Standard Guide for
Use of Spectral Searching by Curve Matching Algorithms with Data Recorded Using Mid-Infrared Spectroscopy[1]

This standard is issued under the fixed designation E2310; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 Spectral searching is the process whereby a spectrum of an unknown material is evaluated against a library (database) of digitally recorded reference spectra. The purpose of this evaluation is classification of the unknown and, where possible, identification of the unknown. Spectral searching is intended as a screening method to assist the analyst and is not an absolute identification technique. Spectral searching is not intended to replace an expert in infrared spectroscopy. Spectral searching should not be used without suitable training.

1.2 The user of this guide should be aware that the results of a spectral search can be affected by the following factors described in Section 5: *(1)* baselines, *(2)* sample purity, *(3)* Absorbance linearity (Beer's Law), *(4)* sample thickness, *(5)* sample technique and preparation, *(6)* physical state of the sample, *(7)* wavenumber range, *(8)* spectral resolution, and *(9)* choice of algorithm.

1.2.1 Many other factors can affect spectral searching results.

1.3 The scope of this guide is to provide a guide for the use of search algorithms for mid-infrared spectroscopy. The methods described herein may be applicable to the use of these algorithms for other types of spectroscopic data, but each type of data search should be assessed separately.

1.4 The Euclidean distance algorithm and the first derivative Euclidean distance algorithm are described and their use discussed. The theory and common assumptions made when using search algorithms are also discussed, along with guidelines for the use and interpretation of the search results.

1.5 The values stated in SI units are to be regarded as standard. No other units of measurement are included in this standard.

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E131 Terminology Relating to Molecular Spectroscopy

E334 Practice for General Techniques of Infrared Microanalysis

E573 Practices for Internal Reflection Spectroscopy

E1252 Practice for General Techniques for Obtaining Infrared Spectra for Qualitative Analysis

E1642 Practice for General Techniques of Gas Chromatography Infrared (GC/IR) Analysis

E2105 Practice for General Techniques of Thermogravimetric Analysis (TGA) Coupled With Infrared Analysis (TGA/IR)

E2106 Practice for General Techniques of Liquid Chromatography-Infrared (LC/IR) and Size Exclusion Chromatography-Infrared (SEC/IR) Analyses

3. Terminology

3.1 *Definitions*—For general definitions of terms and symbols, refer to Terminology E131.

3.1.1 *Euclidean distance algorithm*—the Euclidean distance algorithm measures the Euclidean distance between each library spectrum and the unknown spectrum by treating the spectra as normalized vectors. The closeness of the match, or hit quality index (HQI), is calculated from the square root of the sum of the squares of the difference between the vectors for the unknown spectrum and each library spectrum.

3.1.2 *first derivative Euclidean distance algorithm*—in the first derivative Euclidean distance algorithm the Euclidean distance is also computed, except the derivative of each spectrum is calculated prior to the Euclidean distance calculation.

3.1.3 *hit quality index (HQI)*—a table which ranks the library spectra in the database according to their hit quality values (see 7.5).

3.1.4 *hit quality value*—the spectral search software compares each spectrum in the database to that of the unknown and

---

[1] This guide is under the jurisdiction of ASTM Committee E13 on Molecular Spectroscopy and Separation Science and is the direct responsibility of Subcommittee E13.03 on Infrared and Near Infrared Spectroscopy.

Current edition approved May 1, 2015. Published June 2015. Originally approved in 2004. Last previous edition approved in 2009 as E2310 – 04 (2009). DOI: 10.1520/E2310-04R15.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

assigns a numeric value for each library entry demonstrating how similar the two spectra are.

3.1.4.1 *Discussion*—There are several methods for assigning hit quality values and either a high or low value can be assigned as the best match. Refer to the software manufacturers documentation.

3.1.5 *normalization*—the mathematical technique used to compensate for an intensity difference between two spectra (see 5.1).

3.1.6 *peak searching*—the process whereby the peak table of the spectrum of an unknown material is evaluated against a library of peak tables. Each reference spectrum in the library contains a peak table and the peak table is individually compared to the peak table of the unknown, and assigned a numerical value as to the goodness of fit.

3.1.7 *reference spectrum*—an established spectrum of a known compound or chemical sample.

3.1.7.1 *Discussion*—This spectrum is typically stored in retrievable format so that it may be compared against the sample spectrum of an analyte.

3.1.7.2 *Discussion*—This term has sometimes been used to refer to a background spectrum; such usage is not recommended.

3.1.8 *search algorithm*—the mathematical formula used to make a point-by-point comparison of two spectra.

3.1.9 *spectral library*—a collection of reference spectra stored in a computer readable form, also called a library, database, or spectral database.

3.1.10 *spectral searching*—the process whereby a spectrum of an unknown material is evaluated against a library of digital reference spectra. Each reference spectrum in the library is individually compared to the spectrum of the unknown, and assigned a numerical value as to the goodness of fit. To perform this comparison, each data point in the unknown spectrum is compared to each corresponding point in the reference spectrum.

4. Theory

4.1 *Beer's Law*—One of the basic principles that make spectral searching possible is Beer's Law (see Terminology E131), which states that $A = abc$, where A is the absorbance, a is the absorptivity, b is the sample pathlength, and c is the concentration of the analyte of interest. As long as Beer's Law applies, two spectra of the same material recorded under similar conditions can be made to appear the same by normalization of the data.

NOTE 1—In an ideal case, this is true for transmittance spectra, but there are differences in the spectral peak intensities when reflectance spectra are compared to transmittance spectra.

