Personne ne conteste raisonnablement que la théorie musicale doive reposer sur les principes de l'acoustique physique. En privilégiant les rapports consonants que les sons peuvent entretenir entre eux, l'acoustique adopte, à son tour, un formalisme faisant appel à l'étude des fractions simples.

Les scientifiques réduiraient volontiers la musique (consonante) à des rapports de fréquences sonores s'exprimant en termes de fractions simples mais les musiciens ne se soucient de rien de tout cela : ils ne se fient qu'à leur instinct. D'ailleurs, la théorie musicale ils l'ont apprise dans le meilleur des cas à un âge où ils ne pouvaient guère se soucier d'arithmétique et encore moins de physique.

La théorie musicale pourrait, en principe, tirer parti de ce double parrainage de la science, et proposer un exposé reposant sur des notations rationnelles héritées d'une méthodologie qui a fait ses preuves en sciences. On est loin du compte : on continue d'enseigner la musique sur des bases empiriques et les élèves sont priés de l'apprendre sur le tas sans essayer de comprendre les tenants et les aboutissants de ce qu'on leur enseigne.

En exergue de ses leçons de théorie musicale, dispensées en Sorbonne, le musicologue Jacques Chailley n'a pas manqué de placer cette sentence attribuée à Anaximandre, qui n'a rien perdu de son actualité : "".

Cette chronique repose le problème en amont, commençant par discuter les facteurs essentiels qui différencient la vue de l'ouïe donc l'image du son.

Il est plutôt rare qu'on ait l'occasion de voir de ses propres yeux les plus beaux tableaux du monde. On doit souvent se contenter d'une reproduction. Il n'en va pas autrement en musique où la plupart des œuvres connues ne sont jamais jouées en salles ou alors cela se passe à l'autre bout de la terre. Dans ce cas tout le monde un peu curieux se contente d'un enregistrement. Dans tous les cas une reproduction ne vaut pas l'original mais elle vaut mieux que rien.

Un tableau (une image) occupe le plus souvent (s'il est rectangulaire) deux dimensions d'espace, une en abscisse et une en ordonnée (Ajoutez-y une troisième dimension si le relief de la toile importe, cela ne changera rien à la discussion). Le paramètre temps ne joue aucun rôle en tous cas au cours de la durée qui vous intéresse.

Il arrivera inévitablement, si ce temps est très long, que le tableau se dégrade et que des réfections s'avèrent nécessaires qui ne seront pas de la main de l'auteur. Cette déchéance n'affecte pas l'œuvre musicale qui est fixée dans une partition recopiée à l'infini. On peut perdre un tableau (original), par exemple parce qu'on l'a volé, jamais on ne perdra la 9ème de Beethoven.

Notez que si un tableau a été volé c'est parce qu'un marché très lucratif existe pour les œuvres garanties d'origine. Une copie parfaite ne possède pas cette valeur marchande alors que plus souvent qu'on ne l'imagine quantité d'experts, sinon tous, n'y ont vu aucune différence. Ce genre de mésaventure n'existe pas en musique : la 9ème dont il était question n'existe que le temps de son exécution en salle puis elle disparaît pour renaître, différente, lors du concert suivant. Non seulement on ne peut pas la voler mais on ne peut pas davantage la falsifier ou lui inventer une sœur inconnue des experts. En cela la musique n'a pas de prix, comprenez comme vous voulez.

L'œuvre musicale n'occupe aucune dimension spatiale, uniquement une dimension temporelle et cela ne va pas sans poser quelques problèmes aux auditeurs potentiels. Quand une oeuvre d'une certaine ampleur dure une ou deux heures il n'y a pas d'alternative au fait qu'il faut lui consacrer tout ce temps et si l'on n'a pas tout compris, ce qui est d'ailleurs souhaitable (car une œuvre comprise au premier coup d'oreille a peu de chance de résister à des écoutes répétées), il faudra recommencer. Deux cas sont alors possibles pas toujours exaltants : on a détesté au point qu'aucune réparation ne semble possible (n'abandonnez quand même pas trop vite, c'est ce que vous dites à vos enfants qui se sont trop tôt découragés dans leurs études !) et on a le sentiment qu'on a perdu son temps ou on soupçonne qu'on pourrait comprendre davantage en recommençant l'expérience mais tout cela prend du temps et le répertoire effectif (celui réellement jouée en concert) ou potentiel (celui mille fois plus étendu disponible en enregistrements divers) semble hors de portée dans la durée d'une vie qu'on consacre rarement totalement à la musique.

