

From "output" to "input" of AI voice, what are investors betting tens of millions of dollars on?

白鲸出海

2025-07-30 06:02

Can the once-taken-for-granted voice input function enjoy a second heyday?

On July 16th, voice input startup Willow Voice announced the completion of a \$4.2 million angel round of financing, led by YC. Just a few days ago, on June 25th, another voice input startup, Wispr Flow, also announced the completion of a \$30 million Series A round of financing.

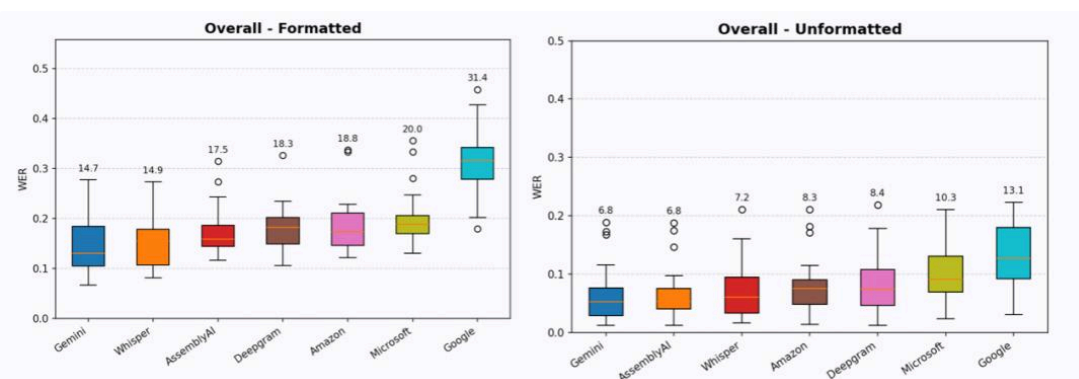
Previously, we had been observing the AI voice track. However, most of the companies that received financing were focused on voice synthesis, that is, "output." For example, ElevenLabs, a leading company in the track, completed a \$250 million Series C round of financing in January this year, with a valuation exceeding \$3 billion.

However, these two recent rounds of financing seem to be sending another signal. Voice startups that focus on "input" are attracting the attention of capital.

Voice input has been around since 2012. Why can it still receive financing?

Willow Voice and Wispr Flow (hereinafter referred to as Willow and Flow for short) focus on ASR technology (Automatic Speech Recognition). The two products are quite similar. **They are both somewhat like "voice input methods." Users only need to press a specific button on their computer or mobile phone to directly transcribe the spoken content into text.**

At first glance, this is a function that we are already familiar with in our daily lives. For example, WeChat launched the "voice-to-text function" on the iOS platform in 2019. Apple even launched the initial version of "Voice Dictation" in iOS 6 in 2012. Moreover, star startups in the AI era such as ElevenLabs and OpenAI also cover the scenario of voice-to-text conversion.



Formatted: The model needs to directly output the correct format, which means not only the vocabulary is correctly recognized, but also the capitalization and punctuation are correctly output. Unformatted: Only the recognition accuracy of the words themselves is considered.

According to the test by VoiceWriter.io, except for the transcription function of Google Cloud, which is slightly worse, there is not much difference among other products. In the scenario of unformatted text, the error rate of most products is below 10%. **This is comparable to the level of humans without professional transcription training. However, in the scenario of formatted text where punctuation and capitalization need to be considered, the performance of voice transcription models is slightly worse, and the average error rate of all products increases by 10%.**

According to Tanay Kothari, the founder of Flow, in a podcast, **although the WER of AI when transcribing unformatted text is already very low**, it doesn't really matter. Because even if a product can achieve an error rate below 1%, it still means there is a wrong word in every few sentences, and users still can't fully trust AI.

Moreover, due to the differences between spoken and written language, even if the model can transcribe exactly what the user says, users won't directly send the output text as a message or save it in a note-taking app. They still need to simplify and correct the text.

Guided by this concept, the difference between Flow and traditional voice-to-text conversion is that it pursues "zero-editing information." In practice, both products add a "text processing" step between "AI directly transcribing content" and "outputting content" to provide users with text that can be directly used. This text processing step has three levels. First, formatting the text output, such as correct sentence segmentation and removing filler words. Second, understanding the context, such as automatically correcting slips of the tongue and recognizing emotions. Third, context recognition, that is, the ability to output text in different styles in different input scenarios such as DMs, emails, and notes.