5. Spectral Data Pre-Treatment

5.1 *Normalization:*

5.1.1 Normalization of spectra compensates for the differences in sample quantity (concentration or pathlength, or both) used to generate the reference spectra in the library and that of the unknown. The spectra are normalized over the complete spectral range of the library. When searching less than the full spectral range of the library, the spectra must be re-normalized over the new range before an accurate comparison can be made. Normalization of a spectrum for library searching is a two step process. First, the minimum absorbance value in the selected spectral range is subtracted from all the absorbances in the same range. The resulting values are then scaled by dividing by the maximum result value in the range. The end result is a spectrum (or a sub-range portion of a spectrum) where the minimum value is zero (0) and the maximum is one (1) absorbance. If the range chosen for normalization has only one or two strong bands and a few medium intensity bands, the range of the spectrum must be reselected or the spectrum will be dominated by the strong bands in the spectrum and the HQI will be insensitive to weaker fingerprint bands necessary for identification of a specific compound. Successful compound identification may require the spectral match exclude the strongest bands, then the normalization will be based on a medium intensity band, and weak fingerprint bands will be emphasized in the HQI.

5.2 *Data Point Matching:*

5.2.1 The algorithms used for searching a spectrum against a library use a calculation that mathematically compares the data points of the spectrum being searched to the data points of the spectra in the library. This requires that the data points in both the sample and library spectra occur at the same frequency. If the data points in the sample and library spectra are not aligned in this manner, then one of the spectra must be mathematically altered (interpolated) to make the data points match. Typically the unknown spectrum being searched is altered to match the data point spacing of the spectra in the library.

5.2.2 Data point matching is commonly accomplished using a linear data point interpolation method. In this method, the slope and offset of a line segment is calculated between the absorbances of every pair of data points in the spectrum. A new set of absorbances is calculated by locating the values that occur on the line segments at positions corresponding to the datapoint frequency of the library spectrum.

6. Conditions or Issues Affecting Results

6.1 Spectral quality is one of the primary conditions or issues that can affect search results. There is no substitute for a carefully recorded spectrum. There are several conditions or issues that affect spectral quality as pertains to spectral searching. These conditions or issues apply to both the spectra used to create the reference database and to the unknown spectrum.

6.2 *Baselines:*

6.2.1 A flat baseline is preferred for the Euclidean distance algorithm as the Euclidean distance algorithm compares each data point in the unknown spectrum to the corresponding data point in the reference spectrum. The effect of an offset or slope in the baseline is interpreted as a difference between the two spectra. Therefore, when a spectrum with a sloping baseline or offset is evaluated using the Euclidean distance algorithm, a simple baseline correction should be used.

NOTE 2—Negative bands can also produce an offset in the baseline as a result of the data normalization process.

6.2.2 The first derivative Euclidean distance algorithm minimizes the effect of an offset or sloping baseline. In this algorithm, the comparison is made between the difference of a pair of adjacent points in the unknown spectrum to the difference between the corresponding pair of adjacent points in the reference spectrum. In effect, this causes the first derivative Euclidean distance algorithm to look only at the differences in the slope of adjacent data points between the two spectra. Fig. 1 shows how the two algorithms view the same two spectra.

NOTE 3—The first derivative algorithm converts a sloping baseline into an offset that is then eliminated by the normalization procedure.

6.3 *Sample Purity:*

6.3.1 The physical state of the sample should be as close as possible to the physical state of the reference materials used to obtain the library. For example, a pure liquid sample would ideally be searched against a library of spectra of only liquid reference materials. A sample which is probably a mixture, such as a commercial formulation, should be compared to a library of commercial formulations.

6.3.2 In some cases the nature of the sample may not be well understood. An unknown sample may be a pure material or a mixture. It may have additional contaminants that will affect its spectrum by adding spurious bands. In addition there are several other sources of spurious spectral features that may appear as either positive or negative bands. Several of these are listed below:[3]

6.3.2.1 Features due to variations in the carbon dioxide or water vapor levels in the optical path,

6.3.2.2 Bands from a mulling agent,

6.3.2.3 Halide salts used as window material and as the diluent for both pellets and diffuse reflection analysis often contain contaminants such as adsorbed water, hydrocarbon, and nitrates. Always use dry halide salts and keep unused halide salts in a desiccator,

6.3.2.4 Water can alter the spectrum of the sample from its dry state. Spectra of inorganic samples with waters of hydration are particularly sensitive to adsorbed water,