La description à l'identique d'une image passe par une pixellisation en x et en y plus une pixellisation chromatique. Le degré de précision requis dépend du pouvoir de résolution de l'œil. Celui-ci est connu, il vaut une minute d'arc en ce qui concerne la distinction entre deux points contigus et 1/8 000 000 de la plage fréquentielle visible par l'œil. Cela demande d'être précisé.

Le pouvoir de résolution spatial de l'œil est de l'ordre de la minute d'arc, soit 1/(57x60) radians. Pour un observateur situé à la distance, d, cela signifie que deux points situés, en face, à distance de d/(57x60) sont à la limite d'être résolus. Un écran standard haute définition comportant 1920x1080 px dépasse déjà cette exigence.

L'œil humain normal est sensible aux ondes électromagnétiques (la lumière) dont la fréquence est comprise entre 400 1012 et 750 1012 Hz. Cette intervalle est étroit car le rapport entre la fréquence visible la plus élevée (correspondant au violet) et la fréquence la plus basse (correspondant au rouge) est inférieur à 2. Autrement dit, la plage des fréquences visibles ne couvre même pas une octave.

En comparaison, l'oreille normale est sensible aux ondes sonores dont la fréquence est comprise entre 16 et 20 000 Hz. Cet intervalle est proportionnellement beaucoup plus étendu puisqu'il dépasse 10 octaves. Cette différence considérable a des conséquences importantes.

Pixellisation de l'espace pictural.

Pixelliser une image spatialement revient à l'assimiler à une grille qui confond les points qui occupent des positions voisines au-delà de la tolérance requise par le pouvoir de résolution de l'œil. Elle peut se faire à pas constant en x comme en y à condition de garantir la précision dans le cas frontal qui est le plus critique. C'est bien ce que l'on fait lorsqu'on assimile l'écran d'un moniteur à une grille comportant 1920x1080 px.

Pixelliser une image chromatiquement revient à affecter un code RGB des couleurs (Red Green Blue) à chaque point défini à l'étape précédente. Etant données les performance de l'œil dans ce domaine on assure la fidélité en réservant 1 octet, soit 8 bits, pour chaque composante RGB. Cela autorise 16 777 216=224 possibilités

Partition de l'octave.

Ce problème essentiel pour la construction des gammes et des modes a occupé plusieurs siècles, débouchant sur des dizaines de solutions toutes différentes. Leur nombre indique clairement qu'aucune ne donnait satisfaction pour une raison qu'on a rapidement identifiée : il est impossible de concilier la justesse d'une gamme avec sa transposabilité.

Si l'on avait confié le problème de la partition de l'octave à un mathématicien qui n'entendait rien à la musique, il n'aurait pas cherché midi à quatorze heures : il l'aurait fait au plus simple soit à pas constant ou, si l'on raisonne en termes de fréquences sonores, f, en répartissant celles-ci en progression géométrique. Formellement cela donnerait : , où :

l'indice, n =0, ±1, ±2, numérote les notes

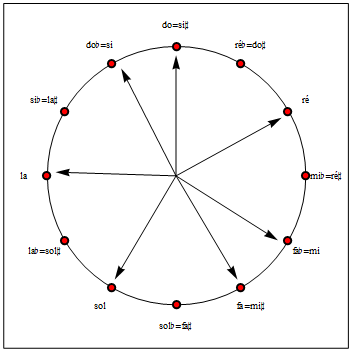
l'entier, N, précise le nombre de notes au sein d'une octave

f0 est la fréquence de la note de référence, appelée diapason

Les valeurs attribuables à N et à f0 ne sont pas de la compétence du mathématicien, il ne dispose d'aucun critère de décision à leur sujet.

