

# AUDIO IN SIGHTLINE

Sightline Help Documentation



## Table of Contents

Table of Contents .....	2
Understanding the Data.....	3
Audio File Format .....	3
Hours and GB.....	4
Audio Metadata .....	4
Suggested Field Requests .....	4
Phonetic Searching .....	5
Phonetic Search Logic .....	5
Phoneme Primer .....	5
Confidence Level .....	5
Building Strong Search Terms .....	6
Sightline Audio Search Logic .....	7
Term Operator: ANY vs ALL.....	7
Wildcard Searching .....	7
Proximity Searching .....	8
Language Pack/Dialect.....	8
Numbers and Acronyms .....	8
Combining Audio Searches with Other Searches .....	8
Loading Audio into Sightline .....	9
Trimming .....	9
Waveform.....	10
Language Packs .....	10
Metadata Fields.....	10

---

## Understanding the Data

Audio is an increasingly common, non-traditional, file type in today's eDiscovery landscape. These audio recordings can take many forms, including: trading desk recordings, mandated stored audio archives, web conferencing/video conferencing archives, recorded webinar archives or recorded training materials, and VOIP phone systems with recorded voicemails embed into emails. This is, partially, the product of increased multimedia business communications and more consumer recording devices. As a result, this data is often targeted by regulators and is generally more compelling as you can hear bad actors use their own spoken words.

There are unique challenges with audio files in eDiscovery. Human speech is highly imperfect with garbled noises, low talkers, misspoken words, coughing, static, background noise, competing talkers, and heavy accents to name a few. There is also open mic "hoot" recordings, or speakerbox audio, that records "open air" and trader calls that are often filled with numbers, acronyms, shorthand and slang verbiage. To compound these challenges, the recording systems that store this data are also imperfect with proprietary formats, compression and limited or inconsistent metadata.

### Audio File Format

The most common, and easiest, format to handle in eDiscovery is the WAV file. The WAV can be easily loaded and played by nearly every audio player and, because it is of high quality with low compression, it can be converted to MP3 with little to no loss of quality. The downside to WAV format, is that because it is not compressed, the file sizes are often large, resulting in low transfer speeds and the necessity for massive storage requirements.

Most phone storage systems should have the ability to export WAV formats, but due to the size limitations mentioned above, corporate entities may opt for other, more compressed, formats such as NMF, AUD, OGG, WMV or AU. While we have developed software to convert many of the file types used in the storage of audio files in corporate systems, the number of variations in file type and compression rates, prohibits us from being able to convert all audio data.

For this reason, it is important to ask in what file type the audio is stored. It may also be useful to find out the compression rate, but this is not usually immediately known. If the file type is something other than WAV or MP3 it would be beneficial to get a sample ahead of the full delivery. By examining the sample, our teams can quickly determine if we can convert and process the audio, and, if we cannot, work with our Development team to add the type and or compression to our existing suite of tools.

---

## Hours and GB

While the standard processing and hosting of audio files is done based on GB size – just like non-audio – the pricing to use Sightline is the same as if the data were put into relativity. Where the pricing differs for audio is in the additional options of Audio Trimming (the removal of dead space in open speaker calls) and Phonetic Indexing (building a language specific index of sounds to be searched) as these options are based on the number of hours of audio. Unlike converting GB to page counts, there is no reliable method for converting GB to audio hours. This is due to the nature of audio and the various file types and compression rates. 100 audio files can be an hour and a half or 1,200 hours depending on the file type and compression rates used for the file type. The best option, as with all areas of audio, is to review a sample of the audio.

## Audio Metadata

Unlike email and efiles, audio files require a separate export of metadata information. The data associated with phone calls, e.g. Start Time, Stop Time, CallerID, Called Number, etc., is not embedded in the audio file and subsequently extracted with processing software such as Nuix. While certain fields can be extracted, e.g. Duration, FileSize, FileExt, they do not contain the information that is most useful in culling audio files via the metadata. Because of the unique nature of audio files, the processing time zone is unimportant as the relevant date fields can only be provided in a separate export and are not dependent on the processing engine's settings.