6.3.2.5 Solvent bands from samples run in solution, and

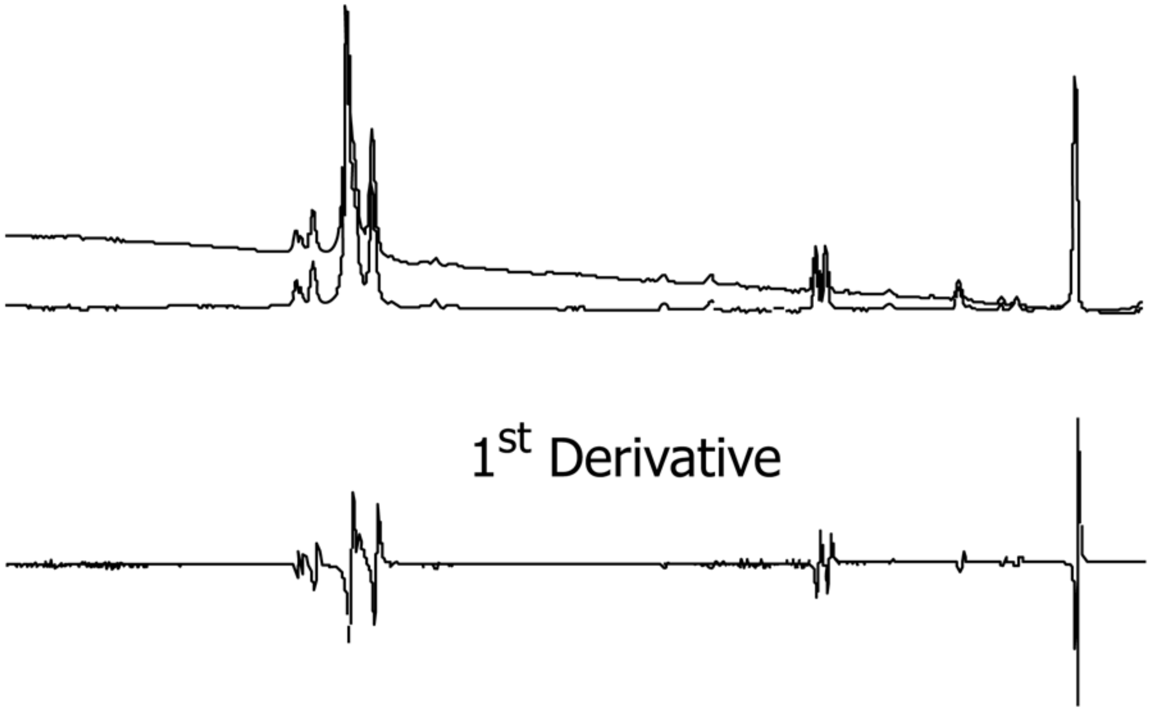
6.3.2.6 Bands from solvents left over from an extraction or from casting a film from a solution.

NOTE 4—Retain spectra of any solvents used, so that bands due to the solvent can be identified in the spectrum of the unknown.

NOTE 5—If the solvent bands in a region of the spectrum cannot be removed from the spectrum (by either re-recording the spectrum, using an uncontaminated sample, or by spectral subtraction using the solvent reference spectrum), then that region of the spectrum should be excluded during a search. It is not sufficient to remove the offending bands digitally by drawing a straight line through the region before the search. The search algorithm will calculate a poor match in this region for any reference spectrum containing features in the region. It should be realized that the removal of the solvent bands may also remove underlying features in the sample spectrum.

6.4 *Absorbance Linearity (Beer's Law):*

6.4.1 A spectrum recorded using good practices (see Practices E334, E1252, E1642, E2105, and E2106) should follow Beer's Law, and so maintain the relative absorbance intensities of its bands, independently of sample thickness. As long as this ratio between the bands is maintained, the spectra can be normalized and a good comparison between spectra can be made. For a spectrum to meet this requirement, each ray of



The bottom two spectra demonstrate the results of the 1st derivative of a spectrum with a sloping baseline as compared to a spectrum with a flat baseline. The two spectra in the bottom trace are almost completely overlapped.

FIG. 1

---

[3] Coleman, Patricia B., *Practical Sample Techniques for Infrared Analysis*, CRC Press, FSBN# 0849342031: 8/26/93.

light of a given frequency must pass through the same amount of sample. There are at least two general cases where this may not happen.

6.4.1.1 One case occurs when there is an uneven thickness of sample in the beam. For example, if the sample is wedge shaped in thickness, or irregular in shape, some rays of light pass through the thin part and some rays pass through the thicker part of the wedge. A similar concern arises when making KBr pellets for analysis. Unless the powder is carefully spread in the pellet die, the pellet can be pressed with a density gradient across the diameter. The sample must also be evenly distributed by thorough mixing of the sample and pellet matrix. This is of particular concern when the beam geometry is smaller than the sample diameter, and is a common problem when using a beam condensing accessory or an infrared transmitting microscope.

6.4.1.2 A second case is when the sample does not completely cover the entire beam cross-section. This occurs with a film that has a void in it, or when a spectrum of a liquid is recorded with an air bubble present in the sample. Some rays of light pass through the sample and some rays pass through the void. The net effect is that fewer molecules are measured than if all the rays passed though the sample resulting in a distortion of the observed relative band intensities.

6.4.2 In the example shown in Fig. 2, when there is no void in the sample and all of the rays pass through an equal thickness of sample, the ratio of intensities between Bands A and B is 2.33. When half of the rays pass through the sample and half of the rays pass though a void with no sample, the ratio of Bands A and B changes to 1.78. See Figs. 2-5.

6.5 *Sample Thickness:*

6.5.1 A spectrum which is acquired with too much sample cannot be properly normalized. When a sample band absorbs all of the available energy at a given frequency, this produces a transmittance value of zero. The resulting absorbance value is infinity, and the normalization step becomes meaningless. In addition, the relative band intensities become highly distorted when normalized to the infinitely absorbing band.

