Multi-way Hierarchic Classification of Musical Instrument Sounds

Alicja A. Wieczorkowska
Polish-Japanese Institute
of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland
alicja@pjwstk.edu.pl

Zbigniew W. Raś, Xin Zhang, Rory Lewis
UNC-Charlotte, Computer Science Dept.
9201 University City Blvd.
Charlotte, NC 28223, USA
ras@uncc.edu

Abstract

Musical instrument sounds can be classified in various ways, depending on the instrument or articulation classification. This paper reviews a number of possible generalizations of musical instruments sounds classification which can be used to construct different hierarchical decision attributes. Each decision attribute will lead us to a new classifier and the same to a different system for automatic indexing of music by instrument sounds and their generalizations. Values of a decision attribute and their generalizations are used to construct atomic queries of a query language built for retrieving musical objects from MIR Database (see http://www.mir.uncc.edu). When query fails, the cooperative strategy will try to find its lowest generalization which does not fail, taking into consideration all available hierarchical attributes. This paper evaluates only two hierarchical attributes upon the same dataset which contains 2628 distinct musical samples of 102 instruments from McGill University Master Samples (MUMS) CD Collection.

1. Introduction

Classification of musical instrument sounds can be performed in various ways [3, 5]. This paper reviews several hierarchical classifications of musical instrument sounds but concentrates only on two of them: Hornbostel-Sachs classification of musical instruments and classification of musical instruments by articulation with 15 different articulation methods (seen as attribute values): blown, bowed, bowed vibrato, concussive, hammered, lip-vibrated, martele, muted, muted vibrato, percussive, picked, pizzicato, rubbed, scraped and shaken. Each hierarchical classification represents a unique decision attribute which leads us to a discovery of a new classifier and the same to a different system for automatic indexing of music by instruments and their certain generalizations. Values of all hierarchical attributes leading to classifiers of high confidence will be used to construct atomic queries for automatic and flexible [1], [2] retrieval of music by instruments and their generalizations.

2. Classification of musical instruments

In this section, we discuss various ways of classification of musical instrument sounds, based on classification of musical instruments [3, 5] and articulation, i.e. the method of playing.

The main classification, based on the Hornbostel and Sachs system (with extensions) is shown in Figure 1. Basic classification includes aerophones (wind instruments), chordophones (string instruments), idiophones (made of solid, non-stretchable, resonant material), and membranophones (mainly drums); idiophones and membranophones are together classified as percussion. Additional groups include electrophones, i.e. instruments where the acoustical vibrations are produced by electric or electronic means (electric guitars, keyboards, synthesizers), complex mechanical instruments (including pianos, organs, and other mechanical music makers), and special instruments (include bullroarers, but they can be classified as free aerophones).

Each category can be further subdivided into groups, subgroups etc. and finally into instruments.

Idiophones subcategories include:

- Struck together (concussion) - claves, clappers, castanets, spoons, and finger cymbals are counted into this group
- Struck - gongs, xylophones, slit drums, steel drums, struck bells, lithophones (stones struck with other stones), metallophones
- Rubbed - musical glasses, a moistened cloth (cuica), a stick, a bow (musical saw)
- Scraped - washboards, cog rattles, animal bones, sticks

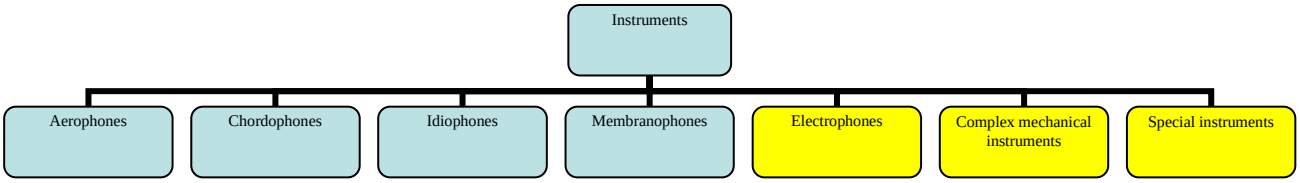

Figure 1. Hornbostel-Sachs classification of musical instruments (extensions on the right, marked in yellow

- Stamped - covered pits, curved boards, hard floors
- Shaken - rattles, jingles, gourd rattles
- Plucked - Jew's harp, thumb piano