Cette formule simple était connue dès le 16ème siècle (les mathématiques de l'époque étaient déjà bien avancées) mais elle a été rejetée au motif qu'elle ne respectait pas les règles fondamentales de la consonance découvertes par les acousticiens.

Pourtant, ironiquement, elle a été adoptée définitivement vers les années 1800 sous l'appellation du "tempérament égal", qui positionne les notes de l'octave à égales distances sur un cercle qui se referme parfaitement. Le choix N=12 (et accessoirement, f0=440 Hz) s'est imposé sans que plus personne aujourd'hui ne proteste mais n'anticipons pas !



D'emblée on note quelques subtilités étranges dans l'appellation des notes : 7 noms distincts seulement (do, ré, mi, fa, sol, la , si, quand on s'attendrait à 12) et plus curieux encore certaines notes dites anharmoniques correspondent à une même fréquence (do#=ré*b*), en particulier une seule et même touche sur un clavier. C'est une des singularités de l'histoire musicale qui s'est construite selon un tout autre schéma que nous allons rappeler et qui n'a plus pu revenir sur des conventions de langage adoptées universellement. Il en est résulté une orthographe musicale aussi impossible à simplifier que l'orthographe française usuelle. Elle en aurait pourtant bien besoin, ne serait-ce que dans la gestion abracadabrante des altérations, mais les musiciens sont d'un avis contraire estimant que c'est très bien comme cela car ces "anomalies" facilitent la lecture des interprètes plus qu'elles ne compliquent la tâche des apprentis musiciens, enfin de la cohorte de ceux qui n'ont pas abandonné leurs études faute d'avoir compris les raisons d'une complexité qui leur échappait.