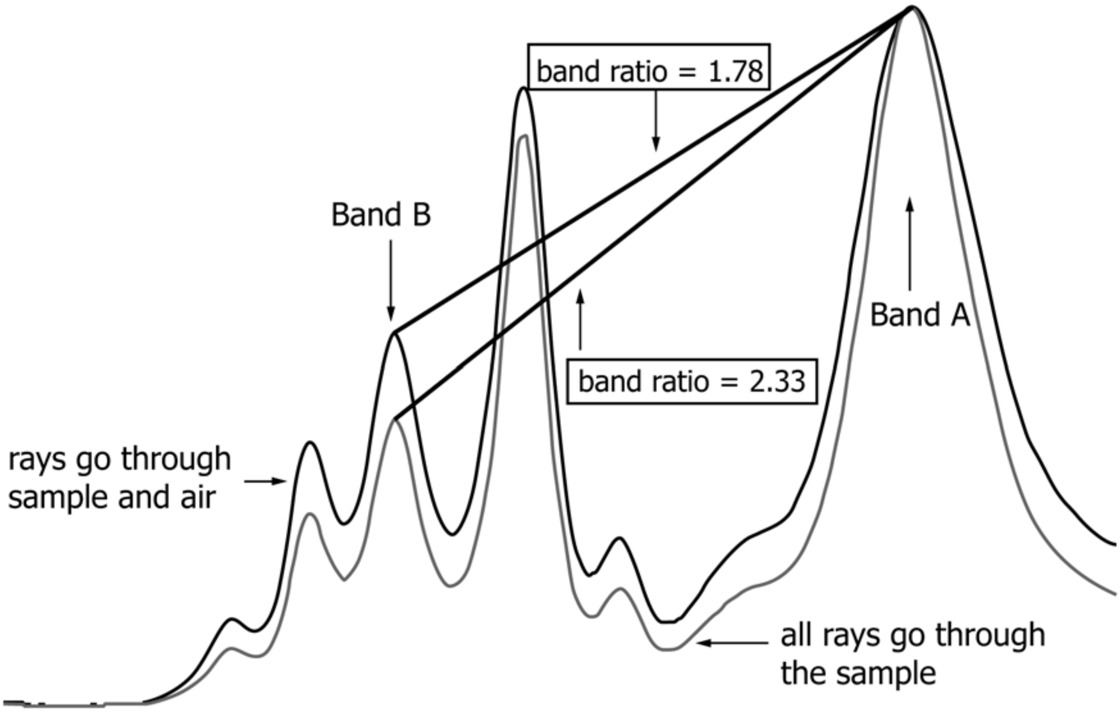
6.5.2 A sample can also be run at a thickness that exceeds the linear range of the detector. Each detector is only linear to some absorbance value. When this is exceeded, the bands at low and high absorbance values will no longer maintain their ratios. As a guide, when practical, ensure that the strongest band of interest has an absorbance of no more than 1.0 absorbance units.

6.5.3 A spectrum that has too great an absorbance can often be detected by the presence of an apparently "flat" region associated with a peak. Often the rest of the spectral bands will still maintain their correct relative intensities. Therefore, eliminating the offending region from a search can produce better results.

6.5.4 Care should be taken not to use regions of the spectrum that are measured at frequencies outside the effective measurement range of the optical components installed in the spectrometer.

6.6 *Sampling Technique and Preparation:*

6.6.1 The analyst should be aware that sample preparation methods can alter the sample or the resulting spectrum. In addition, the measurement technique (optical accessory used)



The ratio between Bands A and B is different for a spectrum in which all of the rays pass through the same sample thickness compared to a spectrum where half the rays pass though air and half the rays pass though the sample.

The two spectra in Fig. 2 have been normalized to give a maximum of 1.0 absorbance at the strongest band, Band A.

FIG. 2

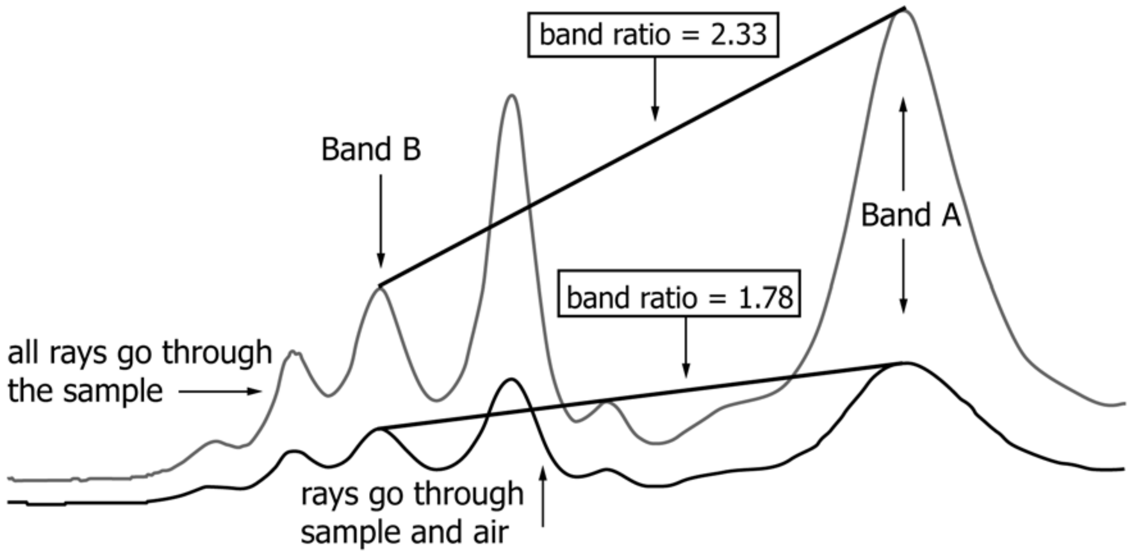
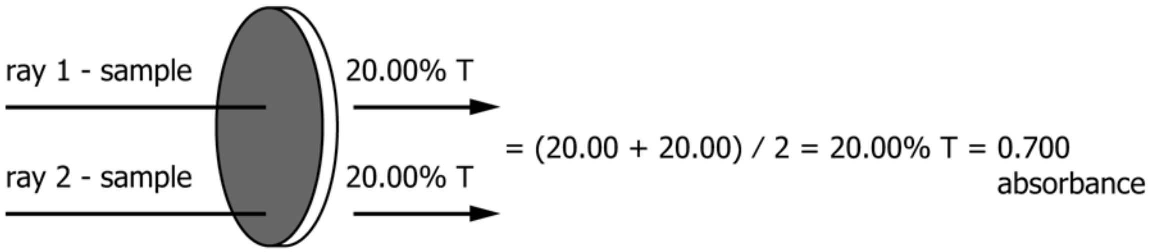


