MIMIC in the OMOP Common Data Model*

A1FIRSTNAME A1LASTNAME, *AES Member* **AND A2FIRSTNAME A2LASTNAME,** *AES Fellow*

(abc@abc.com)   (xyz@abc.com)

Universityxyz, City, Country

This paper discusses the implementation of spectral delay using filters comprising a cascade of many low-order allpass filters and an equalizing filter. The spectral delay filters have chirp-like impulse responses causing a large, frequency-dependent delay that is useful in audio effects processing. An equalizing filter design and a multirate technique, which stretches the allpass filters, impulse response, are introduced.

0 Introduction

Intensive Care Units (ICUs) are particularly sensitive units where demand of care is rising[1] and mortality is up to 30% which is a major health care problem [2]. Studies have shown that intensivists use a limited number of clinical information concepts to guide the decision[3] and that the medical practices are sparse and variable. Knowing that ICU patients' health record are highly detailed including connected devices this is a paradox. The increasing adoption of Electronic Health Records (EHR) systems worldwide makes it possible to capture large amounts of clinical data [4] and big data mining has the potential to play an important role in clinical medicine [5]. Indeed on the basis of broad patient medical informations expectations are to improve clinical outcomes and practices, allow personalized medicine and guide decision thought early warning systems, and also enrol large and multicentric cohort easily while minimizing costs.

By now several commercial or noncommercial, opensource or nonopensource ICU databases have been developed. *CUB-REA* is a semi-automatic collection of 25 years and over 300k ICU patients stays from 30 distinct ICUs in Paris region and covers administrative data only[?, ?]. Yet it is subject of 15 international publication and access is under conditions. *OutcomeREA* is a manual collection of 20 years over 20k UCU patients stays with medium granularity data from 20 distinct ICUs in France. Yet it is subject of 50 international publications and access is under conditions. *eICU* [?] is an automatic collection of 2 years of 1.5M ICU patients stays from 500 ICU in the USA and covers high granularity data[6]. It is a commercial database. *MIMIC* (Medical Information Mart for Intensive Care) is a semi-automatic collection of 10 years

and over 60k ICU patients stays with very high granularity (including EKG) from 1 ICU in Boston. Yet it is subject of 300 international publication. It is a freely-available database. All those databases have their own dedicated model. Their structural model are all based on relational database but all do have tables and columns with different meaning and granularity. As an example MIMIC do have two inputevents tables reflecting its source center changed its EHR. Also their conceptual model are mostly different. For example MIMIC do have ICD9 for condition terminology, while CUB-REA do have both CIM9 and CIM10.

Some studies have shown that using a common data model (CDM) for databases allows multicentric research, allow mining rare disease and catalyse research in general by allowing sharing practices, code, and tools[7]. On the other hand some studies have shown that results are not replicable from one to another database [8] and that keeping the local conceptual model [9] and structure [10] of database for research leads to better outcomes. Last but not least, studies have show that CDM can lead to different results [11]. A dozen of CDM have emerged but we limited the candidate data models to those designed and used for clinical researches, and those freely available in the public domains without restrictions. *OMOP* (Observational Medical Outcomes Partnership Common Data Model) is a CDM designed for multicentric Drug adverse Event and now enlarges to medical, clinical and also genomic use cases. OMOP provides both structural (as as set of relational tables) and conceptual (as a set of standard terminologies) such SNOMED for diagnoses, RxNORM for drug ingredients and LOINC for laboratory results. While OMOP has proven its fiability [12] the fact that concept mapping process is known to have impact on results [13] and that applying the same protocol on different data sources leads to different results [8] reveals the importance of keeping the local codes to allow local analysis. Several example of

---

transforming databases into OMOP have been published [14, 15] and yet OMOP stores 682 milion patients records from all over the world[16]. OMOP had 5 versions, and prones its strong backward compatibility. *i2b2/SHRINE* is a medical cohort discovery tool used in more than 200 hospitals over the world. SHRINE is one of the attempt to federate multiple instances of i2b2. The i2b2 star schema has proven its high flexibility thanks to the modular design of the fact tables allowing storing numerics, characters or concepts. Its single terminology model is a path based hierarchical table does not allow to modelise graph ontology (such snomed). While i2b2 is highly efficient for cohort discovery, it's model wasn't designed for ad-hoc analysis. A bidirectional terminology mapping [17] between local concepts and the SHRINE core set of standard concepts is a prerequirement to participate to a SHRINE network. *HL7-FHIR* (Fast Healthcare Interoperability Resources) is a medical data exchange API specification. FHIR provides a structural CDM that can be materialized as JSON, XML or RDF format. FHIR is flexible and does not specify a standard conceptual model so that each hospital can add extension to implement specific data or share within it's local terminology making each FHIR implementation sensibly divergent. While some research show it as a promising CDM for ad-hoc analysis [18] or cohort discovery [19], its graph nature adds a layer of transformation making usage complicated for data-scientists as well as difficult to create standardized analysis. Finally the model envolves and does not make the asumption of backward compatibilities along the versions.

To demonstrate the feasibility and the opportunity to use a CDM for ICU databases, we choose MIMIC since it is freely accessible and then the result of this work will be accessible for subsequent improvement, analysis, and demonstration. Secondly, MIMIC does have the most broad and high granularity database leading to evaluate the ability of the CDM to ingest such complex dataset. More importantly, MIMIC stated they would think of a CDM[**?**], that would solve some structure and conceptual weaknesses of MIMIC mentionned above and this work should help MIMIC in it's goal to build a international ICU database [6]. Among the other CDM, OMOP looks like the best fit as it allows both multicentric standardised analysis as well as monocentric specific modeling and analysis. Compared to PCORnet [20] OMOP performs better in the evaluation database criteria compared with the other models (and PCORnet in particularly) : completeness, integrity, flexibility, simplicity of integration, and implementability, accommodate the broadest coverage of standard terminologies, provides more systematic analysis with analytic library and visualizing tools, provides easier SQL models. Compared to i2b2/SHRINE, the unilateral mapping methodology of OMOP is more effective than the bidirectional [17] SHRINE mapping. Hence OMOP proposes a broader set of standardised concepts. In terms of structural CDM OMOP is highly rigourous in how data shall be loaded in a particular table when i2b2 is highly flexible with a one general table that solution every data domains. This rigourous approach is necessary for standardi-

sation. Previous work have loaded i2b2 with MIMICiii [6] - however, the concept mapping step have limited the results since i2b2 design do not store the local ontology or information where OMOP design allows to keep the concept mapping unfinished. OMOP has this advantage to keep the terminology mapping step not mandatory by keeping the local codes in a usable format. As compared to FHIR, OMOP performs better as a conceptual CDM as FHIR does not specify which terminology to implement. In terms of structural CDM OMOP relational model can be materialized into csv format and stored in any relational database when FHIR materialised as json needs some processing and more skills to be exploited. OMOP shares the advantages of all above models. Previous preliminary work have been made on translating MIMIC into OMOP [**?**] and still work remains to be affined and upgraded to be furthermore evaluated.