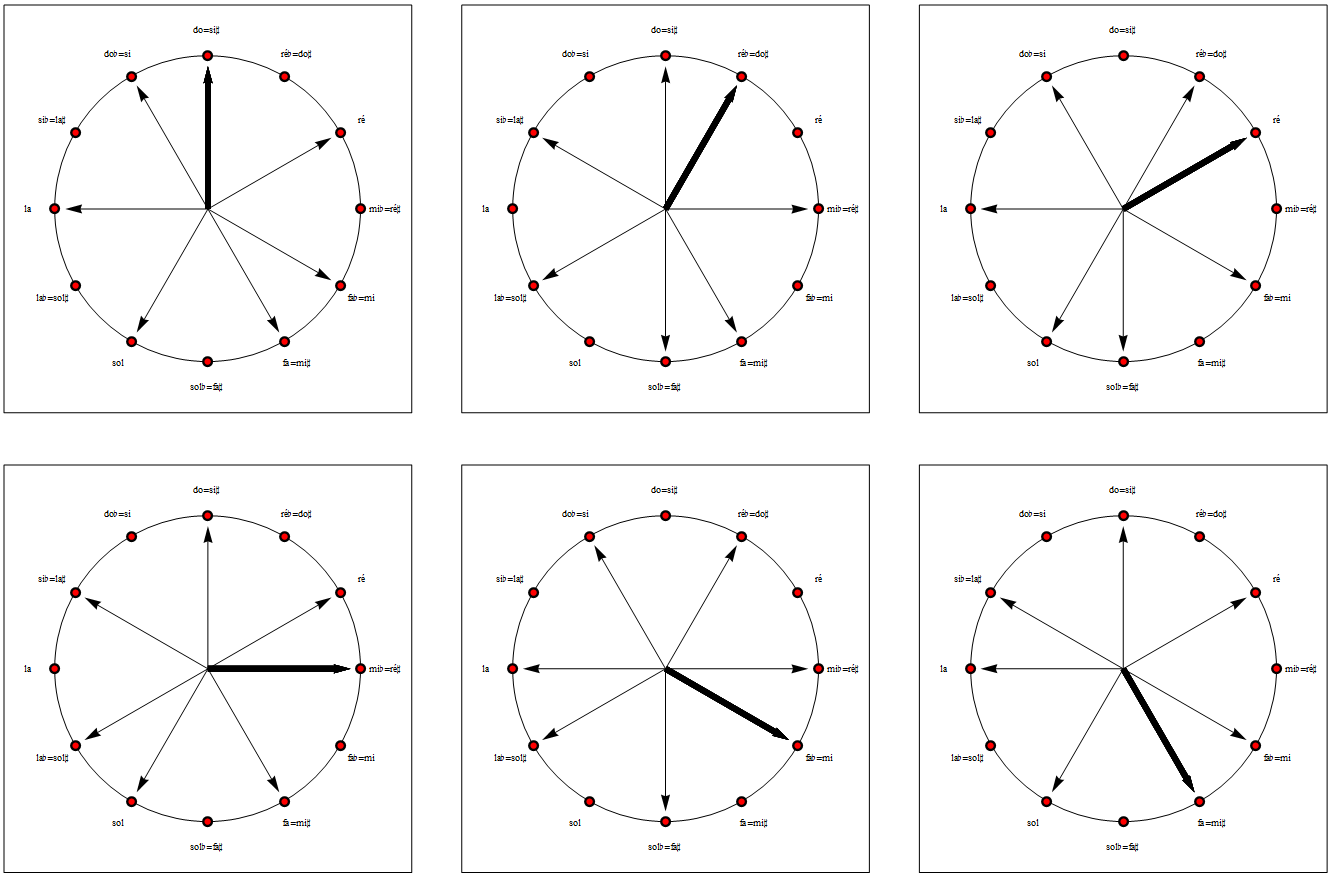
Notion de mode. La musique (de l'époque) classique n'utilise pas les 12 notes disponibles dans un morceau, elle n'en retient qu'un sous-ensemble, pour des raisons évidentes de mélodie et d'harmonie : toutes les notes ne s'accordent pas. Ce n'est pas une obligation et la musique dodécaphonique ne respecte pas ce genre de restriction. Il existe autant de modes possibles que de sous-ensembles de l'ensemble des 12 notes, soit 212=4096 au total, dont la majorité est sans intérêt. En musique classique, les modes les plus connus sont le mode majeur et les modes mineurs. La figure suivante reproduit les 12 modes majeurs basés sur une note de départ, baptisée la tonique (représentée en gras dans chaque diagramme). Il est obtenu en faisant tourner le système des flèches d'un 12ème de tour à chaque graphe. L'ensemble des flèches désigne à chaque fois les notes concernées par le ton majeur retenu, celles que le compositeur utilisera en priorité sauf altération accidentelle.

