a)調音運動の開始に先立って口蓋垂は挙上され、(b)口腔は発話冒頭の音素に対応して変形しはじめる。また(c)発話末では口蓋垂が再び下降する。

大部分の発話では、開始時刻については(a),(b)のいずれか両方、終了時刻については(c)の特徴の実現に注目することで発話区間を決められる。しかし、なかには例外的な扱いを必要とする発話もある。例えば、鼻子音で始まる発話の場合、発話開始時にも鼻腔への通路は閉鎖されない。ただし鼻子音調音時の口蓋垂は典型的な休止状態（呼吸時）よりも高く位置することが多いので、それが参考になることが多い。母音で始まる発話では発話開始時にも声道の顕著な狭窄は存在しないが、口蓋垂の特徴と、声道が図 1 のように中立的な状態にあれば、そこからの変位が始まる時刻を参考にして、開始時刻を決めることができる。

そして、少数ではあるが、本当に発話の開始・終了時刻の決定が困難な発話もある。例えば、発話速度が大きい場合に、音響的なポーズが生じていても、その区間に調音的な休止が生じないことがある。撥音で終わる発話に鼻子音で始まる発話が後続する発話の場合、先行発話末尾から後続発話冒頭にかけて、口蓋垂は挙上されることがなく、また先行発話末の撥音の直後の無音区間において既に後続発話冒頭の子音のための閉鎖ないし狭窄が形成されていることがある。このようなケースでは、音響的な無音区間の中央に境界を設定し、声道の休止状態が認定できなかったことを示すタグ[noRP]を付与している（他のタグについては 3.3 節参照）。

３．２　音素列

検索のためにもうひとつ重要であるのが、発話内容をテキスト化した情報である。本データベースでは、スライドに印刷された文字列とその疑似音素表記（前者を text、後者を phoneme と呼んでいる）を情報として提供する。Text と phoneme の対応関係は大部分単純であるが、いくつか注意すべき点がある。

① 短母音は/a, i, u, e, o/、長母音は/aH, iH, uH, eH, oH/
② カ行直音は/ka, ki, ku, ke, ko/、拗音は/kja, kju, kjo, kje/（直音のキが/kji/でないことに注意）
③ タ行直音子音は/ta, ci, cu, te, to/。拗音は/cja, cju, cjo, cje/
④ ハ行直音は/ha, hi, hu, he, ho/。拗音は/hja, hju, hjo, hje/
⑤ ヤ行直音子音は/ja, ju, jo, je/
⑥ 促音は/Q/、撥音は/N/
⑦ 二重母音と連母音は区別しない
⑧ ファ行「ファ・フィ・フェ・フォ」は/fa, fi, fe, fo/（ハ行のフとは別子音）
⑨ 外来音の「スィ・ティ・トゥ・ズィ・ディ・ドゥ」は/s_i, ti, tu, z_i, di, du/
⑩ 外来音の「フュ」は/fju/, 「デュ」は/dju/

一部の phoneme には補助情報を埋め込んだものがある。同音語である「貝」と「下位」、「帰る」と「飼える」などは、/kai_1/と/kai_2/、/kaeru_1/と/kaeru_2/のように添字で区別されている。また「紹介状」と「消化異常」、「里親」と「砂糖屋」のように形態素境界が異なる疑似的同音語を/sjoHka-izjoH/と/sjoHkai-zjoH/、/sato-oja/と/satoH-ja/のように形態素境界を示す補助記号で区別した例がある。ただし、すべての形態素境界が示されているわけではないことに注意。

3．3　発話の誤り

話者が何らかの理由で発話を間違えることがある。もっとも頻繁に生じるのは、スライドに印刷された文字の読み間違い、特に濁音と半濁音の混同である。MRI 装置内では眼鏡やコンタクトレンズを装着できないため、プラスチック製レンズで臨時に視覚矯正具をつくって、それを装着してもらっている。しかし矯正が完全とは限らず、読み間違えの原因になることがある。また漢字を読み間違えることも稀にある。さらに、発話の途中でリズムが乱れることもある。よくあるのは、語中に短いポーズを含んだ発話である。

こうした誤りには話者自身が気付くことが多い。その場合、話者は同じ発話をその場でもう一度繰り返して発音するよう指示されている。話者が気付かなかった誤りに実験者が気付いた場合は、そのセッションが終了した後に誤りを指摘し、次のセッションで当該スライドを再度収録している（これも同一項目に同一話者による複数の発話が存在する原因になる）。ただし、実験者が誤りを聞き落とすこともあるし、時間の制約が厳しい場合、リズムの乱れのように相対的に軽微な誤りの場合は、あえて再度の収録を行わないこともある。その結果、誤りを含む、あるいは完全ではない発話がデータに含まれることがある。

こうした発話に対する対処方針は以下のとおりである。まず、話者が自分で修正したケースでは、誤った発話はアノテーションの対象外とし、修正した発話にだけ開始・終了時刻を付与する。次に、実験者が誤りに気付いて、次のセッションで正しい発話が収録された場合も、誤った発話は対象外とする。何らかの誤りを含む発話しか収録されていない場合は、発話に以下のタグを付与する。これらのタグはアノテーション作業の段階では発話内容の音素列（3.2 節参照）に続けて記入し、データベースを構築する際の後処理で分離している。

① 発話中に不要なポーズが含まれる発話には[pz]タグを付与する
② 発話のリズムに乱れが感じられる発話には[d]タグを付与する

③ 明らかな読み間違い発話には[err;]タグを付与する。このタグではセミコロンの直後に誤った発話の音素表記を記入する。例えば「フュージョン」が誤って「フュージュン」と発話された場合であれば[err; fjuHzjuN]になる