In order to evaluate the MIMIC-OMOP transformation we propose to answer to remaining question such the difficulty of transforming/maintaining an OMOP dataset from an existing one, how well the initial dataset is integrated and how much data is lost in the process, how clear and simple the model is to be queried simply and efficiently by scientists, how well design it is to be enriched by collaborative work, and finally in what extend OMOP can integrate and makes feedbacks to intensivists in a realtime context. This work is then evaluated through 3 axes: Transformation, Analytics and Contribution. The *first* major contribution of this study is to evaluate OMOP in a real life and well known freely accessible database. The *second* major contribution is to provide a freely accessible dataset in the OMOP format that might be useful for researchers. The *third* major contribution is to provide the OMOP community some useful transformations dedicated to ICU and that can be reused in any OMOP dataset.

1 Material & Method

1.1 Data Transformation

Ajouter les deux schemas de database design

The MIMIC III version 1.4 21 clinical data tables and 5 tables for vocabulary OMOP CDM v5.3.1 which defines 15 standardized *clinical* data tables, 3 *health* system data tables, 2 *health economics* data tables, 5 tables for *derived* elements and 12 tables for standardized *vocabulary*. We didnt use the health economics data tables (not provided by MIMIC)

All the process is available freely on the github website [**?**].

Extract-Transformation-Load (ETL) processes is a methodology that allows migrating data from a source location to a target location. ETL first extract the data from the source location, and then apply transformations such structural modifications or conceptual modifications generally on a dedicated computer and technology. The last step is to load the resulting data into the target location. Transformation program are generally written in a programming language such java, c++ or python. Extract-

Load-Transform (ELT) processes is a slightly different methodology that does not use a dedicated transformation server and limits data transfers. The data is extracted and directly loaded into the target location. It is transformed afterwards in place. ELT allows to factorize both computer resources, and people knowledges. Indeed transformations are then written in a database dialect such SQL, as well as source and target location. Improving database resources will then benefit for both ETLers and end users. Since RDBMS computer resources cannot scale well and does not provide a good support for procedural language, ETL have been for long time used in conjunction with RDBMS. The emergence of distributed platform such hadoop allows to take part of ELT because they both allow to scale well horizontally and write java procedural user defined function that are used in conjunction with SQL queries. In this work, we decided to transform MIMIC into OMOP thought a ELT to limit the programming knowledges needed for code maintenance and to allow end users to participate in this process. PGSQL has been chosen as the database support for ELT because it is the primary support of MIMIC database and allows community to run the ETL on limited resources without licensing need. Finally PGSQL have recently made huge effort to handle data-processing better.

The MIMIC source data have been loaded with the provided scripts into a PGSQL (PGSQL) instance and a subset of 100 patient chosen upon their broad representativity in terms of data have been cloned into a second instance to serve as a light, representative development set. Each tables of the source have been added a global sequence incremented from 0 that serves as primary key and link into the OMOP target tables. The target tables have been created from the provided script except the indexes that would have slow down the data migration with useless computations. The integrity constraints (primary keys, foreign keys, non nullable columns) have been included to apply sanity checks at runtime. Some tricky transformations have been implemented as PGsql functions. The ETL is composed of 22 PGSQL scripts, each extracting information from the source or from concept mapping tables, transforming and loading one OMOP target table. The ordering of those scripts matters and is done sequentially thought a main script. Each ETL part have been tested thought pg-TAP, a unit testing framework for PGSQL. This allows to make sure there is no loss of informations, or code regression during the ETL development or for further updates of the source code. The unit tests are composed of 15 scripts, each checking a particular OMOP target table is loaded correctly - most of the tables are covered and tests covers simple counts, aggregated counts or distribution checks. A SQL script updates the OMOP tables when needed (the list of the modifications is detailed bellow). All character typed columns limited in length have been changed to unlimited since this might cause unpredictable truncation of content, and this has no bad impact on PGSQL's storing size or performance. The *visit_occurrence* and *visit_detail* table have been corrected accordingly to some discussion on the forum. The *note_nlp* table have been extended with some fields accordingly to the documentation online. The character *offset* column have been split into two integers columns because the offset word is a SQL reserved word and it makes sense to fill the resulting *offset_begin* and *offset_end* resulting columns.

The structural transformation have been done in few iteration of several phases. The first phase consists of looping over each MIMIC table and choose for each columns an equivalent location in OMOP. In general, the MIMIC documentation and the OMOP documentation were sufficient informative. In several cases, we needed to get clarification from the MIMIC contributors on the dedicated github repository, or from the OMOP community on the dedicated forum. All choices have been discussed in the MIMIC-OMOP github issues, and can be tracked into the commit log. The resulting table to table work flow is materialized in the repository as bi-directional documentation: from MIMIC table to OMOP table and vice-versa.

By design MIMIC aggregates informations from various systems such the emergency specific tool. Thus the transfer information is spread into multiple table, such *admissions*, *transfers* and *icustays*. OMOP centralizes this information in the *visit_detail*. As a result some hard transformation have been made. We added emergency stays as a normal location for patients throughout their hospital stay. Icustays mimic table was deleted as it is a derived table from transfers table (2) and we decided to assign a new new *visit_detail* pour each stay in ICU (based of the transfers table) whereas mimic preferred to assign new icustay stay if a new admissions occurs ¿ 24h after the end of the previous stay.

While various types of information are stored in the measurement table, the dedicated OMOP concepts types were not enough to distinct them. We added some measurement types. The actual visit_detail doesn't introduce pertinent information and duplicate informations from visit_occurrence table. For admitting_concept_id and discharge_to_concept_id, we extended the dictionary in order to track bed transfers and ward transfers. For visit_type_concept_id we assigned a new concept for any level of granularity necessary for your use case (ward, bed...)

