



---

**“The Reference Standard For Human Voice Capture™”**

# Table of Contents

1.0	About Voice Factory International (VFI)	3
2.0	Voice Talent Services	3
2.1	Talent Management	3
2.2	Talent Selection	4
2.3	Contract Negotiations	4
2.4	Talent Direction	5
2.5	Delivery Time	5
3.0	Voice Database Set Services	5
3.1	Voice Database Set Capture	6
3.2	Database Manual Development	6
3.2.1	Domain-Specific Manuals	6
3.2.2	General Language Manuals	7
3.2.3	Fast Service	7
3.2.4	Voice Preservation	7
4.0	General Recording Services	8
4.1	Pre-Recorded Voice Prompts (PRP)	8
4.2	Interactive Voice Response (IVR)	8
4.3	Automated Speech Recognition (ASR)	8
4.4	Other Applications	8
5.0	Large Scale Scientific Data Collection	10
6.0	Consulting Services	11
7.0	Rates	11
8.0	Facilities	10
8.1	The Anechoic Chamber - Background	10
8.2	Specifications	10
8.3	Design Salients	11

9.0 General Recording Facilities for Audio and Video Production and Large-Scale Data Collection	12
9.1 Equipment/Specifications	12
9.2 Microphones	12
9.3 Microphone Pre-Amplifiers	13
9.4 Audio/Video Recording/Mastering Software and Hardware	13

## **1.0 About Voice Factory International (VFI)**

Voice Factory International (VFI) provides the highest quality capture of the spoken human voice available worldwide for speech industry applications. VFI's [ANECHOIC CHAMBER](#) is unique – no comparable scientific voice capture facility exists for commercial applications. VFI's state-of-the-art facility was designed to exceed the specifications required to create the voice database sets (voice sets) utilized for text-to-speech synthesis (TTS) applications. VFI has turned the art of recording into a science, and has set the reference standard for spoken voice capture.

VFI also offers [CUSTOM DESIGNED SPOKEN-VOICE RECORDING FACILITIES](#) to provide clients with all of their speech recording needs. Whether you need custom-crafted voice sets, individualized pre-recorded prompts (PRP), spoken voice recording for testing Automated Speech Recognition (ASR) products, speech capture for Interactive Voice Response (IVR), or any other spoken voice recording product, VFI delivers custom solutions at off-the-shelf prices more quickly than anyone in the industry.

## **2.0 Voice Talent Services**

### **2.1 Talent Management**

At the heart of every spoken voice product is a distinctive human voice. But, to speech industry clients, the quality of the speaker's voice is only the beginning. Voice talent must not only speak well, but must receive custom training for TTS voice-capture, and PRP and IVR voice recording. Importantly, the same talent must be consistently available to supplement a client's needs even after initial market implementation. VFI's comprehensive talent management services ensure that clients can concentrate on their core business, resting easy in the knowledge that their corporate image is accurately and favorably represented through their speech-enabled products and services.

On occasion long-standing talent used by a client becomes unavailable due to circumstances beyond the client's control. In such cases, VFI can quickly find an ideal replacement with voice attributes that are so similar to the former talent's, that clients can incorporate the new voice without any appreciable perceptual difference. This unique service translates into considerable savings and uninterrupted corporate image.

## **2.2 Talent Selection**

VFI's New York City location, America's "melting pot," ensures access to the world's largest and most diverse talent pool. VFI is uniquely qualified to tap into this unparalleled resource.

VFI's proprietary talent selection and retention processes ensure the shortest time to market from audition, submission of samples to the client, voice database set (voice set) capture, through delivery. VFI has selected hundreds of voices, in dozens of languages and dialects, suitable for speech applications. VFI can provide hundreds of samples to help clients determine the voice(s) they would like to deploy in their applications. Following an initial consultation, VFI will work with each client to create **objective criteria** for talent selection. Within three weeks, an initial submission of twenty-five (25) voice samples will be presented. VFI's proprietary processes for talent selection are designed to create a high likelihood that clients will find ideal voices from the first set of samples submitted for any project. In the event more samples are requested, VFI will promptly submit as many as the client requests.

Within three weeks of the initial consultation, clients will be able to choose the voice, or voices they want to represent their enterprise. This process shortens time to market and saves costs for everybody.

Alternatively, VFI will work with pre-existing talent to ensure the highest quality voice capture and voice recording available.