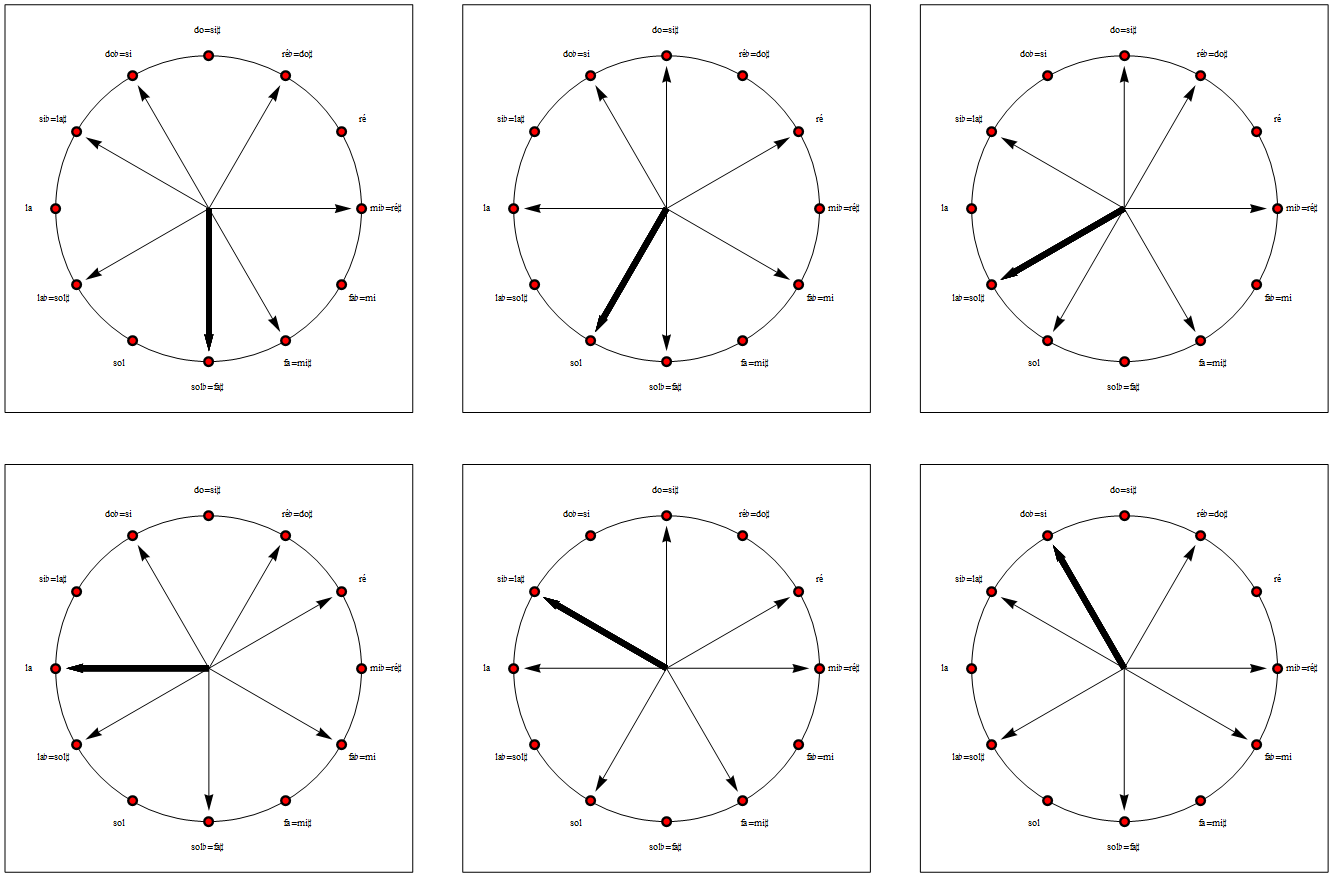
Concentrons-nous sur le mode de mi majeur, c'est la cinquième figure. Outre le mi qui sert de tonique, les notes acceptées par ce mode sont : (mi), fa# ou sol*b*, sol# (ou la*b*), la, si (ou do*b*), do# ou ré*b*, ré# ou mi*b*. Deux règles orthographiques (heureusement compatibles) lèvent les ambiguïtés liées aux enharmonies éventuelles : primo, le même nom de note ne peut figurer deux fois dans l'échelle et secundo on ne mélange pas les dièses et les bémols. Mi étant imposé, ré# doit être préféré à mib, d'où do# s'impose tout comme si, fa# et sol#. Le mode de mi majeur permet de mettre 4 dièses (fa, do, sol, ré) en évidence à la clé :



Les règles orthographiques invoquées peuvent paraître arbitraires (comme toutes les règles conventionnelles) mais elles servent les intérêts des interprètes déjà sérieusement malmenés par l'utilisation d'une portée non linéaire.

Rappelons que la portée est conçue pour ne recevoir au niveau des lignes et des interlignes que l'une ou l'autre des 7 notes de base, peu importe qu'elles soient altérées localement (ou globalement au niveau de l'armure). Cette échelle n'est pas linéaire puisqu'il peut y avoir un nombre variable de demi-tons entre deux positions consécutives. C'est sans doute le moyen de gagner de l'espace pour la représentation des notes en hauteur mais c'est aussi une source de complications dont l'interprète se passerait bien. La mise en évidence à la clé de dièses ou de bémols ne tempère la difficulté qu'à la condition expresse que les règles orthographiques précédentes soient respectées.





Cela dit ce cercle chromatique est d'un usage confortable en particulier au niveau de la transposition qui s'en trouve réglée. Dans la figure ci-avant, les flèches pointent sous des angles qui correspondant aux intervalles du mode majeur, avec une tonique (dessinée en gras) qui pointe provisoirement sur la note do.