The conceptual transformation consists on both loading OMOP with the MIMIC local concepts, and mapping them to the standards concepts. The local codes have been loaded in the concept table starting from 2 billion as specified in the OMOP documentation [**?**]. The MIMIC local codes are then added to the OMOP provided code (starting from 1 to 2 billion) and can be distinguished with the domain_id equals to MIMIC code. As much as possible information from the MIMIC coding table have been loaded in order to allow analytics with the local code without loosing information as compared to the MIMIC original model (worst case we concatenated information as free text in the concept_name column). To deal with concept mapping for each domain a csv file have been built as a support for manual standard mapping. This solution can scale for medical users without database engineering background. We tried to adopt the same methodology in their creation ; some obvious fields are needed : local and standard name, local and

standard id. Moreover evaluation and comments fields are good practices and may help contributers Those csv file are automatically loaded into PGSQL and get the sequence added too. There were four distinct cases. In the *first* case MIMIC is already in the OMOP standard terminology (eg: LOINC laboratory results), the work is already made. In the *second* case the mapping is already provided by OMOP (eg: ICD9/SNOMED-CT) then the data tables have been loaded accordingly. In the *third* case the mapping is not provided, but is small enough to be done manually in few hours (such as demographic status, signs and symptoms). In the *fourth* case the mapping is not provided and the terminology is huge (such admission diagnoses, drugs). Then we decided to only map the subset of the code that are the most represented. In all case a mapping has to be done we have setup a semi-automatic methodology to make auto-suggestion. Many mapping tools exist on the area RELMA provided by LOINC, USAGI provided by OHDSI. Most of those tools are based on linguistic mapping [cite], and the approach have been shown to be the most effective[cite]. Following our prime idea to build low dependency tools, we managed to build a light semi-automatic tool based on PGSQL full-text ranking feature. Once the concept table has been loaded with both standard and local concepts the full text index ranks the top n standard concepts that best match the local codes based on description or label. This work was followed and check by a physician.

- The key table for omop is the concept table. The standard vocabulary of OMOP is mainly based on the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) - A mapping between many classifications and the standard omop ones (ICD-9 and snomed-CT for examples) is already provided with *concept_relationship* table. We have used this to the maximum extent possible (laboratory exams, exit diagnoses and procedures) - concept-driven methodology : as the omop model did we adopt a "concept-driven methodology", domain of each local concept drive the concept to the right table. MIMIC design is functional since tables hold data per medical domain (eg: chartevents table contains information that are displayed as charts in the ICUs). OMOP has a different approach and pool together data by thematic or domain (eg: any data which derived from a measurement goes into the measurement table). The ETL needs then to split some MIMIC tables into multiple domain and consequently OMOP tables. To get more readability in the ETL we decided that the choice on which domain belongs a MIMIC data is driven by the concept it is linked to. The concept driven methodology consists in dispatching the data from a single table by working on the concepts and not the ETL itself.

1.2 Data Analytics

- datathon / organization A total of 150 person and 20 teams from X countries were present for the two days event. 20 projects had been prepared thought a forum. - technical architecture OMOP have been loaded into apache HIVE 1.2.1. into ORC format. Users had access to the ORC

dataset from jupyter notebooks with, python or scala. A SQL webclient allowed teams to write SQL from presto on the same dataset. The hadoop cluster was a based on 5 computers with 16 cores and 220GO ram. The MIMIC-OMOP dataset has been loaded from a PGSQL instance to HIVE thought apache SQOOP 1.4.6. directly into the ORC format. Participants had also access to the physical modeling of the database thought schemaspy to have access to both table/column comments and primary/foreign key materialising tables relations.

1.3 Contribution

MIMIC provides a lot of SQL scripts to calculate derived scores and existing cohorts. Some of them have been translated based on the OMOP data and populates the OMOP cohort tables. Unprecedented derived informations have been introduced and loaded such corrected calcemia, kaliemia, P/F ratio, corrected osmolarity

A set of *general denormalized* tables has been built on top of the original OMOP format wich have the concept_name from the concept table for both standard and local codes. The concept table is a central piece of the OMOP format and as a result it is involved in many joins to get the concept label. A set of *specialized analytics* tables has been built on top of the original OMOP format. The microbiology events is a reorganization of data from measurements for microorganisms and related antibiograms and is inspired from MIMIC *microbiology_event* table.

We tried not to infer results. For examens whereas it's logical to put a specimen for many labevents results (as one sample of blood may be used to multiple exams) we decided to create as many specimen row as laboratory exams because the information is not present in MIMIC. It was the same when date information were not provide (*start/end_datetime* for *drug_exposure*). - chartevents and labevents provide many number field as a string which is not handy for statistical analyse. We provide a standard and easy improval by the community model to extract numerical value from string The MIMIC laboratory results have been restructured to fit in OMOP format. In particular, the numerical value (value_as_number) comes with a mathematical operator (operator_concept_id) and a measurement unit (unit_concept_id). The MIMIC semi-structured raw laboratory data have been structured with a PGSQL function to extract those information. The *note_nlp* table allows to store Natural Language Processing results derived from plain text notes. In order to evaluate this table we provided two information extraction pipelines. The *first* pipeline *section extractor* based on apache UIMA framework splits the notes into sections in order to help analysts to choose or avoid some sections from their analysis. The sections patterns (such "Illness History") have been automatically extracted from texts from regular expressions, automatically filtered by keeping only one with frequency higher than 1 percent and manually filtered to exclude false positives with a total of 1200 sections. The resulting sections patterns candidate have been then manually regrouped into similar 400 groups. The extracted sections have not been

mapped to any standard terminology such LOINC CDO. The reason is the CDO LOINC has decided to stop to maintain and to remove it's sections from its standard arguing it is too difficult to maintain, and this sections are not widely used [**?**]. The *second* pipeline has extracted some numeric values such weight, height, body mass index and cardiac left ventricular ejection fraction within medical notes with a python script. The resulting structured numerical values have been loaded in the measurement or the observation tables.

2 Result

2.1 Transformation

The MIMIC to OMOP conversion was done by two developers (one data engineer and one intensivist) in an estimate amount of 500 hours. This includes the ETL, the git documentation, the concept mapping, the contributions and the unit tests. The ETL (including unit tests, and generation of the ready to load archive) on the 100 patient subset takes five minutes and allows fast cycles of developments. On the full MIMIC dataset the ETL lasts 3 hours. The resulting csv archive is slightly the same size than the original one, and this is also the same once instanced in PGSQL and indexed.

In order to evaluate how well transformed is the data we first adopt the data quality evaluation grid (2.1) developed by Khan and Al [**?**] which is commonly used as a reference [**?**].