発話の誤りは、実際にアノテーションを始めてみて、初めて気づくものも少なくない。そのため誤りに関するタグについては現在も試行錯誤を続けている。

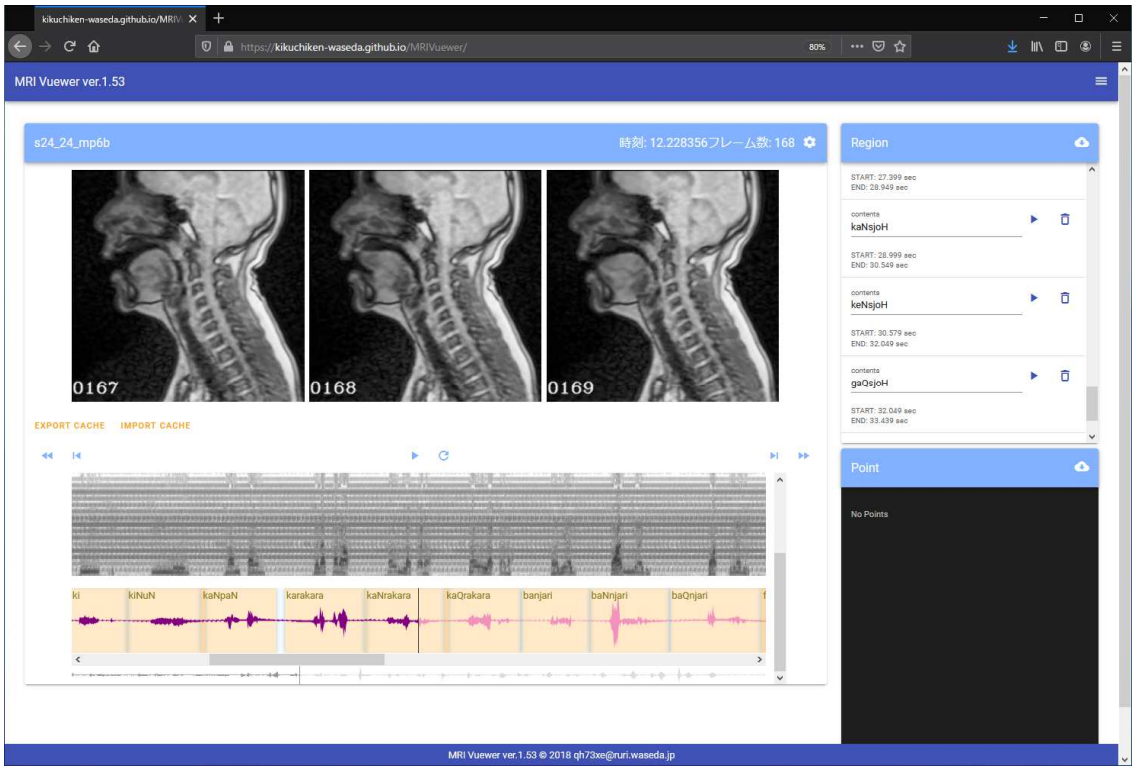


図 2: MRI Vuewer Ver. 1.53 の動作画面（ブラウザは windows 10 の firefox）

3．4　データビューワー

以上 3 種類のアノテーションを実施するためには、収録された連続画像を音声と関連づけながらブラウジングする必要がある。MRI データの標準形式である DCM 形式ファイルの操作環境としては米国国立衛生研究所(NIH)で開発されたオープンソースソフトウェア ImageJ や、ImageJ を含むパッケージである Fiji が有名であるが、これらの環境では音声データを参照することができない。そこで、われわれは独自のデータビューワーを開発した（浅井・菊池・前川 2018）。

このソフト（MRI Vuewer）は Java script で開発されており、ウェブブラウザ上で稼働する。[1] MP4 形式動画ファイルを読み込んで、画面上部に連続する 3 フレームの rtMRI 画像を、画面下部にサウンドスペクトログラムと音声波形を表示する（図 2 参照）。中央の画像が音声波形上のカーソル位置に対応している。

[1] Vuewer の語は開発に利用した Java Sript のライブラリ名（vue.js）に由来する。

ユーザーは、音声波形画面をマウスでドラッグすることで任意の時間区間を指定することができ、その区間にアノテーション用文字列を付与することができる。画面右端上部には、付与された文字列が時間情報とともに表示されている。この文字列の右側の▶ボタンをクリックすると、当該区間の rtMRI 動画と音声が再生される。前節に述べたアノテーション作業はすべて、このデータビューワーを用いて実施している。本ソフトにはさらに特定のフレームを対象に MRI 画像の測定を行う動作モードもある。後述する研究成果のための測定は、この動作モードを利用したものが多い。本ソフトウェアは現在も機能拡張中であるが、暫定安定板が GitHub で公開されている。[2]

3．5　音声器官輪郭抽出

アノテーションに関して最後に現在開発中の重要な技術に触れる。rtMRI 動画を音声学の研究に利用しようとする場合、多くの場合、ユーザーは注目する音声器官の位置やその時間変化を測定することになる。測定対象は、調音音声学の場合、舌輪郭の最高点の位置、舌尖の位置、奥舌面の位置、口蓋垂の位置、舌と口蓋の接触点（始端と終端）等々様々であるが、このような情報はいずれも特定の音声器官と結びついた形で定義されている。従って、もし MRI 画像が声道を構成する種々の音声器官に分割されており、各器官の輪郭があらかじめ抽出されていれば、測定の労力を大幅に軽減できる。この可能性を念頭において、われわれは研究開始当初から、MRI 画像から音声器官の輪郭を自動抽出する技術の確立を重要な研究目標としてきた。現在は、機械学習を用いることで、単一話者であれば、少量の教師データによる学習で人手と同等以上の精度を保った自動抽出を実現できている（Takemoto et al. 2019)。この学習では、顔学習に広く利用されているオープンソースの機械学習ライブラリ dlib を利用している。

図 3: 輪郭抽出された rtMRI 画像の例

図 3 に自動抽出された音声器官の輪郭の例を示す。ここで声道は、舌(赤)、口唇・下顎（黄）、口蓋（緑）、咽頭後壁・披裂部（紫）、および喉頭蓋・声帯（水色）の 5 部位に分割されて推定されており、各部位は有限個の点座標の順序集合として表現されている。このような数値データから、例えば舌輪郭の最高点や最奥点を検索するのは極めて容易である。また舌尖と舌端の相対的な位置関係なども、両部位の境界に関する適当な仮定を導入することができれば、容易に知ることができる。さらに舌輪郭と口蓋の接触の有無や接触部位の情報なども若干の工夫で取得できると考えられる。

音声器官の輪郭自動抽出技術は現在も開発が進められている。ここ 1 年の課題は、学習データ以外の話者のデータを分析した場合の精度を高めることにあったが、最近、声道形状を主成分に従ってタイプ分けすることによって重要な進展が得られた。

[2] https://kikuchiken-waseda.github.io/MRIVuewer/

音声器官の輪郭情報は、調音音声学の研究の利便性を高めるだけでなく、音声科学の基礎研究への貢献も期待できる。多くの話者のデータから平均声道を構成することができるが、平均声道からの逸脱をもって、逆に個々の声道の個性を表現することができる。そこから音声の個人差研究に新しい展望が開ける可能性がある(後藤ほか 2020)。

3．6　データベース構築工程のまとめ

ここまでに説明したデータベースの構築工程を模式化して図 4 に示す。上段の文字列は工程の名称である。中段の矩形群のうち、実線の矩形は各工程で生成されるデータの種別を示しており、角が丸まっている矩形はアノテーション情報である。下段に位置する破線の矩形は、各工程で利用される主要な装置・ツール類を示している。

現在利用可能なデータベースは、mp4 形式動画データと発話時刻情報、音素情報、誤りタグ、メタ情報（主に話者情報）から構成されている。音声器官の輪郭情報は最終的なデータベースには含める予定であるが、データの生成にいま少しの時間が必要である。