Les règles de la consonance.

Pour un physicien, un système est dit harmonique si lorsqu'on le dérange, même localement, de sa position de repos puis qu'on l'abandonne à son sort, il se met à vibrer périodiquement. Si cette vibration se produit dans un milieu lui-même élastique (l'air en l'occurrence) elle se propage immédiatement sous la forme d'une onde longitudinale. Il n'y a aucun déplacement global de matière, simplement les molécules d'air atteintes par l'onde entrent à leur tout en vibration et par collisions successives actionnent les molécules de la tranche d'air suivante. Evidemment si la vibration initiale n'est pas entretenue l'énergie initialement mise en jeu se dilue dans cette onde progressive et celle-ci s'atténue. Si l'on veut que le son atteigne un récepteur distant il faut l'entretenir suffisamment pour qu'il en aille ainsi. Les musiciens ne procèdent pas autrement lorsqu'ils frottent une corde de violon avec leur archet, lorsqu'ils actionnent la colonne d'air d'un tuyau d'orgue ou lorsqu'ils frappent une plaque métallique de xylophone. Bien entendu l'excitation initiale est interrompue dès que l'émetteur estime que le son a assez duré. L'onde sonore est longitudinale (= la vibration s'effectue alternativement d'avant en arrière dans le sens de propagation). Lorsqu'elle frappe les tympans de l'auditeur, leurs membranes sont à leur tour mis en mouvement périodique, toujours à la même fréquence, et une impulsion électrique est transmise au cerveau qui se charge de l'interpréter.

La musique est l'art d'accommoder les sons périodiques. Les bruits non périodiques par nature ne l'intéressent pas. bruit par ses qualités acoustiques que les physiciens ont appris à (re)connaître.

Une gamme est dite (d'inspiration) pythagoricienne si elle ne se préoccupe que d'assurer la justesse des quintes. Tout exposé élémentaire de cette théorie repose sur la suite des quintes ascendantes ramenées à une octave unique mais on peut la compléter par la suite des quintes descendantes ce qui revient à prendre également en compte la justesse des quartes puisqu'une quinte descendante (resp. ascendante) est équivalente à une quarte ascendante (resp. descendante). Comme l'a fait remarquer Jacques Chailley, cet empilement de quintes a quelque chose d'inutilement artificiel. Artificiel car que peut bien signifier le recours à la 12ème quinte supérieure qui (sans la réduction évoquée) se balade 7 octaves plus haut (dans les ultrasons, si l'on part du milieu du clavier !) et inutile car il suffit de remplacer la deuxième quinte ascendante par une quarte descendante pour rester dans la même octave sans réduction particulière. On peut donc construire une gamme pythagoricienne par empilement successif de quintes ascendantes et de quartes descendantes. Sur un plan arithmétique, les deux procédés sont strictement équivalents. portée

On dit que les violonistes - en sont-ils conscients ? - tendent d'instinct à jouer selon cette gamme.

Une gamme pythagoricienne ne se préoccupe pas a priori de la justesse des tierces, elle espère simplement (et elle vérifie a posteriori) que celles-ci ne sonneront pas trop faux. Vérification faite la tierce majeure ?? se positionne un peu trop haut d'où une tension se crée appelant une résolution. Avec l'adoption du tempérament égal, nous avons perdu l'habitude de l'écoute dynamique qui résulte de cette profusion de tensions internes mais jouée à l'ancienne la musique baroque illustre à merveille ce principe. Le meilleur exemple à l'appui de cette thèse se trouve dans la musique de Rameau. On a redécouvert cette musique, au disque, dans les années 1950 et on l'a jouée au tempérament égal sans la moindre précaution historique. Voici ce que cela a pu donner jouée par l'orchestre de Jean-François Paillard (On ne se plaignait pas à l'époque, on était déjà très content d'enfin entendre une note du génie Rameau !). En 20 ans, sous l'impulsion de Nikolaus Harnoncourt, la musicologie a fait un travail énorme en faveur de la musique baroque (Quand fera-t-elle le même travail au service de la musique contemporaine ?). C'est un Rameau décapé qu'on a redécouvert avec ses harmonies d'origine aux micro-dissonances et résolutions millimétrées. Les restaurateurs de tableaux ne procèdent pas autrement qui partent à la recherche des couleurs d'origine (quand elles sont encore présentes …).