- size of MIMIC OMOP, row number for the bigest relation (measurement)

For the prescritions MIMIC table 75% (a verifier) of drugs had a gsn code. The conept_relationship table provide mapping between gsn and RxNorm classsifications. To improve the mapping we then proceeded to a manual mapping

The OMOP-CDM contains 37 data tables. We populated 19 tables. From MIMIC we create a standardized model called MIMIC-OMOP.

Quality evaluation criteria 1) Several articles tried to evaluate CDM quality (8, 9). The criterias developed by Khan and Al, which referenced Moody and Shanks (10) metrics were adapted to our study. Our study doesn't want to evaluate the quality of CDM. But we adapt these criterias to assess our ETL work.

Understandability and simplicity will be evaluated in the analytics parts, in real life application.

2) During the all ETL process we created a lot of unit tests thanks to pgTap library. All are available on our github. All the test passed.

3)

Compleness - semantic mapping

The table 2.1 shows where the informations goes and links between MIMIC tables and OMOP tables. Since OMOP is a conceptual model, a same type of data goes in the same table. The best example may be the measurement table which is field by 4 source tables. Is is because of all the numerics datas should go to this table.

Ajouter schema : MIMIC-OMOP_equivalence.png

As shown, all the MIMIC domain are linked to proper OMOP domain. The semantic mapping was not a problem for our work.

Completeness - structural mapping

Baseline characteristics : comparison MIMIC / MIMIC OMOP (basic statistics)

The following table lists the baseline characterization of the population of OMOP compared with MIMIC.

Hopefully most statistics remains similar between two versions. Still some differences exists. Table 2.1 MIMIC contains 61.532 stays in ICU whereas OMOP contains 71.576 ICU stays. That is a 16% increase in stays due to our ETL methodology as explained in the methods.

The table 2.1 shows the number of laboratory measurement per admissions is increased. This is because the laboratory data from MIMIC chartevents has been extracted and treated as laboratory. All the code to created this statistics are provided here (cf extra : basic_statistics.sql).

Loss of datas

We tried to evaluate the percentage of loaded records from the source database to OMOP. We evaluate the percentage of columns and rows lost in the process as done other studies [15]

Depending on tables 40% to 80% of sources columns which do not fit to OMOP where deleted. The exact removed columns are provided in the appendix (cf extras) Almost all the removed columns are redundant with others or provide derived informations. The main concern are some lost timestamps. For example the MIMIC chartevents tables provide the storetime and charttime columns but OMOP only provide one slot for timestamp to stored. Thus storetime was deleted during the ETL. The erroneous rows 2.1 were deleted in the process (marked with a status column in MIMIC tables inputevents_mv, chartevents, procedureevents_mv, note). As MIMIC team told us that they will remove it in the next release because this data are poor quality we decided to do the same. The following table shows the number of rows with errors.

terminology mapping coverage

These results include automatic and manual mapping.

The unmapped concept are the concept_id = 0 (No mapping concept). To improve this mapping we need collaborative work. We provide our csv mapping files on our github. The terminology mapping has been evaluated by a physician. The zero value for concept_id may appear in very different cases. First case the local concept has no equivalent in the standard set of concept. Second case it has not been yet been mapped and may have an standard equivalent. Third case the value is missing and cannot be mapped. In our opinion while all those cases cannot be used for standard queries, they should have a different concept_id in order to be addressed differently.

- % of domain_id not in adequation with table name - some are logical because observation domain may be measurement table and vice verca

- we have mapped many source concept to one standard concept_id. This is because MIMIC provide a lots of equivalent free text concepts. For example for the body temper-

Table 1. Data Transformation Quality Evaluation Metrics

Data Model Dimension	Descriptions
Completeness - semantic mapping	Domain coverage : coverage of sources domains that are accommodated by the standard OMOP model
Completeness - structural mapping	Data coverage : coverage of sources data concepts that mapped to standard OMOP concept
Integrity	"Meaningful data relationships and constraints that uphold th eintent of the data-s original purpose" Khan and Al
Flexibility	The ease to expand the standard model for new datatypes, concepts
Integration	The capacity of the standard model to use multiples terminology and links its to standard one
Implementability	The stability of the models, the community, the cost of adoption
Understandability	The ease of the standard model to be understood
Simplicity	The ease of querying the standard model - the model should contains the minimum of concepts and relationship

ature MIMIC provide 11 distinct concept (Temperature F, Temperature C (calc), Temp Skin [C], Temperature Fahrenheit, Temp Axillary [F], Temperature C, Temperature F (calc), Temperature Celsius, Temp Rectal [F], Temp Rectal, Blood Temperature CCO (C)). Our mapping links all of it to one single concept called temperature. All the units have been converted to celcius. TODO: give example of generalisation (admission_location_to_concept)

Flexibility =========== OMOP had a 100% match of the data models constraints and relationship. Two important tables are provided with the OMOP models to match relationship : concept_relationship and fact_relationship. The fact_relationship table which is a important part of the OMOP CDM. It is used to represent the relationship between data. We used to link drugs in a solution, for microbiology / antibiograms and for visit_detail and caresite. The SQL following query shows how a microorganism is linked to its antibiogram thanks to fact_relationship

Integration ===========

OMOP Terminology coverage has been previously evaluated as excellent [20]. We used the OMOP mapping for NDC-RxNorm, ICD9-SNOMED, CPT4-SNOMED. It was really helpful because MIMIC provides lots of non standard terminology already mapped by OMOP community We tried to evaluated this OMOP mapping. We check 100 items for each mapping used (NDC, ICD9 and CPT4). ICD9 and CPT4 are correcly mapped to SNOMED. But only 85% of NDC are linked to a correct RxNorm code. In part due to incorrect NDC code (from MIMIC), in part because only 78% of NDC codes are mapped to Rxnorm.

But the OMOP common standard vocabulary, SNOMED-CT, did not cover all ICD-9-CM codes (95%). Moreover, not all ICD-9-CM codes can have one-to-one mapping to SNOMED, some are one-to-many (28%)[?]

Implementability ================ - OMOP available since 9 years - this models and its provided concepts are licence free - the community is large and was very helpful. - Full versions (V6, 7 etc.) are usually released each year (1-Jan) and are not backwards compatible. Minor versions (V5.1, 5.2 etc.) are not guaranteed to be backwards compatible though an effort is made to make sure that current queries will not break. Micro versions (V5.1.1, V5.1.2

etc.) are released irregularly and often, and contain small hot fixes or backward compatible changes to the last minor version. (7) TODO: forum centric + github + themis -¿ confusing

Finally OHDSI provides ACHILLES as data quality tool for OMOP. It has been a common practice to run the tests and use this as a quality evaluation in numerous work [?]

