



关于北京海天瑞声科技股份有限公司
首次公开发行股票并在科创板上市申请文件
第三轮审核问询函的回复

保荐机构（主承销商）



（深圳市福田区中心区中心广场香港中旅大厦）

上海证券交易所：

根据贵所于 2019 年 6 月 6 日印发的上证科审（审核）〔2019〕231 号《关于北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市申请文件的第三轮审核问询函》（以下简称“问询函”）的要求，华泰联合证券有限责任公司（以下简称“华泰联合”或“保荐机构”）作为北京海天瑞声科技股份有限公司（以下简称“海天瑞声”、“发行人”或“公司”）首次公开发行股票并在科创板上市的保荐机构（主承销商），会同发行人及发行人律师北京市天元律师事务所（以下简称“天元律师”）和瑞华会计师事务所（特殊普通合伙）（以下简称“瑞华会计师”）等相关各方，本着勤勉尽责、诚实守信的原则，就问询函所提问题逐项进行认真讨论、核查与落实，并逐项进行了回复说明。具体回复内容附后。

说明：

1、如无特别说明，本回复中使用的简称或名词释义与《北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市招股说明书（申报稿）》（以下简称“招股说明书”）一致。涉及招股说明书补充披露或修改的内容已在招股说明书中以**楷体加粗**方式列示。

2、本回复中若出现总计数尾与所列值和不符合的情况，均为四舍若出现总计数尾与所列值和不符合的情况，均为四舍若出现总计数尾与所列值和不符合的情况，均为四舍五入所致。

3、本回复中涉及的我国、我国经济以及行业的事实、预测和统计，包括本公司的市场份额等信息，来源于一般认为可靠的各种公开信息渠道。本公司从上述来源转载或摘录信息时，已保持了合理的谨慎，但是由于编制方法可能存在潜在偏差，或市场管理存在差异，或基于其他原因，此等信息可能与国内或国外所编制的其他资料不一致。

目 录

问题 1、关于核心技术	4
问题 2、关于共同控制	21
问题 3、关于发行人股份支付	28
问题 4、关于发行人采购成本	37
问题 5、关于发行人的营业收入	46
问题 6、关于发行人毛利率	71

问题 1、关于核心技术

回复材料称，发行人的核心技术具备较高技术壁垒、较难为同行业公司或上下游行业突破。发行人注重在相关领域的核心技术积累构建及核心技术保护，但并不单纯依赖以申请专利的形式对核心技术进行保护。

请发行人说明：（1）发行人已提交的 8 项发明专利申请的主要内容、发明人及截至目前所处的阶段，与发行人核心技术的关系，是否与“并不单纯依赖以申请专利的形式对核心技术进行保护”相矛盾；（2）进一步说明发行人的核心技术具备较高技术壁垒、较难为同行业公司或上下游行业突破的依据；（3）与澳大利亚上市公司 Appen 等可比公司的可比业务在市场地位、技术实力、衡量核心竞争力的关键业务数据、主要财务指标的比较分析；（4）列表说明发行人设计的数据库较同行业竞争对手或下游客户自身的训练数据库之间的差异，如何体现发行人数据库设计的优势；（5）以主要类型项目为例，说明发行人提供定制服务和数据库开发的业务流、数据流和财务流的流转情况。

请保荐机构核查并发表明确意见。

答复：

（1）发行人已提交的 8 项发明专利申请的主要内容、发明人及截至目前所处的阶段，与发行人核心技术的关系，是否与“并不单纯依赖以申请专利的形式对核心技术进行保护”相矛盾；

一、发明专利申请的主要内容、发明人、所处阶段、与核心技术的关系

截至本问询函答复出具日，发行人已提交 8 项发明专利申请，其中 5 项已进入实质审核阶段。上述专利申请中，7 项对应于发行人核心技术下的细分技术及研发成果，1 项虽未涵盖在核心技术范围内，但也应用于发行人的具体业务和服务环节中。

相关专利申请的发明人、所处阶段、主要内容及其对应的核心技术/业务情况具体如下：

序号	申请号	专利名称	专利权人	发明人	申请日	截至目前所处的阶段	主要内容	专利对应的核心技术/业务
1	201810449585.3	语音对齐方法及装置	发行人	邵志明、郝玉峰	2018.05.11	实质审查	提供一种语音对齐方法，通过计算多个语音特征向量的自相关系数，解决了多个音频文件起始点自动对齐的问题，有效提升了数据库质量和数据处理效率。	数据同步技术——音频对齐技术
2	201810827500.0	音素误标注的检测方法和装置	发行人	孟君、廖晓玲、郝玉峰	2018.07.25	实质审查	提供一种音素误标注的检测方法和装置，该技术通过采用预先设置的音素置信度模型，获取音素标注结果的置信度，从而快速、准确地获取误标注音素。	分布式高性能自动校验技术——音素标注正确率校验技术
3	201810883930.4	语音检测方法和装置	发行人	邵志明、曹琼、宋琼、郝玉峰	2018.08.06	实质审查	提供一种特定的语音检测方法和装置，可检出语音数据库中同一人的所有语音，提高了数据检验效率和准确度。	分布式高性能自动校验技术——相似说话人自动筛查技术
4	201810988973.9	韵律标注方法、装置和设备	发行人	孟君、曹琼、廖晓玲、郝玉峰	2018.08.28	实质审查	提供了一种韵律标注方法、装置和设备，该技术基于语音识别模型对音素进行自动切分，计算得到语音中静音段的时长，根据停顿时长预标韵律信息，结合基于文本的韵律预测结果修正预测结果，提高了韵律预标的准确率，提升了韵律标注的效率和准确性。	大数据驱动的高效数据处理技术——基于语音特征的韵律预测技术
5	201910022453.7	中英混语料生成办法、装置、电子设备及存储介质	发行人	杨福星、曹琼、郝玉峰	2019.01.10	申请受理	提供一种中英混合语料的生成方法、装置、设备及存储介质，可根据中文语料生成中英混合语料，有效扩充中英混合语料库，为中英混合语音识别的模型训练提供充足的训练数据，从而提高模型的准确性。	多语种多模态数据库设计技术——混合语言语料设计技术
6	201910072150.6	语料选取处理方法、装置、设备及计算机可读存储介质	发行人	杨福星、曹琼、郝玉峰	2019.01.25	申请受理	提供一种语料选取处理方法、装置、设备及计算机可读存储介质，该方法可根据原始句长分布从原始语料中选取句数、句长、分布符合要求的语料作为初始模型，并对前述模型进行修正处理，得到接近原始句长分布要求的最终句长分布模型，满足语料设计中对于句长分布的要求。	多语种多模态数据库设计技术——多语种音素均衡语料设计技术

7	201910073390.8	音频质量评估方法、装置、电子设备及存储介质	发行人	刘永辉、曹琮、郝玉峰	2019.01.25	申请受理	提供一种音频质量评估方法、装置、电子设备及存储介质。该发明专利提供的方法根据发音人语速预设规则进行工程性的自动化分析，能够有效提高音频质量评估效率，且有效提高检测样本的覆盖率。	实时采集质控技术——语音信号质量检测技术
8	201710539306.8	基于麦克风的音频处理方法和装置	中瑞智（发行人全资子公司）	黄大伟；李虹辰	2017.07.04	实质审查	提供一种基于麦克风的音频处理方法和装置，该方法可根据每一麦克风采集的音频的变化量确定待处理的音频数据；在多个麦克风同时采集音频时可确定当前待收音的麦克风并确定相应待处理的音频数据，为后续播放、语音识别等处理奠定基础。	“数据资源相关的应用服务”相关业务

二、不与“并不单纯依赖以申请专利的形式对核心技术进行保护”相矛盾

（一）发行人构建了综合性的核心技术保护体系

发行人综合考虑核心技术类型、申请专利对核心技术的保护效果等，从制度角度、技术角度、法律保护角度出发，采取多重手段对核心技术进行综合保护。

1、制度角度：制定完善的保密管理制度，防止核心技术泄密

发行人制定了完善的保密管理制度，对包括核心技术机密在内的各类保密信息及材料制定了具体的保密要求，并同步制定了保密教育、泄密惩治机制等相关管理制度，提升员工的保密意识、降低核心技术泄密风险。发行人与全体员工均签署了《保密协议》，与核心技术人员均签署了《竞业禁止协议》，明确员工的保密义务，对包括核心技术在内的信息进行保护，防止技术流失和外泄。

2、技术角度：采取有效的技术手段对核心技术进行保护

对于核心技术体系中与数据资源设计开发相关的系统、平台、软件、开发工具等计算机程序，发行人从代码管理、可执行文件管理两个角度出发，采取技术手段防止核心技术泄密，主要包括：引入 GIT、SVN 等代码管理系统，将核心技术相关代码部署于内部服务器中；对核心代码设置访问权限管理制度，限制访问范围；使用 B/S（浏览器/服务器）架构方案，将关键技术隐藏于服务器端；对于无法使用 B/S 架构的客户端软件，则加入时间戳功能进行生命周期管理，降低特定客户端软件在项目结束后被擅自复用、泄露技术的风险。

3、法律保护角度：申请专利保护、计算机软件著作权等保护措施

发行人对于专利保护可覆盖的、生产研发受专利申请公开的影响可控的核心技术类型，申请专利保护；对于与数据资源设计开发相关的系统、平台、软件、开发工具的计算机程序等，通过登记计算机软件著作权的形式加以保护。

（二）专利申请仅覆盖特定技术，与“并不单纯依赖以申请专利的形式对核心技术进行保护”不存在矛盾

发行人已提交的 8 项发明专利申请所保护的主体仅为专利保护可覆盖的部分特定细分技术、技术研发成果，与“并不单纯依赖以申请专利的形式对核心技

术进行保护”不存在矛盾。

综上，发行人对核心技术采取了综合性保护措施，已提交的 8 项发明专利申请所保护的主体仅为专利保护可覆盖的部分特定细分技术、技术研发成果，与“并不单纯依赖以申请专利的形式对核心技术进行保护”不存在矛盾。

(2) 进一步说明发行人的核心技术具备较高技术壁垒、较难为同行业公司或上下游行业突破的依据；

发行人语音语言学基础研究、多语种多模态数据库设计技术、数据同步技术、大数据驱动的高效数据处理技术及分布式高性能自动校验技术等 5 项核心技术具备较高技术壁垒、较难为同行业公司或上下游行业突破，依据具体如下：

1、上述核心技术积累需要大量时间、资源积累

上述 5 项核心技术主要发源并根植于发行人多年积累的语种/方言覆盖能力、数据资源开发经验和资源，上下游企业、同行业公司在短期内难以较易突破，具体如下：

发行人多年自主开展语音语言学基础研究，并辅以与语言学家合作，在数据资源设计、开发方面积累了较多语音语言学基础研究成果，在语种/方言覆盖能力上形成了突出优势。语种/方言覆盖能力也体现在发行人的多语种多模态数据资源设计技术中，使该项技术也具备较强的技术壁垒。

此外，发行人在国内起步较早，在各类语种/方言、应用场景、领域的的数据资源相关的设计、采集、处理（标注）、质检方面具备先发优势、经验优势，使得发行人有能力将积累的资源、经验应用于其数据同步技术、大数据驱动的高效数据处理技术及分布式高性能自动校验技术之中，使得这些技术具备较强的技术壁垒。

慧听科技、标贝科技等境内同行业企业也专业从事数据资源开发，并掌握一定的类似技术，但与发行人相比，在语种/方言覆盖能力、经验资源积累、数据量处理积累方面相比发行人存在一定差距，弥补前述差距需要其投入较多时间、资源，难以在短期内突破发行人通过上述优势构建的核心技术壁垒。

下游行业主要为人工智能技术层及应用层相关企业，其因自身算法模型开发、训练所需，亦掌握一定的语音语言学基础研究知识、数据库设计技术、数据处理技术、数据质检技术等，但相关技术通常侧重于其算法模型所覆盖的特定语种/方言、应用领域，与发行人等专业从事数据资源开发的企业相比，因在产业链专业化分工侧重性不同，其在数据资源开发处理领域的专业化程度、全面性、数据处理量、实践经验、流程把控能力等方面存在一定差距，难以在短期内突破发行人通过上述优势构建的核心技术壁垒。

上游供应商主要为人力资源服务公司、技术服务公司等，发行人主要向其采购生数据采集、标记服务。前述企业主要参与发行人数据资源开发过程的生数据采集（使用发行人提供的采集工具进行录音、拍照、编写文字等）和数据标记的具体执行（使用发行人自主研发的标注软件、工具集和平台对已分配的数据包进行数据属性及特征标记等）等非核心环节，不掌握发行人所具备的数据资源开发相关核心技术，难以突破发行人构建的核心技术壁垒。

2、核心技术的应用构建了发行人的产品/服务优势

发行人将上述核心技术应用于数据资源开发等产品服务生产、提供流程中，使最终提供的产品/服务在数量、语种/方言及场景领域覆盖能力等方面具备较为突出的优势，具体如下：

项目	海天瑞声	Appen	慧听科技	标贝科技
出售的自有产权数据库数量	500 余个	85 个 (包含 73 个语音类数据库数量及 12 个自然语言类数据库；其他类别数据库数量未公开披露)	23 个	23 个
数据库种类	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言、音乐	智能语音、计算机视觉、自然语言、音乐
语种/方言	130 余个	180 个	18 个	4 个

数据来源：截至 2019 年 6 月，Appen、慧听科技、标贝科技官网及公开披露信息。

发行人数据库较同行业竞争对手、下游客户自身的训练数据库之间的差异情况及其体现的发行人数据库设计优势请参见本问询函答复 1（4）。

3、核心技术的应用使得发行人广受认可

发行人通过运用上述核心技术提升了自身的产品服务品质, 获取了合作伙伴、行业机构的认可。发行人所获合作伙伴、行业机构奖项具体如下:

序号	荣誉	颁发机构	颁发时间
1	2019 中国信息技术人工智能行业优秀产品	中国信息协会	2019 年
2	2019 中国信息技术人工智能数据服务领军企业	中国信息协会	2019 年
3	菜鸟网络“驼峰计划”突出贡献奖	菜鸟网络	2019 年
4	腾讯 2019 年服务类供应商“杰出贡献奖”	深圳市腾讯计算机系统有限公司	2019 年
5	信息化影响中国·2018 年人工智能行业最佳产品	中国信息协会	2018 年
6	2018“中国声谷”杯智能语音创新大赛行业先锋奖	中国语音产业联盟	2018 年
7	2017 中国信息技术年度人工智能数据服务领域领军企业奖	中国信息协会	2017 年
8	Oriental COCOSDA 2016 ITN Best Paper Award	Oriental COCOSDA	2016 年
9	百度 API Store 2015 年度大奖评选最佳合作伙伴	百度 API Store	2015 年

(3) 与澳大利亚上市公司 Appen 等可比公司的可比业务在市场地位、技术实力、衡量核心竞争力的关键业务数据、主要财务指标的比较分析;

发行人的同行业竞争对手主要包括 Appen、慧听科技及标贝科技。根据前述公司的公开披露信息, 发行人与其在市场地位、技术实力、衡量核心竞争力的关键业务数据和财务指标等的比较情况如下:

项目	海天瑞声	Appen	慧听科技	标贝科技
市场地位				
成立年份	2005 年	1996 年	2011 年	2016 年
市场地位概述	我国领先的人工智能数据资源产品服务专业提供商, 是我国最早从事数据资源产品服务研发销售的企业之一	较早从事数据资源开发的数据资源产品服务提供商, 经营历史较长, 规模、体量较大	-	-
行业机构评价	参见本问询函答复 1(2)	-	“慧听中文重口音语音数据库”曾获中国语音产业联盟“2015 中国语音创新产品”	中国语音产业联盟会员单位

技术实力、关键业务相关指标情况				
技术实力概述	海天瑞声自主开发了一体化人工智能数据处理技术支撑平台，在基础研究、平台工具、数据资源开发三个维度下均积累核心技术，有机搭建起发行人的人工智能数据资源开发一站式解决方案，将多项具体核心技术整合为发行人特有的核心技术体系。	未公开披露	采用全程质量监控流程，执行完善的标注流程，配合保密管理手段，提供质量上乘的数据服务。	拥有语音合成模型和算法，通过算法+专业的人工数据处理方式，为客户提供优质的语音合成服务。拥有 TOBI 标注体系，通过自主研发的 TTS 评测系统，为客户提供高质量的数据服务。
出售的自有产权数据库数量	500 余个	85 个 (包含 73 个语音类数据库数量+12 个自然语言类数据库；其他类别数据库数量未公开披露)	23 个	23 个
数据库种类	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言	智能语音、计算机视觉、自然语言、音乐	智能语音、计算机视觉、自然语言、音乐
语种/方言覆盖能力	130 余个	180 个	18 个	4 个
正在申请/已授权的专利数量	8 项中国境内专利 (均正在申请中)	2 项 (均已授权)	2 项中国境内专利 (均正在申请中)	2 项中国境内专利 (均正在申请中)
计算机软件著作权数量	98 项	未公开披露	12 项	10 项
其他经营指标				
员工数量	127 人	513 人	未公开披露	未公开披露
合作伙伴/主要客户	大型科技公司，如阿里巴巴、腾讯、百度、微软、三星等；人工智能企业，如科大讯飞、商汤科技、云知声、海康威视等；科研机构，如中国科学院、清华大学等	微软等大型科技公司、汽车厂商及政府	未公开披露	大型科技公司，如阿里巴巴、腾讯、百度、网易、京东等；人工智能企业，如出门问问、旷视等；科研机构，如北京大学等。
主要财务指标 (2018 年度/2018 年末)				
营业收入	19,265.8 万元	36,428.9 万澳元	未公开披露	未公开披露