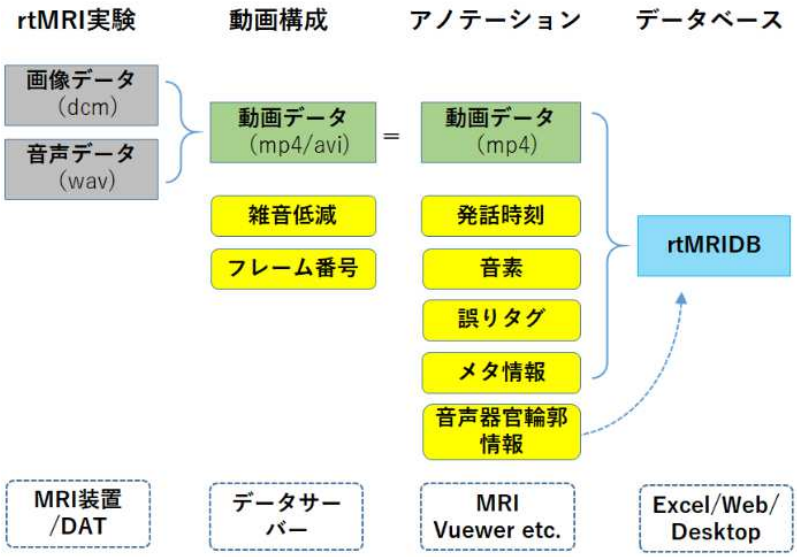


図 4: データベースの開発工程

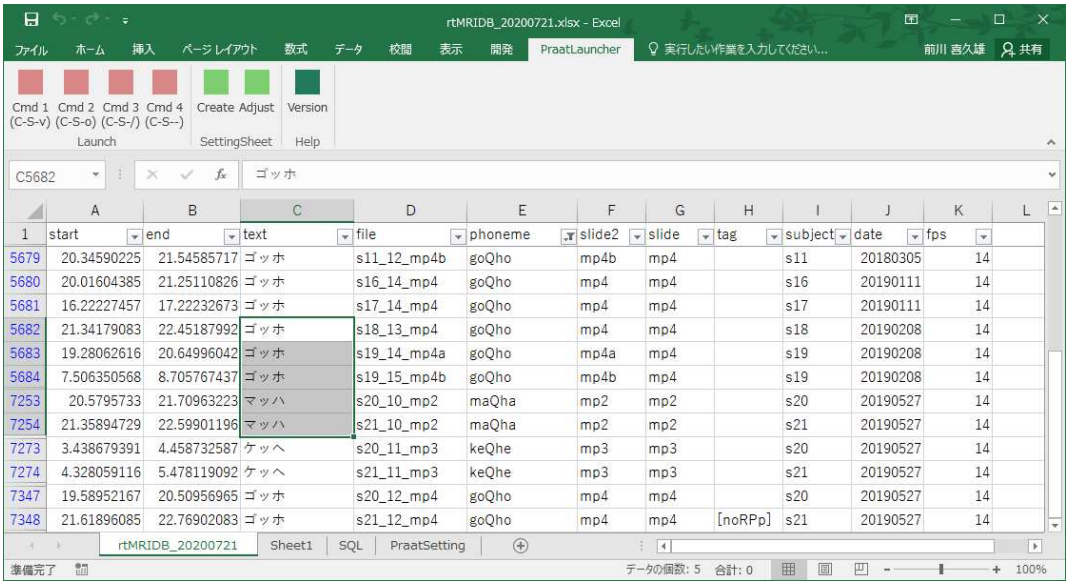
4．データベース検索環境

本節では rtMRI 動画データの検索環境を紹介する。本稿執筆時点で、われわれが利用しているのは、表計算ソフト Excel の開発言語 VBA とオープンソースの動画処理ソフトウェア ffmpeg を利用して実装した検索環境である。[3]

図 5 に検索結果画面を示す。音素列/Qh/（ハ行子音直前の促音）を含む発話を Excel のフィルタ機能を用いて検索した結果であり、51 件がヒットしている。この状態で rtMRI 動画を視聴したいサンプルに対応する行を選択し、指定されたキー操作（例えば CTRL + SHIFT + [）を行うと、選択された発話群がそれらを含む mp4 形式動画ファイル群から切り出されたうえ、1 個の mp4 形式動画ファイルに結合されて、指定された動画ビューワーで再生される。図 5 では 5682 行以下の 5 サンプルが選択された状態にある。現在の仕様で

[3] 開発は本稿の第二著者が行った。

は、選択サンプル（選択行）数が 20 を超えると警告が発せられるが、警告を無視してそのまま処理を実行することもできる。



	start	end	text	file	phoneme	slide2	slide	tag	subject	date	fps
5679	20.34590225	21.54585717	ゴッホ	s11_12_mp4b	goQho	mp4b	mp4		s11	20180305	14
5680	20.01604385	21.25110826	ゴッホ	s16_14_mp4	goQho	mp4	mp4		s16	20190111	14
5681	16.22227457	17.22232673	ゴッホ	s17_14_mp4	goQho	mp4	mp4		s17	20190111	14
5682	21.34179083	22.45187992	ゴッホ	s18_13_mp4	goQho	mp4	mp4		s18	20190208	14
5683	19.28062616	20.64996042	ゴッホ	s19_14_mp4a	goQho	mp4a	mp4		s19	20190208	14
5684	7.506350568	8.705767437	ゴッホ	s19_15_mp4b	goQho	mp4b	mp4		s19	20190208	14
7253	20.5795733	21.70963223	マッハ	s20_10_mp2	maQha	mp2	mp2		s20	20190527	14
7254	21.35894729	22.59901196	マッハ	s21_10_mp2	maQha	mp2	mp2		s21	20190527	14
7273	3.438679391	4.458732587	ケッヘ	s20_11_mp3	keQhe	mp3	mp3		s20	20190527	14
7274	4.328059116	5.478119092	ケッヘ	s21_11_mp3	keQhe	mp3	mp3		s21	20190527	14
7347	19.58952167	20.50956965	ゴッホ	s20_12_mp4	goQho	mp4	mp4		s20	20190527	14
7348	21.61896085	22.76902083	ゴッホ	s21_12_mp4	goQho	mp4	mp4	[noRPp]	s21	20190527	14

図 5：Excel 上の検索環境

図 6：検索結果の再生画面の例

図 6 に結合されたサンプルの再生画面を示す。画面の右下に表示されている文字列は、そのフレームがどの動画ファイルのどの位置から切り出されたかを示す情報である。また画面左下にはもとの動画ファイルのフレーム番号が示されている。結合された動画ファイルを分析するユーザーは、これらの情報を参照することで、自分が現在処理しているフレームの出自を確認できる。

この検索系を利用することでユーザーは多数の話者の調音運動を容易に比較検討することができる。調音音声学の研究にとどまらず、音声学教育・日本語音声の教育にも有益である。また分析に必用な発話だけを抽出した動画ファイルを作成できることで、MRI Vuwer 等による測定作業の効率も格段に向上する。

現在、われわれは、ここに説明した検索系と同等の機能をもったデスクトップアプリとウェブアプリの開発を進めている。これらを用いれば、Excel のライセンスやフリーソフト類のインストール作業なしに rtMRI データベースを利用できる環境が提供できる。

５．言語研究での活用例

本節では、本データベースの言語研究における活用例を紹介する。データベース／コーパスの開発では、構築中のデータを利用することで、データないし設計の問題を把握することが必須のプロセスである。本データベースの場合も、数名分のデータが利用可能になった 2018 年から試行的な分析を開始しており、その一部は主に国際会議論文の形で公開されている。本節ではそれらを簡単に紹介する。以下に紹介する研究はすべて、観察の主観性に起因する先行研究の問題を rtMRI データによって解決した研究である。興味のある読者は直接各論文にあったっていただきたい。