Achilles for quality assessment ============================ A many previous authors, we used Achilles software to evaluate the data quality(4). It is open-source analytics software produced by OHDSI (6). This tool is used for data characterization, data quality assessment (Achilles Heel), and the visualization of observational health data (6). ACHILLES calculates summary statistics and includes a unique function for checking data quality, named Achilles Heel. Other team used this tool to practice data quality assess(4). Achilles Heel issued x errors and y warnings.

- 18h 50k patients: this testifies the model needs structural optimisations - difficult pour ajoute fr. - extension achilles how to ? - comparison with other paper about error/warnings. (Korean Yoon 36-¿28 errors) TODO: minimize achille errors (12-¿? errors)

2.2 Analytics

- French Paris hospital organized a datathon with MIMIC-OMOP 25 teams, 160 participants had 48 hours to undertake a clinical project using the OMOP MIMIC database thought 15000 requests. They had the opportunity to create mixed teams : clinicians brought the questions which need data mining, along with their expertise of the data ; data scientists judged the technical feasibility and eventually implement the various analysis needed This datahon had tested OMOP model in real statistical condition. A datathon was organised in collaboration with the MIT.(3)

- AP-HP calculation clusters, able to access to the data pre-loaded in Jupyter environments, where will be installed the most popular tools and libraries in R and Python, with Hadoop Spark

schema big data platform

Table 2. MIMIC to OMOP data flows

OMOP tables	Number of rows	MIMIC tables
PERSONS	46520	patients, admissions
DEATH	14849	patients, admissions
VISIT_OCCURRENCE	58976	admissions
VISIT_DETAIL	271808	transfers, service
MEASUREMENT	366272371	chartevents, labevents, microbiologyevents, outputevents
OBSERVATION	6721040	admissions, chartevents, datetimevvents, drgcodes
DRUG_EXPOSURE	24934758	prescriptions, inputevents_cv, inputevents_mv
PROCEDURE_OCCURRENCE	1063525	cptevents, procedureevents_mv, procedure_icd
CONDITION_OCCURRENCE	716595	admissions, diagnosis_icd
NOTE	2082294	notevents
NOTE_NLP	16350855	noteevents
COHORT_ATTRIBUTE	2628838	callout
CARE_SITE	93	transfers, service
PROVIDER	7567	caregivers
OBSERVATION_PERIOD	58976	patients, admissions
SPECIMEN	39874171	chartevents, labevents, microbiologyevents

MIMIC-OMOP dataset was 10 GO sized as ORC format. On the other side the same dataset is sized 400 GO on a PGSQL instance. Generating the indexes ORC automatically generates indexing. However these are much light than the PGSQL btree indexes specified by the OMOP DDL.

Writing standard queries (ie: with standard concepts) need to be familiar with the concept_relationship design and also master graph nature of some terminology such SNOMED-CT in order to grab all potential codes that might be related to the one analysts think of primarily. This complexity is inherent of terminology complexity and the closure table **??** design handle perfectly.

2.3 Contribution

The *denormalized derived* tables appears to improves the computations costs and the verbosity of the SQL queries. In addition the resulting tables are much more human readily with the label of the concepts directly in the fact table. Hence a bit of denormalization improves greatly the experience of the data scientist by adding some redundancy in the data while not breaking existing SQL queries. The drawback is it would make it more tricky to update and keep the dataset consistent - which is non applicable because OMOP is generally a static dataset. The *microbiology derived* table simplifies the experience for data-scientists.

We provided many derived values. Community is welcome to improve it - From noteevents : weight, height, LVEF - From measurement : SOFA, IGSII, F/P, corrected Ca / K, BMI, corrected osmolarity

We also provided materialized views with denormalized structures : microbiology tables This is a more ICU centric data structures that may help for analytics.
- ethnicity_concept_id : only two strange concept_name Hispanic or non_Hispanic

3 Discussion

The computation time of the ETL on the PGSQL instance on a modest personal computer is compatible with a community work where collaborator can clone the source code and setup a development instance to reproduce or improve the work. The choice of the ELT based mostly in SQL code lets end users with SQL background only to review and enhance the work. As a result, the targeted community is as broad as possible and we expect translational profiles to get involved.

The datathon has shown that distributed platforms with commodity hardware provides SQL tools allowing OLAP analysis with great performances that overcome OLTP RDBMS weaknesses. Hence it takes advantage of SQL language analytics features such grouping, windowing, joining and mathematic functions that often lack in NOSQL databases.

It is important that OMOP keeps a level of normalisation in order to simplify the ETL and make it consistent. However once done, it is judicious to give access to data-scientist to more denormalized tables and more specialized tables. Multiple concerns exists about OMOP performances and optimization. However there will never be a perfect multi-use case table, and this is the reponsibility of the data scientist to build his own tables, simplified, spe-

Table 3. Baseline characteristics MIMIC versus OMOP

items	MIMIC	OMOP
Persons (Number)	46.520	46.520
Admissions (Number)	58.976	58.976
Icustays (Number)	71.576	61.532
Gender, Female (Number, %)	20.399	20.399 (43 %)
Age (Mean)	64 years, 4 months	64 ans, 4 months
0-5	8110	8110
6-15	1	1
16-25	1434	1434
26-45	5962	5962
46-65	17375	17375
66-80	15793	15793
>80	10301	10301
Emergency	42071	42071
Elective	7706	7706
Surgical patients	19246	19246
Length of stay, hospital (median)	6.46 (Q1-Q3 : 3.74 -11.79)	6.59 (Q1-Q3 : 3.84 - 11.88)
Length of stay, ICU (median)	2.09 (Q1-Q3 : 1.10 - 4.48)	1.87 (Q1-Q3 : 0.95 - 3.87)
Mortality, ICU (Number, %)	5814 (9%)	5815 (9%)
Mortality, hospital (Number, %)	4511 (7%)	4559 (6%)
Lab measurements per admissions (mean)	478	678
Procedures per admissions (mean)	4.6	4.6
Drugs per admissions (mean)	82.8	82.8
Exit dignosis per admissions (mean)	11.0	11.0

cialized for his research and answer efficiently and clearly his needs.

Derived data integrates quite well in OMOP. We made use of note_nlp to store information derived from notes, measurement to store numerical information and cohort_attributes to store scores. However it is still unclear if derived data should be stored per domain or if it should be stored in dedicated derived tables. We found out that there is a lack of tables to track provenance and description of such data.