营业收入增长率	61.80%	117.43%	未公开披露	未公开披露
净利润	6,714.2 万元	4,172.8 万澳元	未公开披露	未公开披露
综合毛利率	64.77%	37.32%	未公开披露	未公开披露

数据来源：截至 2019 年 6 月，Appen、慧听科技、标贝科技官网及公开披露信息；国家知识产权局中国及多国专利审查信息查询平台（<http://cpquery.sipo.gov.cn/>）、中国版权保护中心 CPCC 微平台等公开信息查询渠道及第三方机构可查询信息。

对上述比较情况分析如下：

与 Appen 相比，发行人在营收规模、净利润规模、员工数量等指标上小于 Appen，在体量、数据资源覆盖的语种/方言数量上具备一定劣势；但在我国客户的覆盖程度以及自有知识产权数据库数量上具备一定优势。与慧听科技、标贝科技相比，发行人在数据资源产品覆盖的语种/方言个数、自有知识产权数据库数量、计算机软件著作权数量、在申请专利数量方面均高于公开信息可查询的慧听科技、标贝科技的相关数量，优势明显；除在数据资源种类上暂未涵盖音乐类数据资源外发行人不存在明显劣势。

在市场地位方面，Appen 是较早从事数据资源开发的数据资源产品服务提供商，经营历史较长，规模、体量等相比发行人均具备优势；发行人是我国最早从事数据资源产品服务研发销售的企业之一，在我国市场具备领先地位。多年来，发行人深耕数据资源服务领域，伴随了众多国内客户在人工智能领域特别是智能语音领域的开拓、成长，为其持续提供了全球语种语音数据资源的高质量的本土服务，降低了对国外同类数据资源的依赖。

在主要财务指标方面，慧听科技、标贝科技未公开披露其财务数据信息；发行人在营业收入、净利润规模及营业收入增长率方面相比 Appen 存在一定劣势，综合毛利率则高于 Appen。发行人主营业务毛利率高于 Appen，主要是发行人高毛利率的业务占比高于 Appen 所致，具体包括以下方面：

首先，以自有产权数据库产品的数量对比而言，目前发行人数据库产品数量超过 500 个，Appen 官网展示其数据库产品数量为 85 个，明显低于发行人，因此 Appen 该类毛利率可达到或接近 100%的数据库产品业务对其整体毛利率的提升作用预计小于发行人。

其次，就数据资源定制服务而言，根据 Appen 公开披露资料显示，其大部分收入（报告期内平均占比超过 80%）来源于文本相关度标注业务。发行人同类业务的毛利率水平仅为 20-30%，显著低于其他智能语音类和计算机视觉类业务。报告期内，发行人已很少从事该类业务，而 Appen 收入结构中低毛利率的文本相关度标注业务占比仍旧较高，也使得发行人毛利率水平高于 Appen。

除上述业务结构原因外，Appen 作为海外公司，其主营业务所需的人员、场地均主要分布在美国和澳大利亚，劳务成本和人工成本与我国境内相比较为高，也在一定程度上拉低了其的毛利率。

因此，发行人主营业务毛利率高于 Appen 具备合理性。

综上，与 Appen 相比，发行人在市场地位、体量规模、数据资源覆盖的语种/方言数量、收入利润规模等方面存在一定劣势，但数据库产品资源积累等业务结构的差异使发行人的毛利率高于 Appen；而与慧听科技、标贝科技相比，发行人在自有数据库积累、数据资源覆盖的语种/方言数量、计算机软件著作权数量、在申请专利数量等方面具备显著优势，除在数据资源种类上暂未涵盖音乐类数据资源外不存在明显劣势。

（4）列表说明发行人设计的数据库较同行业竞争对手或下游客户自身的训练数据库之间的差异，如何体现发行人数据库设计的优势；

对于数据资源行业而言，数据库的设计能力主要体现在不同种类数据库的数量、覆盖语种/方言、场景、采集方案复杂度等方面。

一、与同行业竞争对手数据库¹的差异及体现的数据库设计优势：

项目	海天瑞声	Appen	慧听科技	标贝科技
1、智能语音类数据库				
数据库数量	407	73	19	15
覆盖语种/方言数量	77	46	18	4
录音场景类别	5	5	4	2
采集设备种类	8	3	5	3

¹ 注：Appen、慧听科技、标贝科技的数据库数量等相关信息仅包含其官方网站或公开信息披露所列举的数据库相关信息；发行人的数据库信息为截至 2018 年 12 月 31 日之情况。

采集通道覆盖数	5	3	4	1
2、自然语言类数据库				
数据库数量	72	12	2	4
覆盖语种/方言数量	33	9	2	1
3、计算机视觉数据库				
数据库数量	37	-	2	3

与竞争对手在公开信息渠道列示披露的数据库产品相比，发行人设计的数据库产品的差异之处主要体现在：

1、数据库数量更多

发行人在智能语音、自然语言及计算机视觉领域均具备较为丰富的数据库产品储备，在数据库数量上具备显著优势。

2、覆盖语种/方言数量更多

发行人的智能语音数据库覆盖的语种/方言数量较多，具备显著优势；自然语言类数据库产品覆盖的语种方言数量略低于 Appen，并优于慧听科技、标贝科技。

3、采集方案复杂度更高

发行人智能语音数据库的录音场景类别与 Appen、慧听科技相当并优于标贝科技；采集设备种类显著高于竞争对手；采集通道覆盖数相比竞争对手存在一定优势。综合而言，发行人的智能语音数据产品在采集方案复杂度相关指标上更为突出，可提供更为复杂的采集方案。

上述数据库差异表明，与竞争对手相比发行人有能力设计覆盖多语种/方言、多场景、多领域的，采集方案更为复杂的数据库产品，体现了发行人在数据库设计领域的全面性、专业性优势。

二、与下游客户自身的训练数据库之间的差异及体现的数据库设计优势

发行人的下游客户主要为人工智能应用层、技术层的各类机构，其自身的训练数据库通常服务于与其算法模型或产品应用相关的特定语种/方言及应用场景，而发行人在多场景、多领域、多语种/方言数据资源的全面性、综合性设计能力

方面更有优势。

此外，下游客户也常提供其获取的各类生数据给予发行人进行专业化处理，形成可供算法模型训练使用的数据资源，这也反映了发行人在数据处理方面的全面性、专业性优势。

三、与开源训练数据库之间的差异及体现的数据库设计优势

发行人的有采集的单一化标注数据库主要用于训练语音识别模型。对于语音识别算法模型而言，使用不同训练数据库将实现不同的模型训练效果（识别率），进而决定算法模型所属语音识别引擎的优劣。

根据发行人的测试结果，以发行人的 King-ASR-216 数据库为例，其与开源数据库相比可提供更好的语音识别算法模型训练效果（更高的识别率、更低的字错误率），反映了发行人在设计优质训练数据库、优化算法模型训练效果方面的能力和优势。

前述算法模型训练对比情况具体如下：

1、模型训练情况概述

发行人从中文开源语音识别数据库（下称数据库 1²，共 160 小时）、自有产品库 King-ASR-216（随机选取 160 小时）中各随机选取 150 小时数据作为训练数据，分别对同一语音识别开源软件 Kaldi 进行训练得到模型 1 及模型 2，并采用不同的测试集测试模型 1、模型 2 的识别率，以验证数据库质量差异。

其中，在测试数据集中：对于模型 1、模型 2，集内测试集分别为数据库 1、King-ASR-216（随机选取 160 小时）扣除用以训练的 150 小时数据后剩余的 10 小时数据；集外测试集 1、2 均为与数据库 1、King-ASR-216（随机选取 160 小时）无关的开源数据集。

2、模型训练效果对比：

测试数据集	测试结果（字错误率） ^注
-------	-------------------------

² 数据库 1 说明：来源：<http://openslr.org/33/>。开源数据库 1 选取自行业通用的语音识别开源数据集网站 openslr.org，该网站由语音识别领域行业专家、知名语音识别开源软件 Kaldi 的主要开发者 Daniel Povey 运营；此处选取的开源数据库 1 为该网站上所提供的最大的中文语音识别开源数据集。

	模型 1 ^(注 2) (使用发行人数据库产品 King-ASR-216 训练数据训练所得)	模型 2 (使用开源数据库 1 ^(注 3) 训练所得)
集内测试集	11.4%	14.7%
集外测试集 1 ³	21.9%	30.3%
集外测试集 2 ⁴	21.0%	26.8%

注：字错误率=(插入错误字数+替换错误字数+删除错误字数)/的总字数*100%。字错误率是语音识别算法模型识别效果的反向提示指标，字错误率越低则语音识别模型的识别效果越好。

面对不同的测试数据（集内测试集、集外测试集 1、集外测试集 2），使用发行人数据库产品 King-ASR-216 训练所得的模型 1 均获得了明显更优的测试结果，体现发行人提供的语音数据库产品具备更优的产品质量，可提供更佳的语音识别算法模型训练效果，充分验证了发行人在设计优质训练数据库、优化算法模型训练效果方面的能力和优势。

(5) 以主要类型项目为例，说明发行人提供定制服务和数据库开发的业务流、数据流和财务流的流转情况。

1、发行人业务流、数据流和财务流的主要内容

(1) 主要业务流包括：签署合同/订单或确定数据库产品开发决策、数据库设计、数据采集、数据处理、质检、交付验收。

(2) 主要数据流包括：数据库设计环节生成设计文档，数据采集环节生成生数据，生数据或客户提供的数据经数据处理环节生成标签数据，再通过质检环节后形成可交付的成品数据。上述流程涉及对外采购数据服务的为：数据采集环节的生数据采集，以及数据处理环节的数据标记。

(3) 发行人通过业务管理系统中进行合同及订单管理、项目预算、采购支出申请、供应商对账、支出审批、项目验收和决算，并与财务系统衔接。相应地，在财务系统中，在项目支出发生时确认存货（定制服务）或研发费用（数据库开发），在项目验收时确认收入并结转成本（定制服务）。

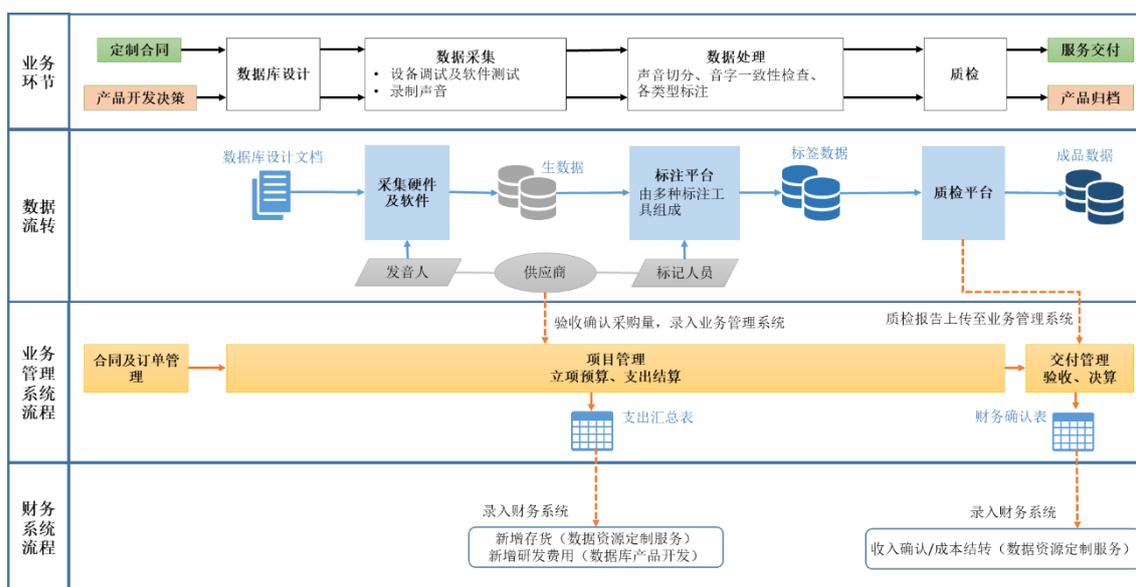
³ 集外测试集 1: <http://openslr.org/38/>，来源于行业通用的语音识别开源数据集网站 openslr.org 的中文开源语音识别数据集，时长 10 小时。

⁴ 集外测试集 2: <http://openslr.org/18/>，来源于行业通用的语音识别开源数据集网站 openslr.org 的中文开源语音识别数据集，时长 6.24 小时。

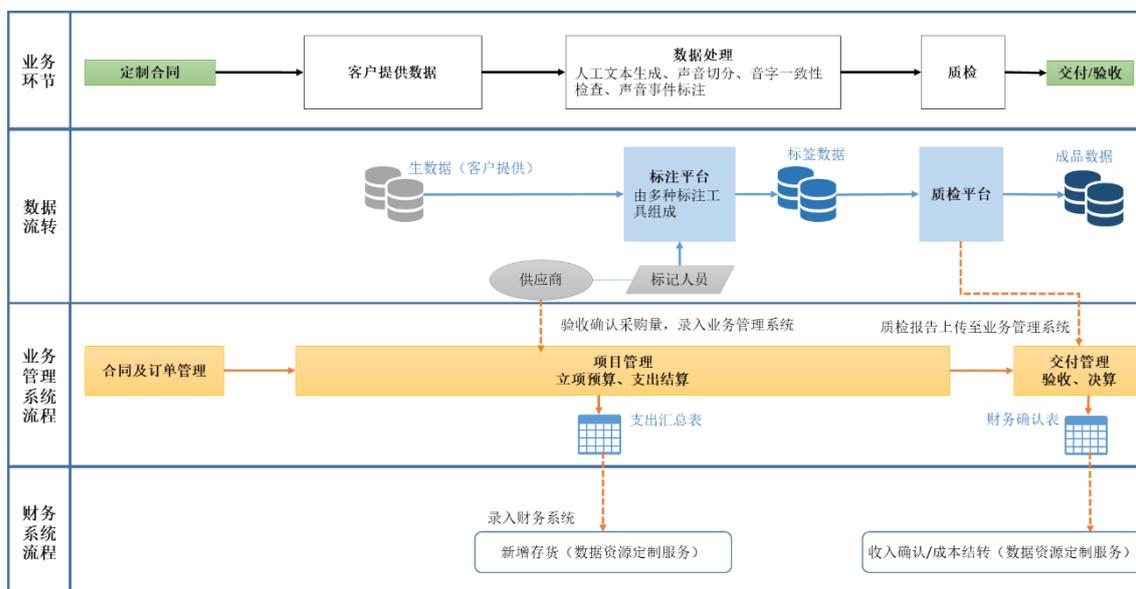
2、发行人业务流、数据流和财务流的流程图

以智能语音为例，对发行人提供定制服务和数据库开发的业务流、数据流和财务流的流转情况说明如下：

(1) 有采集的单一化标注项目、多样化标注项目：



(2) 无采集的单一化标注项目：



*注：部分项目由标记人员在客户提供的标注平台上完成。

以有采集的单一化标注项目为例，对发行人主要业务环节、数据流转、业务管理系统流程、财务流程，以列表形式进行如下进一步详细说明：

业务环节	数据流转	业务管理系统流程	财务流程
<p>对于定制服务：签订合同/获得订单</p> <p>对于数据库产品：确定产品开发决策</p>	<ul style="list-style-type: none"> ◆ 无 	<ul style="list-style-type: none"> ◆ 根据合同分解履约订单，一一对应建立项目，并将项目信息录入业务管理系统；对于数据库产品开发项目，根据产品开发决策建立项目并将项目信息录入业务管理系统 ◆ 核对履约订单与书面合同的核心内容一致性包括客户名称、销售量、销售单价等，确保录入业务管理系统内的项目信息真实、准确 ◆ 项目立项（履约订单与项目之间是一一对应关系，确保可按项目对收入、支出进行独立归集和核算） ◆ 项目预算申请及审核，为成本完整性、准确性的重要控制点（一般项目为三级审批制，重大项目为五级审批制）： <ul style="list-style-type: none"> ✓ 核对预计采集数据量、标记数据量^注，与履约订单的交付量之间进行核对，保证预算采购量的真实、完整、准确 ✓ 对比同类项目间的采集产出比、标注产出比，保证预计采集数据量、标记工作量的合理、完整 ✓ 对比业务采购价格体系，保证预算中采集、标记单价的合理性 ✓ 测算预算毛利率，复核项目盈利能力的合理性 	<ul style="list-style-type: none"> ◆ 无
数据库设计	<ul style="list-style-type: none"> ◆ 生成“数据库设计文档”，包括预设文本、发音人特征分布表（如年龄段、性别、口音区等），以及发音词典。 ◆ 预设文本是发音人的说话素材，每个发音人的预设文本均不同，设计多少个发音人，就有多少个预设文本 	<ul style="list-style-type: none"> ◆ 无 	<ul style="list-style-type: none"> ◆ 无
数据采集	<ul style="list-style-type: none"> ◆ 发音人根据设计的文本及场景，在发行人提供的设备及软件上进行录音，发音人的原始声音生成“生数据” ◆ 每个发音人对应一组声音文件 	<ul style="list-style-type: none"> ◆ 与供应商进行生数据采集与标记工作验收，确认生数据采集量、标记工作量^注及其价格 ◆ 把实际生数据采集量、标记工作量录入业务管理系统，进行支出申请 ◆ 系统根据支出申请生成验收结算单，并发送给供应商，供应商确认生数据采集量、标记工作量及对应价格和金额 	<ul style="list-style-type: none"> ◆ 核算财务岗根据“支出汇总表”每月将项目支出录入用友财务管理系统 ◆ 全部录入完毕后，运营财务岗将财务系统中记录的项目支出导出，与“支出汇总表”进行反向核对 ◆ 根据核对无误的结果，确
数据处理	<ul style="list-style-type: none"> ◆ 标记人员在发行人标注 		