Through a preliminary comparative test of Flow, Willow, and OpenAI Whisper, it was found that **OpenAI Whisper's output only reaches the first level, Flow and Willow can reach the second level, and none of them can reach the third level.**

The goal is great, but the reality falls short

From a product logic perspective, Flow and Willow actually complete the process from "spoken language input" to "written language output." Since it mainly focuses on written language, its usage scenarios are more inclined towards **office scenarios**.

This dictation is incredible – I just talk into my Mac. It removes all my *uhms* and *ahs* and even automatically formats bullets when it realizes I'm dictating a list. (This exact sentence may have been dictated, as well.)

AMMAAR RESHI
Head of Design, ElevenLabs

I use this transcription app every day. You can set one key, hold it down, and talk – it transcribes and formats that audio into text (including summary bullets, quotes, or making minor corrections if you repeat yourself) and auto-pastes into the app.

BEN TOSSELL
Founder, Ben's Bites

In a16z's year-end AI product inventory, Ammaar Reshi, the chief designer of ElevenLabs, and entrepreneur Ben Tossell both recommended Flow. Judging from their recommendations, they use it almost every day. | Image source: a16z

Since voice input has a greater impact on the surroundings compared to keyboard input, it is not very suitable for ordinary office workers sitting at their desks. So, according to the sharing of Flow's founder, they initially targeted **Silicon Valley VCs, entrepreneurs, and executives** who receive a large amount of information, have a need for efficient input, and mostly have their own offices or often work outside the office.

FLOW FOR



Leaders

Unblock teams, build faster with voice



Students

Take better notes, write faster, study smarter



Developers

Speak more context, get better results



Creators

Capture content ideas anytime, anywhere



Lawyers

Dictate case notes and memos on the go

CASE STUDIES FOR



Gaurav the Advisor

Fast, fluent replies for busy advisors



Greg the Writer

Speaking chapters into life

Typical user analysis on Flow's official website | Image source: Flow's official website

After the initial user growth through VCs, entrepreneurs, and executives, Flow started to reach more users with needs through Product Hunt, such as students, code developers, creators/writers, lawyers, and consultants. Like VCs, entrepreneurs, and executives, these users also have a need to process a large amount of text or input long texts, and their work locations are generally flexible. They also often process text outside the office.

model).

Test 1: To-do List scenario

Scenario description: On the way to the office by car, a team leader needs to sort out the important tasks of the day and record them in a note-taking app.

Colloquial content: Um... Today, first, I need to update the icons on the main page and then send a launch notice before 3:30 pm. Second, I need to have a review meeting with the team at 4 pm. Also, send last week's daily report to John. Third, before 5 pm, organize the user feedback summary document. Finally, before 7 pm, send next week's schedule to the product team.

Output requirements: The key information should be correct, and the to-do items should be presented in points automatically.

Outputs of different products:

Wispr Flow:

今天首先要更新主页面的图标，然后在三点半前发上线通知。

第二，四点钟要与团队开复盘会，还有把上周的日报发给 Join。

第三，五点前要把用户反馈的文档整理了。

七点之前要把下周的排期发给产品。

Willow: 今天首先要更新主页面的图标,然后在三点半发上线通知。

第二四点钟要与团队开复盘会,还有把上周的日报发给转。

第三五点钟前要把用户反馈汇总文档整理了。

最后晚上七点钟前要把下周排期发给产品。

OpenAI Whisper: 今天首先要更新产品主页面的图标,然后在三点半前发上线通知,第二四点钟要与团队开复盘会,还有把上周的日报发给John,第三五点钟前要把用户反馈汇总文档整理了,最后晚上七点钟前要把下周的排期发给产品。今天先说到这里,谢谢大家收看,我们下期再见。

Evaluation: In this scenario, none of the three products missed core information such as time and tasks. Flow and Willow both segmented the content according to the marker words like "first/second/third/finally" in the original text, looking more like a To-do List. In terms of punctuation and format, Flow did a better job.

OpenAI's Whisper performed the worst overall. Although it added punctuation, it didn't segment the content and added redundant text at the end.