An other missing aspect is some quality tables to assess and measure the quality of data. MIMIC had some column to keep track of corrupted information. It would be of interest to be able to keep the messy data and allow research on data cleaning and data quality and avoid removing information.

Last but not least, as stated in the introduction a good CDM for ICU would allow near realtime early warning systems and model inference on fresh data. OMOP is clearly designed to provide static dataset and does not have mecanisms for realtime ingestion, and data version control - it is not a datawarehouse. Making analysis on statics datasets is essential in order to have reproducible results. However when the algorithm needs to be moved at bed side, there is a need for a freshness of data, and a way to identify the patient that OMOP won't easily provide. That being said a solution such FHIR is a great way to implement realtime inference from EHR data and that's how FHIR and OMOP are complementary that yet have been investigated [?] but needs further optimizations.

During this work the OMOP forum was very active. Working groups. It is a challenge to manage such large community from all moderator, contributors and from a user perspective. It appears it is not doable for most of people to get involved. The forum is full of details and information. It contrasts with the implementation guide that suffer from not being as well detailed. We think the OMOP community would greatly benefit from systematic and synthetic synchronisation between forum, mailing lists, github and end user documentation.

The datathon real life test has revealed the strong need to make the physical data model including comments on columns and table accessible and we found out that the open-source tool schema spy was of good help. In addition limit the misunderstanding by delivering SQL examples and we found out that the git repository is the best place for documenting and interact with the community.

OMOP has been designed to build static

Table 4. Row level Data lost

Relations	Error Percentage
inputevents_mv	10,00%
chartevents	0.04%
procedureevents_mv	3,00%
Note	0.04%

4 Conclusion

The MIMIC to OMOP transformation has needed efforts that remains reasonable. It is and will ever be a work in progress as the standard concept mapping is a quite infinite process with constant improvement. Fortunately the released version is ready for research and already offers the same perimeter of data as the original MIMIC version and even more with the derived data.

As seen OMOP model efficiency has some weaknesses as it looks to put the cursor more on consistency than on performances. However we have shown that it's easy to overcome the issues and enhance OMOP with a set of design or technology optimization and dedicated structure that in the end remains standards and shareable because derive from the original model.

As compared to the original MIMIC data model, working on OMOP offers the opportunity to write standard code and analysis that might benefit from or to other users internationally. The MIMIC-OMOP database is available online on physionet as well as the original MIMIC database. All the existing work is publicly available on github [?] and has been designed to be easily reviewed, copied or enriched easily accordingly to OMOP or MIMIC open-source philosophy by any end-user aware of SQL.

Future work on evaluation of the existing concept mapping though practical research studies on both local and standard coding will be made. In addition we expect to enhance the OHDSI USAGI concept mapping tool to allow international concept mapping suggestion to transform other foreign ICU databases. Finally investigation on how to articulate FHIR and OMOP to get best of two worlds (patient level versus multicentric level informations) and improve care practically at bed side will be done.

4.1 Chirp-Like Impulse Responses and Group Delay

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable

changes are obtained that are similar to the spectral delay effect [23], [24].

4.2 Contributions
4.2.1 Chirp-Like Impulse Responses and Group Delay

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

$$A(z) = \frac{a_1 + z^{-1}}{1 + a_1 z^{-1}}, \tag{1}$$

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay.[1] Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

$$\tau_{g,max} = \begin{cases} \tau_g(0) = \frac{1-a_1}{1+a_1}, \text{when } a_1 \leq 0 \\ \tau_g(\pi) = \frac{1+a_1}{1-a_1}, \text{when } a_1 > 0. \end{cases} \tag{2}$$

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio

---

[1] This point is emphasized by Loewer, see esp. p. (610).

Table 5. Terminology Mapping coverage

Omop tables (domain)	Total_records	% Mapped_records	Total_concepts_source	% Mapped_concepts_source
PERSONS	93040	100,00%	43	100,00%
VISIT_OCCURRENCE	58976	100,00%	34	100,00%
VISIT_DETAIL	396930	100,00%	28	100,00%
MEASUREMENT				
OBSERVATION				
DRUG_EXPOSURE	24934751	37,00%	7410	53,00%
PROCEDURE_OCCURRENCE	1063525	99,00%	2218	98,00%
CONDITION_OCCURRENCE	716595	92,00%	6984	95,00%
CARE_SITE	144	100,00%	58	100,00%
SPECIMEN	39874171	70,00%	92	77,00%

effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

1) Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay.

2) Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22].

3) In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses.

4) When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

1) Green–function determined experimentally and published.

2) Black–function determined using similarity searches and published.

3) Red–function determined using similarity searches and determined in this study.

4) Blue–O-antigen structure unknown. Function determined using similarity searches and proposed in this study.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].
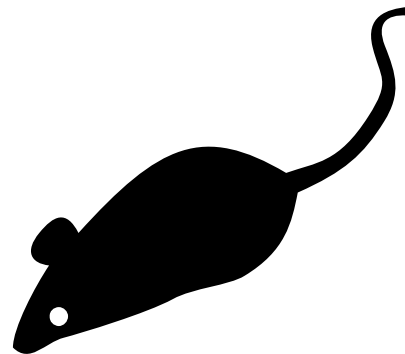
Fig. 1. The spectral delay filter consists of M allpass filters and an equalization filter.

Table 6. Active sites and allosteric sites of the GNE
MNK enzyme

Excerpt No.	Genre	Spatial Mode	Corrlation
1	Pop	FB	94%
2	Classical	FB	33%
3	Jazz	FF	76%
4	Arabian	FF	41%
5	GNE	H220	45%
6	GNE	H45	93%
7	MNK	G416	74%
8	MNK	D413	72%
9	MNK	R420	94%
10	MNK	N516	91%

Note. This table does not include sentence enhancement statutes.
This table does not include sentence enhancement statutes.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay.

$$\tau_{\mathrm{g}}(\omega) = -\frac{d\phi(\omega)}{d\omega}.$$

Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect.

a) Green–function determined experimentally and published.
b) Black–function determined using similarity searches and published.
c) Red–function determined using similarity searches and determined in this study.
d) Blue–O-antigen structure unknown. Function determined using similarity searches and proposed in this study.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22].

Example 1. In this paper, we investigate audio effects processing using high-order allpass filters that consist of

many cascaded low-order allpass filters. These filters have long chirp-like impulse responses.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies.