	<p>工具上，对“生数据”进行检查和标记，生成“标签数据”</p> <ul style="list-style-type: none"> 对“标签数据”进行标记质量抽查 	<ul style="list-style-type: none"> 人工复核采集、标注量，确保系统中数据服务采购量的完整、准确，重点包括以下信息的交叉比对： <ul style="list-style-type: none"> 设计文档数量对比生数据采集量 生数据文件个数对比生数据采集量 标记数据量对比生数据采集量 生数据采集量、标记量对比供应商验收结算单记载数量 生数据采集量、标记量对比预算 采集产出比、标注产出比对比预算 采集、标记单价对比预算 支出申请审批通过后，系统生成“支出汇总表” 	<p>认存货或研发费用</p> <ul style="list-style-type: none"> 根据核对无误的结算金额向供应商付款
质检	<ul style="list-style-type: none"> 对“标签数据”进行全面检查，形成质检结论，使“标签数据”成为通过质检的、可交付的“成品数据” 	<ul style="list-style-type: none"> 提交质检报告至业务管理系统 	<ul style="list-style-type: none"> 无
交付/验收/决算	<ul style="list-style-type: none"> 向客户交付“成品数据”，客户对其进行验收确认 	<ul style="list-style-type: none"> 获取客户验收报告 将验收信息录入业务管理系统 提交项目决算，即结项申请 复核算收、决算信息，保证验收金额的真实、准确，项目支出的准确、完整。复核内容主要包括： <ul style="list-style-type: none"> 验收量对比验收报告 验收量对比履约订单中的销售量 验收单价对比履约订单中销售单价 对比预算，整体复核项目决算中覆盖的环节、各环节的数量是否完整 决算支出对比预算支出，偏离±10%需提交项目决算分析表 验收与决算审批通过后，系统生成“财务确认表”，生成收入、成本。 	<ul style="list-style-type: none"> 核算财务岗根据“财务确认表”将收入、成本信息录入财务管理系统。 全部录入完毕后，运营财务人员将财务系统中记录的收入、成本导出，与“财务确认表”进行反向核对 根据核对无误的结果，确认营业收入、营业成本

注：标记数据量指需要进行标记的数据量；标记工作量指进行标记所花费的工作小时。

请保荐机构核查并发表明确意见。

针对上述事项，保荐机构执行的核查程序如下：

1、保荐机构获取了发行人提交发明专利申请的相关资料及国家知识产权局中国专利公布公告（<http://epub.sipo.gov.cn/>）公示信息，对发行人管理团队及研发团队成员进行了访谈，了解发行人专利申请相关情况、其与核心技术的关系、核心技术的积累历程、技术保护措施及技术壁垒情况。

2、通过公开信息查询、客户访谈、行业专家访谈等手段核查发行人及其同行业主要竞争对手的市场地位、技术实力、关键业务数据、主要财务指标相关信息。

3、通过公开信息查询等方式查询同行业主要竞争对手、下游行业的数据库产品情况并与发行人数据库产品情况进行对比分析；获取发行人数据集与开源数据集对比测试结果。

4、访谈发行人业务、财务相关人员，了解发行人定制服务及数据库开发的具体过程；并登陆发行人的业务管理系统，核查其上述流程中业务流、数据流和财务流的具体流转情况。

为使投资者更好地理解发行人的业务模式及业务流、数据流、财务流的流转情况，发行人已在招股说明书“第六节 业务和技术\一、发行人主营业务、主要产品或服务的情况\（三）主要经营模式\2、生产或服务模式”补充披露上述本问询函答复 1（5）相关内容。

经核查，保荐机构认为：

1、发行人针对各项核心技术采取了综合性保护措施，已提交的 8 项发明专利申请所保护的对象仅为专利保护可覆盖的部分特定细分技术、技术研发成果，与“并不单纯依赖以申请专利的形式对核心技术进行保护”不存在矛盾。

2、发行人的 5 项核心技术具备较高技术壁垒、较难为同行业公司或上下游行业突破的依据主要为：上述技术需要投入大量时间、资源；发行人通过运用上述核心技术构建了产品/服务优势、获得了合作伙伴、行业机构的广泛认可。

3、与 Appen 相比，发行人在市场地位、体量规模、数据资源覆盖的语种/方言数量、收入利润规模等方面存在一定劣势，但数据库产品资源积累等业务结构的差异使发行人的毛利率高于 Appen；而与慧听科技、标贝科技相比，发行人在自有数据库积累、数据资源覆盖的语种/方言数量、计算机软件著作权数量、在申请专利数量等方面具备显著优势，除在数据资源种类上暂未涵盖音乐类数据资源外不存在明显劣势。

4、与竞争对手在公开信息渠道列举的数据库产品相比，发行人设计的数据

库产品的差异之处主要体现在：数据库数量更多、覆盖语种/方言数量更多、采集方案复杂度更高，表明发行人有能力设计覆盖多语种/方言、多场景、多领域的，采集方案更为复杂的数据库产品，体现了发行人在数据库设计领域的全面性、专业性优势。与下游客户自身的训练数据库相比，发行人在多场景、多领域、多语种/方言数据资源的全面性、综合性设计能力方面更有优势。与开源数据库相比，发行人的数据库产品可提供更佳的语音识别算法模型训练效果，充分验证了发行人在设计优质训练数据库、优化算法模型训练效果方面的能力和优势。

问题 2、关于共同控制

请发行人进一步说明：结合发行人最近2年内公司章程、协议或其他安排以及发行人历次股东大会（股东出席会议情况、表决权过程、审议结果、董事会提名和任命等）、董事会（重大决策的提议和表决过程等）、监事会及发行人经营管理的实际运作情况等，以列表形式说明发行人是否由贺琳、唐涤飞二人共同控制。

请保荐机构及发行人律师核查并发表明确意见。

答复：

（1）发行人最近 2 年内公司章程、协议或其他安排

根据发行人最近2年内公司章程规定，发行人股东（大）会、董事会审议事项不存在表决权的特殊安排，关于发行人实际控制权不存在特殊约定。公司章程规定，股东（大）会审议表决事项时，股东以其所代表的有表决权的出资金额/股份数额行使表决权，出席会议股东所持有效表决权过半数或三分之二以上同意方可通过表决事项；董事会审议表决事项时，一人一票，董事会作出决议必须经全体董事过半数通过。

发行人股东之间包括贺琳、唐涤飞之间不存在一致行动协议或共同控制的其他安排。

(2) 最近 2 年内发行人历次股东大会（股东出席会议情况、表决权过程、审议结果、董事会提名和任命等）、董事会（重大决策的提议和表决过程等）、监事会及发行人经营管理的实际运作情况

最近2年内，发行人公司治理结构健全且运行良好，发行人股东（大）会、董事会、监事或监事会及经营管理根据相关法律法规规定以及发行人《公司章程》等内部治理制度实际运作，关于发行人实际控制权并不存在特殊安排。最近2年内，贺琳直接及通过中瑞安控制发行人的股份比例在40%以上，实际控制发行人的股份数量远高于唐涤飞，所享有的表决权足以对股东大会的决议、人员的任免、重大事项提案的提出和审议通过产生重大影响。

1、股东（大）会

最近2年内，关于历次股东（大）会，全体股东包括贺琳、唐涤飞均亲自或授权代表出席会议并按其各自意愿投票表决，全体股东一致同意通过表决事项（关联股东回避表决除外）；关于股东（大）会审议议题，其中选举股东代表监事、监事会工作报告等由监事或监事会审议通过并提交股东（大）会审议，其他股东（大）会审议议题均由董事会审议通过并提交股东（大）会审议；股份公司董事由海天瑞声有限董事会或发行人董事会提名委员会提名，具体人选由贺琳提出，有限公司董事贺琳、唐涤飞、志鹏由贺琳提名。

会议届次	会议时间	主要审议议题	出席情况	表决情况
海天瑞声有限股东会	2017年4月20日	(1) 杭州银杏数、杭州士兰、天津金星、上海兴富对海天瑞声有限投资 (2) 中瑞立将部分出资转让给上海兴富、上海丰琬	全体股东出席	全体股东一致同意
海天瑞声有限股东会	2017年7月13日	(1) 梳理公司自2015年至今的分红事项 (2) 对公司截至2017年6月30日可供分配利润中的1,500万元进行分配	全体股东出席	全体股东一致同意
海天瑞声有限股东会	2017年11月9日	公司整体变更为股份有限公司	全体股东出席	全体股东一致同意
股份公司创立大会	2017年11月9日	(1) 设立北京海天瑞声科技股份有限公司并制定股份有限公司章程 (2) 选举股份有限公司第一届董事会董事 (3) 选举股份公司第一届监事会股东代表出任的监事 (4) 聘请审计机构	全体股东出席	全体股东一致同意

2017 年第一次临时股东大会	2017 年 12 月 15 日	(1) 制定《股东大会议事规则》等规范治理制度 (2) 第一届董事会董事、监事会监事报酬事项 (3) 设立董事会专门委员会	全体股东出席	全体股东一致同意
2018 年第一次临时股东大会	2018 年 2 月 8 日	(1) 2018 年财务预算 (2) 公司使用自有资金进行现金管理 (3) 变更公司法定代表人并修改公司章程	全体股东出席	全体股东一致同意
2017 年度股东大会	2018 年 6 月 12 日	(1) 2017 年度董事会工作报告 (2) 2017 年度监事会工作报告 (3) 2017 年度财务决算报告 (4) 2017 年度利润分配方案 (5) 聘任审计机构	全体股东出席	全体股东一致同意
2018 年第二次临时股东大会	2018 年 11 月 30 日	(1) 修改公司经营范围 (2) 更换公司独立董事	全体股东出席	全体股东一致同意
2019 年第一次临时股东大会	2019 年 2 月 13 日	(1) 2019 年度财务预算 (2) 使用自有资金进行现金管理	全体股东出席	全体股东一致同意
2019 年第二次临时股东大会	2019 年 3 月 19 日	(1) 修订公司章程、《股东大会议事规则》等规范治理制度 (2) 制定公司上市后适用的《股东大会议事规则》等规范治理制度	全体股东出席	全体股东一致同意
2019 年第三次临时股东大会	2019 年 3 月 22 日	(1) 申请首次公开发行人民币普通股股票并在科创板上市 (2) 公司上市后未来三年股东分红回报规划 (3) 制定《北京海天瑞声科技股份有限公司章程(草案)》	全体股东出席	全体股东一致同意
		(4) 确认公司报告期内关联交易事项		关联方贺琳、唐涤飞、中瑞安、中瑞立回避表决, 其余股东一致同意
2018 年年度股东大会	2019 年 4 月 20 日	(1) 2018 年度董事会工作报告 (2) 2018 年度监事会工作报告 (3) 2018 年度财务决算报告 (4) 2018 年度利润分配方案 (5) 聘任审计机构	全体股东出席	全体股东一致同意

2、董事会

最近2年内, 关于历次董事会, 全体董事包括贺琳、唐涤飞均亲自出席会议

并按其各自意愿投票表决，全体董事一致同意通过表决事项（关联董事回避表决除外）；董事会审议议题由董事长贺琳、总经理唐涤飞、董事会专门委员会根据其权限提出，涉及重大决策的由董事长贺琳提出；总经理、董事会秘书由董事长贺琳提名，其他高级管理人员由总经理唐涤飞提名。

会议届次	会议时间	主要审议议题	提案主体	出席情况	表决情况
2017 年第 1 次董事会	2017 年 4 月 17 日	(1) 杭州银杏数、杭州士兰、天津金星、上海兴富对海天瑞声有限投资 (2) 中瑞立将部分出资转让给上海兴富、上海丰琬	董事长贺琳	全体董事出席	全体董事一致同意
2017 年第 2 次董事会	2017 年 5 月 25 日	2016 年年度财务报告	总经理唐涤飞	全体董事出席	全体董事一致同意
2017 年第 3 次董事会	2017 年 7 月 13 日	(1) 梳理公司自 2015 年至今的分红事项 (2) 对公司截至 2017 年 6 月 30 日可供分配利润中的 1,500 万元进行分配	董事长贺琳	全体董事出席	全体董事一致同意
2017 年第 4 次董事会	2017 年 11 月 6 日	(1) 公司整体变更设立股份有限公司 (2) 制定股份有限公司章程 (3) 选举股份有限公司第一届董事会董事 (4) 聘请审计机构 (5) 召开股东会会议及股份有限公司创立大会会议	董事长贺琳	全体董事出席	全体董事一致同意
第一届董事会第一次会议	2017 年 11 月 9 日	(1) 选举贺琳为公司第一届董事会董事长 (2) 聘任唐涤飞为公司总经理 (3) 聘任吕思遥为公司董事会秘书	董事长贺琳	全体董事出席	全体董事一致同意
		(4) 聘任李科、程显峰为公司副总经理 (5) 聘任吕思遥为公司财务负责人 (6) 聘任郝玉峰为技术总监	总经理唐涤飞		
第一届董事会第二次会议	2017 年 11 月 30 日	(1) 制定《股东大会议事规则》等规范治理制度 (2) 第一届董事会董事、监事会监事的报酬事项 (3) 第一届高级管理人员报酬事项 (4) 聘任审计部负责人 (5) 设立董事会专门委员会 (6) 召开 2017 年第一次临时股东大会会议	董事长贺琳	全体董事出席	全体董事一致同意
第一届董事会第三次会议	2017 年 12 月 18 日	(1) 董事会专门委员会议事规则 (2) 董事会专门委员会组成人员	董事长贺琳	全体董事出席	全体董事一致同意
第一届董事会第四次会议	2018 年 1 月 23 日	(1) 2018 年财务预算	总经理唐涤飞	全体董事出席	全体董事一致同意
		(2) 公司使用自有资金进行现金管理	总经理唐涤飞		

会议届次	会议时间	主要审议议题	提案主体	出席情况	表决情况
次会议		(3) 聘任公司副总经理	提名委员会（总经理唐涤飞提名具体人选）		
		(4) 《高级管理人员薪酬与考核管理办法》	薪酬与考核委员会		
		(5) 变更公司法定代表人并修改公司章程	董事长贺琳		
		(6) 召开 2018 年第一次临时股东大会	董事长贺琳		
第一届董事会第五次会议	2018 年 3 月 9 日	向香港海天瑞声增资	战略委员会（董事长贺琳提出具体方案）	全体董事出席	全体董事一致同意
第一届董事会第六次会议	2018 年 4 月 13 日	2017 年度母公司单体审计报告	审计委员会	全体董事出席	全体董事一致同意
第一届董事会第七次会议	2018 年 5 月 22 日	(1) 2017 年度董事会工作报告	董事长贺琳	全体董事出席	全体董事一致同意
		(2) 2017 年度总经理工作报告	总经理唐涤飞		
		(3) 2017 年度财务决算报告	审计委员会		
		(4) 2017 年度利润分配方案	董事长贺琳		
		(5) 2017 年度合并审计报告	审计委员会		
		(6) 2017 年度内部控制评价报告	审计委员会		
		(7) 聘任审计机构会计师事务所	审计委员会		
		(8) 召开 2017 年年度股东大会	董事长贺琳		
第一届董事会第八次会议	2018 年 8 月 15 日	公司 2018 年半年度财务报告	审计委员会	全体董事出席	全体董事一致同意
第一届董事会第九次会议	2018 年 11 月 15 日	(1) 修改公司经营范围	董事长贺琳	全体董事出席	全体董事一致同意
		(2) 更换公司独立董事	提名委员会（董事长贺琳提名具体人选）		
		(3) 召开 2018 年第二次临时股东大会	董事长贺琳		
第一届董事会第十次会议	2019 年 1 月 28 日	(1) 2019 年财务预算	总经理唐涤飞	全体董事出席	全体董事一致同意
		(2) 聘任公司首次公开发行股票并上市保荐机构	董事长贺琳		
		(3) 聘任公司技术总监	提名委员会（总经理唐涤飞提名具体人选）		
		(4) 使用自有资金进行现金管理	总经理唐涤飞		
		(5) 召开 2019 年第一次临时股东大会	董事长贺琳		

会议届次	会议时间	主要审议议题	提案主体	出席情况	表决情况
第一届董事会第十一次会议	2019年3月4日	(1) 修订公司章程、《股东大会议事规则》等规范治理制度 (2) 制定公司上市后适用的《股东大会议事规则》等规范治理制度 (3) 召开2019年第二次临时股东大会	董事长贺琳	全体董事出席	全体董事一致同意
第一届董事会第十二次会议	2019年3月6日	(1) 申请首次公开发行人民币普通股股票并在科创板上市	战略委员会（董事长贺琳提出具体方案）	全体董事出席	全体董事一致同意
		(2) 公司上市后未来三年股东分红回报规划	董事长贺琳		全体董事一致同意
		(3) 制定《北京海天瑞声科技股份有限公司章程（草案）》	董事长贺琳		全体董事一致同意
		(4) 确认公司报告期内关联交易事项	董事长贺琳		关联董事贺琳、唐涤飞回避表决，其他董事一致同意
		(5) 召开2019年第三次临时股东大会	董事长贺琳		全体董事一致同意
第一届董事会第十三次会议	2019年3月18日	2018年度母公司单体审计报告	审计委员会	全体董事出席	全体董事一致同意
第一届董事会第十四次会议	2019年3月31日	(1) 2018年度董事会工作报告	董事长贺琳	全体董事出席	全体董事一致同意
		(2) 2018年度总经理工作报告	总经理唐涤飞		
		(3) 2018年度财务决算报告	总经理唐涤飞		
		(4) 2018年度利润分配方案	董事长贺琳		
		(5) 批准报出2016-2018年度审计报告	审计委员会		
		(6) 2018年度内部控制评价报告	审计委员会		
		(7) 聘任审计机构	审计委员会		
		(8) 召开2018年年度股东大会	董事长贺琳		