Corporate phone audio storage systems will export these fields in a CSV, LOG or other spreadsheet type file. It is important to discuss what fields are going to be exported to better understand what type of information will be available in the system. Here, again, a sample would be beneficial to all parties involved to set expectations and understand pitfalls early in the case.

If it is expected that there will be multiple deliveries of audio data, it is also important to work with the local staff exporting the data to ensure consistent standards in delivery including format, fields, field order and the proper population of this information. With audio review, if the data is not provided, we cannot ensure our ability to properly implement previously devised workflows.

### Suggested Field Requests

At a minimum, an export of metadata should include the Start and Stop times of the call and the name of the file. It is possible to calculate the stop time using the extracted duration, however, this is a calculation done outside of the system and may not be 100% accurate.

In addition to Start and Stop times and the filename, it may also be useful to have the following fields; CallerID, Called Number, Call Direction (incoming, outgoing), and Call Type (speaker, handset, etc.). These fields may have different names depending on the system, but they should exist in some format.

While these fields are not necessary for phonetic searching, they can be useful in data reduction and understanding the context of the calls.

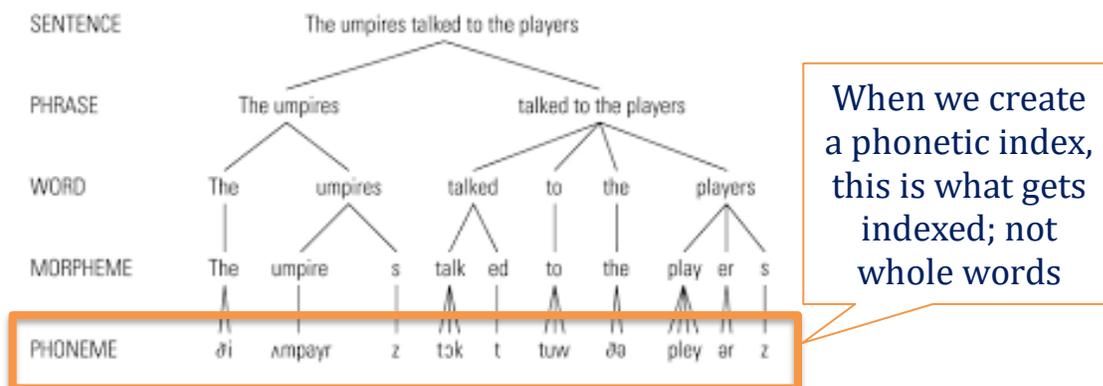
## Phonetic Searching

In addition to the above samples, it is also advantageous to receive the proposed search terms ahead of time. There are two reasons for this; one, most times the search terms start out as the same terms used to search text and will need to be formatted and adjusted for phonetic searching and two, the search terms do take some additional time to set up in Sightline.

## Phonetic Search Logic

Sightline uses Nexidia to index the audio files for searching. Nexidia is unique in that it builds an index of the phonemes (the smallest part of speech) to assemble an index of sounds, rather than words.

### Phoneme Primer



Words are combinations of phoneme “building blocks.” When we create a phonetic indexing, there is no “guessing” of words that match the sounds in the audio file like there is for speech-to-text.

### Confidence Level

Since Nexidia builds its index from phonemes, search hits are returned based on the closeness of the phonemes in the terms being searched to the phonemes in the index. This results in a sliding scale of hits using a Confidence Level. The Confidence Level is a mathematical percentage of the similarity of the searched phonemes to the returned phonemes.

If, for example, you were to search for the term, “hold the door” the system might return audio files that contain the exact phrase, “hold the door” with a high Confidence Level but might also return audio files that contain the phrase, “hose the floor” with a lower Confidence Level. These two phrases sound alike, so both could be returned, but the exact match will most likely have a higher percentage match.

The default Confidence Level is 70% in Sightline. This percentage provides the higher accuracy hits, while also including some lower scored hits to pull in potential matches.

### **Building Strong Search Terms**

When crafting your audio searches there are some general best practices that Consilio has developed. First and foremost, it is advised to not use monosyllabic search terms. Since the index is based on sounds, searching for a monosyllabic term will return any word or phrase that has that sound in it. Search phrases are far superior in finding relevant hits.