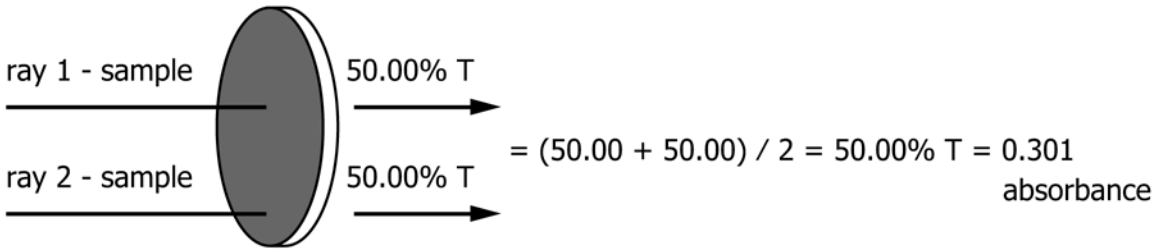
Fig. 3 contains the same spectra as Fig. 2 except the spectra in Fig. 3 are not normalized.

The ratio between Bands A and B is different for a spectrum in which all of the rays pass through the same sample thickness compared to a spectrum where half the rays pass though air and half the rays pass though the sample.

FIG. 3



transmittance for band A



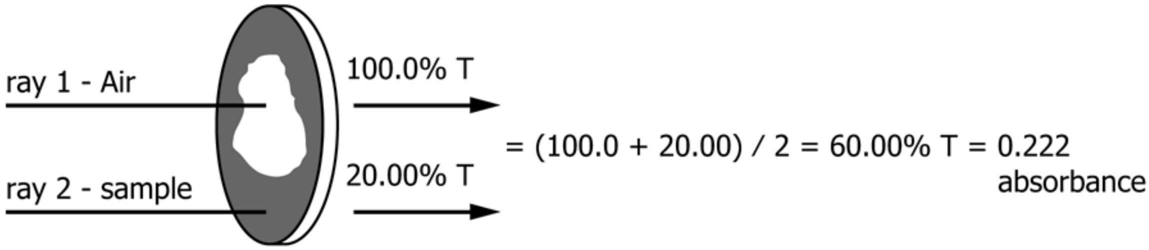
transmittance for band B

For the case where all of the rays pass through an equal thickness of sample:
Band A transmittance through sample = 20.00 %T = 0.700 absorbance.
Band B transmittance through sample = 50.00 %T = 0.301 absorbance.
Relative absorbance of Band A versus Band B = A / B = 2.33.

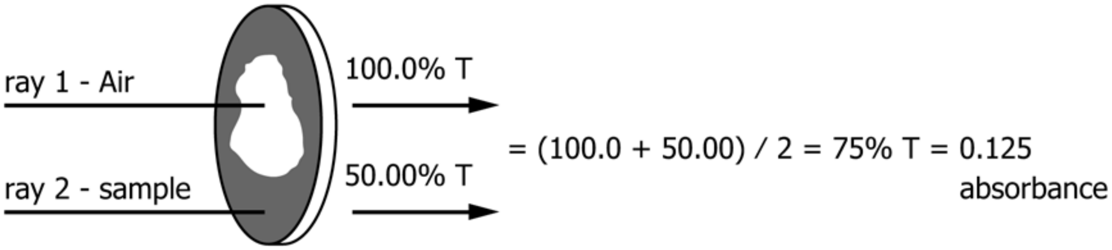
FIG. 4

can have an effect on the spectrum as well. These effects can be very minor and have no impact on the search results, or they may be drastic enough to preclude a good match for a compound that is in the database. See Section 7.

6.6.2 The most severe effects will occur when trying to compare transmittance spectra, such as a KBr pellet, with reflectance spectra, such as internal reflection spectroscopy (IRS) or diffuse reflection. In the reflectance technique, the



transmittance for band A



transmittance for band B

For the case in which half of the rays pass through an idealized void:
Band A and B transmittance through the void = 100.0 %T = 0.000 absorbance.
Avg transmittance of Band A = (20.00 % + 100.0 %)⁄2 = 60.00 %T = 0.222 absorbance.
Avg transmittance of Band B = (50.00 % + 100.0 %)⁄2 = 75.00 %T = 0.125 absorbance.
Avg relative absorbance of Band A versus Band B = A / B = 1.78.

FIG. 5

effective pathlength varies with the frequency of light while for transmittance spectra the effective pathlength is constant. Spectra collected using IRS or ATR are best compared against a database of ATR spectra. If a database of IRS or ATR spectra is not available, the spectra are corrected mathematically to make the relative band intensities appear more like the transmittance spectra.[3] See Practices E573.