3、监事会

最近2年内，贺琳、唐涤飞并未担任公司监事，关于历次监事会，全体监事均亲自出席会议并按其各自意愿投票表决，全体监事一致同意通过表决事项。

4、经营管理的实际运作情况

最近2年内，发行人经营管理的实际运作情况为：

（1）贺琳担任公司董事长，具体负责公司战略规划制订、重大决策研究、重要外部关系管理、融资管理、公司治理机制完善等重大事项，并对高级管理人员的履职情况进行监督，对重大经营管理事项提出意见。除了上述贺琳向董事会提出的提案以外，关于发行人重大经营管理事项，包括发行人2016年组建美国销售团队、2017年加大计算机视觉业务发展力度等，均由贺琳主导决策。

（2）唐涤飞担任公司总经理，主要负责公司日常经营管理事项，包括日常业务运营、项目管理、人员管理等。发行人管理团队，除了总经理唐涤飞之外，发行人副总经理李科主要分管业务开拓和产品开发，技术总监郝玉峰主要分管语言拓展和数据质检，技术总监黄宇凯主要分管技术研发，副总经理、财务负责人、董事会秘书吕思遥主要分管财务管理、公共关系管理及证券事务。上述高级管理人员分工合作，共同管理公司日常经营管理事务。

（3）贺琳为发行人创始股东，在发行人公司创立、业务和技术形成、发展上起到重大作用。在公司发展过程中，鉴于贺琳所持发行人的股权比例及贺琳作为创始股东、董事长在公司的重要地位，贺琳对公司业务和技术发展方向、重大事项决策判断等一直具有重要影响力。具体内容见第二轮审核问询函回复问题2（1）所述。

请保荐机构及发行人律师核查并发表明确意见。

针对上述事项，保荐机构及发行人律师执行的核查程序如下：

- 1、审阅发行人最近2年内公司章程和内部治理制度、历次股东（大）会、董事会、监事会相关会议资料；
- 2、对贺琳、唐涤飞就其会议提案及表决、职责分工等事项进行访谈；
- 3、获取发行人股东包括贺琳、唐涤飞关于不存在一致行动协议或安排的确认文件；
- 4、审阅发行人的工商资料，核查设立以来的董事、高级管理人员及其变更情况，核查贺琳、唐涤飞持有股权的相关情况。

经核查，保荐机构和发行人律师认为：

1、最近2年内发行人公司章程对股东大会、董事会审议事项不存在表决权的特殊安排，关于发行人实际控制权不存在特殊约定；发行人股东之间包括贺琳、唐涤飞之间不存在一致行动协议或共同控制的其他安排。

2、最近2年内，发行人公司治理结构健全且运行良好，贺琳实际控制的股份所享有的表决权足以对股东大会的决议、人员的任免、重大事项提案的提出和审议通过产生重大影响。关于历次股东（大）会，全体股东包括贺琳、唐涤飞均亲自或授权代表出席会议并按其各自意愿投票表决，全体股东一致同意通过表决事项（关联股东回避表决除外）；关于股东（大）会审议议题，其中选举股东代表监事、监事会工作报告等由监事或监事会审议通过并提交股东（大）会审议，其他股东（大）会审议议题均由董事会审议通过并提交股东（大）会审议；股份公司董事由海天瑞声有限董事会或发行人董事会提名委员会提名，具体人选由贺琳提出，有限公司董事贺琳、唐涤飞、志鹏由贺琳提名。关于历次董事会，全体董事包括贺琳、唐涤飞均亲自出席会议并按其各自意愿投票表决，全体董事一致同意通过表决事项（关联董事回避表决除外），董事会审议议题均由董事长贺琳、总经理唐涤飞、董事会专门委员会根据其权限提出，涉及重大决策的由董事长贺琳提出；总经理、董事会秘书由董事长贺琳提名，其他高级管理人员由总经理唐涤飞提名。关于历次监事会，全体监事均亲自出席会议并按其各自意愿投票表决，全体监事一致同意通过表决事项。

3、贺琳、唐涤飞按照分工对公司进行管理，贺琳作为董事长对高级管理人员的履职情况进行监督，对经营管理重大事项提出意见。贺琳作为创始股东在发行人公司创立、业务和技术形成、发展上起到重大作用，贺琳对公司业务和技术发展方向、重大事项决策判断等一直具有重要影响力。

因此，发行人并非由贺琳、唐涤飞二人共同控制，发行人实际控制人为贺琳。

问题 3、关于发行人股份支付

回复材料显示：（1）张蕊因职务调整为审计部经理，2016年度张蕊向创世

联合转让13.3333万元实缴出资额，占中瑞安全部出资额的8.89%，对应海天瑞声有限的出资比例为1.60%，对应的海天瑞声有限出资金额为9.6万股，转让后张蕊持有中瑞安全部出资额的5.00%；(2)唐涤飞原持有中瑞立21.05%的实缴出资额，2016年3月唐涤飞向创慧科瑞转让上述实缴出资额，转让时创慧科瑞的股东分别为唐涤飞和贺琳，2016年12月贺琳将其持有创慧科瑞的股权转让给志鹏。

请发行人进一步说明：(1)张蕊向创世联合转让实缴出资额的原因及合理性，未在2015年度按照13.89%的出资比例计算股份支付费用的原因及依据，是否符合企业会计准则的相关规定；(2)唐涤飞转让中瑞立实缴出资额、贺琳转让创慧科瑞股权一系列交易安排的原因及合理性；(3)中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况。

请保荐机构和申报会计师核查并发表明确意见。

答复：

(1)张蕊向创世联合转让实缴出资额的原因及合理性，未在2015年度按照13.89%的出资比例计算股份支付费用的原因及依据，是否符合企业会计准则的相关规定；

一、张蕊向创世联合转让实缴出资额的原因及合理性

发行人在2015年8月开始筹划股权激励事宜时，综合考虑了激励对象的入职年限、岗位分工、贡献程度等因素来确定核心员工激励份额。张蕊加入海天瑞声时间较早，历任海天瑞声财务经理、综合部总监等职务。截至2015年末，公司财务负责人岗位处于空缺状态。因此，在设计持股平台的股东结构及持股比例时，综合考虑张蕊的入职年限、岗位分工、贡献程度，以及未来拟担任财务负责人职务的情形，决定授予张蕊认缴中瑞安138.8889万元，占平台内出资比例为13.89%，对应海天瑞声有限的出资比例为2.50%。张蕊实缴出资20.8333万元。

2016年8月，随着新投资人的进入以及公司业务快速发展，发行人的上市准备工作也开始提上日程，各项工作对财务负责人的能力提出了更高的要求。考虑到张蕊在拟上市公司的会计核算体系建设、内部控制体系搭建等方面缺乏一定

的经验，发行人董事会决定重新选聘请财务负责人，并对张蕊的职务进行调整，同时同步调整了向张蕊授予的股权。实施方案上，由张蕊先行转让给普通合伙人创世联合，后续授予明确的激励对象时再由创始联合转出。

经张蕊本人同意，其于 2016 年 8 月向创世联合转让 88.8889 万元的出资（其中实缴 13.3333 万元），占中瑞安全部出资额的 8.89%，对应海天瑞声有限的出资比例为 1.60%。

综上，张蕊向创世联合转让实缴出资额是在发行人公司业务发展及完善公司内部治理的背景下，基于张蕊岗位调整等因素综合考虑而进行的。经公司、张蕊协商一致，对张蕊授予股权激励份额调整为向创世联合认缴 50 万元出资额（其中实缴 7.5 万元），占中瑞安全部出资额的 5%，对应海天瑞声有限的出资比例为 0.9%。综合考虑所有激励对象的入职年限、岗位分工、贡献程度等因素，对张蕊授予股权激励份额进行调整以及调整后的授予份额具备合理性。

二、2015 年未按 13.89% 的出资比例计算张蕊股份支付费用的原因及依据

发行人在 2016 年进行以前年度财务核算进行规范时，对 2015 年的股份支付事项进行了追溯确认。

鉴于在股份支付事项追溯确认时，相关管理人员的职责已经明确到位，持股平台的股权结构已经调整到位，因此，发行人按照调整到位的出资金额确认 2015 年股份支付费用，对张蕊按照持有中瑞安 5% 的出资份额（对应海天瑞声有限的出资比例为 0.90%）确认股份支付费用。

综上，发行人于 2016 年及时对激励方案进行了调整并整体追溯确认、调整了 2015 年末的股权激励，因此，从调整后的报表来看，2015 年未按持有中瑞安 13.89% 的出资比例计算张蕊股份支付费用。对张蕊确认的股份支付金额与其向发行人提供的服务相匹配，符合企业会计准则的相关规定。

三、假设 2015 年按 13.89% 的出资比例模拟测算张蕊股份支付费用的影响

2015 年 12 月，发行人授予张蕊认缴中瑞安 138.8889 万元，占持股平台中瑞安的出资比例为 13.89%，对应海天瑞声有限的出资比例为 2.50%，对应发行人的出资金额 15 万元。经张蕊同意，其于 2016 年 8 月向创世联合转让 88.8889 万

元的出资（其中实缴 13.3333 万元），占中瑞安全部出资额的 8.89%，对应海天瑞声有限的出资比例为 1.60%，对应发行人的出资金额 9.6 万元。

2015 年按 13.89% 的出资比例模拟测算张蕊股份支付费用的影响如下：

1、模拟测算股份支付费用计算表：

计算公式	项目	金额（元）
①	转让注册资本	96,000.00
②	每 1 元注册资本对应的公允价值	22.18
③	受让成本/每 1 元注册资本	1.00
④=①×（②-③）	测算确认股份支付费用	2,033,280.00

2、对发行人当期财务报表影响：

影响科目	2015-12-31/ 2015 年	影响金额 （万元）	2015-12-31/2015 年 （模拟确认后）
资本公积	2,427.23	203.33	2,630.56
盈余公积	72.44		72.44
未分配利润	-789.00	-203.33	-992.33
管理费用	3,278.81	203.33	3,482.13
净利润	-1,225.14	-203.33	-1,428.47
扣除股份支付后 净利润	1,202.09		1,202.09

根据上述测算可得，若按张蕊最初持有中瑞安 13.89% 的出资比例计算股份支付费用，则将增加发行人 2015 年管理费用 203.33 万元，减少 2015 年净利润 203.33 万元，但不会对 2015 年末的净资产造成影响。

四、小结

综上，张蕊向创世联合转让实缴出资额是在发行人公司业务发展及完善公司内部治理的背景下，基于张蕊岗位调整等因素综合考虑进行的；前述激励份额调整考虑了激励对象的入职年限、岗位分工、贡献程度等因素，具备合理性。未在 2015 年度按照 13.89% 的出资比例计算股份支付费用具备合理性，符合企业会计准则的相关规定。假设 2015 年度按照 13.89% 的出资比例计算张蕊的股份支付费用，仅影响报告期期初的所有者权益结构，对报告期内利润表项目不构成影响。

（2）唐涤飞转让中瑞立实缴出资额、贺琳转让创慧科瑞股权一系列交易安排的原因及合理性

一、唐涤飞转让中瑞立实缴出资额的背景

中瑞立系作为员工持股平台而设立。设立之初，拟以贺琳、唐涤飞共同投资设立创慧科瑞作为中瑞立的 GP。2015 年 12 月 18 日，中瑞立设立时，创慧科瑞尚未办理完毕工商登记手续，贺琳、唐涤飞经协商后决定暂以唐涤飞作为 GP，待创慧科瑞设立后，唐涤飞再将相应的出资额转让给创慧科瑞。中瑞立设立时，唐涤飞以货币认缴出资 210.5263 万元并实际缴纳 21.0526 万元，占中瑞立出资总额的 21.05%，占穿透后海天瑞声股权比例为 2%。2015 年 12 月 23 日，创慧科瑞设立时，贺琳、唐涤飞分别持有其 50%、50%的股权。2016 年 3 月，唐涤飞向创慧科瑞转让上述 210.5263 万元的出资（其中实缴 21.0526 万元出资）。

二、贺琳向志鹏转让创慧科瑞股权的背景

贺琳与志鹏于 2015 年 12 月 15 日正式签署《顾问服务协议》，约定贺琳聘请志鹏作为顾问协助其向投资人转让海天瑞声有限股权，贺琳向志鹏转让海天瑞声有限 1%股权作为服务费用。

经贺琳与志鹏协商一致，2016 年 12 月贺琳、志鹏签署《转让协议》，贺琳将所持创慧科瑞 50%股权（占穿透后海天瑞声的股权比例为 1%）转让给志鹏。

三、小结

综上，唐涤飞转让中瑞立实缴出资额、贺琳转让创慧科瑞股权等一系列交易安排符合实际情况，具有合理性。上述股权转让价款均已结清，不存在争议或纠纷。

（3）中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况

一、中瑞立历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况

中瑞立于 2015 年 12 月作为员工持股平台而设立，除设立时各合伙人合计实

缴 100 万元出资外，不存在其他出资情形。自设立以来，中瑞立存在两次股权转让情形。中瑞立出资或股权转让的价格、定价依据、资金来源和支付情况、完税具体情况如下：

事项	出资或股权转让情况		出资/股权转让价格 (万元)	定价依据	资金来源	出资/股权转让 价款支付情况	完税情况	
	合伙人	认缴/实缴出资 (万元)						
2015 年 12 月，设立	贺琳 (LP)	421.0526/42.1053	42.1053	出资原值	自有资金	已实缴部分出资	不涉及	
	唐涤飞 (GP)	210.5263/21.0526	21.0526					
	黄大伟 (LP)	115.7895/11.5789	11.5789					
	宋琼 (LP)	94.7368/9.4737	9.4737					
	廖晓玲 (LP)	94.7368/9.4737	9.4737					
	辛晓峰 (LP)	31.5790/3.1579	3.1579					
	葛兴彤 (LP)	31.5790/3.1579	3.1579					
	小计	1,000/100	100					—
2016 年 3 月，第一次股权转让	转让方	受让方	转让的认缴/实缴出资额 (万元)	出资/股权转让价格 (万元)	定价依据	资金来源	出资/股权转让 价款支付情况	完税情况
	唐涤飞	创慧科瑞	210.5263/21.0526	21.0526	出资原值	股东贺琳、唐涤飞以自有资金向创慧科瑞提供借款	已支付完毕	不涉及
	贺琳	陈琪	94.7368/9.4737	9.4737	出资原值	自有资金	已支付完毕	不涉及
2017 年 6 月，第二次股权转让 ^注	贺琳	创慧科瑞	83.5914/8.3591	—	—	—	—	不涉及
		黄大伟	45.9752/4.5975					
		宋琼	37.6161/3.7616					
		廖晓玲	37.6161/3.7616					
		陈琪	37.6161/3.7616					
		辛晓峰	12.5387/1.2539					
		葛星彤	12.5387/1.2539					

注：2017 年 6 月，第二次股权转让的具体情况为：中瑞立将所持海天瑞声有限部分股权出资转让给上海兴富、上海丰琬，全体合伙人同意将股权转让价款全部分配给贺琳，并相应减少贺琳对中瑞立的出资份额。为使除贺琳外的其他合伙人通过中瑞立间接持有海天瑞声有限的股权保持不变，贺琳将所持部分中瑞立出资转让给其他合伙人，不需支付对价。

二、中瑞安历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况

中瑞安于 2015 年 12 月作为员工持股平台而设立，除设立时各合伙人合计实缴 150 万元出资外，不存在其他出资情形。自设立以来，中瑞安存在三次股权转让情形。中瑞安出资或股权转让的价格、定价依据、资金来源和支付情况、完税具体情况如下：

事项	出资或股权转让情况		出资/股权转让价格 (万元)	定价 依据	资金来源	出资/股权转让价款 支付情况	完税 情况	
	合伙人	认缴/实缴出资 (万元)						
2015 年 12 月， 设立	程显峰 (LP)	361.1111/54.1667	54.1667	出资 原值	自有资金	已实缴部 分出资	不涉 及	
	李科 (LP)	277.7778/41.6667	41.6667					
	郝玉峰 (LP)	194.4444/29.1667	29.1667					
	张蕊 (LP)	138.8889/20.8333	20.8333					
	贺琳 (GP)	27.7778/4.1667	4.1667					
	小计	1,000/150	150	—	—	—	—	
2016 年 3 月，第 一次股 权转让	转让方	受让方	转让的认缴/实 缴出资额 (万元)	出资/股权 转让价格	定价 依据	资金来源	出资/股权 转让价款 支付情况	完税 情况
	贺琳	创世联合	27.7778/4.1667	4.1667	出资 原值	创世联合自有 资金及股东贺 琳以自有资金 提供的借款	已支付完 毕	不涉 及
2016 年 8 月，第 二次股 权转让	程显峰	吕思遥	72.2222/10.8333	10.8333	出资 原值	自有资金	已支付完 毕	不涉 及
	创世 联合	吕思遥	11.1111/1.6667	1.6667	出资 原值	自有资金	已支付完 毕	不涉 及
	创世 联合	李科	27.7778/4.1667	4.1667	出资 原值	自有资金	已支付完 毕	不涉 及
	张蕊	创世联合	88.8889/13.3333	13.3333	出资 原值	创世联合自有 资金及股东贺 琳以自有资金 提供的借款	已支付完 毕	不涉 及
2019 年 1 月，第 三次股 权转让	程显峰	创世联合	288.8889/43.3334	43.3334	出资 原值	股东贺琳以自 有资金向创世 联合提供借款	已支付完 毕	不涉 及