Determine the native language or location of the Custodians being searched to account for any accents or foreign language. Nexidia uses Language Packs to build the indices. If, for instance, the caller is based in the UK, we would load the International English Language Pack. If the user speaks Spanish, we would load the Spanish Language Pack. You can load multiple Language Packs to a database, but each will build its own separate index that will need to be searched and billed separately.

If there are any proper nouns in the search term list, find the common pronunciations of each term. Company names like, “Consilio” will need to be searched phonetically to ensure the correct pronunciation is captured, e.g. “con-sill-ee-oh”. This is the same for names like, “Smyth” which can be pronounced, “Smith” or “Sm-eye-th.”

When crafting phonetic searches, it is important to always search the exact term in addition to any phonetic variations, e.g. Libor should be searched as both “Libor” and “Lie-bore.” You should put a hyphen between the sounds of a single word. The hyphen joins the sounds together for the search. If you do not use the hyphen, the system will treat the space as a pause, which may affect the results.

It is possible that business specific jargon or acronyms may need to be searched. It is important to find out the correct pronunciation of the jargon term or terms being searched so you may craft the search accordingly. Additionally, if there are any acronyms in your searches you will need to find out if they are said as a word (like NASA or OSHA) or spoken as letters (like FBI or CIA). If they are said as a word, they should be phonetically spelled out (e.g. “nah-suh”). If they are spoken as letters they will need to be hyphenated (e.g. “f-b-l”) or spelled phonetically (e.g. “eff-bee-eye”).

Just like any text-based searching, you should first look for any obvious flood terms. For audio this could be any monosyllabic terms as mentioned above or words that are spoken regularly on the calls. If the name of the company was Dewey, Cheatem and Howe, you would not want to search for the term Cheatem, as any time someone answered the phone, they would most likely say the company name.

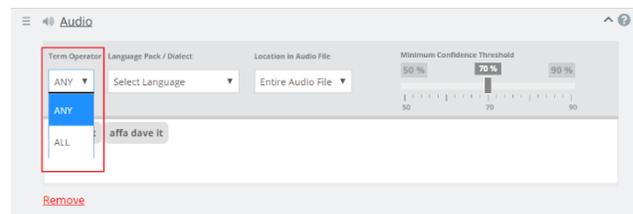
Lastly, it is useful to know the environment in which the calls were recorded. If the calls were recorded on a trading floor or in a board room, there could be multiple talkers conversing over one another. There could also be a lot of background noise to anticipate. In these instances, you may want to consider lowering the Confidence Level.

## Sightline Audio Search Logic

The audio search in Sightline is user friendly and straight forward so long as you understand how the index is built. You can search for a word by simply typing in that word, or the phonetic spelling of the word, into the search window; e.g. you can type “affidavit” or “affa-dave-it” into the search. Always use a hyphen for phonetic spelling of a single word. If you have more than one term, or a phrase, they can be entered as separate words with a space between them.

### Term Operator: ANY vs ALL

In addition to typing out your terms into the search window, you are also given the option for run the search as an ANY search or an ALL search. ANY will return any audio file in with one or more of the listed terms is found. ALL will return any audio file in which every term listed is present in the file.

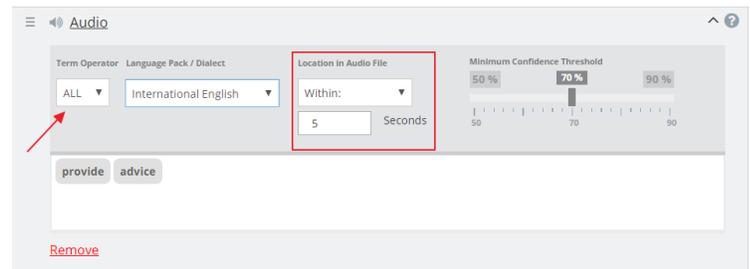


### Wildcard Searching

Since Nexidia builds its index on phonetic building blocks there is no wildcard logic that needs to be used. Running a search for, “tech” will return any audio file that contains that sound, including technology, technologist and nanotech. Think of searching in audio as running all your text-based search terms as bracketed wildcard searches; “tech” in audio searching is similar to searching “\*tech\*” in text-based searching.