## **2.3 Contract Negotiations**

Depending on the client's objectives, VFI can provide exclusive or non-exclusive rights to voice talent. By policy, VFI contracts directly with talent. In the majority of cases there are no agents to deal with, and union issues are always handled by VFI, thereby ensuring that all talent is delivered on a fee-simple basis. The end result is the most favorable commercial rate for voice talent.

Additionally, because VFI is acutely aware of the need for talent availability after the initial implementation of any voice solution, VFI will custom-tailor individual contracts with voice talent to ensure their timely availability for supplemental work. This guarantees that clients will never be left in the lurch when mission-critical voice application updates are necessary. Individual arrangements are made on a case-by-case basis, with each client's objectives paramount.

In cases where a client has pre-existing talent under contract, VFI can ensure the seamless integration of any supplemental voice capture, ensuring a hassle-free experience for both the client and the client's voice talent.

## **2.4 Talent Direction**

Recording for TTS requires assiduous attention to the technical requirements of voice capture. Equally important is the human element. In practice, different speech synthesis engines work best with voices specifically suited to their design. Following the initial choice of talent candidates, VFI works with TTS engineers for the synthesis platform of choice to ensure that the specified voice(s) will provide application-compatible material.

Subsequently, VFI's experts work with talent every step of the way to enhance vocal performance, with the target synthesis engine specifically in mind. Variables such as meter, consistency, and accuracy of pronunciation are constantly monitored during and after production. If any corrections are needed, they are made prior to delivery. This optimization process produces the best possible end results, and streamlines the entire process.

## **2.5 Delivery Time**

Time is money. Because reducing delivery time reduces opportunity costs and translates to shorter times to market, VFI guarantees the fastest turnaround time in the industry. Clients can expect a **standard set of twenty-five voice samples** for evaluation no later than three weeks following initial consultation by VFI with the client.

## **3.0 VOICE DATABASE SET SERVICES**

VFI possesses the world's only facility specifically designed and operated to provide end-to-end voice capture and consulting services to the speech industry.

VFI has the capacity to produce over 200 voice database sets (voice sets) per year. The production of a voice set calls for the precise, phase-linear, reverberation-free recording of a pre-trained human voice articulating custom-scripted words and syllables to capture a full set of context-sensitive speech units (phonemes, diphones, and triphones) for a given language or dialect. The importance of a completely reverb-free environment, completely isolated from any extraneous sound, is critical for the successful capture of speech for TTS applications. VFI's anechoic chamber provides such an environment.

Following capture of a voice set, a procedure called "tagging" is performed to permit the voice set to be converted to a voice font. The voice font is then operated by a voice synthesizer in real-time. The process is called "concatenative synthesis," distinguishable from the more elementary production of a "pre-recorded prompt" (PRP). VFI also provides world-class recording for the production of PRP, IVR and Automated Speech Recognition (ASR) using the same talent under contract for TTS applications. This assures seamless results and consistency in voicing for combined TTS/PRP/IVR/ASR applications.

By employing VFI's all-inclusive voice capture and recording services, clients can realize substantial cost savings, while being assured that they are receiving the best the industry has to offer in this highly specialized and demanding field.

### **3.1 Voice Database Set Capture**

VFI has unparalleled experience in capturing custom voice database sets (voice sets) for the development of custom voice fonts for TTS synthesizers. VFI has captured thousands of hours of voice samples specifically for TTS applications and has extensive samples in dozens of languages and dialects, adaptable to a myriad of applications. By partnering with VFI to create custom voice sets at off-the-shelf prices, clients can capitalize on VFI's...

- [ANECHOIC CHAMBER](#): the only one of its kind available for large-scale, commercial applications
- Experience working with speech synthesis R & D teams to maximize platform-specific prosody and pronunciation during voice set capture
- Ability to carry out specialized voice capture that meets the most stringent technical specifications, attaining levels of resolution that surpass industry standards

### **3.2 Database Manual Development**

#### **3.2.1 Domain-Specific Manuals**

Until now, "domain-specific," or industry-specific development of manuals for TTS implementation has been an extremely costly, time-consuming proposition. Today, thanks to VFI's unrivalled experience and proprietary methodology for the development of domain-specific TTS manuals, clients can realize unheard-of savings for even the most challenging applications, not only in hard dollars, but also in time-to-market. Among others, VFI has developed domain-specific manuals for the following industries:

- GPS-based navigation
- Medical referencing and prescription instructions
- Financial reports and business briefings
- Legal notices and insurance explanations
- Shipping instructions and delivery guides
- Meteorology reports and forecasts
- Transit arrivals and onboard announcements
- Inventory pull-lists and material control assessments
- Name/place infrastructure for call centers
- Time/date in voice mail applications
- Reservation confirmations/flight progress reports
- Entertainment applications

### **3.2.2 General Language Manuals**

VFI also produces baseline manuals for any language or dialect. As with domain-specific manuals, VFI's unparalleled experience developing foreign language manuals, combined with our proprietary methodology, means that clients receive the best service, at the best price, for custom-designed material.

### **3.2.3 Fast Service**

VFI offers the fastest turnaround times in the industry for voice selection (3 weeks), manual development (3 weeks) and voice set capture (3 weeks). VFI can deliver a completely customized voice set to the client in an average of nine weeks in the following standard languages:

- **English**, any American, British or other variant or dialect
- **Middle Eastern Languages**, such as Arabic, Hebrew, and Urdu
- **Asian Languages**, such as Japanese, Chinese (Mandarin or Cantonese), Korean and Hindi
- **European Languages**, such as French, Spanish, Portuguese, Italian, German, Russian, and Polish

For less common languages and dialects, turnaround time is calculated on a case-by-case basis.

### **3.2.4 Voice Preservation**

From time to time, TTS vendors receive requests for voice capture from the general public. Often, such an inquiry is made in the context of an individual's impending loss of voice function due to illness or death. These people anxiously desire to convert their voices into TTS voice fonts so they can continue to be "heard" as they sound now. This demand has rarely, if ever, been accommodated, in spite of the fact that it will likely grow over time as the cost of converting voice recordings into voice fonts falls dramatically.

In the future, production efficiencies and falling costs will enable anyone to synthesize his or her voice at an accessible price. VFI's production expertise as TTS voice database suppliers allows us to answer this demand today.

VFI will provide studio and production time on a cost plus-basis to preserve voices for eventual creation of voice fonts. We encourage you to ask about this important service.

## **4.0 GENERAL RECORDING SERVICES**

VFI also provides general recording services for systems integrators, for industry and governments, and even for individual consumers. Companies and government agencies partner with VFI to make information available through speech in a variety of circumstances.

### **4.1 Pre-Recorded Voice Prompts (PRP)**

Together with VFI's voice set capture capabilities, and through its state-of-the-art [GENERAL RECORDING FACILITY](#), VFI assures clients of seamless integration of voice applications that employ TTS, IVR, PRP, and ASR. This ensures that clients can secure flexible and consistent voice attributes for every application across the full range of currently available and yet-to-be-introduced platforms.

### **4.2 Interactive Voice Response (IVR)**

VFI understands that Interactive Voice Response (IVR) systems are an integral part of a client's corporate image. Professionally recorded voices by VFI can support your corporate brand and help create and maintain long-lasting customer relationships in a cost-effective, time-sensitive manner.

### **4.3 Automated Speech Recognition (ASR)**

Voice data collection quality is a key factor in maximizing the performance of speech recognition engines. VFI's ability to fulfill voice capture and modeling specifications with great exactitude make it an excellent choice for collecting new voice data in any language. Our extensive library of pre-collected voice data can be used to improve existing Hidden Markov Models (HMMs) and neural net baselines via superior coverage and comprehensive environmental noise modeling.

VFI also offers custom voice data capture, as well as real-world ambient noise capture. These services are priced on a per project basis at highly competitive rates. By availing themselves of VFI's capabilities, ASR clients can obtain invaluable data for performance maximization, realizing cost savings that can be channeled to other business needs.

### **4.4 Other Applications**

Speech enabled applications and documents can make data readily accessible to the visually impaired, ensuring compliance with the requirements of the American with Disabilities Act (ADA).

Currently, most firms utilize existing staff to comply with ADA requirements in an ad-hoc manner. This is an expensive, time-consuming burden resulting in wasted human capital and inadequate compliance.

By partnering with VFI, clients are free to focus on their core competencies and are assured of seamless, integrated compliance programs.