三、创世联合历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况

创世联合于 2015 年 12 月拟作为中瑞安的普通合伙人而设立，设立时注册资本为 1 万元，2018 年 1 月增加至 30 万元。自设立以来，创世联合不存在股权转让情形。创世联合出资的价格、定价依据、资金来源和支付情况、完税具体情况

如下：

事项	出资或股权转让情况		出资价格 (万元)	定价依据	资金来源	出资/股权转让 价款支付情况	完税 情况
	股东	出资金额 (万元)					
2015年12月，设立	贺琳	1	1	出资原值	自有资金	已实缴全部出资	不涉 及
2018年1月，增资	贺琳	29	29	出资原值	自有资金	已实缴全部出资	不涉 及
小计		30	30	—	—	—	—

四、创慧科瑞历次出资或股权转让的价格、定价依据、资金来源和支付情况、完税情况

创慧科瑞于2015年12月拟作为中瑞立的普通合伙人而设立，设立时注册资本为1万元，2018年6月增加至40万元。自设立以来，创慧科瑞存在一次股权转让情形。创慧科瑞出资的价格、定价依据、资金来源和支付情况、完税具体情况如下：

事项	出资或股权转让情况		出资价格 (万元)	定价依据	资金 来源	出资/股权转 让价款支付 情况	完税 情况	
	股东	出资金额 (万元)						
2015年12月， 设立	贺琳	0.5	0.5	出资原值	自有 资金	已实缴全部 出资	不涉 及	
	唐涤飞	0.5	0.5					
2018年6月， 增资	唐涤飞	19.5	19.5	出资原值	自有 资金	已实缴全部 出资	不涉 及	
	志鹏	19.5	19.5					
小计		40	40	—	—	—	—	
2017年1月， 股权转让	转让 方	受 让 方	转让的出 资 金 额 (万元)	出 资 / 股 权 转 让 价 格	定 价 依 据	资 金 来 源	出 资 / 股 权 转 让 价 款 支 付 情 况	完 税 情 况
	贺琳	志鹏	0.5	0.5	出资原值	自有 资金	已支付完毕	不涉 及

五、小结

1、除中瑞立2017年6月第二次股权转让以外，中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让均按照出资原值确定价格；中瑞立2017年6月第二次股权转让系因中瑞立将所持海天瑞声有限的股权转让价款全部分配给贺琳，为使除贺琳外的其他合伙人通过中瑞立间接持有海天瑞声有限的股权保持不变，贺琳将所持部分中瑞立出资转让给其他合伙人，不需支付对价。

2、中瑞立、中瑞安、创世联合和创慧科瑞历次实缴出资已缴纳完毕，资金来源为合伙人/股东的自有资金；历次股权转让价款已支付完毕，资金来源为受让方自有资金或受让方股东以自有资金提供的借款；

3、中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让不涉及缴纳个人所得税。

请保荐机构和申报会计师核查并发表明确意见。

针对上述事项，保荐机构及申报会计师执行的核查程序如下：

1、对贺琳、张蕊进行访谈，了解股权激励份额调整的原因；审阅中瑞安关于该次出资变更的工商登记档案，获取张蕊向创世联合转让出资份额的协议及价款支付凭证；获取张蕊的个人履历、及其出具的关于持有发行人股份真实性的确认文件。

2、审阅中瑞立、创慧科瑞的工商登记档案资料；获取唐涤飞向创慧科瑞转让中瑞立出资额相应价款支付凭证；获取贺琳与志鹏签署的《顾问服务协议》以及贺琳向志鹏转让创慧科瑞股权转让价款支付凭证；核查确认中瑞立受让海天瑞声股权及股份支付确认情况；分别对贺琳、唐涤飞、志鹏进行访谈，确认上述交易安排的背景；获取贺琳、唐涤飞、志鹏关于持有发行人股份真实性的确认文件。

3、审阅中瑞立、中瑞安、创世联合和创慧科瑞的工商登记档案资料；获取前述机构的实际缴纳出资凭证，审阅其财务报表。获取中瑞立、中瑞安、创慧科瑞的出资/股权转让相关的价款支付凭证，核实个人所得税缴纳事项，并与相关方就转让原因、价款支付情况进行访谈。对贺琳、唐涤飞进行访谈，获取股东借款相关往来凭证，核查创世联合、创慧科瑞设立及受让出资的资金来源。获取中瑞安、中瑞立合伙人及创世联合、创慧科瑞股东出具的关于持有出资/股权真实性的确认文件。

经核查，保荐机构及申报会计师认为：

1、张蕊向创世联合转让实缴出资额是在发行人公司业务发展及完善公司内部治理的背景下，基于张蕊岗位调整等因素综合考虑进行的；前述激励份额调整考虑了激励对象的入职年限、岗位分工、贡献程度等因素，具备合理性。未在 2015

年度按照 13.89%的出资比例计算股份支付费用具备合理性，符合企业会计准则的相关规定。假设 2015 年度按照 13.89%的出资比例计算张蕊的股份支付费用，仅影响报告期期初的所有者权益结构，对报告期内利润表项目不构成影响。

2、综上，唐涤飞转让中瑞立实缴出资额、贺琳转让创慧科瑞股权等一系列交易安排符合实际情况，具有合理性。

3、（1）除中瑞立 2017 年 6 月第二次股权转让以外，中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让均按照出资原值确定价格；中瑞立 2017 年 6 月第二次股权转让系因中瑞立将所持海天瑞声有限的股权转让价款全部分配给贺琳，为使除贺琳外的其他合伙人通过中瑞立间接持有海天瑞声有限的股权保持不变，贺琳将所持部分中瑞立出资转让给其他合伙人，不需支付对价。（2）中瑞立、中瑞安、创世联合和创慧科瑞历次实缴出资已缴纳完毕，资金来源于合伙人/股东的自有资金；历次股权转让价款已支付完毕，资金来源于受让方自有资金或受让方股东以自有资金提供的借款；（3）中瑞立、中瑞安、创世联合和创慧科瑞历次出资或股权转让不涉及缴纳个人所得税。

问题 4、关于发行人采购成本

回复材料显示：（1）发行人被采集人关系网络或发行人员工的人际关系网络寻找合适的境外数据被采集人，采集完成后通过 Speech Ocean Ltd.向境外数据被采集人支付数据服务费用；（2）对于境内数据服务提供商，根据组织和管理被采集人员、标记人员的数量规模和管理难度，发行人与供应商分别约定采用 10-11.5% 的服务费率，采购定价具有公允性和合理性，处于境内同类公司的市场报价区间；（3）北京零点、北京天创共盈科技发展有限公司、北京中惠商客科技有限公司由北京博睿、恩泽尔实际控制人王驰受托统一进行经营管理，信阳市晨曦电子科技有限公司由信阳联智实际控制人洪琪受托统一进行经营管理；（4）发行人通过一体化人工智能数据处理技术支持平台对标注服务实施管理。

请发行人：（1）说明报告期内发行人寻找境外数据采集人的方式及供应商的具体名称、采购金额和被采集人的具体情况，该等的供应商的服务费率与

SpeechOcean 的毛利率的比较情况及差异原因；(2) 发行人报告期各期前五大劳务服务企业的服务费率及变动情况，结合采购额变化、人员数量规模和各采集项目的具体情况，进一步说明发行人如何确定相关的服务费率，发行人如何量化劳务服务企业的管理难度并反映在服务费率中；(3) 进一步说明报告期内是否出现返工的情况，发行人与劳务服务企业的服务费用调整机制，因返工而导致的服务费用调整金额、原因和依据；(4) 发行人向同一人控制或管理的多家劳务服务企业进行采购的原因及合理性，发行人未直接向王驰和洪琪直接持股的企业进行采购劳务服务的原因及合理性。

请保荐机构和申报会计师提供发行人主要的业务信息系统的情况，对发行人一体化人工智能数据处理技术支持平台等核心业务系统进行信息系统专项核查并出具专项核查报告，对系统控制、运营日志和运营数据的真实性、准确性和完整性进行核查、对财务数据和业务数据的衔接和是否存在运营数据缺失或被违规修改的情形进行核查，并对与信息系统相关的内部控制有效性发表明确意见。

请保荐机构和申报会计师对上述事项进行核查并发表明确意见。

答复：

(1) 说明报告期内发行人寻找境外数据采集人的方式及供应商的具体名称、采购金额和被采集人的具体情况，该等的供应商的服务费率与 SpeechOcean 的毛利率的比较情况及差异原因

报告期内，初期发行人通过被采集人关系网络或发行人员工的人际关系网络寻找境外数据被采集人。随着业务规模增长，发行人 2016 年即已开始逐步转变为由具备资源优势的人力资源服务企业或劳务服务企业等供应商负责被采集人的匹配和组织，其通过境外华人群体资源为发行人提供境外生数据采集服务。该等境内外供应商数量较多，共计 62 个，其中主要涉及的供应商为除信阳联智、睿祥文义以外的发行人报告期各期前五大数据服务供应商，即广州善世、芜湖善慧、北京博睿、北京零点、北京冠华等境内供应商和恩泽尔、TWIN STUDIOS、THINKWILD STUDIOS, S.L.、SAKURA JAPAN LLC、CHUSHOUKIGYOU, CONSULTING, CENTER., LTD.等境外供应商。报告期内，为发行人寻找境外数据被采集人的供应商涉及的采购金额和采集人数情况如下：

年度	采购金额（万元）	采集人数（人）
2018 年度	1,364.55	13,506
2017 年度	726.97	9,583
2016 年度	652.33	12,550

报告期内，发行人境外数据被采集人的采购金额逐年上升，是由于发行人外语种业务快速发展，同时下游客户对本土采集的要求也逐步提高，因此发行人境外采集的规模也逐年提升；境外数据被采集人的单价也呈现上升趋势，主要是多样化标注语音等单个被采集人价格较高的外语种业务增长所致。

考虑到境外数据被采集人的稀缺度、寻找难度以及不同国家地区的商业环境等实际情况，上述主要供应商为发行人寻找境外数据被采集人的服务费率在 6-11.5% 之间。SpeechOcean 的毛利率为 9.09%，处于前述范围内，不存在明显差异。

(2) 发行人报告期各期前五大劳务服务企业的服务费率及变动情况，结合采购额变化、人员数量规模和各采集项目的具体情况，进一步说明发行人如何确定相关的服务费率，发行人如何量化劳务服务企业的管理难度并反映在服务费率中

报告期各期，发行人前五大劳务服务企业的采购金额和服务费率情况如下：

2018 年度			
序号	供应商名称	采购金额（万元）	服务费率
1	广州善世人力资源服务有限公司	1,755.67	11.5%
	芜湖善慧人力资源服务有限公司	1,541.91	11.5%
2	北京博睿驰信人力资源有限公司	1,681.57	11.5%
	恩泽尔商贸有限公司	249.09	6%
3	北京冠华英才国际经济技术有限公司	905.61	10%、11.5%
4	TWIN STUDIOS	229.97	10%
5	THINKWILD STUDIOS, S.L.	129.85	9%
2017 年度			
序号	供应商名称	采购金额（万元）	服务费率
1	芜湖善慧人力资源服务有限公司	1,064.36	11.5%
	广州善世人力资源服务有限公司	658.34	11.5%

2	北京博睿驰信人力资源有限公司	529.80	11.5%
	恩泽尔商贸有限公司	244.78	6%
	北京零点起航营销策划有限公司	55.33	10%
3	北京冠华英才国际经济技术有限公司	539.82	部分采购由 10%提高至 11.5%
4	SAKURA JAPAN LLC	189.68	10%
5	信阳市联智网络科技有限公司	89.08	10%
2016 年度			
序号	供应商名称	采购金额（万元）	服务费率
1	广州善世人力资源服务有限公司	948.82	10%提高至 11.5%
2	北京零点起航营销策划有限公司	401.13	10%
	北京博睿驰信人力资源有限公司	267.87	10%提高至 11.5%
	恩泽尔商贸有限公司	34.20	6%
3	北京冠华英才国际经济技术有限公司	644.25	10%
4	北京睿祥文义技术服务有限公司	224.63	10%
5	CHUSHOUKIGYOU,CONSULTING, CENTER.,LTD.	109.33	9%

发行人支付给数据服务提供商的服务费率为生数据采集、标记服务采购金额的 6-11.5%，服务费率有所差异的原因是税务成本和管理难度的差异所导致的。

对于境内数据服务提供商，由于不同的供应商在各类型项目中组织和管理被采集人员、标记人员的数量规模和管理难度有所差异，因此发行人与供应商分别约定采用 10-11.5%的服务费率，存在小幅差异。近年来，行业快速发展，发行人业务在规模上有显著增长，拓展方向上也发生了一系列变化，因此对部分主要供应商的要求有所提升，使得部分主要供应商的服务费率由 10%提高到 11.5%。2016 年，广州善世提供劳务服务涉及人数超过 4 万人，管理难度较大，北京博睿主要负责为发行人自然语言和计算机视觉类业务提供服务，细分方向众多，所需人员多样性较强，因此提高了服务费率。2017 年，北京冠华部分业务涉及发行人多样化标注语音，由于发行人该类业务增长明显，对发音人资源的广度要求逐步提高，因此提高了服务费率。

对于境外数据服务提供商，考虑到境外数据被采集人的稀缺度、寻找难度、其所处的不同国家和地区的商业环境、税务成本等实际情况，经发行人与供应商

协商约定采用 6-10% 的服务费率，其中发行人长期境外供应商恩泽尔报告期内服务费率为 6%，未发生变化。其他境外数据服务提供商地处日本、欧洲等地区，人工成本、日常管理成本均较高，服务费率也相对较高。

(3) 进一步说明报告期内是否出现返工的情况，发行人与劳务服务企业的服务费用调整机制，因返工而导致的服务费用调整金额、原因和依据

发行人在标注过程中对标记人员进行技术指导，并对数据质量进行定期和不定期检查，检查不合格的数据，发行人要求数据服务提供商及时修改数据，以保证项目进度。该等数据修改均处于标记人员提供劳务的过程中，是发行人对供应商进行采购管理的一项内部控制措施，此时发行人尚未对其提供的劳务服务进行验收并与劳务服务企业进行结算，因此不涉及服务费用调整。

报告期内，发行人对所采购的劳务服务进行验收后，即代表数据质量已经合格，因此不存在验收并结算后需要返工的情形。

(4) 发行人向同一人控制或管理的多家劳务服务企业进行采购的原因及合理性，发行人未直接向王驰和洪琪直接持股的企业进行采购劳务服务的原因及合理性

发行人主要供应商中，存在多家劳务服务企业受同一人控制或管理的情况，主要是由于供应商实际控制人出于针对不同地域特征划分区域管理等因素而形成的，属于其个人商业考虑。例如，广州善世和芜湖善慧受同一人控制，分别负责华南地区和华东地区业务；北京博睿和恩泽尔受同一人控制，分别负责境内和境外业务，发行人向其进行采购时需要根据所属业务的地域特点与供应商不同主体签约，具备合理性。

同时，发行人除直接向王驰直接持股的北京博睿、恩泽尔和洪琪直接持股的信阳联智采购劳务服务外，还存在向王驰受托经营的北京零点采购劳务服务的情形。保荐机构对北京零点唯一股东进行了访谈，经核查，北京零点股东因家庭原因难以确保足够的精力和时间继续经营北京零点，由于公司尚有员工和部分业务，为妥善安排，北京零点股东与长期相识的北京博睿股东王驰达成委托经营协议，由王驰统一对北京零点进行经营管理，后因王驰经营成果较好，该等委托经营关

系一直持续至今。

发行人已合并计算和披露了向同一人控制或管理的劳务服务企业进行采购的金额等情况。由于该等劳务服务供应商提供的服务主要为根据客户需求进行劳务人员的匹配、组织和管理，其签约主体本身的安排对于发行人采购不构成实质影响。

请保荐机构和申报会计师提供发行人主要的业务信息系统的情况，对发行人一体化人工智能数据处理技术支持平台等核心业务系统进行信息系统专项核查并出具专项核查报告，对系统控制、运营日志和运营数据的真实性、准确性和完整性进行核查、对财务数据和业务数据的衔接和是否存在运营数据缺失或被违规修改的情形进行核查，并对与信息系统相关的内部控制有效性发表明确意见。

保荐机构和申报会计师针对公司报告期信息安全治理环境、信息系统管理和信息系统运行情况的测试，结合对公司业务模块运营模式的理解，从主营业务收入、主要成本等对其业务运营、财务报告有较大影响的业务和管理流程出发，通过穿行相关业务及管理流程，选取其业务主体及管理流程，筛选出支撑这些业务流程运营的主要信息系统，信息系统具体情况如下：

序号	系统名称	系统介绍
1	一体化数据处理技术支持平台	是由公司自主研发的技术支持平台，整合并贯通了数据库设计、数据采集、数据处理、质量检测、质量控制和数据安全管理等各个环节，嵌入了项目管理、质量控制、数据安全的流程管理和工具，主要支持业务人员进行人工智能数据资源开发相关工作的软件系统。
2	业务管理平台	是由公司自主研发的，以支撑公司项目管理工作、串联公司主要业务及财务流程，解决公司业务内部管理分析、对外信息披露、业务财务映射符合财务原则和标准、业务数据细分及内部控制管理各环节数据统计口径不一致、不及时、不准确等痛点问题为目标，加强合同及项目管理，提高日常工作效率，进行更准确的项目核算而开发的管理系统。该平台于2017年1月1日上线使用，实现了公司合同及项目管理等基本功能。
3	用友财务系统	是公司外购、主要由财务部使用的财务管理软件，其以总账系统为核心，包括往来、现金、项目、税务、工资管理、固定资产等模块，负责支撑公司会计核算、财务管理和财务决策等职能。