５．１　発話末の撥音

Maekawa(2019)では発話末に位置する日本語撥音/N/の調音位置を東京方言男性話者 3 名の rtMRI データを用いて分析した。日本語の語末撥音は口蓋垂鼻音であると記述されることが多いが、実際には主に直前母音の調音位置の影響によって、調音位置は硬口蓋から口蓋垂まで、大幅かつ組織的に変動する。撥音直前の母音種別と話者の情報だけを与えれば、統計モデルによって撥音の調音位置は正確に予測することができること、また構築された予測モデルには高い汎化力があることも示した。

５．２　ラ行子音

前川(2019)では東京語のラ行子音/r/の調音を分析した。この子音は International Phonetic Association (1990)では voiced postalveolar flap と記述されているが、これには従来から異論が多い。男性話者 7 名による語頭/r/の rtMRI データを検討した結果、調音位置は/s/と同じ範囲に広がっており、postalveolar に限定されていないこと、また、声道閉鎖時の舌尖と舌端の位置関係から推測すると典型的な flap 調音は観察されず、大部分は tap 調音であることが判明した。結論として日本語の/r/は voiced alveolar tap or flap と記述すればよい。これは IPA における歯茎弾き音の定義そのものである。

５．３　ワ行子音

Maekawa (2020)では東京語のワ行子音/w/の調音を分析した。この子音の音声記号としては IPA の[w]が用いられることが多いが、[w]は両唇軟口蓋の二重調音子音である。rtMRI データで東京方言話者男女 11 名が発音した単独モーラ「ワ」における/w/の調音運動を観察したところ、軟口蓋の挙上が認められる発話は皆無であった。南カリフォルニア大学が公開している音声学者による[w]の調音の rtMRI 動画と比較すると明らかな相違が認められる。日本語のワ行子音は有声両唇接近音であると考えられ、IPA の[w]や[ɰ]（有声軟口蓋接近音）での転機には注記を要する。

５．４　モンゴル語母音調和

Saito, Yurong & Maekawa (2019)ではモンゴル語バーリン方言の母音調和を分析した。現代モンゴル語母音調和がどのような特徴に関する調和であるかについては諸説が対立しており定説がない。この研究では舌根と咽頭壁の距離と舌の最高点の位置を検討した。両者ともに母音調和と相関した差が観察されたが、前者の方により明瞭な差が観察されたことからモンゴル語の母音調和は舌根位(ATR: advanced tongue root)に関する調和である可能性が高いことを報告した。この研究はモンゴル語における舌根位の変化を初めて可視化した研究となった。

6．おわりに

本稿では、われわれが開発を進めてきたリアルタイム MRI 動画日本語調音運動データベースの概要を示し、その特徴を論じた。最後に世界における rtMRI データの整備状況に触れておきたい。現在、欧米では我々のデータよりも撮像速度の速い rtMRI データの取得が実現しており、一部では 100fps を超える撮像速度も実現している（Lim et al. 2018; Ramanarayanan et al. 2018）。これらに比較すると我々のデータは撮像速度の点で一時代前の水準に留まっていることは否めない。

ただし高速撮像を実現するためには k-space からの画像再構成（付録参照）に様々なサンプリング上の工夫を凝らす必要があり、それが画質の低下を招くことが多い。われわれのデータは、撮像速度と引き換えにムラのない高い画質を実現していると主張するのは半ば負け惜しみであるが、実際に調音音声学の分析に活用した経験からすると、画質が分析の質に大きく影響することも事実である。

次にデータ共用の姿勢について言えば、欧米のグループにはあまり実績がない。南カリフォルニア大グループのホームページではデータが公開されているが(Narayanan et al. 2011)、デモンストレーション以上の目的をもった公開ではない。本格的なデータ公開を目指していたフランスの研究グループ（Douros et al. 2019）も MRI データを取得した病院との間で権利問題が発生して公開が難航しているとのことである（私信による情報）。

われわれのデータベースの一部は 2020 年度中の公開を予定しており、その後も適宜拡張を予定している。組織的に収集された多人数の rtMRI データがデモンストレーション目的を超えて研究用に公開されるのは、おそらくこれが世界初の機会になると思われる。

謝辞：本研究は、日本学術振興会科学研究費（17H02339 および 20H01265,いずれも代表者は前川）、国立国語研究所「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」（代表者:浅原正幸）、および令和 2 年度人間文化研究機構・機構長裁量経費により実施しました。

文 献

Blaimer, M., F. Breuer, M. Mueller, R. B. Heidemann, M. A. Greiwold & P. M. Jakob (2004). “SMASH, SENSE, PILS, GRAPPA – How to Choose the Optimal Method”. *Topics of Magnetic Resonance Imaging*, 15(4), 223-236.

Breuer, F. A., P. Kellman, M. A. Griswold & P. M. Jakob (2005). “Dynamic Autocalibrated Parallel Imaging Using Temporal GRAPPA (TGRAPPA)”. *Magnetic Resonance in Medicine*, 53, 981-985.

Douros, Ioannis K., Jacques Felblinger, Jens Frahm, Karyna Isaieva, Arun A. Joseph, Yves Laprie, Freddy Odille, Anastasiia Tsukanova, Dirk Voli & Pierre-André Vuissoz (2019). “A multimodal real-time MRI articulatory corpus of French for speech research”. *Proc. INTERSPEECH 2019*, Graz, 1556-1560 (DOI: 10.21437/Interspeech.2019-1700).

Frahm, J., A. Haase & D. Matthaei (1986). “Rapid NMR Imaging of Dynamic Processes using the FLASH Technique”. *Magnetic Resonance in Medicine*, 3(2), 321-327.

Griswold, M. A., P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer & A. Haase (2002). “Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA)”. *Magnetic Resonance in Medicine*, 47, 1202-1210.

Haase, A. (1990). “Snapshot FLASH MRI. Applications to T1-, T2-, and chemical shift Imaging”. *Magnetic Resonance in Medicine*, 13, 77-89.

Haase, A., D. Matthaei, R. Bartkowski, E. Dühmke & D. Leibfritz (1989). “Inversion Recovery Snapshot FLASH MR Imaging”. *Journal of Computer Assisted Tomography*, 13(6), 1036-1040.

Huettle, S. A, A. W. Song & G. MaCarthy (2004). *FUNCTIONAL Magnetic Resonance Imaging*. Sinauer Associates, Inc. (3. Basic Principles of MR Signal Generation, 4. Basic Principles of MR Signal Formation, 5. MR Contrast Mechanisms and Pulse Sequence)

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet.* Cambridge University Press.

Kitamura, Tatsuya, Hironori Takemoto, Kiyoshi Honda, Yasuhiro Shimada, Ichiro Fujimoto, Yuko Syakudo, Shinobu Masaki, Kagayaki Kuroda, Noboru Oku-uchi & Michio Senda (2005). "Diffrence in vocal tract shape between upright and supine postures: Observation by an open-type MRI scanner". *Acoustical Science and Technology*, 26 (5), 465-468.