- Green–function determined experimentally and published.
- Black–function determined using similarity searches and published.
- Red–function determined using similarity searches and determined in this study.
- Blue–O-antigen structure unknown. Function determined using similarity searches and proposed in this study.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

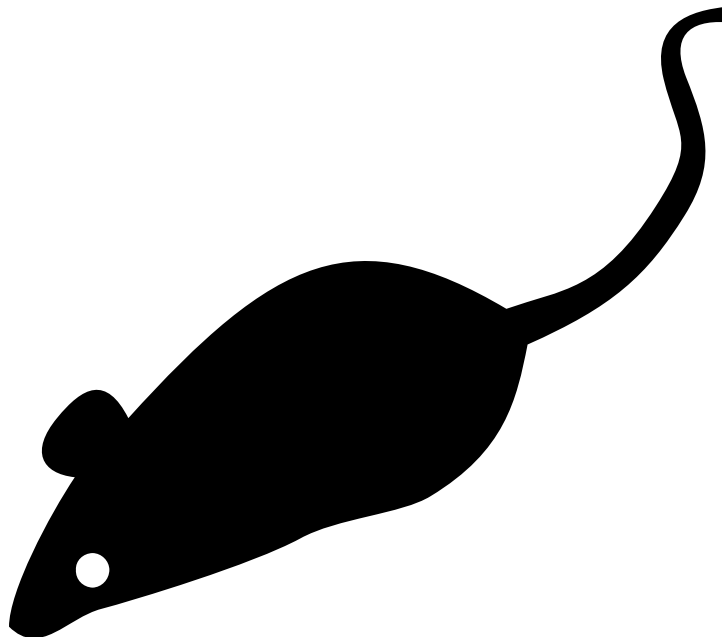Green–function determined experimentally and published.

Fig. 2. This paper is organized as follows. In Section 1, we discuss the group delay of a cascade of first-order allpass filters and its relation to the chirp-like impulse response of the spectral delay filter. Furthermore, a multirate method to stretch the impulse response of the spectral delay filter is proposed. Section 2 discusses the amplitude envelope of the impulse response and suggests a design method for the equalizing filter. Section 3 presents application examples using the spectral delay filter. Section 4 concludes this paper.

Black–function determined using similarity searches and published.

Red–function determined using similarity searches and determined in this study.

Blue–O-antigen structure unknown. Function determined using similarity searches and proposed in this study.

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

5 SUMMARY

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that con-

sist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

6 CONCLUSION

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22]. In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24]. Note that articles might have a digital object identifier [25].

7 ACKNOWLEDGMENT

8 REFERENCES

[1] D. C. Angus, M. A. Kelley, R. J. Schmitz, A. White, J. Popovich, "Caring for the critically ill patient. Current and projected workforce requirements for care of the critically ill and patients with pulmonary disease: can we meet the requirements of an aging population?" *JAMA*, vol. 284, no. 21, pp. 2762–2770 (2000 Dec).

[2] E. Azoulay, C. Alberti, I. Legendre, C. B. Buisson, J. R. Le Gall, "Post-ICU mortality in critically ill infected patients: an international study," *Intensive Care Med*, vol. 31, no. 1, pp. 56–63 (2005 Jan).

[3] J. L. Vincent, "Is the current management of severe sepsis and septic shock really evidence based?" *PLoS Med.*, vol. 3, no. 9, p. e346 (2006 Sep).

[4] M. K. Ross, W. Wei, L. Ohno-Machado, ""Big data" and the electronic health record," *Yearb Med Inform*, vol. 9, pp. 97–104 (2014 Aug).

[5] Y. Zhang, S. L. Guo, L. N. Han, T. L. Li, "Application and Exploration of Big Data Mining in Clinical Medicine," *Chin. Med. J.*, vol. 129, no. 6, pp. 731–738 (2016 Mar).

[6] C. Chronaki, A. Shahin, R. Mark, "Designing Reliable Cohorts of Cardiac Patients across MIMIC and eICU," *Comput Cardiol (2010)*, vol. 42, pp. 189–192 (2015).

[7] J. J. Gagne, "Common Models, Different Approaches," *Drug Saf*, vol. 38, no. 8, pp. 683–686 (2015 Aug).

[8] D. Madigan, P. B. Ryan, M. Schuemie, P. E. Stang, J. M. Overhage, A. G. Hartzema, M. A. Suchard, W. DuMouchel, J. A. Berlin, "Evaluating the impact of database heterogeneity on observational study results," *Am. J. Epidemiol.*, vol. 178, no. 4, pp. 645–651 (2013 Aug).

[9] H. Morgenstern, B. Rafaely, "Spatial Reverberation and Dereverberation Using an Acoustic Multiple-Input Multiple-Output System," *J. Audio Eng. Soc*, vol. 65, no. 1/2, pp. 42–55 (2017 Jan.Feb.), [Online]. Available: https://doi.org/10.17743/jaes.2016.0063.

[10] O. H. Klungel, X. Kurz, M. C. de Groot, R. G. Schlienger, S. Tcherny-Lessenot, L. Grimaldi, L. Ibanez, R. H. Groenwold, R. F. Reynolds, "Multi-centre, multi-database studies with common protocols: lessons learnt from the IMI PROTECT project," *Pharmacoepidemiol Drug Saf*, vol. 25 Suppl 1, pp. 156–165 (2016 Mar).

[11] Y. Xu, X. Zhou, B. T. Suehs, A. G. Hartzema, M. G. Kahn, Y. Moride, B. C. Sauer, Q. Liu, K. Moll, M. K. Pasquale, V. P. Nair, A. Bate, "A Comparative Assessment of Observational Medical Outcomes Partnership and Mini-Sentinel Common Data Models and Analytics: Implications for Active Drug Safety Surveillance," *Drug Saf*, vol. 38, no. 8, pp. 749–765 (2015 Aug).

[12] J. M. Overhage, P. B. Ryan, C. G. Reich, A. G. Hartzema, P. E. Stang, "Validation of a common data model for active safety surveillance research," *J Am Med Inform Assoc*, vol. 19, no. 1, pp. 54–60 (2012).

[13] C. Reich, P. B. Ryan, P. E. Stang, M. Rocca, "Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases," *J Biomed Inform*, vol. 45, no. 4, pp. 689–696 (2012 Aug).

[14] C. Maier, L. Lang, H. Storf, P. Vormstein, R. Bieber, J. Bernarding, T. Herrmann, C. Haverkamp, P. Horki, J. Laufer, F. Berger, G. Honing, H. W. Fritsch, J. Schuttler, T. Ganslandt, H. U. Prokosch, M. Sedlmayr, "Towards Implementation of OMOP in a German University Hospital Consortium," *Appl Clin Inform*, vol. 9, no. 1, pp. 54–61 (2018 01).