Membranophones include the following subcategories:

- Cylindrical drum
- Conical drum
- Barrel drum
- Hourglass drum
- Goblet drum - for example darabukke
- Footed drum
- Long drum
- Kettle or pot drum
- Frame drum - tambourine, bodhran (Celtic), paddle drum
- Friction drum
- Mirliton/kazoo - this is not a drum - this group includes kazoo, comb and waxed paper, zobo

Chordophones subcategories include:

- Zither - simple zither, long zither, plucked board zither (psaltery, harpsichord, virginal, spinet), and struck board zither (dulcimer, clavichord, piano)
- Lute (plucked) - mandolins, guitars, ukuleles
- Lute (bowed) - viols (fretted neck), fiddles, violin, viola, cello, double bass, and hurdy-gurdy (no frets)
- Harp - bow or arched harp (the neck is bent like a bow over the resonator), angle harp (the neck runs straight at angle over the resonator), frame harp
- Lyre
- Bow

Aerophones are classified according to the mouthpiece used to set air in motion to produce sound (blow hole, whistle, single and double reed, lip vibrated):

- End-blown flute
- Side-blown flute - transverse flute
- Nose flute
- Globular flute - simple referee whistle, ocarina
- Multiple flutes
- Panpipes
- Whistle mouthpiece - recorder, flageolet
- Single reed - examples: clarinet, saxophones
- Double reed - examples: oboe, shawm bassoon, contrabassoon
- Air chamber - bagpipes; concertinas, accordions, harmonicas, pipe organs (considered free-reed instruments because there is a reed for each pipe or note desired)
- Lip vibrated (trumpet or horn), called brass (trumpets, French horn, trombones)
- Free aerophone - examples: bullroarers, spinning humming tops, buzzing discs

Aerophones subcategories are also called woodwinds or brass, but this criterion is not based on the material the instrument is made of, but rather on the method of sound production. In woodwinds, the change of pitch is mainly obtained by the change of the length of the column of the vibrating air. Additionally, overblow is applied to obtain second, third or fourth harmonic to become the fundamental. In brass instruments, overblows are very easy because of wide bell, and therefore overblows are the main method of pitch changing.

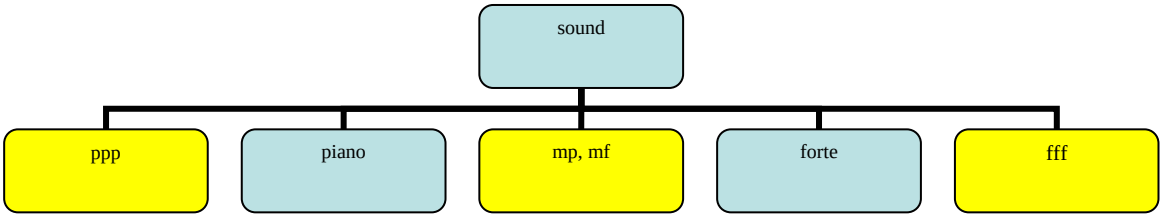

Figure 2. Classification of musical instrument sounds according to the dynamics (main levels marked in blue, additional levels marked in yellow

Musical instrument sounds can be also classified according to the dynamics, i.e. how loudly they are played. Exemplary method of classification is shown in Figure 2.

However, the problem may appear how to classify sustained sounds with dynamics changing in time: crescendo, diminuendo. The solution is to classify very short windows, and then each part of sound can be classified into a different class of dynamics.

Sounds can be also classified according to the articulation. It can be performed in three ways: (1) sustained or non-sustained sounds, (2) muted or not muted sounds, (3) vibrated and not vibrated sounds.

This classification may also be difficult, since the vibration may not appear in the entire sound; some changes may be visible, but no clear vibration. Also, brass is sometimes played with moving the mute in and out of the bell.

According to the contents of the spectrum, the musical instrument sounds can be classified into the following three types: (1) harmonic spectrum, (2) continuous spectrum, or (3) mixed spectrum.

Most of music instrument sounds of definite pitch have some noises/continuity in their spectra.

According to MPEG-7 classification [4], there are four classes of musical instrument sounds: (1) Harmonic, sustained, coherent sounds - well detailed in MPEG-7, (2) Nonharmonic, sustained, coherent sounds, (3) Percussive, nonsustained sounds - well detailed in MPEG-7, (4) Noncoherent, sustained sounds.