Additionally, speech-enabled data can be invaluable to tens of millions of people with cognitive disorders such as dyslexia, thereby increasing the productivity of often exceptionally gifted performers.

Companies and government agencies can partner with VFI to create voice-enabled:

- Annual Reports
- Instructional Materials
- Operation and Maintenance Manuals
- Safety and Emergency Procedures
- Policy and Procedures Manuals
- Sales Presentations
- Web-enabled Databases

## **5.0 LARGE-SCALE SCIENTIFIC DATA COLLECTION**

VFI is uniquely equipped and qualified to conduct large-scale data collection. VFI's facility and staff offer production services to the scientific community carrying out fundamental research in virtually any recording medium. Clients are able to carry out exploratory work or establish baseline measurements using VFI's facilities, with the contemporaneous capture of high-resolution data in real-time, including...

- Digital audio production
- Digital video production
- Thermal imaging
- Psychometric data collection
- Biometric data collection
- Galvanic skin response
- Respiratory patterns
- Electrocardiogram
- Brain-wave Mapping
- Electroencephalography

For example, VFI recently collaborated with a client formed by former Bell Labs researchers and developers to devise a biometric data collection methodology for a Department of Defense (DOD) study. This study used a combination of thermal imaging, digital video imaging, and digital audio recording to create a viable model for evaluation of a subject's emotional state. VFI delivered results within four weeks of the initial bid.

## 6.0 CONSULTING SERVICES

VFI offers consulting services to the scientific community for any project requiring large-scale data collection. These services are offered on a per-project basis. Please email us at [consulting@vfinyc.com](mailto:consulting@vfinyc.com) and someone will contact you within two business days to discuss your needs.

## 7.0 RATES

VFI's prices are the most competitive in the industry. For standard pricing, please refer to VFI's rate card or call one of our representatives to discuss your requirements and obtain a quote. You may also email us at [sales@vfinyc.com](mailto:sales@vfinyc.com)

## 8.0 FACILITIES

### 8.1 The Anechoic Chamber - Background

There are only a handful of true anechoic chambers designed for capture of the human voice. All of them are in research institutions and therefore unavailable for commercial applications – except one. It is fair to say that the anechoic chamber designed by VFI in New York City ranks among the finest in the world, complying fully with stringent specifications imposed by AT&T Labs for their Natural Voices™ TTS products. VFI's best-in-class anechoic chamber was constructed and brought on line in 2001. VFI staffs the facility with trained specialists in spoken voice recording, editing, engineering/mastering, and voice talent procurement and direction.

### 8.2 Specifications

It cannot be emphasized enough just how demanding are the specifications achieved by VFI's anechoic chamber. The most advanced state-of-the-art chambers at universities and research laboratories worldwide typically cannot achieve flat response below 30 Hz. VFI's facility supports a totally phase-linear recording environment from 2 Hz and up. The noise floor inside VFI's chamber is minus 120 db, well beyond the capture range of even the most sophisticated state-of-the-art recording equipment. To achieve these standards, VFI's anechoic chamber features a floating room contained within **three other enclosures**, the outermost being the building itself.

### 8.3 Design Salients include:

- The foundation of the ANECHOIC CHAMBER is constructed within a reinforced concrete retaining wall containing six tons of sand, supporting a

- sub-floor system designed to absorb and cancel ultra-low frequency vibrations.
- The innermost floor on which the inner chamber is built floats on 40 high-tension steel springs.
  - Throughout the entire structure, no two materials of the same kind come directly in contact. Because each particular material vibrates at a specific rate and transfers vibrations in certain frequencies, VFI's construction is designed to break up vibrations and prevent their transfer.
  - All surfaces are constructed at oblique angles; no right angles exist inside the chamber. This ensures the reverb-free, noiseless environment crucial for TTS applications.
  - Eight feet of high-density, acoustic-grade fiberglass, plus nine inches of acoustic-grade Celetex<sup>®</sup> are housed in between the outer and inner walls of the structure.
  - Two feet of anechoic wedge Sonex<sup>®</sup> are affixed to every wall and ceiling of the chamber.
  - Three four-inch air spaces separate the walls and sound insulating materials. Each wall is faced with three inches of sheetrock for a total 1 foot of sheetrock.
  - Sealing the VFI chamber's joints and contact points required over five hundred pounds of acoustic caulking.
  - To silence heating, ventilation and air conditioning (HVAC), the VFI chamber incorporates a five-ton unit that delivers conditioned air through four hundred feet of ductwork that gradually increases in size and feeds into a primary baffling system (8 ft x 12 ft x 3 ft), which in turn feeds conditioned air through one hundred feet of ductwork, into another baffle, and finally into the chamber. Air returns are similarly constructed. There is absolutely no detectable noise in the chamber, even when the most sensitive microphone is held up directly to the vents.
  - To enter the chamber, one must go through two soundproof metal doors weighing eight hundred pounds each.