6.6.3 The best sampling technique to use for an unknown sample is the technique that was used to run the sample in the database. Unfortunately, with an unknown we cannot first find the technique and then run the sample. The best approach is to know the technique or techniques used to run the spectra in the database and try to use the same technique whenever possible. As this will not always be possible, the user should try to be aware of how the technique used to generate the unknown spectrum differs from the technique used to generate the database and how that can affect the search results.

6.6.4 In the case of transmission data, the variations between spectra acquired as a KBr pellet, cast film, melt, mulls (without mulling agent bands), or neat are generally not as great as to keep a good match out of the top ten or twenty results. Keep in mind, as with any generalities, there are always exceptions. The more common differences that will occur will be differences in the bandwidth. The more crystalline the material is, the sharper the bands will be. A rough approximation of the crystallinity of the techniques is, in order of decreasing crystallinity: annealed solid, crystalline melt (a sample that has been melted and returned to its crystalline form before analysis), mull, KBr, cast film, liquid melt, and neat or solution.

6.6.5 Whenever possible the user should avoid techniques that have an indeterminate pathlength, such as diffuse reflection. These data do not maintain the relative intensities of the bands in a spectrum. This has adverse effects on the search algorithm because inaccurate normalization of the data will result and the user must be prepared to look at many more results to find a good match, or similar compound. The Kubelka-Munk calculation may provide some modification of the relative band intensities, but DRIFTS spectra should always be compared (where possible) to spectra collected using a similar technique.

6.6.6 Similar care should be employed when searching spectra of unknown samples collected using the photoacoustic (PAS) technique as varying interferometer scan speed directly affects the depth of sample penetration and correspondingly the linearity and relative intensity of the absorption bands for the unknown compound as compared to the spectral library. Care must be taken in the interpretation of search results for PAS spectra.

6.7 *Physical State of the Sample:*

6.7.1 Changes in the phase of the sample (solid, liquid, and gas) or crystalline phase of a solid can also have a drastic effect on search results. A low molecular weight compound in the vapor phase presents a very different spectrum from the same compound in the condensed phase.

6.7.2 One important effect that alters with the sample state is hydrogen bonding. A liquid such as methanol shows a very broad absorption for the O-H stretch. In the vapor state there is less hydrogen bonding broadening this band, resulting in a band that is sharper and at a substantially higher wavenumber than the corresponding band in the condensed state. Substituent atoms in a molecule may also influence the extent of hydrogen bonding as will the molecular concentration in a matrix.

6.8 *Wavenumber Range:*

6.8.1 The wavenumber range over which a search is conducted has an effect on the search results. As a general guideline, searching over the entire spectral range of the database is the simplest and often effective method of searching. If the range of the unknown spectrum is greater than the range of the database, then the software should truncate the spectrum to match the database.

6.8.1.1 If the spectral range of the unknown is less than that of the database, then the database must be truncated. Truncating the database will usually result in a somewhat longer search time because the database must be re-normalized with the new range before the search can be performed. This operation should be handled automatically by the search software.

6.8.1.2 While it is possible to add straight-line "data" to the end of an unknown spectrum to extend its range to match that of a database, it is not advisable to do so, even though this may result in a faster search.This provides false information to the search algorithm, by not providing bands that may exist in the region that was extended.

6.8.2 Often one or more regions of the spectrum may be excluded from a search. This technique can be used in place of spectral subtraction to remove bands due to impurities or solvents. It may also be used to force the algorithm to ignore some regions of the spectrum and focus on other regions of the spectrum. This can be an effective search strategy and experience shows that this technique will often improve the results of the search. However, the greater the wavenumber range(s) the user eliminates from the search, the less information the algorithm has to discriminate between results. One technique that often proves useful is to focus on the fingerprint region below 2500 cm^{-1}.

6.9 *Spectral Resolution:*

6.9.1 There are two issues of concern with respect to resolution. One is the resolution of the database and the other is the resolution of the unknown spectrum. In most cases, however, resolution does not play a significant role in the search results.

6.9.2 When the unknown spectrum is recorded at a resolution better than the database, the unknown spectrum can be de-resolved automatically by software to match the data point spacing of the database. The results should be the same as if the unknown spectrum had been recorded at the lower resolution.

6.9.3 If the unknown is acquired at a resolution that is worse than that of the database, the software will generally interpolate the unknown spectrum to match the data point spacing of the spectra in the database. This is much faster than de-resolving all the spectra in the database. It is, however, not a fully valid technique and can cause a degradation of the search results. The greater the difference in resolution between the unknown spectrum and the database spectra, the greater the possible degradation of the results.

6.9.4 In general, the better the resolution of the database and the sample spectrum, the better the search results. Increasing the resolution of the database does, however, produce diminishing returns in terms of time and space requirements. For example, it is meaningless to collect spectra of resolution greater than the average spectral bandwidth of the liquid or solid sample, generally, 4 cm^{-1}. See Practice E1252.

6.10 *Choice of Algorithm:*

6.10.1 There is no single algorithm that works best in all cases. Almost any algorithm can match a quality sample spectrum to that of the correct compound if it is in the database. Problems occur when there are mixtures; poorly run spectra; the unknown is not in the database; or the unknown is not run by the same technique as the spectra in the database.