This also can be misleading, since pizzicato is not clearly present in this classification, as harmonic, non-sustained sound.

3. Comparison and the need of different models for musical sounds classification

In this section we outline a formal framework for evaluation and comparison of different classifications of musical sounds. Musical instruments are represented as leaves of a hierarchical decision attribute, denoted in our case by d, and its different types and subtypes are represented as internal nodes of d. We can also cluster musical instruments in a number of ways and the same generate many possible hierarchical structures each defining a new decision attribute. In our database, musical instruments are represented as sample musical sounds described by a large number features, denoted by A, including MPEG7 descriptors and other/non-MPEG7 descriptors in the acoustical perspective of view, where both spectrum features and temporal features are taken.

The goal of each classification is to find descriptions of musical instruments or their classes (values of attribute d) in terms of values of attributes from A. Each classification results in a classifier which can be evaluated using standard methods like bootstrap or cross-validation. In this paper we use ten-fold cross-validation.

Let us assume that $S = (X, A \cup \{d\}, V)$ is a decision system, where d is a hierarchical attribute. We also assume that $d_{[i_1,...,i_k]}$ (where $1 \leq i_j \leq m_j$, $j = 1, 2..., k$) is a child of $d_{[i_1,...,i_{k-1}]}$ for any $1 \leq i_k \leq m_k$. Clearly, attribute d has $\Sigma\{m_1 \cdot m_2 \cdot ... \cdot m_j : 1 \leq j \leq k\}$ values, where $m_1 \cdot m_2 \cdot ... \cdot m_j$ shows the upper bound for the number of values at the level j of d. By $p([i_1, ..., i_k])$ we denote a path (d, $d_{[i_1]}$, $d_{[i_1,i_2]}$, $d_{[i_1,i_2,i_3]}$,..., $d_{[i_1,...,i_{k-1}]}$, $d_{[i_1,...,i_k]}$) leading from the root of the hierarchical attribute d to its descendant $d_{[i_1,...,i_k]}$.

In this section, we initially concentrate on classifiers built by rule-based methods (for instance: LERS, RSES, PNC2) and next on classifiers built by tree-based methods (for instance: See5, J48 Tree, Assistant, CART, Orange).

Let us assume that R_j is a set of classification rules extracted from S, representing a part of a rule-based classifier $R = \bigcup\{R_j : 1 \leq j \leq k\}$, and describing all values of d at level j. The quality of a classifier at level j of attribute d can be checked by calculating $Q(R_j) = \frac{\sum\{sup(r) \cdot conf(r) : r \in R_j\}}{\sum\{sup(r : r \in R_j\}}$, where $sup(r)$ is the support of the rule r in S and $conf(r)$ is its confidence. Then, the quality of the rule-based classifier R can be checked by calculating $Q(\bigcup\{R_j : 1 \leq j \leq k\}) = \frac{\sum\{Q(R_j) : 1 \leq j \leq k\}}{k}$.

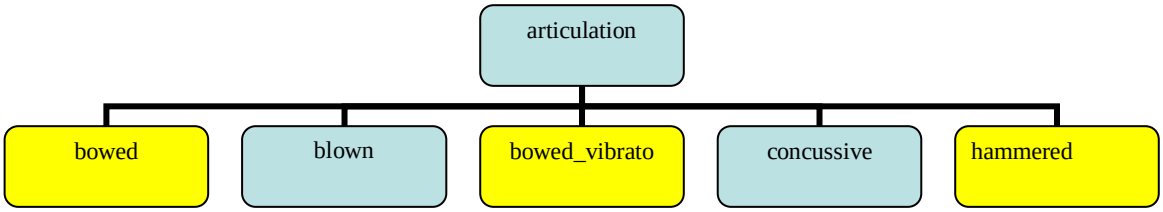

Figure 3. Articulation of musical instrument sounds in hierarchical classification