Thus, for scientific work as well as for commercial applications, speech clients can be assured that VFI's anechoic chamber and support services are the only ones in the world capable of delivering truly clean recordings of the human voice.

## **9.0 General Recording Facilities for Audio and Video Production and Large-Scale Data Collection**

Voice Factory International's (VFI) state-of-the-art audio recording studios are specifically designed for recording the spoken human voice. For this reason, they are far better suited to recording the spoken voice than are traditional music or mixed-use recording studios. Unlike off-the-shelf sound booths found at multi-use

recording studios, VFI's audio booths are custom-built and engineered to produce a naturally rich ambience that enhances and captures every nuance of the spoken word.

Unlike multi-use studios that route the signal through unnecessary switching circuitry and effects generators, VFI's proprietary equipment configurations utilize the shortest signal paths possible. They are designed for the sole purpose of spoken voice recording. The results are an astounding purity of tone and optimal rendering of ambient space.

VFI's microphones have been specifically selected to ensure the highest-quality recording of the spoken human voice. Our producers and engineers assiduously match a particular microphone to best capture the attributes of each voice. This results in the most realistic, accurate recording of any spoken text.

## **9.1 Equipment / Specifications**

### **9.2 Microphones**

Any great recording has at its core a great microphone. VFI's microphones are hand-picked to deliver the best possible results under any circumstance.

- **B&K 4006** – This one-of-a-kind microphone was custom-made for use in an anechoic environment by the renowned firm of Brüel & Kjær. It is an omni-directional condenser microphone with a flat frequency response from 2 Hz to 30 KHz.
- **Beyerdynamic M88 “Classic” Limited Edition** – Only 999 of these outstanding, individually numbered microphones were made. VFI has six of them. This hypercardioid microphone operates on a pressure gradient and features a dynamic transducer with a frequency response from 30 – 20,000 Hz.
- **Electro-Voice RE 20** – A staple in both broadcast and voice-over studios, the RE 20 is one of the most versatile microphones in the industry. This dynamic, cardioid microphone has a smooth, warm tone, and flat frequency response from 45 – 18,000 Hz.
- **Shure SM7A** – The SM7A, commonly referred to, as the “Johnny Carson Microphone” is a unidirectional, dynamic microphone meticulously designed for professional audio recording. It has exceptionally smooth, flat response from 50 – 20,000 Hz and outstanding tonal qualities.
- **Beyerdynamic MC833** – The MC833 stereo cardioid microphone is equipped with three true condenser systems. As such, it can achieve true MS and XY capability without an external matrix. It achieves a frequency

response of 20 – 20,000 Hz, and boasts a phenomenal signal-to-noise ratio of 70 dB relative to 1 Pascal.

- **Beyerdynamic M160** – The M160 is renowned by audio recording professionals for its warm, natural sound. One of the few ribbon microphones still in production, its dynamic, hypercardioid, double-ribbon design features; among its other attributes, it features a noise attenuation of 25 dB at 110°. The M160 has a frequency response of 40 – 18,000 Hz.

### 9.3 Microphone Pre-Amplifiers

- **Aphex 1100 “Class A” Tube Preamplifiers:** VFI uses only the Aphex 1100 “Class A” tube microphone pre-amplifiers. Beyond achieving mere technical excellence through proprietary circuitry, they deliver a sound that is uniquely involving, compelling, and “real”. They feature a 24-bit output resolution and have selectable output resolutions of 44.1, 48, and 96 kHz.

### 9.4 Audio/Video Recording/Mastering Software and Hardware

- **ProTools / DigiDesign:** VFI uses the industry standard ProTools software and hardware by DigiDesign for digital recording and audio recording mastering.
- **Final Cut Pro:** Final Cut Pro is a leading software solution for professional digital video editing.