保荐机构和申报会计师对所识别出支撑的这些业务流程的主要信息系统进

行了信息系统一般性控制测试，评估公司信息系统管理水平及主要信息系统对业务运行的支撑能力和存在的风险，以确定信息系统稳定性和业务财务相关数据运行环境和来源的可靠性。

保荐机构和申报会计师采取必要的评估程序和方法对报告期内支撑公司主要业务运营和财务入账流程的信息系统及相关业务数据及财务数据进行了专项核查，涉及公司信息科技治理、主要信息系统一般性控制管理，并对存在于公司主要信息系统中的相关业务运营数据及财务数据进行了验证和分析，以评估和验证公司相关信息系统在运行管理和应用控制方面的有效性。保荐机构和申报会计师分别出具了关于发行人主要信息系统的专项核查报告。

经核查，保荐机构、申报会计师认为：公司目前对主要信息系统的控制总体有效，未发现重大或重要的控制缺陷。同时综合整体的风险排查和信息系统应用控制测试、业务运营数据及财务数据验证和分析结果，公司报告期内的主营业务和财务报告流程中涉及的信息系统处理及应用控制总体有效，未见明显异常。

除上述信息系统专项核查外，保荐机构和申报会计师针对数据服务费相关采购支出的真实性、准确性和完整性，主要执行了以下核查程序：

1、获取了发行人的采购业务相关管理制度，了解了发行人数据服务采购的管理流程，并选取样本，就其重要内部控制环节（立项和预算管理、采购申请及审批、合同签订、质检验收、过程控制、供应商对账、结算付款等）执行控制测试，核查采购内控管理的有效性；

2、针对发行人报告期内数据服务费执行采购核查程序，从账面记录的数据服务费采购出发，获取采购框架合同、采购订单资料、经双方盖章确认的验收结算单、收到的发票、支付采购结算款项的银行回单，检查供应商名称、验收服务的内容、验收时间、结算数量、结算单价、结算金额、支付金额等信息与发行人记录是否一致，核查采购交易的真实性、准确性；

3、抽取发行人部分业务项目，从发行人后台系统导出采集、标注环节的人员服务记录，与验收结算单所记载的采购服务数量核对，核查采购交易的真实性、准确性、完整性；实地参与了采集项目的采集过程、对标记人员进行了访谈；

4、分析了发行人报告期内向数据服务采购数量、采购单价的变动原因及其合理性，以及采购数量和产出情况的配比关系及其合理性，核查发行人数据服务费采购是否存在异常波动，并调查分析造成波动的原因，对其合理性进行分析复核；

5、获取了发行人董事、监事、高级管理人员及相关财务人员的银行账户资金流水，核查是否存在与主要数据服务供应商及其关键人员存在资金往来并为发行人代垫成本费用的行为，核查发行人成本费用的完整性；

6、对发行人报告期内主要数据服务供应商进行函证，核查采购金额和往来余额的真实性和准确性，报告期内各年发函金额占当年采购金额的比例分别达到79.25%、78.41%和84.77%，所发函证均取得回函，回函金额不存在差异；

7、对发行人报告期内主要数据服务供应商进行了实地走访，所走访的数据服务供应商的对应各年数据服务采购金额占各年数据服务采购总金额的比例超过90%，访谈过程中，主要就双方合作时间、合作方式、提供服务类型、各年交易规模、发行人采购占其同类销售占比、验收结算方式、与外部劳务人员签订协议情况等情况进行详细了解，核查采购交易的真实性、合理性，以及采购模式的可持续性；

8、获取了发行人报告期内主要数据服务供应商的工商资料和报告期内税务合规证明，并穿透获得其股东和其他关键人员、关联公司名单，检查与发行人、发行人关联方及其董事、监事、高级管理人员等关键人员是否存在关联关系；

9、获取了发行人报告期内主要数据服务供应商出具的说明和提供的财务数据，核查其与发行人之间采购定价是否公允，采购交易是否符合商业实质并为供应商带来商业收益，确认双方之间及双方股东、董事、监事、高级管理人员、关键经办人员之间是否存在关联关系，是否存在提供劳务服务以外的其他交易和资金往来并为发行人代垫成本费用的行为，核查发行人成本费用的完整性；

10、查询了互联网上发布的采集、标注服务价格信息，与发行人向终端被采集人和标记人员的采购单价进行比较，核查发行人采购单价是否公允；复核了发行人对数据服务供应商类似的市场同类人力资源服务公司和劳务公司管理服务

费的市场询价结果，核查发行人主要数据服务供应商服务费率的定价公允性；

11、获取了实际控制人曾经控制的关联方 SpeechOcean Ltd.的审计报告、账务记录、银行流水等财务资料，对报告期内发行人与其之间的关联销售和采购交易进行了穿透核查，核查了相关合同和银行流水，对重要的最终销售客户进行了实地走访或视频访谈；核查了 2015 年度关联销售和采购的具体内容；测算了报告期内 SpeechOcean Ltd.的现金流转情况，并对相关数字的勾稽关系进行了复核；获取并查阅了境外律师出具的关于 SpeechOcean Ltd.注销事宜的法律意见书；

12、获取了发行人实际控制人及其配偶、发行人总经理控制、担任董事或高级管理人员、具有重大影响的企业（包括注销或转让的）及其董事、监事、高级管理人员出具的说明，确认该等企业及其董事、监事、高级管理人员与发行人供应商之间是否在交易、资金往来或其他利益安排，是否存在为发行人代垫成本费用的行为，核查发行人成本费用的完整性；

经核查，保荐机构、申报会计师认为：

发行人针对数据服务采购业务建立了较为完善的内部控制并得到了有效执行；发行人及其关联方与数据服务供应商之间不存在关联关系；报告期内，发行人与数据服务供应商采购业务真实、采购数据准确、完整，采购定价公允；发行人对数据服务费采购的核算真实、准确、完整，不存在其他相关方为发行人代垫成本费用的情形，不存在体外支付成本费用的情形。

请保荐机构和申报会计师对上述事项进行核查并发表明确意见。

经核查，保荐机构、申报会计师认为：报告期内为发行人寻找境外数据被采集人的供应商的服务费率与 SpeechOcean 的毛利率不存在明显差异；发行人根据劳务服务供应商的税务成本和管理难度差异，与供应商协商确定服务费率，相关服务费率变动合理；发行人报告期内对所采购的劳务服务进行验收后，即代表数据质量已经合格，不存在验收并结算后需要返工的情形；发行人向同一人控制或管理的多家劳务服务企业进行采购的原因合理，发行人除直接向王驰直接持股的北京博睿、恩泽尔和洪琪直接持股的信阳联智采购劳务服务外，还存在向王驰受托经营的北京零点采购劳务服务的情形，原因合理。

问题 5、关于发行人的营业收入

申报材料显示：（1）发行人已按语种披露了智能语音类数据资源定制服务和数据库产品的销量和销售价格；（2）发行人存量数据库的销售收入逐年上升；（3）发行人应收账款和收入函证的回函差异较大。

请发行人：（1）进一步区分不同项目特征（采集、单一或多样化标注等）和语言类型（中文、外语、小语种和方言等）补充披露智能语音类定制服务和数据库的销售量和销售价格并分析变动原因、不同特征和语言类型项目之间销售价格差异原因；（2）计算机视觉类定制服务收入增长较快，进一步区分不同项目特征（手写体、光学字符识别和行车相关视频等）补充披露计算机视觉类定制服务的销售量和销售价格并分析变动原因、不同特征和语言类型项目之间销售价格差异原因；（3）说明报告期各期各类型数据库的库存量、新增量和结存量，并结合各类型数据库的生命周期进一步分析各期销售和未销售的对应数据库数量变动情况、销售数据库的开发成本和销售价格。

请保荐机构和申报会计师对上述事项进行核查并发表明确意见，进一步说明回函不符的事项履行的核查程序，列表说明具体调节过程，包括期初余额、当期确认收入、当期收到的预收账款和应收账款当期和期后回款的对应关系，调节事项是否经发行人客户确认。

答复：

（1）进一步区分不同项目特征（采集、单一或多样化标注等）和语言类型（中文、外语、小语种和方言等）补充披露智能语音类定制服务和数据库的销售量和销售价格并分析变动原因、不同特征和语言类型项目之间销售价格差异原因

发行人已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/3、主要产品和服务的销量和价格分析/（1）智能语音领域”中对以下内容进行了补充披露：

①智能语音类数据资源定制服务

报告期内，公司智能语音类数据资源定制服务的销量和价格情况如下：

业务细分 类别	语种	2018 年度		2017 年度		2016 年度	
		销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)
有采集的 单一化标 注语音	中文	38,286.00	355.14	16,759.01	448.14	6,533.00	553.11
	外语种	20,055.00	722.76	8,028.03	872.46	11,091.40	716.31
	小计	58,341.00	481.51	24,787.04	585.57	17,624.40	655.82
无采集的 单一化标 注语音	中文	88,196.01	302.46	41,646.21	312.41	50,191.57	291.28
	外语种	3,067.41	955.13	1,886.61	572.92	2,268.81	653.94
	小计	91,263.42	324.40	43,532.82	323.70	52,460.38	306.96
多样化标 注语音	中文	783.86	17,110.43	313.70	27,392.84	237.06	20,957.44
	外语种	536.47	51,796.84	463.54	25,327.38	228.74	21,508.92
	小计	1,320.33	31,204.04	777.24	26,161.01	465.80	21,228.26

A、销量变动情况分析

报告期内销量整体呈快速增长的趋势，与公司收入规模的变动趋势一致。随着智能语音行业的高速发展，有采集的单一化标注语音和多样化标注语音销量显著增长，带动数据资源定制服务业务规模的持续扩大。无采集的单一化标注语音2017年销量出现下滑，是由于个别主要客户当年项目启动较晚且验收周期较长，因此当年完成验收实现销售的销量较低。2018年，无采集的单一化标注语音增长较快，也受益于前述原因。总体而言，无采集的单一化标注语音销量在报告期内仍然保持增长。

B、价格变动情况分析

公司智能语音类数据资源定制服务的价格变动主要受到行业市场竞争、资源稀缺程度、客户或项目的特殊要求等原因的影响，具体情况如下：

a、有采集的单一化标注语音

中文业务单价逐年下降，主要原因是行业快速发展使得中文数据资源日益丰富，且发行人业务相对成熟，同时市场竞争较为充分，因此价格相应下降所致。

外语种业务2017年单价较高，2016年和2018年较为平稳。2017年单价较高是由于发行人开展了一些低噪声环境、车载环境等场景以及法语、德语的项目，报价较高，拉升了2017年的平均单价。英语类项目的价格在报告期内较为平稳，该类别涵盖全球多个国家和地区的英语，如新加坡、英国、加拿大等英语母语国家，以及西班牙、意大利等非英语母语国家，资源相对分散，发达国家物价水平较高，导致成本较高，因而销售价格较高。其他语种覆盖的语种超过20个，由于每种语言特点均不相同，客户对特定语种的需求紧迫程度、发音人资源稀缺程度在不同年份也有所不同，因此会造成一定的价格波动。

语种	价格（元/小时）		
	2018 年度	2017 年度	2016 年度
英语	869.51	875.40	836.82
其他语种	664.79	870.70	690.18

外语种业务单价明显高于中文业务，是由于外语种资源相对稀缺，采集和标注难度高于中文所致，具备合理性。

b、无采集的单一化标注语音

中文业务单价较为平稳。外语种业务由于整体收入规模较小（各期收入不超过300万元，占各期智能语音类数据资源定制服务收入比例约2-4%），因此易受单个项目影响，2018年单价大幅上升，是由于发行人当年首次开展印度及巴基斯坦地区语种项目，由于语种稀缺，报价较高。

外语种业务单价明显高于中文业务，是由于外语种资源相对稀缺，标注难度高于中文所致，具备合理性。

c、多样化标注语音

多样化标注语音项目采用小规模受过专业发音训练的发音人，通常一个项目最终选定一个发音人。采集过程中发音人录音时间长达数月、对语音质量要求高、不同发音人量级差别大，后续处理精度高，通常需要处理至音素级别，因此按小时计算的销量远低于单一化标注语音项目，同时单位价格远高于单一化标注语音项目，符合该类细分业务的业务特性。

由于多样化标注语音项目的定制化程度最高，因此其不同项目之间的单价差异更明显，报告期各期平均单价的合理波动区间更大。

中文业务2017年单价提升，主要原因是发行人开展了中文搭配其他语种混读、温柔型风格等项目，由于难度较高且资源稀缺造成单价较高；同时2017年是下游智能音箱行业的集中研发时期，各智能音箱厂商研发周期有限，需求紧迫，发行人议价空间较大；2018年单价明显下降，是由于开展了部分无需进行音素边界标注的多样化标注语音项目，成本相对较低，因此价格较低。

外语种业务2018年单价大幅上升，主要原因是随着下游语音合成技术应用性的提升，客户对发音人要求越来越高，例如对其音色的代表性、发音的专业性等方面都提出了更高的要求。2018年，发行人根据客户需求启动了欧美多个语言的合成项目，此类项目需求较一般合成项目要求更高：对于录音棚，要求更低的背景噪声和更高的信噪比；对于发音人，要求发音人具备五年以上播音、配音及主持相关经验，每个项目要求提供至少五十个候选发音人，同时对于候选发音人通过100个该语种母语用户的主观评测进行选择；对于录音过程，除配备录音工程师外，要求同时配备录音指导，录音指导需具备音乐相关领域本科以上学历，五年以上音乐导演或配音导演工作经历，以及熟练的英语沟通能力。这类要求均超出以往年度一般合成项目的要求，因此使得项目难度和成本显著提升，同时价格也相应显著上涨。

相比单一化标注语音项目的价格受语种影响较大，多样化标注语音项目的价格主要取决于发音人价格和标注类型，而与语种关联性较低，中文与外语种之间不存在固有的价格差异。发音人主要取决于专业性、知名程度、当地物价水平等因素，即使是同一语言，个体差异也可能较大。多样化标注中的主要标注类型音素边界标注是针对音素（如“a”，“o”，“b”，“h”等），所有语言均可用国际音标表示其发音，标注成本与语言相关度较弱。

②智能语音类数据库产品

报告期内，公司智能语音类数据库产品的销量和价格情况如下：

业务细分	语种	2018年度	2017年度	2016年度
------	----	--------	--------	--------

类别		销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)	销量 (小时)	价格 (元/小时)
有采集的 单一化标 注语音	中文	68,536.93	354.79	41,159.09	473.30	23,881.88	374.47
	外语种	47,099.49	720.45	27,063.42	868.52	26,961.93	765.26
	小计	115,636.42	503.73	68,222.51	630.09	50,843.81	581.70
多样化标 注语音	中文	50.20	12,514.05	114.47	16,405.95	133.76	16,307.24
	外语种	45.50	28,520.32	183.77	23,649.12	75.98	29,045.19
	小计	95.70	20,124.13	298.24	20,869.06	209.74	20,921.67

A、销量变动情况分析

有采集的单一化标注语音销量的总体变动趋势与公司业务发展趋势一致，在报告期内实现了明显增长。2018年，公司更多地将多通道采集技术运用到有采集的单一化标注语音数据库产品的开发中，该类型数据库产品所包含的小时数量相对于单通道数据库产品成倍增加，进一步带动了销量的增长。

多样化标注语音销量在2017年外语种业务的带动下有所增长，在2018年则出现明显下降，是由于随着下游相关技术的发展和应用场景的拓展，个性化程度日益提高，该趋势下，多样化标注语音定制化程度高的特性日益显著，下游用户普遍转向采购公司的数据资源定制服务，因此降低了数据库产品的销量。

B、价格变动情况分析

公司智能语音类数据库产品的价格变动主要受到市场同类数据库产品稀缺程度和发行人数据库产品的具体定位及拓展方向等原因，具体情况如下：

a、有采集的单一化标注语音

中文和外语种业务均呈现2017年单价较高，2016年和2018年较为平稳的情况。中文业务2017年单价较高的原因在于发行人判断自由对话数据对语音识别算法模型的训练将起到促进的作用，于当年开发了部分自由对话类数据库产品，该类数据库价格较高。同时，发行人2017年逐步开展重口音、方言类数据库产品开发和销售业务，该类数据库在市场上稀缺性较高，因此价格较高。外语种业务2017年单价较高的原因为发行人当年开发并销售了多个东南亚语种的数据产品，例如印度各种方言、朝鲜语、阿拉伯语等，由于该类语种在市场上的

稀缺性较高，因此价格较高。

外语种业务单价明显高于中文业务，是由于国内外语种资源相对稀缺，采集和标注难度高于中文所致，具备合理性。

b、多样化标注语音

如前述原因，多样化标注语音项目按小时计算的产出量远低于有采集的单一化标注语音项目，同时单位价格远高于有采集的单一化标注语音项目。

此外，多样化标注语音数据库产品受到发音人定位的不同，价格变动比较明显，报告期各期平均单价的合理波动区间较大，因此上述中文和外语种业务的价格整体处于合理波动区间。

(2) 计算机视觉类定制服务收入增长较快，进一步区分不同项目特征（手写体、光学字符识别和行车相关视频等）补充披露计算机视觉类定制服务的销售量和销售价格并分析变动原因、不同特征和语言类型项目之间销售价格差异原因