The quality of a tree-based classifier can be given by calculating its quality for every node of a hierarchical decision attribute d. Let us take a node $d_{[i_1,...,i_k]}$ and the path $p([i_1,...,i_k])$ leading to that node from the root of d. There is a set of classification rules $R_{[i_1,...,i_m]}$, uniquely defined by the tree-based classifier, assigned to a node $d_{[i_1,...,i_m]}$ of a path $p([i_1,...,i_k])$, for every $1 \leq m \leq k$. Now, we define $Q(R_{[i_1,...,i_m]})$ as $\frac{\sum\{sup(r)\cdot conf(r): r \in R_{[i_1,...,i_m]}\}}{\sum\{sup(r): r \in R_{[i_1,...,i_m]}\}}$. Then, the quality of a tree-based classifier for a node $d_{[i_1,...,i_m]}$ of the decision attribute d can be checked by calculating $Q(d_{[i_1,...,i_m]}) = \prod\{Q(R_{[i_1,...,i_j]}) : 1 \leq j \leq m\}$. In our experiments, presented in Section 4 of this paper, we use *J48 Tree* as the tool to build tree-based classifiers. Also, their performance on level m of the attribute d is checked by calculating $Q(d_{[i_1,...,i_m]})$ for every node $d_{[i_1,...,i_m]}$ at the level m. Finally, the performance of both classifiers is checked by calculating $Q(\bigcup\{R_j : 1 \leq j \leq k\})$ (the first method we proposed).

Learning values of a decision attribute at different generalization levels is extremely important not only for designing and developing an automatic indexing system of possibly highest confidence but also for handling failing queries. Values of a decision attribute and their generalizations are used to construct atomic queries of a query language built for retrieving musical objects from MIR Database (see http://www.mir.uncc.edu). When query fails, the cooperative strategy [1], [2] may try to find its lowest generalization which does not fail. Clearly, by having a variety of different hierarchical structures available for d we have better chance not only to succeed but succeed with a possibly smallest generalization of an instrument class.

4. Experiments

As we mentioned in the previous section, our goal is to construct and evaluate different hierarchical classifications of musical instrument sounds and on the basis of obtained results suggest which of them look promising to form a base for our music automatic indexing system. The process of building hierarchical classifications of musical instruments is sometime very difficult and often we need to make not easy decisions in order to construct an attribute. Clearly, we have to avoid subjective decisions. Otherwise, we are forced to build ontology layer associated with a decision attribute. For example, we all agree that a trumpet can be played sustained. But the piano is very difficult to play that way, unless we decide that any damping of a sound means not sustained. For example, even though the sound will seem to be sustained when we press the loud pedal down, it is in fact damping, although slowly. So, can we also say that negligible damping over time long enough (for example, more than one second) means it is sustained?

Dynamics (see Figure 2) could constitute another base for constructing a classifier, with loudness at the root of the classification tree, and then for example only 3 branches: *ppp*, *mf/mp*, *fff*, classified on the basis of normalized energy or other criteria, but this can be misleading. However, since our database does not contain instances of the same instrument played for example $forte$ (loud) and $piano$ (soft), although there are examples of piano soft and loud in MUMS CD-library [6]), then we may consider to investigate such a classifier in our future research. The next level may include instrument+dynamics, for example trumpet+loud. We may also include some intermediate levels.

We may also have a separate classifier for spectrum. In the top node it should be binary - noise only or not. The next level (if it is not only noise): harmonic or not. For example, *bells* have inharmonic sound, i.e. spectrum, the frequencies of the elements of the spectrum are not harmonic, i.e. f_n frequency does not have to be nf_1 (example: harmonic spectrum may have 100Hz, 200Hz, 400Hz, 500Hz and so on; non-harmonic may have 100, 240, 310, 600, 650 etc.).

For simplicity reason, this paper only refers to two hierarchical schemas:

Schema 1 (Family) is described by the following attributes: instrument (for example French horn) with 108 types as its values, instrument+articulation (for example French horn muted) with 152 values, level one of Hornbostel and Sachs classification (for example aerophone) with 5 values, level two of Hornbostel and Sachs classification (for example lip-vibrated) with 19 values.