Lim, Yongwan, Yinghua Zhu, Sajan Goud Lingala, Dani Byrd, Shrikanth Narayanan & Krishna Shrinivas Nayak (2018). "3D dynamic MRI of the vocal tract during natural speech". *Magnetic Resonance in Medicine*, 2018, 1-10 (DOI: 10.1002/mrm.27570).

Maekawa, Kikuo (2019). "A real-time MRI study of Japanese moraic nasal in utterance-final position." *Proc. ICPhS 2019*, Melbourne, 1987-1991.

Maekawa, Kikuo (2020). "Remarks on Japanese /w/". *ICU Working Papers in Linguistics (ICUWPL)*, 10, 45-52. (http://id.nii.ac.jp/1130/00004625/)

Masaki, Shinobu, Mark K. Tiede, Kiyoshi Honda, Yasuhiro Shimada, Ichiro Fujimoto, Yuji Nakamura & Noburo Ninomiya (1999). "MRI-based speech production study using a synchronized sampling method". *Journal of the Acoustical Society of Japan (E)*, 20 (5), 375-379.

McGibney, G., M. R. Smith, S. T. Nichols & A. Grawley (1993). "Quantitative Evaluation of Several Partial Fourier Reconstruction Algorithms used in MRI". *Magnetic Resonance in Medicine*, 30, 51-59.

McRobbie, D. W., E. A. Moore, M. J. Graves & M. R. Prince (2007). *MRI From Picture to Proton – Second Edition*. Cambridge University Press. (14. A heart to heart discussion: cardiac MRI)

Narayanan, Shrikanth, Erik Bresch, Prasanta Ghosh, Louis Goldstein, Athanasios Katsamanis, Yoon Kim, Adam Lammert, Michael Proctor, Vikram Ramanarayanan & Yinghua Zhu (2011). "A multimodal real-time MRI articulatory corpus for speech research". *Proc. INTERSPEECH 2011*, Florence, 837-840.

Pruessmann, K. P., J. Weiger, M. B. Scheidegger & P. Boesiger (1999). "SENSE: Sensitivity Encoding for Fast MRI". *Magnetic Resonance in Medicine*, 42, 952-962.

Ramanarayanan, Vikram, Sam Tilsen, Michael Proctor, Johannes Töger, Louis Goldstein, Krishna S. Nayak & Shrikanth Narayanan (2018). "Analysis of speech production real-time MRI". *Computer Speech & Language*, 52, 1-22 (doi.org/10.1016/j.csl.2018.04.002).

Saito, Yoshio, Yurong & Kikuo Maekawa (2019). "An Investigation into Modern Mongolian Vowel Harmony Using Real-time Magnetic Resonance Imaging." *Proc. ICPhS 2019*, Melbourne, 1431-1434.

Shiller, Douglas M., David J. Ostry & Paul L. Gribble (1999). "Effects of gravitational load on jaw movements in speech". *The Journal of Neuroscience*, 19 (20), 9073-9080.

Sodickson, D. K. & W. J. Manning (1997). "Simultaneous Acquisition of Spatial Harmonics (SMASH)： Fast Imaging with Radiofrequency Coil Arrays". *Magnetic Resonance in Medicine*, 38, 591-603.

Takemoto, Hironori, Tsubasa Goto, Yuya Hagihara, Sayaka Hamanaka, Tatsuya Kitamura, Yukiko Nota & Kikuo Maekawa (2019). "Speech organ contour extraction using real-time MRI and machine learning method." *Proc. INTERSPEECH 2019*, Graz, 904-908 (DOI: 10.21437/Interspeech.2019-1593.).

浅井拓也・菊池英明・前川喜久雄(2018).「調音運動動画アノテーションシステムの開発」日本音響学会 2018 年秋季研究発表会講演論文集, 1235-1238.

籠宮隆之(2004).「音声収録作業の概要」『日本語話し言葉コーパス』添付電子マニュアル (https://pj.ninjal.ac.jp/corpus_center/csj/manu-f/recording.pdf).
国立国語研究所(1978).『X 線映画資料による母音の発音の研究：フォネーム研究序説』国立国語研究所報告 60, 秀英出版(X 線映画は https://mmsrv.ninjal.ac.jp/x-sen/で公開).
国立国語研究所(1990).『日本語の母音，子音，音節 ： 調音運動の実験音声学的研究』国立国語研究所報告 100, 秀英出版.
後藤翼・天野沢海・竹本浩典・北村達也・能田由紀子・前川喜久雄(2020).「rtMRI 動画から抽出した発話器官の輪郭データに基づく平均声道の生成と分析」日本音響学会 2020 年秋季研究発表会講演論文集,掲載頁未定.
能田由紀子・北村達也・籠宮隆之・竹本浩典・前川喜久雄(2019).「磁気センサシステムをもちいた計測による座位･仰臥位･腹臥位における舌運動の差異の検討」日本音響学会 2019 年春季研究発表会講演論文集, 823-824.
前川喜久雄(2019)「日本語ラ行子音の調音：リアルタイム MRI による観察」日本音声学会第 33 回全国大会予稿集, 98-103.

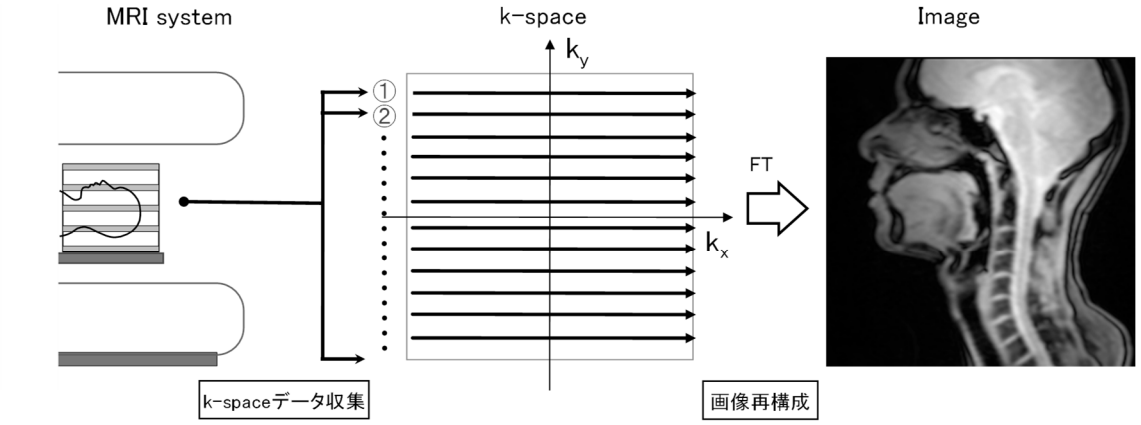
付録：本研究に用いられた MRI 動画撮像技術

正木信夫(ATR)，島田育廣(ATR-Promotions)

A1．MRI 動画撮像技術の発話研究への応用（背景）

ここでは、本研究で用いられた MRI 動画撮像技術について述べる。その前提として、MRI の画像生成の原理（Huette, Song & MaCarthy 2004）を理解していることが望ましいが、本稿を理解する上では、以下の 3 点を押さえていればいい（図 A1）。