Test 2: Memo scenario with professional terms

Scenario description: Before a brokerage's earnings report review meeting, as a brokerage analyst, the user needs to briefly summarize the highlights of the earnings report, form a written memo, and share it with other team members.

Colloquial content: "Um... I just read the earnings report. Although XX had year-on-year growth this quarter, the quarter-on-quarter growth was negative. And the proportion of its subscription revenue is increasing, mainly due to the contributions of XX and XX. Also, we need to take a look

Output requirements: The key information should be correct, professional terms should be used correctly, and the tone should be formal.

Outputs of different products:

Wispr Flow: 我刚看了财报, XX这季度虽然同比是有增长的, 但环比是负的。然后, 他的订阅收入在上升, 占比也在上升, 主要是XX和XX的贡献。

另外, 他跟阿里的可转债得看看是不是有**摊包风险**的问题。我建议把重点放在产品结构和付费动能这块。营收增速还是偏保守吧。

Willow: 我刚看了XX的财报。XX这季度虽然同比是有增长的, 但环比是负的。"然后他订阅收入占比在上升, 主要是XX和XX的贡献。另外**他跟阿里可展债得看看**, 是否有**贪包风险**的问题。我建议重点放在产品结构和付费动能这一块, **营修增速**还是偏保守。

OpenAI Whisper: 呃,我刚看了那个财报,XX这季度虽然同比是有增长的,但环比是负的,然后他那个订阅收入占比在上升,主要是那个XX和XX的贡献吧,另外他跟阿里的那个**科展站**得看看了,就是是不是有**贪包风险**的问题,我建议把重点放在产品结构和付费动能这块,营修营收增速还是偏保守的。**谢谢观看 欢迎订阅我的频道**

Note: The red words are the ones with errors. The text is generated by AI and is only for testing purposes, having nothing to do with reality.

Evaluation: In the memo scenario with some professional terms, all three products made mistakes in the term "dilution risk." Willow and Whisper also made quite a few other errors. Even after I manually added the term "dilution risk" in Flow, Flow still didn't output it correctly. Overall, none of the three products can handle more professional scenarios well, but Flow did a slightly better job.

In addition, none of the three products corrected "colloquial" expressions such as "the quarter-on-quarter growth was negative," nor did they do any logical sorting. When I input the content, I split the "revenue" part into the first and third sentences, and none of the three products combined the similar content together.

Test 3: Email reply scenario to a customer

Scenario description: At the airport, the user needs to reply to a customer's inquiry email and provide suggestions.

Colloquial content: Hello. I see that you want to optimize your sales process. I think your current problem is quite typical. There isn't a very systematic screening mechanism after the initial leads come in, which leads to low efficiency when the sales team follows up. In this kind of situation, several of our previous customers have also encountered it. We usually suggest standardizing the lead scoring criteria or introducing a relatively lightweight CRM system. I'll organize a previous case of ours in the next few days and send it to you later to see if it's helpful.

Output requirements: Automatically segment the content, present it in an email format, and use a formal writing style.

Outputs of different products:

看到你们说想优化销售流程，我感觉其实你们现在的问题还挺典型的。就是前期线索进来之后没有特别系统的筛选机制，导致后面销售在跟进时效率挺低的。

像这样的情况，我们之前有几个客户也遇到过。一般会建议统一一下线索评分的标准，一个比较轻量的CRM系统啥的也可以。

我这两天再把我们之前做的一个案例给你整理一下，届时你看看是不是有参考价值。

Willow:

你好，

看到你们说的想优化销售流程，我感觉你们现在的问题还挺典型的：前期线索进来之后没有特别系统的筛选机制，导致后面销售在跟进的时候效率挺低的。像这种情况，我们之前几个客户也遇到过。一般会先统一一下线索评分的标准或者引入一个比较轻量的CRM系统。

这个是我这两天再把我们先做的一个案例整理一下，到时候你再看看是不是有参考价值。

OpenAI Whisper:

您好,看到你们说想优化销售流程,我觉得你们现在问题还挺典型,就是前期线索进来之后没有特别系统的筛选机制,然后导致后面销售在跟进的时候效率挺低的,像这种情况,我们之前几个客户也遇到过,一般会先统一一下线索评分的标准,或者引入一个CRM,我这两天再把我们之前做的一个案例给你整理一下,到时候你看看是不是有参考价值。

Evaluation: Both Flow and Willow started the "Hello" on a new line according to the email format. Flow did a better job in terms of segmentation. None of the three products could well modify the colloquial expressions in the input. Only Flow changed "later" to "at that time." Overall, the writing style of the email is still very colloquial, and users need to manually modify it before sending.