由于计算机视觉类业务各细分类别差异较大，例如手写体应用于手写输入，光学字符识别应用于拍照翻译等功能，行车相关视频应用于自动驾驶，且计价单位不同，因此不同类型项目的销售价格之间无法横向比较。

发行人已在招股说明书“第八节 财务会计信息与管理层分析/九、经营成果分析/（一）营业收入分析/3、主要产品和服务的销量和价格分析/（2）计算机视觉和自然语言领域”中对以下内容进行了补充披露：

报告期内，公司计算机视觉类数据资源定制服务中手写体、光学字符识别、行车相关视频业务的销量和价格情况如下：

业务细分 类别	2018 年度		2017 年度		2016 年度	
	销量	单位价格 (元)	销量	单位价格 (元)	销量	单位价格 (元)
手写体	2,740	494.90	3,875	1,238.89	795	2,444.21
光学字符 识别	47,419.86	65.13	13,238.04	48.40	1,811.69	18.16
行车相关	1,188,754.20	4.77	-	-	-	-

业务细分 类别	2018 年度		2017 年度		2016 年度	
	销量	单位价格 (元)	销量	单位价格 (元)	销量	单位价格 (元)
视频						

注：手写体、光学字符识别、行车相关视频的销量单位分别为千词组、百张、十个目标。

①手写体

报告期内，公司手写体业务2018年销量相较2017年下降，主要是受到客户2018年项目启动时间较晚的影响，截至2018年末，手写体业务仍有共计22,704千词组的较大规模尚未验收的在产品结存，因此整体在手订单量保持了持续增长。平均销售单价逐年下降，主要是由于该类业务的语种差异所导致的。2016年，发行人手写体业务除英语外还包括西班牙语、意大利语等小语种，市场资源稀缺，采集难度较大，因此价格在报告期最高。2017年，语种涉及日语、英语、德语等，采集难度相对有所降低。2018年，该项业务的主要语种为中文，采集难度明显降低，因此价格也显著下降。

②光学字符识别

2016年，公司光学字符识别业务规模很小，收入仅为3.29万元，处于探索阶段，业务较为简单，无采集环节仅进行标注，因此单价较低。2017年和2018年，公司在光学字符识别业务方面实现了较快拓展，介入采集环节且采集量显著增长，销量快速增长的同时，单价也持续提升。

③行车相关视频

公司行车相关视频业务是2018年新拓展业务。随着人工智能重要应用领域自动驾驶在技术和产业方面的快速发展，下游企业对于行车数据的需求不断扩大，公司借助这一契机积极拓展行车相关视频业务，2018年实现营业收入566.81万元。

(3) 说明报告期各期各类型数据库的库存量、新增量和结存量，并结合各类型数据库的生命周期进一步分析各期销售和未销售的对应数据库数量变动情况、销售数据库的开发成本和销售价格

一、智能语音领域

报告期内，发行人各年新增的智能语音数据库产品数量分别为 76 个、49 个和 24 个，各年末库存结余量分别为 334 个、383 个和 407 个，各年新增数据库产品在报告期内实现销售的数量以及未实现销售的数量情况具体如下：

单位：个

年份	各期新增数量（产成数据库个数）	期末库存量	各期产成的数据库中，报告期内实现销售的数据库个数			各期产成的数据库中，报告期内未实现销售的数据库个数		
			2016 年	2017 年	2018 年	2016 年	2017 年	2018 年
2013 年及以前	161	161	40	68	47	121	93	114
2014 年	50	211	11	9	14	39	41	36
2015 年	47	258	5	17	25	42	30	22
2016 年	76	334	44	23	33	32	53	43
2017 年	49	383	-	49	12	-	-	37
2018 年	24	407	-	-	24	-	-	-
合计	407	-	100	166	153	307	241	254
销售数据库的平均开发成本（万元/个）			27.39	25.14	26.62	-	-	-
销售数据库的平均销售价格（万元/个）			26.53	22.68	28.52	-	-	-

报告期各期新增的智能语音数据库产品大部分可在当期实现首次销售，其中 2016 年新增智能语音数据库产品 76 个，实现销售 44 个，有 32 个未在当期实现销售，主要原因是发行人在该年度致力于业务拓展，加大了智能语音数据库产品新增储备所致。

总体而言，发行人的智能语音数据库产品具备较长的生命周期。2018 年实现销售的智能语音数据库产品覆盖了 2013 年及以前、2014 年至 2018 年各年产成的数据库产品，其中最早产成年份为 2005 年。产成年份较早的数据库产品在报告期内也仍实现了较好的销售情况，以 2013 年及以前年度产成的 161 个数据库产品为例，在上述数据库产品中，分别有 40 个、68 个、47 个在 2016 年、2017 年、2018 年实现了销售。发行人智能语音数据库产品具有较长生命周期的原因，一方面是由于发行人具备了一定的行业前瞻性和洞察力，其开发的智能语音数据库产品符合行业发展趋势，能够持续满足下游客户需求，另一方面则是由于人工智能下游技术应用行业不断涌入新的市场参与者，对于较为基础的通用型数据库产品也仍保持了一定的需求。

报告期内实现销售的智能语音数据库产品对应的平均开发成本和平均销售价格总体较为稳定，平均开发成本分别为 27.39 万元、25.14 万元和 26.62 万元，平均销售价格分别为 26.53 万元、22.68 万元和 28.52 万元。由于数据库产品可重复授权销售，因此平均销售价格低于平均开发成本并不一定导致发行人开发数据库产品出现亏损。

二、计算机视觉领域和自然语言领域

报告期内，发行人各年新增的计算机视觉数据库产品个数分别为 3 个、19 个和 11 个，各年末的库存结余量分别是 7 个、26 个和 37 个，各年新增计算机视觉数据库产品在报告期内实现销售的数量以及未实现销售的数量情况具体如下：

单位：个

年份	各期新增数量（产成数据库个数）	期末库存量	各期产成的数据库中，报告期内实现销售的数据库个数			各期产成的数据库中，报告期内未实现销售的数据库个数		
			2016 年	2017 年	2018 年	2016 年	2017 年	2018 年
2015 年及以前	4	4	-	1	4	4	3	-
2016 年	3	7	3	-	-	-	3	3
2017 年	19	26	-	11	1	-	8	18
2018 年	11	37	-	-	11	-	-	-
合计	37	-	3	12	16	34	25	21
销售数据库的平均开发成本（万元/个）			5.68	11.87	11.35	-	-	-
销售数据库的平均销售价格（万元/个）			6.09	15.65	7.45	-	-	-

报告期内，发行人各年新增自然语言数据库产品个数分别为 14 个、19 个和 16 个，各年末的库存结余量分别为 37 个、56 个和 72 个，各年新增自然语言数据库产品在报告期内实现销售的数量以及未实现销售的数量情况具体如下：

单位：个

年份	各期新增数量（产成数据库个数）	期末库存量	各期产成的数据库中，报告期内实现销售的数据库个数			各期产成的数据库中，报告期内未实现销售的数据库个数		
			2016 年	2017 年	2018 年	2016 年	2017 年	2018 年
2015 年及以前	23	23	10	3	5	13	20	18
2016 年	14	37	14	4	5	-	10	9
2017 年	19	56	-	17	1	-	2	18

2018 年	16	72	-	-	16	-	-	-
合计	72	-	24	24	27	48	48	45
销售数据库的平均开发成本 (万元/个)			5.16	3.36	6.61	-	-	-
销售数据库的平均销售价格 (万元/个)			16.73	10.60	13.72	-	-	-

发行人报告期各期新增的计算机视觉和自然语言数据库产品大部分可在当期实现首次销售。由于发行人计算机视觉和自然语言数据库产品大部分为报告期内产成，年限较短，其生命周期情况尚难以判断，2015 年及以前产成的数据库产品在 2018 年部分实现了销售，仍保持在生命周期之内。

报告期内实现销售的计算机视觉数据库产品对应的平均开发成本分别为 5.68 万元、11.87 万元和 11.35 万元，平均销售价格分别为 6.09 万元、15.65 万元和 7.45 万元。2017 年和 2018 年平均开发成本上升的原因是随着行业快速发展，计算机视觉数据库在数据规模、数据类型丰富度等方面均有了明显提升，从而提高了开发成本。平均销售价格在 2017 年随着平均开发成本上升而提升，2018 年下降的原因是由于发行人为进一步拓展计算机视觉类业务的客户资源，在定价策略上有所改变，通过较低定价吸引客户所致。由于数据库产品可重复授权销售，因此平均销售价格低于平均开发成本并不会导致发行人开发数据库产品出现亏损。

报告期内实现销售的自然语言数据库产品对应的平均开发成本分别为 5.16 万元、3.36 万元和 6.61 万元，平均销售价格分别为 16.73 万元、10.60 万元和 13.72 万元。由于自然语言数据库产品涉及的细分类别较多，其平均开发成本和平均销售价格的波动趋势并不固定。

请保荐机构和申报会计师对上述事项进行核查并发表明确意见

经核查，保荐机构、申报会计师认为：发行人智能语音类数据资源定制服务和数据库产品的销售量和销售价格变动原因合理，不同特征和语言类型项目之间销售价格差异原因合理，不存在异常；发行人计算机视觉类数据资源定制服务的销售量和销售价格变动原因合理，计算机视觉类业务各细分类别差异较大，且计价单位不同，销售价格无法横向比较的原因具备合理性；发行人的智能语音数据库产品总体而言具备较长的生命周期，计算机视觉和自然语言数据库产品大部分

为报告期内产成，年限较短，其生命周期情况尚难以判断，2015年及以前产成的数据库产品在2018年部分实现了销售，仍保持在生命周期之内。

请保荐机构和申报会计师进一步说明回函不符的事项履行的核查程序，列表说明具体调节过程，包括期初余额、当期确认收入、当期收到的预收账款和应收账款当期和期后回款的对应关系，调节事项是否经发行人客户确认。

一、进一步说明针对回函不符事项履行的核查程序

报告期内，客户未直接在回函中对发行人应收账款和营业收入确认的情况较多，主要原因系发行人确认收入根据客户验收报告，而客户确认采购依据发行人开具的发票，双方核算的口径不一致而导致。针对回函不符事项，保荐机构、申报会计师主要履行了以下核查程序：

1、了解发行人与客户之间的验收、结算和对账流程，并对业务环节中的关键控制点进行测试；

2、通过函证取得客户可以确认的信息（包括发票与付款信息），与发行人账面记载的应收账款金额、营业收入金额、开具发票金额、客户验收金额和收款金额进行核对，分析其勾稽关系的合理性，并对差异情况进行调节；

3、执行收入核查程序，以营业收入为起点，检查销售合同、项目执行过程中与客户的邮件沟通记录、产品或服务交付记录、客户验收报告、开票记录和银行回款流水等，就上述资料中的客户名称、合同条款、业务内容、数据数量、数据交付时间、客户验收时间、开票时间和收款时间等信息进行匹配和穿行测试，检查营业收入和应收账款确认的真实性和准确性。

4、对发行人主要客户进行了实地走访，了解客户和发行人之间的合作模式、合作业务规模和结算方式。通过实地走访，未发现发行人与客户之间销售业务存在异常情形。

二、列表说明具体的调节过程，包括期初余额、当期确认收入、当期收到的预收账款和应收账款当期和期后回款的对应关系，调节事项是否经发行人客户确认。

根据最新的回函情况，报告期内应收账款和营业收入的函证汇总情况更新如下：

单位：万元

年度/时点	报表科目	发函情况			回函情况 ¹			期后回款 ² (截至2019年4月30日)
		账面金额	发函金额	发函比例	回函对应的发函金额	回函金额	回函差异	
2018-12-31 /2018年	应收账款	3,096.75	2,911.30	94.01%	2,411.53	1,115.45	1,296.08	2,073.17
	营业收入	19,265.77	17,724.02	92.00%	13,684.32	12,861.77	822.54	—
2017-12-31 /2017年	应收账款	2,707.57	2,535.84	93.66%	2,134.69	1,471.89	662.80	2,438.69
	营业收入	11,907.09	10,652.10	89.46%	9,629.78	9,451.94	177.84	—
2016-12-31 /2016年	应收账款	1,772.88	1,644.45	92.76%	1,644.45	1,945.75	-301.31	1,657.03
	营业收入	8,422.86	7,334.49	87.08%	7,121.11	7,223.29	-102.18	—

注1：发行人第二轮审核问询函回复第8题中列示的“营业收入回函金额”，是发行人根据客户的回函金额及客户的应收账款期初与期末差异，推算出的客户采购金额；本轮回函情况中营业收入回函金额是截至本轮审核反馈提交时客户认可的发票金额，并作为调节过程的基础；

注2：期后回款金额指截至2019年4月30日，报告期各期末应收账款对应的期后回款金额。

应收账款与期后回款的对应关系核查如下：

发行人开具的发票与销售合同编号存在对应关系，客户的回款与发行人开具的发票存在对应关系。针对应收账款的期后回款的核查，保荐机构、申报会计师检查了销售合同编号与开具发票的对应关系、客户回款金额与发票的对应关系、应收账款与销售合同的对

应关系。经核查，期后回款逐一匹配至相应的应收账款。

报告期内，营业收入、应收账款函证的回函差异主要集中在前五大营业收入客户（按品牌）和前五大应收账款对象（按品牌）。根据审核问询要求，将前五大营业收入客户（按品牌）和前五大应收账款对象（按品牌）函证情况按照法人主体列示如下：

(一) 前五大营业收入客户（按品牌）的函证汇总情况

单位：万元

年度	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
2018年	阿里巴巴	淘宝（中国）软件有限公司	2,019.42	2,019.42	1,399.64	619.78	
		阿里巴巴（中国）网络技术有限公司	1,602.97	1,602.97	1,398.16	204.81	
		浙江天猫技术有限公司	1,418.37	1,418.37	1,195.87	222.50	
	三星	Samsung Electronics Co., Ltd.	1,679.68	1,679.68	—	—	未回函
		北京三星通信技术研究有限公司	603.15	603.15	61.64	541.51	
		Samsung R&D Institute India - Bangalore Pvt Ltd.	153.99	153.99	311.73	-157.74	
	腾讯	腾讯科技（深圳）有限公司	978.68	978.68	1,354.46	-375.78	
		深圳市腾讯计算机系统有限公司	390.84	390.84	424.54	-33.69	
		腾讯科技（成都）有限公司	168.54	168.54	162.59	5.95	
	微软	微软（中国）有限公司	957.75	957.75	1,904.58	-946.83	
		Microsoft Corporation-USA	341.98	341.98	341.98		
	百度	百度国际科技（深圳）有限公司	334.26	334.26	350.39	-16.13	
		北京百度网讯科技有限公司	375.27	375.27	284.01	91.26	
		百度时代网络技术（北京）有限公司	90.29	90.29	—	—	未回函
		百度在线网络技术（北京）有限公司	88.47	88.47	88.47		
	2017年	腾讯	深圳市腾讯计算机系统有限公司	417.77	417.77	374.53	43.24

年度	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
		腾讯科技（上海）有限公司	88.41	88.41	74.67	13.74	
		腾讯科技（深圳）有限公司	1,263.87	1,263.87	1,030.28	233.59	
	三星	SAMSUNG ELECTRONICS CO., LTD.	113.96	113.96	113.96		
		北京三星通信技术研究有限公司	1,439.12	1,439.12	1,558.19	-119.07	
	阿里巴巴	阿里云计算有限公司	526.90	526.90	387.14	139.76	
		淘宝（中国）软件有限公司	241.61	241.61	317.30	-75.69	
		广州神马移动信息科技有限公司	110.26	110.26	110.26		
		浙江天猫技术有限公司	118.77	118.77	81.89	36.88	
		阿里巴巴（北京）软件服务有限公司	63.68	63.68	63.68		
		支付宝（中国）网络技术有限公司	45.44	45.44	—	—	未回函
	微软	Microsoft Corporation-USA	477.99	477.99	477.99		
		微软（中国）有限公司	580.87	580.87	826.27	-245.40	
	百度	北京百度网讯科技有限公司	585.29	585.29	585.29		
2016年	阿里巴巴	淘宝（中国）软件有限公司	361.67	361.67	361.67		
		阿里云计算有限公司	160.01	160.01	160.01		
		阿里巴巴（北京）软件服务有限公司	93.75	93.75	93.75		
		阿里巴巴（中国）网络技术有限公司	10.16	10.16	10.16		
		北京阿里巴巴云计算技术有限公司	9.98	9.98	9.98		
		支付宝（中国）网络技术有限公司	166.60	166.60	166.60		

年度	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
		高德信息技术有限公司	175.81	175.81	175.81		
	三星	北京三星通信技术研究有限公司	863.80	863.80	948.39	-84.59	
	微软	微软（中国）有限公司	389.74	389.74	525.50	-135.76	
		Microsoft Corporation-USA	237.27	237.27	237.27		
		Microsoft Corporation 印度	198.80	198.80	198.80		
	SpeechOcean Ltd.	SpeechOcean Ltd.	716.94	716.94	716.94		
	百度	北京百度网讯科技有限公司	526.78	526.78	526.78		