6.10.2 The Euclidean distance algorithm is usually a good first choice. It will often find compounds similar to the unknown even when the unknown is not in the database. A drawback to this algorithm is that poor baselines can severely limit the ability of this algorithm to find a good match. As a general rule the Euclidean distance algorithm will give more weight to broad bands than to sharp bands. This is because the algorithm is essentially trying to match areas under the curve. This can cause the algorithm to minimize a spectroscopically significant band, like a sharp nitrile band, while emphasizing a broad band like an O-H stretch.

6.10.3 The first derivative Euclidean distance algorithm is much less sensitive to sloping baselines or to an offset in the baseline than the Euclidean distance algorithm. Compared to the Euclidean distance algorithm, this algorithm is more sensitive to sharp bands and less sensitive to broad bands. This also makes the algorithm more sensitive to band positions than the Euclidean distance algorithm. This can cause the algorithm some difficulty in finding a spectrum in a database if there are impurities or solvent in the unknown, or if the sampling technique used has caused some of the bands to shift in the spectrum of the unknown. The first derivative Euclidean distance algorithm is less capable at finding compounds similar to the unknown when the unknown is not present in the database.

7. Guidelines for the Analysis of the Search Results

7.1 The most important guideline to follow in analyzing a search result is to *never* accept the first hit (match) in a search result without comparing the spectrum of the unknown with that spectrum and that of several of the other hits. Anything that causes the unknown to appear different than the reference (baseline, absorbance non-linearity, impurities) can cause the best match to appear lower in the results list. Always look at several hits.

7.2 Look at more than just the first hit. The Euclidean distance algorithm and the first derivative Euclidean distance algorithm will generate a HQI value for each spectrum compared. This HQI is an attempt to rank the spectra in the database according to the algorithms' determination of how well a database spectrum matches the unknown spectrum. Every entry in the database will match to some degree, whether it is a good or a bad match. Just because a spectrum has one of the highest HQI values does not automatically make it a good match. For example, if the reference database has only one spectrum in it, then no matter what unknown is searched against the database the first hit will always be of that spectrum. Always keep in mind that the algorithm is performing pattern matching and knows nothing about spectroscopy or the history of the unknown sample.

7.2.1 The above example of a single entry in the database, is an exaggerated example of what will happen if the unknown is not in the database. If the first hit does look like a good match the user should still look at several more matches to confirm at least the chemical class of the first hit. When using a well represented database the user will likely find several examples within the top ten or twenty hits that represent the same chemical class or functionality of the first hit. This can be used to improve the confidence level of the first hit. Often a better match can be found as the user looks through the other "matched" spectra.

7.3 A common assumption when doing spectral searching is that spectral similarity equates to chemical structural similarity. Although this is often true, it is not an absolute fact. Often a definition of structural similarity can change. Consider the example of a spectrum with a strong and broad band such as an O-H stretch. This would cause the Euclidean distance search algorithm to weight the results very heavily toward compounds containing that band. This could minimize the contributions of the other more discriminating bands in the spectrum. If the primary concern is whether the unknown spectrum contained an alcohol group, then the results from the Euclidean distance search algorithm will be structurally similar. However, if the user were more interested in other regions of the spectrum and the alcohol group was of no interest, then the search results may not be structurally similar.

7.4 It is also possible that several very similar spectra could exist in the hit list. Fig. 6 shows the top four spectra of a search result. All of the top four hits are carbonates. It would be easy to conclude, by looking only at the first spectrum in the hit list, that the unknown was manganese carbonate. However, from examination of the subsequent hits, it can be seen that there are several carbonates in the list. Because the algorithm tends to look at area under the curve, the algorithm may count differences in the baseline or in the amount of water in the sample as more of a difference than any change in the position of the two small bands around 860 cm^{-1} and 700 cm^{-1}, critical for proper identification. Fig. 7 shows a similar effect for organic compounds. Although some differences are still visible in the example shown in Fig. 7, as the number of CH_2 groups increases, it becomes increasingly more difficult to discriminate between similar spectra as a result of the normalization of the spectra to the C-H stretch around 3000 cm^{-1}. Normalization of the spectra to the C-H bending mode at 1450 cm^{-1} would increase the sensitivity of the search algorithm to the fingerprint region.

7.5 *HQI—Reporting and Analyzing Results:*

7.5.1 The actual values given for the hits shown in the HQI are less important than the distance between successive HQI values. Both the actual values and the difference between matches are less important than actual visual comparison of the spectra. The visual comparison should include the comparison of all the peaks in both the unknown and the reference spectra. A good spectral match will include both peak positions and relative intensities of the spectrum of the unknown as compared to the reference spectra.

7.5.2 For the Euclidean distance algorithm using a scale of 1000 as the best and 0 as the worst, a HQI value of around 940 or better can be a good indication of a reasonable match. This number will vary with different algorithms. A good match can also be found with a much lower numbers. Even HQI values as low as the 600s or 700s can be obtained when the correct spectrum is matched in the database. Many things can cause a low hit quality value, such as a poor baseline, a high noise