After the experience, although the quality of Flow and Willow is acceptable, they are still quite far from their goal of "zero editing." In the recognition of professional terms and in formal writing styles, the output doesn't meet the standards. I also tested English transcription, and the results were similar, with transcription errors also occurring.

Conclusion

Although the test results show that Flow and Willow still have a certain gap from "zero editing" in scenarios such as professional terms and formal writing styles, according to media reports, Flow has a very high user stickiness and payment rate. As of now, Wispr Flow has officially announced that its monthly user growth rate has exceeded 50%, the 6 - month retention rate of active users has reached 80%, and the payment rate is as high as 19%. Its annual revenue (from July 2024 to July 2025) has reached \$3.8 million.

The fact that they haven't achieved the "zero editing" goal doesn't stop users from paying. This is because products like Flow, although they can't completely eliminate the need for user intervention, still provide a differentiated experience compared to before.

Concept

I recently tried my hands at vibe coding, a term coined by Andrej Karpathy. For this, I used Cursor AI, and for dictation, I used Wispr Flow. A few key things to keep in mind while going for vibe coding:

- Your AI dictation tool is very, very important. In my case, Wispr Flow did a great job.
- If the AI dictation is poor, the entire flow of vibe coding gets disturbed.
- Your LLM is also quite crucial. If the LLM is weak, you are going to bang your head.

Initially, I was a little perplexed between **Wispr Flow and superwhisper**- the two major tools for AI dictations out there. But later, I chose Wispr Flow because of a couple of reasons:

- Wispr Flow is available for both Mac and Windows, while superwhisper is just for Mac.
- The error rate for Wispr Flow is any day better than superwhisper.
- Punctuation handling is better for Wispr Flow
- Latency-wise, Wispr Flow is any day better.

Do let me know which tools you are using that are better than Cursor AI and Wispr Flow.

Many users on Reddit and Product Hunt said that the experience of using Flow to interact with ChatGPT or for Vibe Coding is very good. | Image source: Reddit

According to comments on Reddit and Product Hunt, in less formal input scenarios, Flow performs much better than other products and is quite satisfactory. For example, some users use Flow to interact with Cursor in natural language (as shown in the picture above) for Vibe Coding. In this scenario, users only need to press a key on the Mac keyboard and can directly talk. Although Flow still has some deficiencies in the third - level conversion to written language for specific scenarios, it performs significantly better than other products led by OpenAI's Whisper in the first two levels of formatting and context understanding.

Flow's high user stickiness and payment rate also reflect that reducing the "friction" in human - machine interaction through voice input and achieving efficiency improvement may be a feasible solution. Although neither Flow nor Willow has achieved the goal of "zero editing" in all scenarios at present, with the further improvement of large - model capabilities and the accumulation of data, there will probably be great improvements in the future.

According to the founder of Flow, if "voice input" can reach a level that users can trust in the future, it won't be long before "voice input" replaces the keyboard and becomes a new paradigm of human - machine interaction (voice operating system). "Real - world efficiency improvement + the possibility of subverting the old paradigm in the future" may be the reason why VCs are willing to invest real money in voice input.

Reference articles:

1. With an 80% retention rate and a 19% payment rate, why did this AI voice keyboard secure \$56 million in financing? By Crow Intelligence
2. This AI - native voice input method, Flow, raised \$30 million in Series A financing. Its experience outperforms WeChat and Sogou input methods. By Uncle Huang of AI Products
3. The Best Speech Recognition API in 2025: A Head - to - Head Comparison

The data comes from third -

该文观点仅代表作者本人，36氪平台仅提供信息存储空间服务。

For media inquiries, interview requests, business collaboration and Recruit global ecological partners, please contact mohaiying@36kr.com.

Partner: startuprad.io

36kr Europe (eu.36kr.com) delivers global business and markets news, data, analysis, and video to the world, dedicated to building value and providing business service for companies' global expansion.

© 2024 36kr.com. All rights reserved.