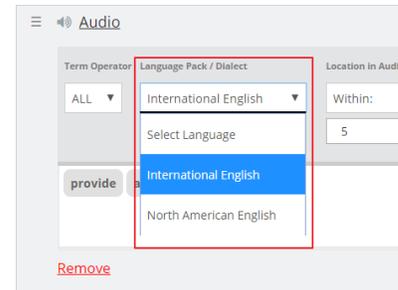
## Proximity Searching

Sightline allows for basic proximity searches to be run. When you add more than one term to the search box as an “ALL” search, you can use the “Location in Audio” dropdown to choose the number of seconds that can exist between the terms.



## Language Pack/Dialect

As discussed previously, Nexidia builds its phonetic index using Language Packs. Multiple Language Packs can be used in the same database; however, they must be run separately. If multiple Language Packs are used, you will need to select the appropriate Language Pack from the dropdown.



## Numbers and Acronyms

When searching for numbers, they will need to be spelled out phonetically; e.g. “100” should be searched as “one hundred.” Please also take note of how people generally speak numbers. Most times people do not say, “zero” and instead say “oh”; e.g. “90210” might be spoken, “nine oh two one oh” rather than “nine zero two one zero.”

Acronyms should also be spelled out phonetically. If you are trying to find someone talking about the FBI, you would need to search for, “f-b-i” or “eff-bee-eye.” As previously discussed, if the acronym is also a word, like NASA or OSHA, you should spell out the word, e.g. “nah-suh” and “oh-sha.”

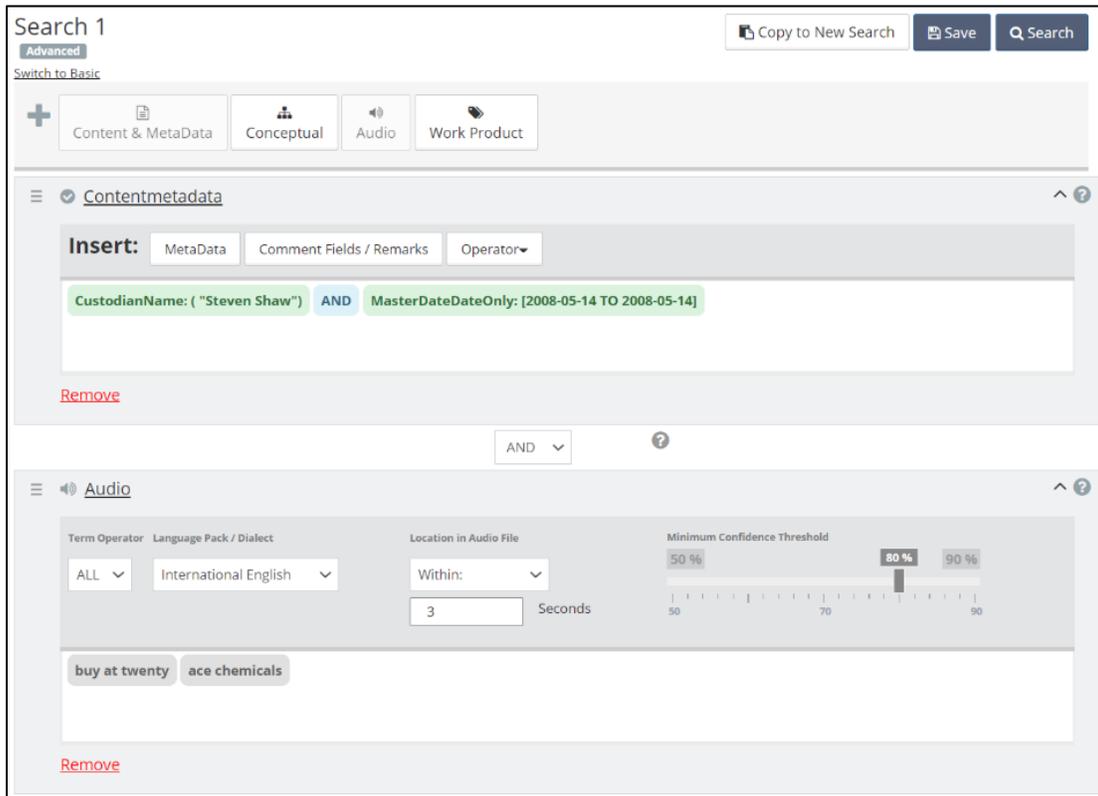
## Combining Audio Searches with Other Searches