Timbre	$A1$	$A2$	$B1$	$B2$
Alto Flute	94.91%	97.43%	99.9%	99.97%
Alto Trombone	99.85%	99.93%	99.6%	99.87%
Bach Trumpet	99.85%	99.93%	99.6%	99.87%
Bass Drum	99.9%	99.95%	100%	100%
Bass Flute	99.85%	99.93%	99.60%	99.87%
Bassoont	99.8%	99.9%	99.6%	99.87%
Bass Trombone	95.55%	97.75%	99.6%	99.87%
Bells	100.00%	100%	99.9%	99.97%
Bflatclarinet	96.82%	98.4%	99.6%	99.87%
Blocks Temple	100%	100%	99.9%	99.97%
Blocks Wood	100%	100%	99.9%	99.97%
Bongo	99.90%	99.95%	100%	100%
Bright Tambourine	99.9%	99.95%	99.8%	99.93%
Burma Temple Bells	99.9%	99.95%	99.8%	99.93%
Cabasa	99.90%	99.95%	99.80%	99.93%
Castanets	100%	100%	99.9%	99.97%
Cello	99.8%	99.9%	99.5%	99.83%
Celloharmonics	100%	100%	99.5%	99.83%
Cencerros	99.9%	99.95%	99.8%	99.93%
Chimes Bamboo	99.9%	99.95%	99.8%	99.93%
Claves Fx	100%	100%	99.9%	99.97%
Conga Closedtone	99.9%	99.95%	98.5%	99.5%
Conga Opentone	99.9%	99.95%	100%	100%
Conga Pop	99.9%	99.95%	98.5%	99.5%
Conga Slide	99.9%	99.95%	98.5%	99.5%
Congatumba Compasa	99.9%	99.95%	98.5%	99.5%
Congatumba Mambo	99.9%	99.95%	98.5%	99.5%
Congatumba Rhumba	99.9%	99.95%	98.5%	99.5%
Contrabassoon	99.8%	99.9%	99.6%	99.87%
Cowbells	99.9%	99.95%	99.8%	99.93%
Crotales	99.9%	99.95%	99.8%	99.93%
Ctrumpet	99.85%	99.93%	99.6%	99.87%
Cuica	99.9%	99.95%	0.00%	50%
Cuica Slide	99.9%	99.95%	0.00%	50%
Cymbal Chinese	99.9%	99.95%	99.8%	99.93%
Cymbal Finger	99.9%	99.95%	99.8%	99.93%
Cymbal Orchestra	99.9%	99.95%	99.8%	99.93%
Cymbal Turkish	99.9%	99.95%	99.8%	99.93%
Doublebass	99.8%	99.9%	99.5%	99.83%
Drum Brake	99.9%	99.95%	0.00%	50%

Table 1. Conditional accuracy of some individual instruments - classifier represented by Schema 1 (B1 - multiplication, B2 - average) and classifier represented by Schema 2 (A1 - multiplication, A2 - average)

At the root of the hierarchical classifier we have Sachs/Hornbostel classification, level 1. At its second level we have Sachs/Hornbostel classification, level 2. Instrument names constitute its third level, and the last level which might be considered, in our future testing, is instrument+ariculation.

Schema 2 (Articulation) refers to articulation (blown, bowed, bowed-vibrato,concussive, hammered, lip-vibrated, martele, muted, muted-vibrato, percussive, picked, pizzicato, rubbed, scraped, shaken; see Figure 3 - it represents only a part of the hierarchical tree).

We used a database of 10512 music recording sound objects which contains 2628 distinct musical samples of 102 instruments from McGill University Master Samples (MUMS) CD Collection. MUMS objects have been widely used for research on musical instrument recognition all over the world. We implemented and tested two classifiers for two different hierarchical classification schemes upon the same dataset. In experiment for *Schema 1*, there were 630 objects from the idiophone instrument family, 242 objects from the membranophone instrument family, 744 objects from the chordophone instrument family, and 1012 objects from the aerophone instrument family. In experiment for *Schema 2*, there were 15 different articulation methods (attribute values): blown, bowed, bowed vibrato, concussive, hammered, lip-vibrated, martele, muted, muted vibrato, percussive, picked, pizzicato, rubbed, scraped and shaken. Both classifiers were 10-fold cross-validated. We used *J48 Tree* of WEKA to build them. Multiplication $Q(d_{[i_1,...,i_m]})$, for every node $d_{[i_1,...,i_m]}$ at the level m of attribute d, was used to calculate the conditional accuracy of a classifier for the class of instruments associated with $d_{[i_1,...,i_m]}$. We also applied a mean value to compare the performance of the classification for the purpose of flexible query answering, where the best abstract category can be identified even if the estimation fails in its child level. For simplicity reason, we use abbreviation $A1$ for *Multiplication (Schema 2)*, $A2$ for *Average (Schema 2)*, $B1$ for *Multiplication (Schema 1)*, and $B2$ for *Average (Schema 1)*.