①MRI の画像生成は「k-space データ収集」と「画像再構成」の 2 つのプロセスで成り立っている。
②「k-space データ収集」とは、k-space と呼ばれる 2 次元空間に、強力な静磁場と高速で変化する傾斜磁場でつくられる多様な磁場環境の中で電磁波の照射によって観測される、体内のプロトン分布に関する情報を収集（格納）することである。
③「画像再構成」とは②で得られた k-space の 2 次元データに 2 次元フーリエ変換を施すことでプロトンの体内分布に関する画像を作り出すことである。



図A1：MRIの画像再構成の原理

MRI はもともと動画撮像には不向きな技術である。これは十分な空間分解能を実現するために必要な k-space データを、観測対象物の動態分析に必要となる時間分解能を確保しながら取得することが難しい、という技術的な制約に起因する。例えば、256 x 256 ピクセル程度の空間分解能を確保した 10fps のフレームレート（時間分解能）を持つ動画を得ることを考えてみる。この場合、k-space では 128 x 80 程度の 2 次元データが必要であるが、これだけの各フレームのデータをフレームレートの逆数の時間（1/10 s = 100ms）内に行い、さらにそれを同じ周期で繰り返すことが難しいということである。

そのような制約があるにもかかわらず、心臓の拍動や肺の呼吸運動のような動きを観測するために、MRI 動画の撮像技術は MRI の実用機が世の中に出た 1980 年代の後半から開発されてきた。注意すべきはここで対象とした運動が周期的な動作であるということである。当時の撮像プロトコル開発者は、周期的運動であるという特性を逆に利用したのである。すなわち、観測対象の運動周期と同期してスキャンを行うことで、k-space 上のデータを時間分散的に収集するプロトコルを開発した（McRobbie et al. 2007）。発話研究においてもこ

れを応用して、外部トリガーを用いた周期的な繰り返し発話による動画撮像が行われた。この方法では、被験者はMRIスキャンのタイミングに合わせて1秒から2秒程度の同じ発話を 100 回程度繰り返した。スキャンタイミングと被験者の発話を同期させるために、被験者にはリズミカルな同期音を提示した。この方法は、すべての被験者に適応可能ではなかったが、適切なタイミングで均一な発話ができる被験者においては安定した高品質の動画を得ることができた。(Masaki et al. 1999)

しかし、発話動態研究において、被験者に繰り返し発話を強いることは、動態観測の研究対象の範囲を狭めることに他ならない。例えば吃音者にとっては、同じ発話を指定されたインターバルで繰り返すことは困難である。たとえ吃音者でなくても同じ発話を繰り返す中では発話のタイミングや調音器官の運動軌道がずれることもあり、その不均一性は動画の品質を左右する。このような事情から、十分な空間分解能の画質と十分な時間分解能のフレームレートの両方を兼ね備え、繰り返し無しに1度の発話でデータ採取が可能な、「リアルタイム撮像」の実現は、MRI を用いる動画観測を行う発話研究者にとって、待望の技術であった。

A2. MRI動画のリアルタイム撮像に向けた技術開発

リアルタイム撮像の技術は主に3つの側面から開発が行われてきた。第1は高速シーケンスの開発、第2はフレームレートを上がるための採取データを削減する技術の開発、そして第3は k-space に格納するデータを複数フレームのデータ採取時間の中で行う技術開発である。ここではこれら3つの技術について具体的に説明する。

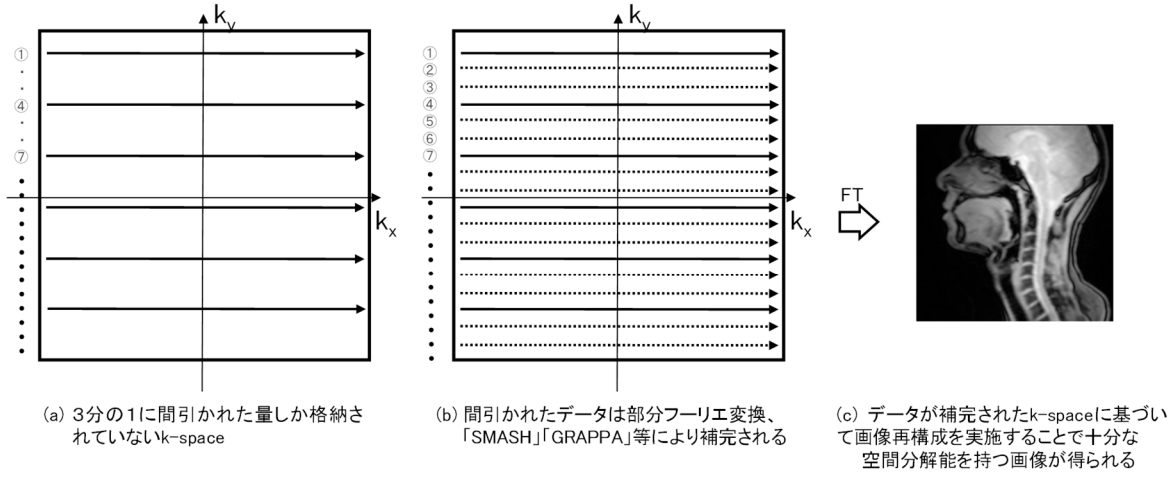
なお本稿ではそれぞれの技術開発における改良点について、MRI 動画の高速化への貢献という視点に立って、その目的と改良のポイントをわかりやすく解説することを目指した。そのため、改良の原理を簡略化して表現した部分があり、MRI の専門家からみると不十分な説明となっている可能性があることを否めないことをお断りしておく。

(1)「k-space データ収集」を短時間で行うシーケンスの開発

MRI は強磁場内でスピン運動するプロトンから発生する電磁波を観測することでそのプロトン密度の空間分布を濃淡により画像化する技術である。1980 年代の初期の MRI 実用化において開発されたのは、電磁パルス(RF パルス)によって励起されたプロトンの横緩和(スピン-スピン緩和)を手掛かりにプロトン分布を観測し画像を再構成する Spin echo と呼ばれるシーケンスであった。しかしこの方式では、プロトンのスピンを 90° 倒して横緩和をスタートさせる RF パルス(90° パルス)と、その後一旦ばらけたスピンの位相をエコー信号採取時に再度同期させるための RF パルス(180° パルス)を照射する必要がある。このように2回の RF パルス照射が必要となることから、最初の RF パルス照射からデータ採取までにかかる時間(TE)が長く、結果として動画のフレームレートを決める繰り返し時間(TR)も長くなる。そこで、2つめの 180° パルスの代わりに、傾斜磁場を反転させることでエコー信号を発生させる Gradient echo 系のシーケンス「FLASH(fast low angle shot)」が開発され、TR の短縮、ひいては MRI 撮像時間の短縮が図られた(Frahm, Haase & Matthaei 1986)。これは、体動の影響を少なくする、というような臨床の現場の必要があって開発されたものであるが、MRI 動画撮像においてはそのフレームレート向上に大きく貢献した。