[15] F. FitzHenry, F. S. Resnic, S. L. Robbins, J. Denton, L. Nookala, D. Meeker, L. Ohno-Machado, M. E. Matheny, "Creating a Common Data Model for Comparative Effectiveness with the Observational Medical Outcomes Partnership," *Appl Clin Inform*, vol. 6, no. 3, pp. 536–547 (2015).

[16] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. Wong, P. R. Rijnbeek, J. van der Lei, N. Pratt, G. N. Noren, Y. C. Li, P. E. Stang, D. Madigan, P. B. Ryan, "Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers," *Stud Health Technol Inform*, vol. 216, pp. 574–578 (2015).

[17] A. J. McMurry, S. N. Murphy, D. MacFadden, G. Weber, W. W. Simons, J. Orechia, J. Bickel, N. Wattanasin, C. Gilbert, P. Trevvett, S. Churchill, I. S. Kohane, "SHRINE: enabling nationally scalable multi-site disease studies," *PLoS ONE*, vol. 8, no. 3, p. e55811 (2013).

[18] A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, P. J. Liu, X. Liu, M. Sun, P. Sundberg, H. Yee, K. Zhang, G. E. Duggan, G. Flores, M. Hardt, J. Irvine, Q. V. Le, K. Litsch, J. Marcus, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. Howell, C. Cui, G. Corrado, J. Dean, "Scalable and accurate deep learning for electronic health records," *CoRR*, vol. abs/1801.07860 (2018), URL `http://arxiv.org/abs/1801.07860`.

[19] H. Morgenstern, B. Rafaely, "i2b2 implemented over SMART-on-FHIR," *J. Audio Eng. Soc*, vol. 65, no. 1/2, pp. 42–55 (2017 Jan.Feb.), [Online]. Available: https://doi.org/10.17743/jaes.2016.0063.

[20] M. Garza, G. Del Fiol, J. Tenenbaum, A. Walden, M. N. Zozus, "Evaluating common data models for use with a longitudinal community registry," *J Biomed Inform*, vol. 64, pp. 333–341 (2016 12).

[21] D. Preis, "Phase Distortion and Phase Equalization in Audio Signal Processing—A Tutorial Review," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 774–779 (1982 Nov.).

[22] J. S. Abel, D. P. Berners, "MUS424/EE367D: Signal Processing Techniques for Digital Audio Effects," (2005), unpublished Course Notes, CCRMA, Stanford University, Stanford, CA.

[23] C. Roads, "Musical Sound Transformation by Convolution," presented at the *Int. Computer Music Conf.*, pp. 102–109 (1993).

[24] C. Roads, *The Computer Music Tutorial* (MIT Press, Cambridge, MA), 1st ed. (1996).

[25] H. Morgenstern, B. Rafaely, "Spatial Reverberation and Dereverberation Using an Acoustic Multiple-

Input Multiple-Output System," *J. Audio Eng. Soc*, vol. 65, no. 1/2, pp. 42–55 (2017 Jan.Feb.), [Online]. Available: https://doi.org/10.17743/jaes.2016.0063.

APPENDIX

Filtering an audio signal with an allpass filter does not usually have a major effect on the signal's timbre. The allpass filter does not change the frequency content of the signal, but only introduces a phase shift or delay. Audibility of the phase distortion caused by an allpass filter in a sound reproduction system has been a topic of many studies, see, e.g., [21], [22].

$$\phi(\omega) = -\omega + 2\arctan\left(\frac{a_1 \sin\omega}{1 + a_1 \cos\omega}\right) \qquad (1)$$

In this paper, we investigate audio effects processing using high-order allpass filters that consist of many cascaded low-order allpass filters. These filters have long chirp-like impulse responses. When audio and music signals are processed with such a filter, remarkable changes are obtained that are similar to the spectral delay effect [23], [24].

NOMENCLATURE

a_c = condensation coefficient condensation coefficient condensation coefficient

TLR = Toll-like receptor

PAMPs = pathogen-associated molecular patterns condensation coefficient condensation

---

THE AUTHORS

A1firstname A1lastname
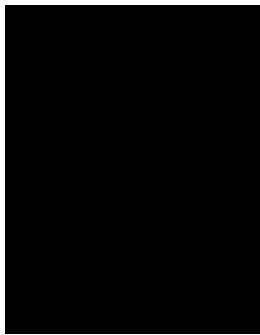
A2firstname A2lastname

A1firstname A1lastname is professor of audio signal processing at Helsinki University of Technology (TKK), Espoo, Finland. He received his Master of Science in Technology, Licentiate of Science in Technology, and Doctor of Science in Technology degrees in electrical engineering from TKK in 1992, 1994, and 1995, respectively. His doctoral dissertation dealt with fractional delay filters and physical modeling of musical wind instruments. Since 1990, he has worked mostly at TKK with the exception of a few periods. In 1996 he spent six months as a postdoctoral research fellow at the University of Westminster, London, UK. In 2001-2002 he was professor of signal processing at the Pori School of Technology and Economics, Tampere University of Technology, Pori, Finland. During the academic year 2008-2009 he has been on sabbatical and has spent several months as a visiting scholar at the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, Stanford, CA. His research interests include musical signal processing, digital filter design, and acoustics of musical instruments. Prof. Välimäki is a senior member of the IEEE Signal Processing Society and is a member of the AES, the Acoustical Society of Finland, and the Finnish Musicological Society. He was the chairman of the 11th International Conference on Digital Audio Effects (DAFx-08), which was held in Espoo, Finland, in 2008.

●

A2firstname A2lastname is a consulting professor at the Center for Computer Research in Music and Acoustics (CCRMA) in the Music Department at Stanford University where his research interests include audio and music applications of signal and array processing, parameter estimation, and acoustics. From 1999 to 2007, Abel was a co-founder and chief technology officer of the Grammy Award-winning Universal Audio, Inc. He was a researcher at NASA/Ames Research Center, exploring topics in room acoustics and spatial hearing on a grant through the San Jose State University Foundation. Abel was also chief scientist of Crystal River Engineering, Inc., where he developed their positional audio technology, and a lecturer in the Department of Electrical Engineering at Yale University. As an industry consultant, Abel has worked with Apple, FDNY, LSI Logic, NRL, SAIC and Sennheiser, on projects in professional audio, GPS, medical imaging, passive sonar and fire department resource allocation. He holds Ph.D. and M.S. degrees from Stanford University, and an S.B. from MIT, all in electrical engineering. Abel is a Fellow of the Audio Engineering Society.