It is possible to combine Audio Searches with both Content & Metadata Searches and Work Product Searches. Conceptual Searching is not allowed with Audio Searching. To combine searches, you simply need to click the search you wish to add to your Audio search from the Advanced Search window.



***Please note, that the Audio Search must be last when combining search logic.***

In the below example, the User is searching for any audio files from Custodian Steven Shaw that were made on May 14, 2008 and contain the phrases “Buy at Twenty” and “Ace Chemicals” spoken within 3 seconds of each other at no less than 80% Confidence Level. The Audio search portion must be last when combining searches for the search hits to be properly highlighted in the audio files.



The screenshot shows the 'Search 1' interface with the following configuration:

- Search Type:** Advanced (with 'Switch to Basic' option)
- Search Categories:** Content & Metadata, Conceptual, Audio, Work Product
- Content Metadata Section:**
  - Insert:** Metadata, Comment Fields / Remarks, Operator
  - Criteria:** CustodianName: ("Steven Shaw") AND MasterDateDateOnly: [2008-05-14 TO 2008-05-14]
  - Action:** Remove
- Audio Section:**
  - Term Operator:** ALL
  - Language Pack / Dialect:** International English
  - Location in Audio File:** Within: 3 Seconds
  - Minimum Confidence Threshold:** 80% (slider between 50% and 90%)
  - Search Terms:** buy at twenty, ace chemicals
  - Action:** Remove

## Loading Audio into Sightline

Due to the nature of audio data, there are several steps that must be taken when loading audio data. These steps are outlined below.

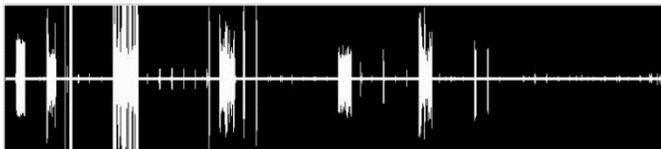
### Trimming

Consilio has developed technology to trim quiet space out of the audio files. This works by analyzing the files for decibel levels. At a designated decibel level, for an indicated amount of time, the silence can be trimmed from the file leaving only the speaking. This will save time and space as the files will be trimmed in size and in length.

*Note: if the time of day is important to the review, trimming cannot be used as you would be unable to determine the time of day from the trimmed file.*

## Waveform

The waveform you see in the system is generated and loaded separately from the audio file itself. This is part of the standard workflow of loading audio files.



If the data will not be trimmed of silence, the waveform can be used to visualize periods of silence and use the slider bar to skip to those areas.

## Language Packs

As previously discussed, Nexidia uses Language Packs to build the indices. You must decide when the data is loaded as to which Language Pack, or Packs, will be used. You may select multiple Language Packs to be used, however there would be a separate charge for each.

## Metadata Fields

The metadata for audio comes in two forms; the data extracted from the audio file itself by the processing software, and the additional fields provided in an export from the phone logging system. To ensure proper loading of the data, it is useful to always view the metadata export prior to loading to check the headers and field order.

There are three default fields in Sightline for Audio files, they are: AudioOriginalDuration, AudioPlayerReady and AudioTrimmedDuration. The first and last fields relate to the trimming of audio discussed above. The second field controls the audio player in the DocView of Sightline. As a best practice, the fields provided in the export for overlay should retain their original name but be prefaced with “Audio” like the existing fields; e.g. “CallerID” would be loaded as “AudioCallerID”.