Une gamme est dite (d'inspiration) zarlinienne si elle se préoccupe autant de la justesse des tierces que de celles des quintes (donc des quartes). Un tel système sonnera fatalement plus juste, d'où une écoute moins dynamique que la précédente, plus confortable pour ceux qui redoutent les tensions harmoniques mais moins épicées pour ceux qui les aiment.

Il n'y a pas de miracle, ajouter une contrainte supplémentaire (la justesse des tierces) entraîne inévitablement une démultiplication des notes et des complications accrues, dans les faits inextricables, face aux problèmes de transposition. C'est ce qui a finalement causé la perte du système zarlinien qui était pourtant naturellement supérieur "sur le papier". On dit que les chanteurs ont naturellement tendance à chanter selon ce système qui leur permet davantage de justesse. Il est vrai que l'organe vocal est d'une souplesse telle qu'il autorise des choix de notes impossibles pour un trombone.

Il est à peu près établi (pour ne pas dire que c'est une certitude) qu'en écrivant les deux livres du "Clavier bien tempéré" Bach ne poursuivait nullement l'objectif de démontrer la parfaite viabilité du tempérament égal. Celui-ci était connu des théoriciens mais il n'était pas appliqué ce qui aurait demandé qu'on révolutionne la facture des instruments qui n'interviendra que plus tard. D'ailleurs, quel intérêt y aurait-il eu à écrire un prélude en ut dièse, plutôt qu'en ut, quand les deux sonneraient à l'identique à un demi-ton d'écart ? On admet aujourd'hui que Bach voulait montrer que même en tempérament inégal et à condition d'accorder un peu différemment le clavecin, il était possible d'écrire dans chacune des 7 tonalités rejetées à l'époque car jugées impraticables, celles surchargées en dièses (5, 6 et 7) et en bémols (4, 5, 6 et 7). L'origine de la confusion provient de ce qu'il n'existe qu'un seul système à tempérament rigoureusement égal alors qu'il existe une infinité de systèmes à tempérament inégal, exigeant des accordages différents et finalement abandonné sous la pression des facteurs d'instruments qui ne pouvaient remplir leur mission.

Pour situer les valeurs nécessaires à ce compromis, la [psychoacoustique](https://fr.wikipedia.org/wiki/Psychoacoustique" \o "Psychoacoustique) a étudié les limites temporelles et fréquentielles de l'[audition humaine](https://fr.wikipedia.org/wiki/Audition_humaine), qui sont celles du spectre sonore :

* L'évènement sonore le plus court que les personnes puissent isoler dure au moins 10 ms[2](https://fr.wikipedia.org/wiki/Spectre_sonore#cite_note-2).
* Le domaine des fréquences auxquelles l'oreille humaine est plus ou moins sensible s'étend d'environ 16 [Hz](https://fr.wikipedia.org/wiki/Hertz) à environ 16 000 Hz[3](https://fr.wikipedia.org/wiki/Spectre_sonore#cite_note-3).
* Dans le meilleur des cas, l'oreille humaine peut arriver à distinguer des sons qui différent en fréquence d'environ deux [savarts](https://fr.wikipedia.org/wiki/Cent_et_savart), soit une proportion de 1 à 1,002. Cette performance se dégrade progressivement jusqu'à la proportion de 1 à 1,25 aux limites de l'aire audible, du côté des basses, des aigües, et des sons de faible intensité[4](https://fr.wikipedia.org/wiki/Spectre_sonore#cite_note-4).

|  |  |
| --- | --- |
| Unison | 1:1 |
| Octave | 2:1 |
| Fifth | 3:2 |
| Fourth | 4:3 |
| Major Third | 5:4 |
| Minor Third | 6:5 |
| Major Sixth | 5:3 |