臨床現場の MRI 撮像において広く使われていた「FLASH」シーケンスではあったが、短い TR の宿命としてコントラスト不足が指摘されていた。これに対する対策として Inverse Recovery 系のシーケンス「Turbo FLASH」が開発された(Haase et al. 1989)。このシーケ

ンスでは、Inverse pulse を Preparation pulse として印加し、これにより生ずる緩和過程の期間内に、いわば「FLASH」を高速化した「Snapshot FLASH」（Haase 1990）を入れ込むことでコントラストの改善を図ったのである。しかしこの方法は Preparation pulse を印加して Snapshot FLASH を実施するまでに、一定の緩和時間は待つ必要があり、MRI 動画撮像への応用においては必ずしもフレームレートの改善には繋がっていない。しかし、「Turbo FLASH」でありながら Preparation pulse を印加しない条件を設定することができる。これを用いると緩和時間を待たずして各フレームの「Snapshot FLASH」を開始することができ、「Turbo FLASH」でも「FLASH」と同程度のフレームレートで撮像することが可能である。本実験の一部では、この「Preparation pulse 抜きの Turbo FLASH」が採用されている（本稿「A3.（2)」を参照)。



図A2：k-spaceデータ収集におけるデータの間引きと間引かれたデータの補完

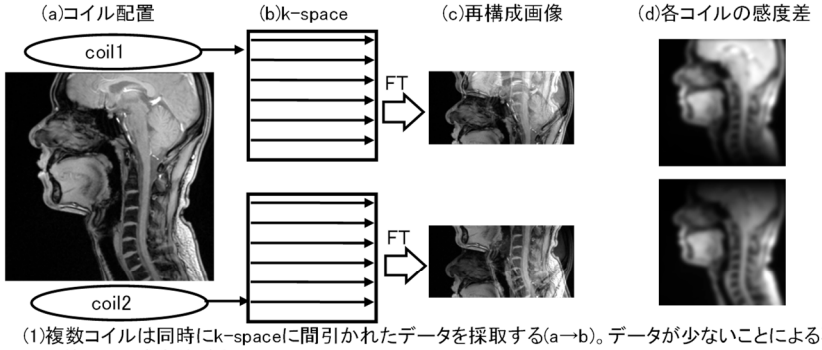
(2)「k-space データ収集」におけるデータを間引いてフレームレートを上げると同時に、データ削減に伴う画質劣化を抑制するデータ補完技術の開発

例えば、256 x 256 の空間分解能を持つフレームレート 15fps の動画撮像をする場合、約 70ms で 1 フレーム分の k-space データを採取する必要がある。FLASH を用いた場合、データ採取の間隔を TR=2.5ms とすることができるが、十分な空間分解能を確保するためには、80 行分のデータを 80 回の RF パルス照射によって採取して k-space 空間を埋める必要がある。しかしこれでは 1 フレームのデータ採取に 200ms 程度かかり、5fps 程度の時間分解能しか得られない。そこで、格納するデータ数を 3 分の 1 程度に間引いて、k-space 上では均等に採取されたデータが並ぶようにする（図 A2(a)）。これにより 66.7ms の周期で k-space データを採取することから、15fps の動画を実現できる。

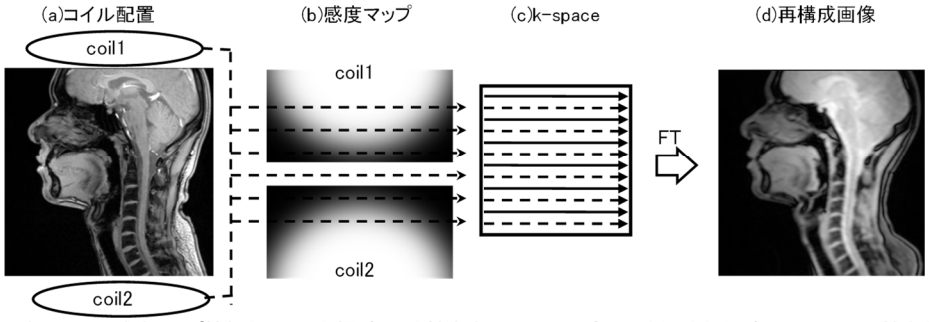
しかし、このまま画像再構成を行った場合、データ数が少ないために十分な空間分解能が得られないばかりか、画像に折り返し歪みが生じてしまう。これを回避するためには、「画像再構成」の前に k-space データ収集時に間引かれたデータを補う必要がある。この補完にはさまざまな方法が提案されているが、例としては、部分フーリエ変換（partial Fourier reconstruction）を用いるもの（McGibney et al. 1993)、マルチチャンネルシステムの複数コイルの感度分布情報等を利用して k-space 上の欠損情報を補う「SMASH」（Sodickson & Manning 1997）や「GRAPPA」（Griswold et al. 2002）などが提案されている（図 A 3)。

このようにして補完された k-space データにより画像再構成を行うことで十分な空間分解能を持つ画像を得ることができる（図 A2（b）及び(c)）。

なお、マルチチャンネルシステムを用いて画像再構成をする手法として、k-space 上ではなく画像処理段階で感度分布情報を用いる「SENSE」(Pruessmann et al. 1999）という手法があることも付記しておきたい（「SMASH」「GRAPPA」「SENSE」については Blaimer et al. (2004)にまとめられている)。



(1)複数コイルは同時にk-spaceに間引かれたデータを採取する(a→b)。データが少ないことによる再構成では折り返しが発生し(c)、コイル配置等による感度差が生じる(d)。



(2)各コイルの感度マップ情報を利用し欠損データを補完する(a→c)。当初から採取されたデータ(⟶)と補完されたデータ(⇢)全部を用いて画像再構成を行うことで十分な空間分解能を持つ画像を得る(c→d)。