Table 1 shows the accuracy of classification of instruments by our two classifiers. For some of the instruments, the performance of the classifier associated with *Schema 1* (attributes *B1, B2*) is better than the one associated with *Schema 2* (attributes *A1, A2*), such as Alto Flute, Bass Trombone and b-Flat Clarinet; for some instruments the performance of a classifier associated with *Schema 2* is better, such as violin, Cuica, Cuica slide, Drum Brake and Drum Log; for other instruments, the difference between the performances of the two classifiers was insignificant.

Table 3 shows the classification accuracy, for classes of instruments (second level of the tree representing decision attribute), by the first classifier. Table 4 shows the average

	$A1$	$A2$	$B1$	$B2$
Overall	97.45%	98.71%	99.4%	99.8%

Table 2. Overall conditional accuracy on individual instrument level - classifier represented by Schema 1 (B1 - multiplication, B2 - average) and classifier represented by Schema 2 (A1 - multiplication, A2 - average)

	$multiplication$	$average$
aero-free	99.9%	99.95%
aero-lip-vibrated	99.6%	99.65%
aero-side	99.9%	99.95%
aero-single-reed	99.6%	99.65%
chrd-composite	99.5%	99.7%
chrd-simple	99.6%	99.8%
idio-concussion	99.9%	99.95%
idio-rubbed	99.9%	99.95%
idio-scraped	99.9%	99.95%
idio-shaken	99.9%	99.95%
idio-struck	99.8%	99.85%
mem-conical	98.5%	98.5%
mem-cylindrical	100%	100%

Table 3. Conditional accuracy of instrument classes (level 2) - classifier represented by Schema 1

classification accuracy for both classifiers at all levels of the corresponding decision attribute.

The overall accuracies of our two different schemes are listed in Table 2. The classification represented by *Schema 1* is significantly better than by *Schema 2* in terms of the average accuracies of all the timbres, given the fact that in the *Schema 1* there are three different levels of estimation probabilities for the computation of the total estimation.

By cross checking the two schemes, we observed that the timbre estimation of instruments had higher accuracy than that of instruments from other families by the articulation classification schema. Also, among the musical objects played by different articulations, the sounds played by lip-vibration tended to be less correctly recognized by the family classification schema.

Classifier 1, Family-Level-1	99.83%
Classifier 1, Family-Level-2	99.76%
Classifier 1, Family-Level-3	99.94%
Classifier 2, Articulation-Level-1	99.77%
Classifier 2, Articulation-Level-2	99.73%

Table 4. Average confidence

5. Summary

This paper describes the initial research on hierarchical classification of musical instrument sounds with respect to families of instruments and articulation. The obtained hierarchical classifiers yield very good results.

6. Acknowledgments

This work is supported by the National Science Foundation under grant IIS-0414815, and also by the Research Center at the Polish-Japanese Institute of Information Technology, Warsaw, Poland.

References

[1] Gaasterland, T. (1997) Cooperative answering through controlled query relaxation, in *IEEE Expert*, Vol. 12, No. 5, 48-59

[2] Godfrey, P. (1993) Minimization in cooperative response to failing database queries, in *International Journal of Cooperative Information Systems*, Vol. 6, No. 2, 95-149

[3] E. M. Hornbostel, C. Sachs: Systematik der Musikinstrumente. Ein Versuch. Zeitschrift fuer Ethnologie, **46**, (4-5) (1914) 553–90. Internet: http://www.uni-bamberg.de/ppp/ethnomusikologie/HS-Systematik/HS-Systematik

[4] ISO/IEC JTC1/SC29/WG11: MPEG-7 Overview. (2004) Available at http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

[5] LinguaLinks Library: Musical Instruments subcategories. Internet ((2006): http://www.sil.org/LinguaLinks/Anthropology/ExpnddEthnmsclgyCtgrCltrlMtrls/MusicalInstrumentsSubcategorie.htm

[6] F. Opolko, J. Wapnick: MUMS – McGill University Master Samples (1987) CD's