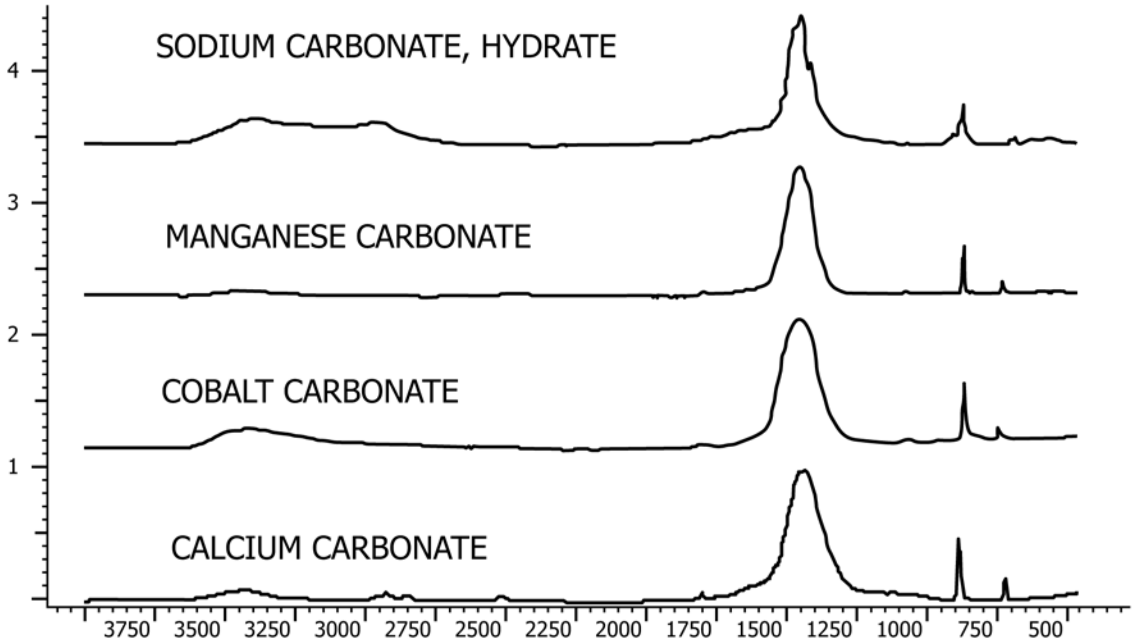


FIG. 6 Four Similar Looking Spectra of Carbonates

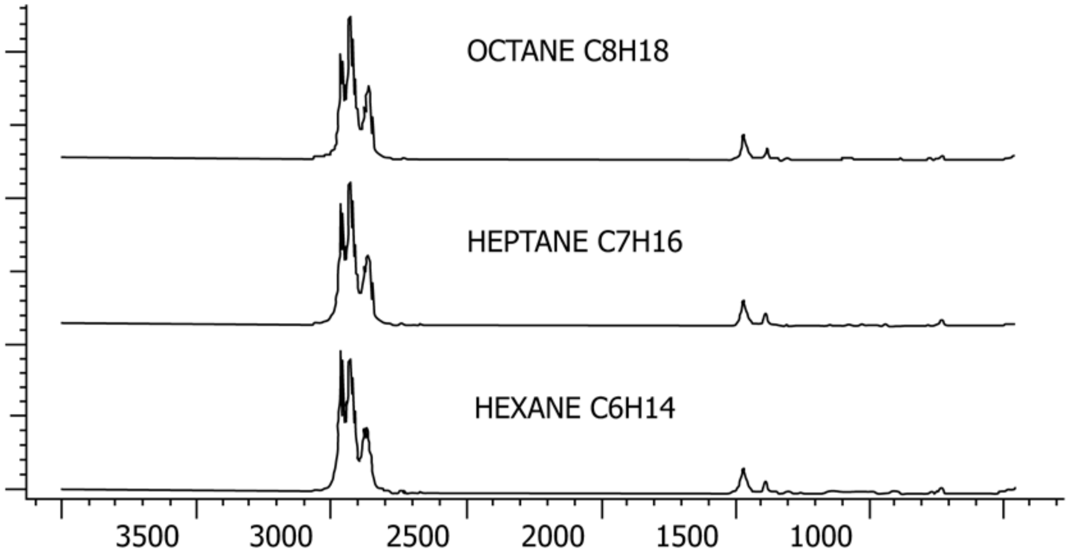


FIG. 7 Three Similar Looking Spectra of Alkanes

level, or the presence of impurities. These situations contribute additively to the lowering of the hit quality value. Hence, the actual value of the HQI ranking is not critical and should only be used as a guide.

7.5.3 The break in (gap between) HQI numbers can often be used as an indication of how well a spectrum may match. For the Euclidean distance algorithm using a scale of 1000 as the best and 0 as the worst, a break in the HQI of about 30 to 60 points is a good indication of a reasonable match. This break will not always occur, even for a good match. This break is dependent on the number and types of materials in the reference database. The break in the HQI can be used as a guide but cannot replace visual comparison of the spectra in the hit list to the unknown spectrum.

7.5.4 The numbers given above are for the Euclidean distance algorithm using a scale of 1000 as the best possible match and zero as the worst possible match. Software that uses a different scale for the best and worst matches should still give the same relative HQI described above. Algorithms other than the Euclidean distance algorithm can give greatly differing numbers for the HQI and the break in the HQI for a good match. HQI values for a first derivative Euclidean distance search are generally lower than the Euclidean distance search.

8. Keywords

8.1 infrared; search algorithms; spectral search; spectroscopy

ASTM International takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM International Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, at the address shown below.

This standard is copyrighted by ASTM International, 100 Barr Harbor Drive, PO Box C700, West Conshohocken, PA 19428-2959, United States. Individual reprints (single or multiple copies) of this standard may be obtained by contacting ASTM at the above address or at 610-832-9585 (phone), 610-832-9555 (fax), or service@astm.org (e-mail); or through the ASTM website (www.astm.org). Permission rights to photocopy the standard may also be secured from the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, Tel: (978) 646-2600; http://www.copyright.com/