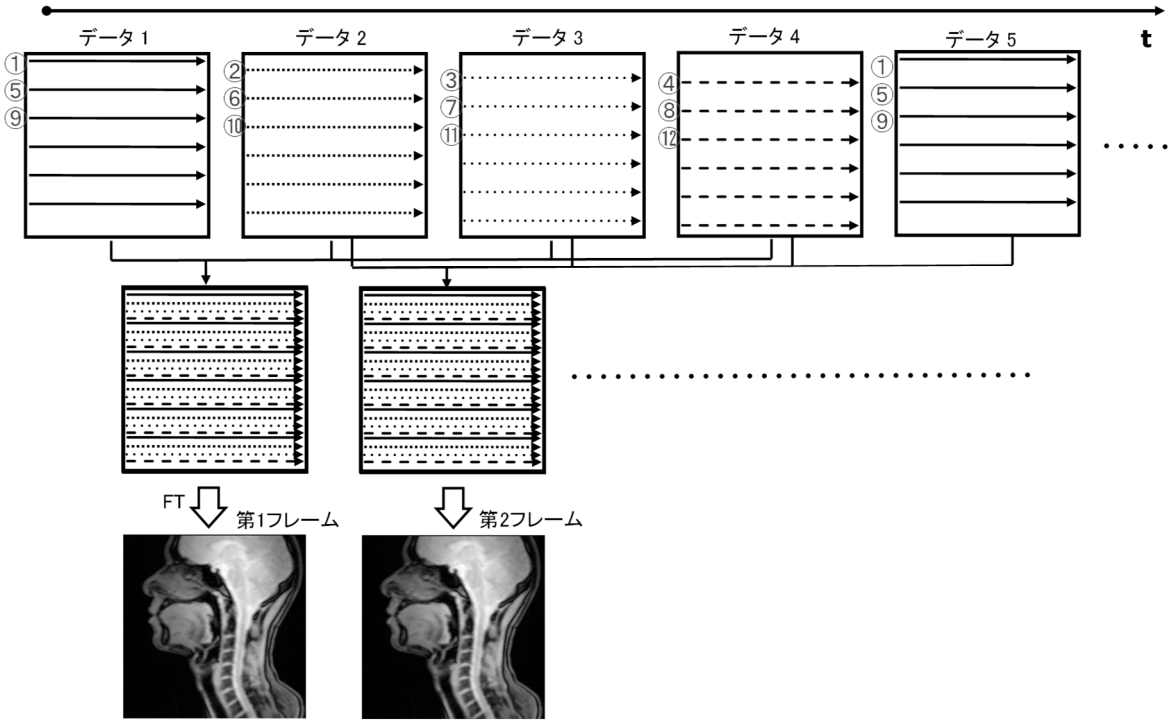
図A3：複数コイルのによるk-spaceデータの補完と画像再構成（２コイルの例）

(3)「k-space データ収集」を複数フレームに亘るデータ採取の中で時間分散的に実施することで、見かけの時間分解能を向上させる技術の開発

これは、1 フレーム分の k-space データ収集を複数フレームに相当する時間をかけて収集して画像再構成を行う手法である。図 A4 は 1 フレームの画像再構成に必要な k-space データを、4 フレーム分の時間の中で分散的に採取する例を示している。すなわち、第 1 回目のデータ採取では「①⑤⑨・・・」の番号のある行のデータのみ採取し、他のデータは採らないこととする。同様に、第 2 回目のデータ採取では「②⑥⑩・・・」のデータを格納する。このようにして採取した各フレームの k-space データを用い、実際の動画用の画像再構成を行うが、例えば動画の第 1 フレームの再構成には上記の第 1 ～第 4 のデータ採取で得られたものを用い、第 2 フレームの再構成には第 2 ～第 5 のデータ採取で得られたものを用いる。このように k-space データ補完技術を時間方向に拡張した方法の代表格が TGRAPPA（Breuer et al. 2005）であるが、その中では、図 A3 に示したようなマルチチャンネルシステムの複数コイルから得られる情報に基づく k-space データ補完技術も組み込まれており、画質を確保する工夫がなされる。

この場合、再構成された動画のフレームレートは、データ取得のタイミングと同じレート

であると言えるが、再構成された画像自体には連続する複数回のデータ採取（この例では4回のデータ採取）で得られたデータ構成されていることになるので、得られた動画の各フレームの画像は、前後のフレームの影響を受けている、ということを考慮しておく必要がある。



図A4： k-spaceデータの収集における時系列的に隣接する複数のデータによる補完

3. 本研究で用いた撮像パラメータ

本研究では2種類のパラメータ条件で撮像を行った。ひとつは15fpsを目指した条件であり、もうひとつは30fpsを目指したものである。

(1)15 fpsを目指したパラメータ設定（FLASH +GRAPPA）

この動画撮像ではFLASHシーケンス（Frahm, Haase & Matthaei 1986）を採用し、フレームレート向上の手法としてGRAPPA（Griswold et al. 2002）を用いた。ここでは前述「2.(2)」で図A2を用いて説明したように、1フレームの画像再構成に必要なk-spaceデータを1/3に間引いて収集した。具体的には1フレームあたり84行のk-spaceデータが必要であるところ、その1/3の28回のみ行うことで速度を3倍に上げている。撮像パラメータの表（表A1）の「Acceleration Factor：3」はこの高速化の倍数を示している。これにより1フレームあたり70msでのデータ採取を実現し、フレームレートは14.2857fpsとなった。

k-space内で間引かれたデータを補間する方法として採用されたGRAPPAでは、マルチコイルから得られるデータからk-spaceの中での各コイルの貢献度を考慮した補間法が採用されている。各コイルの未観測データへの貢献度の評価は、動画採取の直前のプレスキャンのデータを用いて行われる。このようにして得られた128 x 84のデータを用いて各フレームの画像再構成を行い、空間分解能を256 x 256ピクセルとする動画像を得た。撮影は512フレーム分のデータが格納できる領域を確保し、35.8秒間のデータを採取した。

(2)30 fps を目指したパラメータ設定（TURBO FLASH +TGRAPPA）

この動画撮像では上記「2.(1)」で説明した「Preparation pulse 抜きの Turbo FLASH」を採用し、フレームレート向上の手法として TGRAPPA（Breuer et al. 2005）を用いた。TGRAPPA では前述「2.(3)」で図 A4 を用いて説明したように、1 フレーム分の画像再構成に必要となる 80 行の k-space データを 20 行ずつ 4 回のデータ採取に分散して取得している。このため、見かけ上、データ採取の速度が 4 倍になっている。撮像パラメータの表（表 1）の「Acceleration Factor : 4 」はこの高速化の倍数を示している。これにより 20 行のデータ採取が 36.8ms で行われることとなり、フレームレートとしては 27.1739fps を実現している（注：前述の「FLASH」の TR=2.5ms では 1 フレーム内の 1 行分のデータ採取時間を示しているのに対し、この「Preparation pulse 抜きの Turbo FLASH」の TR=36.8ms は 1 フレーム内の全 20 行分のデータ採取時間を示しているので注意されたい）。

このように取得した k-space データを用い、まず第 1 フレームの再構成では最初の連続する 4 回のデータ採取で蓄積された 1 フレーム分の 128 x 80 のデータを用いて 256 x 256 ピクセルの画像が再構成された。次のフレームの再構成は 1 スキャンごとのデータをシフトさせて実施された。撮像時、513 フレーム分のスキャンをデータ採取時間 18.88 秒で行った。

表 A1：本実験で用いた撮像条件

目標フレームレート	15fps	30fps
実際のフレームレート	14fps（14.2857fps）	27fps（27.1739fps）
Sequence	FLASH	Turbo FLASH（*1）
TR	2.5ms	36.8ms
TE	0.98ms	0.91ms
Flip angle	8deg	5deg
Slice thickness	10mm	10mm
FOV	256mm x 256mm	256mm x 256mm
Scan matrix	128 x 84	128 x 80
Reconstruction matrix	256 x 256	256 x 256
Acceleration method	GRAPPA	TGRAPPA
Acceleration factor	3	4
No. of measurements	512frames	513frames
Scan time	35.8s (*2)	18.88s (*3)

*1: より正確には「Preparation pulse 抜きの Turbo FLASH」である（本文「2.(1)」参照）
*2: 2.5ms x 84 /3 => 70ms/slice x 512 => 35.8s
*3：36.8ms/slice x 513 => 18.88s