(二) 前五大应收账款对象（按品牌）函证汇总情况

单位：万元

时点	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
2018年 12月31日	阿里巴巴	阿里巴巴（中国）网络技术有限公司	217.11	217.11		217.11	
		浙江天猫技术有限公司	274.96	274.96	136.48	138.48	
		淘宝（中国）软件有限公司	752.11	752.11	587.65	164.47	
	腾讯	深圳市腾讯计算机系统有限公司	213.39	213.39	未确认	213.39	
		腾讯科技（成都）有限公司	101.84	101.84	未确认	101.84	
		腾讯科技（深圳）有限公司	139.04	139.04	未确认	139.04	
	Amazon	Amazon Web Services, Inc.	351.94	351.94	—	—	未回函
	百度	北京百度网讯科技有限公司	289.80	289.80	未确认	289.80	

时点	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
	深圳市天道日新科技有限公司	深圳市天道日新科技有限公司	110.00	110.00	110.00		
2017年 12月31日	腾讯	腾讯科技（上海）有限公司	14.71	14.71	未确认	14.71	
		深圳市腾讯计算机系统有限公司	52.16	52.16	4.00	48.16	
		腾讯科技（深圳）有限公司	769.76	769.76	186.58	583.19	
	阿里巴巴	支付宝（中国）网络技术有限公司	10.79	10.79	—	—	未回函
		浙江天猫技术有限公司	39.10	39.10	39.10		
		阿里云计算有限公司	181.16	181.16	181.02	0.14	
		淘宝（中国）软件有限公司	95.84	95.84	122.45	-26.61	
	乐视	乐乐创新智能科技（北京）有限公司	209.88	209.88	—	—	未回函
	百度	北京百度网讯科技有限公司	202.76	202.76	202.76		
富士康	云智汇（深圳）高新科技服务有限公司	191.52	191.52	191.52			
2016年 12月31日	阿里巴巴	阿里巴巴（中国）网络技术有限公司	10.77	10.77	10.77		
		支付宝（中国）网络技术有限公司	16.30	16.30	16.30		
		阿里云计算有限公司	39.62	39.62	39.62		
		高德信息技术有限公司	121.55	121.55	121.55		
		淘宝（中国）软件有限公司	153.70	153.70	153.70		
		阿里巴巴（北京）软件服务有限公司	14.28	14.28	14.28		
	腾讯	腾讯科技（深圳）有限公司	311.63	311.63	218.21	93.42	
	百度	北京百度网讯科技有限公司	205.96	205.96	205.96		

时点	客户品牌	客户名称	账面金额	发函金额	回函金额	回函差异	备注
	乐视	乐乐创新智能科技（北京）有限公司	115.85	115.85	115.85		
	联想	联想（北京）有限公司	110.00	110.00	110.00		

（三）回函差异调节过程

1、关于营业收入的调节过程

发行人按照项目交付与验收进度为依据确认对客户的营业收入，而客户通常按照收到发行人的发票作为确认当期采购的情形，双方核算的差异导致营业收入的回函存在差异，具体分析如下：

类型	发行人（以验收为依据）			客户（以发票为依据）			是否需要调节
	数据验收时间	开具发票时间	确认收入	数据验收时间	取得发票时间	确认采购	
情形 1	当期	当期	√	当期	当期	√	否
情形 2	当期	前期	√	当期	前期	×	是
情形 3	当期	尚未	√	当期	尚未	×	
情形 4	尚未	当期	×	尚未	当期	√	
情形 5	前期	当期	×	前期	当期	√	

注：情形 1 即当发行人确认收入的时间与客户确认采购的时间一致时，无需调节。除情形 1 外，其余 4 种情形均需要对营业收入进行调节。

对发行人营业收入回函差异的调节步骤为：以客户回函的金额为基础，结合验收与开具发票各种情形进行调节，具体如下：

	项目	备注
营业收入回函差异调节过程	客户回函的发票金额	—
	减：当期开票、前期验收	情形 5
	减：当期开票、尚未验收	情形 4
	加：当期验收、前期开票	情形 2
	加：当期验收、尚未开票	情形 3
	调节后的营业收入金额	—

2、关于应收账款的调节过程

发行人报告期内每年末对客户的应收账款根据年初余额、本年含税营业收入、本年收款、本年只开票税金影响等项目的逻辑关系计算而来，其中开具发票和本年收款经过客户回函认可，本年营业收入根据客户回函情况进行专门调节。只开票税金影响金额较小，主要通过复核未验收已开票项目（包括开票并收预收款和只开票两种类型）金额进行确认。

基于此，对于发行应收账款回函差异的主要通过检查年初余额、本年含税营业收入、本年收款、本年只开票税金影响与年末应收账款余额的会计核算逻辑等式的一致性进行核查。

3、前五大品牌客户回函差异的调节过程

(1) 营业收入回函差异调节过程

①2018年回函差异调节过程

单位：万元

品牌	主体	发函金额	回函金额	调节过程				调节后金额	调节结论
				当期开票 前期验收	当期开票尚 未验收	当期验收 前期开票	当期验收 尚未开票		
逻辑关系		①	②	③	④	⑤	⑥	⑦=②-③-④+ ⑤+⑥	⑧
阿里巴巴	淘宝（中国）软件有限公司	2,019.42	1,399.64	64.52	70.81		755.12	2,019.42	一致
	阿里巴巴（中国）网络技术有限公司	1,602.97	1,398.16				204.82	1,602.97	一致
	浙江天猫技术有限公司	1,418.37	1,195.87	36.89	141.51		400.90	1,418.37	一致
三星	北京三星通信技术研究有限公司	603.15	61.64	19.34	33.39		594.28	603.15	一致
	Samsung R&D Institute India - Bangalore Pvt Ltd.	153.99	311.73		157.74			153.99	一致
腾讯	腾讯科技（深圳）有限公司	978.68	1,354.46	352.71	62.10	39.03		978.68	一致
	深圳市腾讯计算机系统有限公司	390.84	424.54	35.39	15.30		17.00	390.84	一致
	腾讯科技（成都）有限公司	168.54	162.59		34.12		40.07	168.54	一致
微软	微软（中国）有限公司	957.75	1,904.58	23.19	1,179.88	250.58	5.66	957.75	一致
百度	百度国际科技（深圳）有限公司	334.26	350.39		60.74	41.88	2.72	334.26	一致

品牌	主体	发函金额	回函金额	调节过程				调节后金额	调节结论
				当期开票 前期验收	当期开票尚 未验收	当期验收 前期开票	当期验收 尚未开票		
	北京百度网讯科技有限公司	375.27	284.01		13.25		104.51	375.27	一致

②2017年回函差异调节过程

单位：万元

品牌	主体	发函金额	回函金额	调节过程				调节后金额	调节结论
				当期开票 前期验收	当期开票尚 未验收	当期验收 前期开票	当期验收 尚未开票		
逻辑关系		①	②	③	④	⑤	⑥	⑦=②-③-④+ ⑤+⑥	⑧
腾讯	深圳市腾讯计算机系统有限公司	417.77	374.53				43.24	417.77	一致
	腾讯科技（上海）有限公司	88.41	74.67				13.74	88.41	一致
	腾讯科技（深圳）有限公司	1,263.87	1,030.28	289.39	84.91	25.64	582.24	1,263.87	一致
三星	北京三星通信技术研究有限公司	1,439.12	1,558.19		647.16	452.90	75.20	1,439.12	一致
阿里巴巴	阿里云计算有限公司	526.90	387.14	40.32	95.63	52.46	223.25	526.90	一致
	淘宝（中国）软件有限公司	241.61	317.30	150.26	28.23		102.79	241.61	一致
	浙江天猫技术有限公司	118.77	81.89				36.89	118.77	一致
微软	微软（中国）有限公司	580.87	826.27	70.44	345.77	168.01	2.81	580.87	一致

③2016年回函差异调节过程

单位：万元

品牌	主体	发函金额	回函金额	调节过程				调节后金额	调节结论
				当期开票 前期验收	当期开票 尚未验收	当期验收 前期开票	当期验收 尚未开票		
逻辑关系		①	②	③	④	⑤	⑥	$⑦=②-③-④+⑤+⑥$	⑧
三星	北京三星通信技术研究有限公司	863.80	948.39		637.08	479.25	73.24	863.80	一致
微软	微软（中国）有限公司	389.74	525.50	2.83	236.71	103.77		389.74	一致

(2) 应收账款回函差异调节过程

单位：万元

时点	客户 品牌	主体	期初应收账 款	本期销售		只开票 税金影 响	本期收款	期末应收 账款	逻辑关系检 查	核对结论	期后回款
				营业收入	含税 销售额						
逻辑关系			①	②	③	④	⑤	⑥	$⑦=①+③+④-⑤-⑥$	—	—
2018年 12月31 日	阿里巴 巴	淘宝（中国）软件有限公司	95.84	2,019.42	2,149.43	-0.70	1,492.46	752.11		核对一致	733.20
		阿里巴巴（中国）网络技术有限 公司		1,602.97	1,764.11		1,547.00	217.11		核对一致	180.47
		浙江天猫技术有限公司	39.10	1,418.37	1,503.48		1,267.62	274.96		核对一致	274.96

时点	客户品牌	主体	期初应收账款	本期销售		只开票税金影响	本期收款	期末应收账款	逻辑关系检查	核对结论	期后回款
				营业收入	含税销售额						
逻辑关系			①	②	③	④	⑤	⑥	⑦=①+③+④-⑤-⑥	——	——
	腾讯	腾讯科技（深圳）有限公司	769.76	978.68	1,068.65	6.21	1,807.76	139.04		核对一致	12.99
		深圳市腾讯计算机系统有限公司	52.16	390.84	421.68		226.78	213.39			213.39
		腾讯科技（成都）有限公司		168.54	178.65		8.30	101.84			101.84
	百度	北京百度网讯科技有限公司	202.76	375.27	413.07	-8.92	317.11	289.80		核对一致	159.37
2017年 12月31日	腾讯	深圳市腾讯计算机系统有限公司		417.77	482.72	0.91	431.47	52.16		核对一致	52.16
		腾讯科技（上海）有限公司		88.41	99.50		84.79	14.71		核对一致	14.71
		腾讯科技（深圳）有限公司	311.63	1,263.87	1,387.15	-2.76	926.26	769.76		核对一致	769.76
	阿里 巴	阿里云计算有限公司	39.62	526.90	556.04	-0.33	414.17	181.16		核对一致	181.16
		淘宝（中国）软件有限公司	153.70	241.61	256.11	22.37	336.34	95.84		核对一致	95.84
2016年 12月31日	腾讯	腾讯科技（深圳）有限公司		318.51	357.27	4.36	50.00	311.63		核对一致	311.63

问题 6、关于发行人毛利率

发行人已区分不同项目类型和特征说明了智能语音类业务销量、在产品结存和新增数据库产品时长与采集服务数量、标注服务数量之间的配比关系，多样化标注语音的标注产出比显著高于单一化标注语音。

请发行人：（1）结合多样化标注语音的不同项目特征进一步说明多样化标注语音的标注产出比的变动原因、远高于其他项目的原因及合理性，相关处理技术如何提高标注的工作效率；（2）结合不同项目特征和语言类型的项目情况，进一步列示单位价格、单位成本中采集和标注服务费的变动情况并对毛利率波动进行分析。

请保荐机构和申报会计师进行核查并发表明确意见。

答复：

（1）结合多样化标注语音的不同项目特征进一步说明多样化标注语音的标注产出比的变动原因、远高于其他项目的原因及合理性，相关处理技术如何提高标注的工作效率

一、多样化标注语音标注产出比远高于其他项目的原因及合理性

多样化标注语音标注产出比远高于其他项目的原因因为标注精细度和复杂度远远超过单一化标注语音项目。

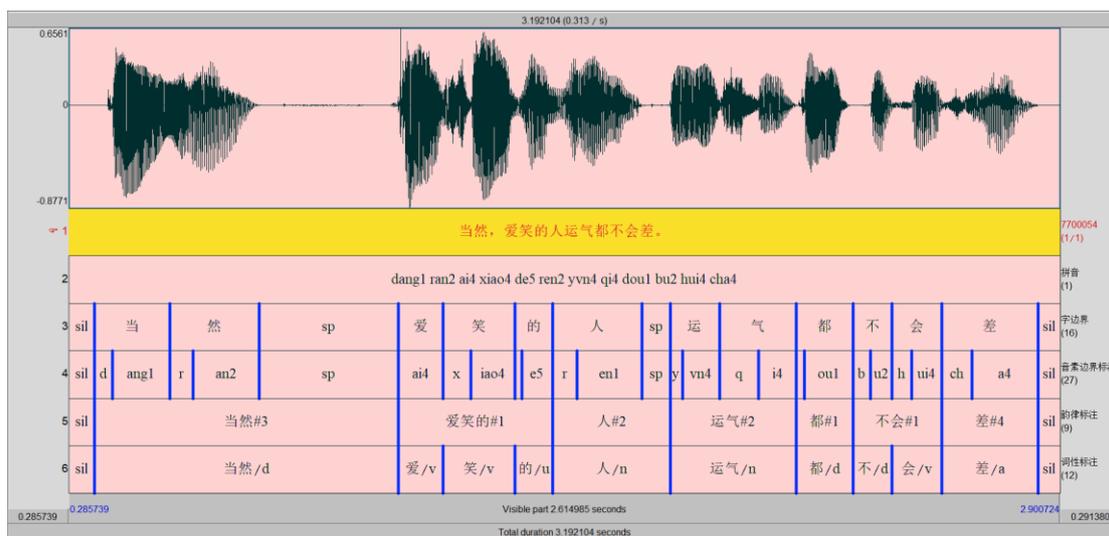
单一化标注语音要求较为简单，有采集的单一化标注语音，通常是发音人对照预设文本进行朗读，因此只需要完成语音与文字一致性的检查、语音边界标注；无采集的单一化标注语音，通常需要先进行人工文本生成后，再进行前述标注即可。

而对于多样化标注语音，主要应用于语音合成算法模型训练，为了使合成的声音更加接近于人声，需要对语音进行各种细节描述，因此需要对数据进行多样化标注，包括音段（音素、音素边界）和超音段（韵律）标注信息，精细和复杂程度远高于单一化标注语音，因此标记同样时间长度的语音文件，所需要的时间更长、环节更多，导致多样化标注语音的标注产出比远高于单一化标注语音。

标注内容	标注说明	有采集的单一化标注语音	无采集的单一化标注语音	多样化标注语音
语音边界标注	标记语音的起止点	√	√	√
人工文本生成	将语音内容听写成文字	√(如果无预设文本)	√	√
语音属性标注	标记发音人性别、方言区		√	
声音事件标注	标记笑声、关门声等非语音类型		√	
音素标注	根据实际发音,标注真实的音段信息,含音素、音节边界、重音/声调			√
音素边界标注	根据实际发音,逐个标示出每个音素在语音上的起止位置			√
韵律标注	听声音判断每个音节末尾的韵律边界类型,分别用#1~#4表示韵律词、二级韵律短语、一级韵律短语、语调短语			√
词性标注	标注每个语法词词性信息			√
标注产出比 ^注		约为 3	约为 12	约为 50-60

注：标注产出比即每小时音频产出所需的标注工作小时数

以多样化标注语音为例，其标注内容与结果如下所示：



因此，多样化标注语音由于其标注内容复杂，环节较多，因此标注产出比远高于其他项目，具备合理性。

二、多样化标注语音标注产出比的变动原因及相关处理技术情况

多样化标注语音标注产出比的变动主要体现为 2018 年单位产出的标注量下降。主要原因是随着 2018 年发行人音素边界自动预测、基于语音特征的韵律预测等高效数据处理技术的成熟和普遍应用，发行人数据处理工作效率得到提高，标注过程中的人工工作量有效减少，使得单位产出的标注量有所下降。

通过技术辅助数据处理，发行人标注工作效率提升主要体现在以下两个方面：

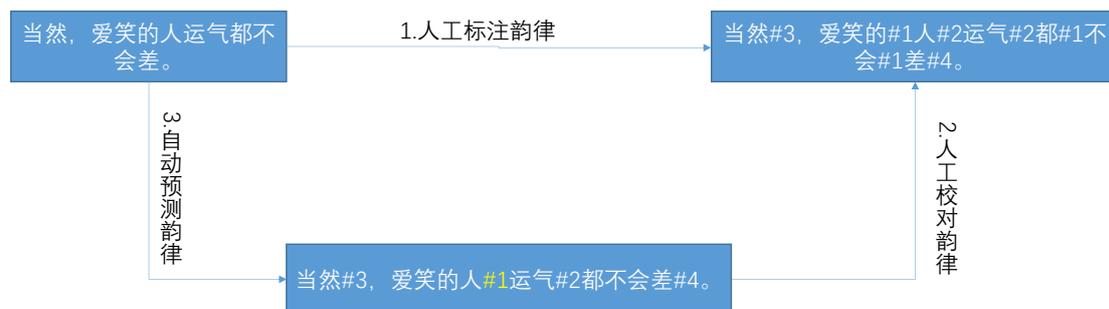
1、发行人利用音素标注正确率校验技术，发现并定位人工标注错误，以帮助快速修改错误，提高数据质量与工作效率。

音素标注的正确性是衡量多样化标注语音数据库质量的重要指标，通常需要通过标注人员通过语音听辨确认音素（音素为根据语音的自然属性划分出来的最小语音单位，以 1,000 句文本为例，按照每句话平均 15 个字，每个字包括声、韵母，音素量为 $1,000 \times 15 \times 2 = 30,000$ 个）。由于人工标注结果难免存在错误，而从大量数据中发现并修正错误极为耗时。

发行人音素标注正确率校验技术，利用语音识别中 GMM-HMM 三音子声学模型，对每个音素样本进行置信度打分，针对不同置信度区间设计不同的音素样本检查校对策略，将校对重点进行有效集中，在减少整体校对工作量的同时，提高了音素标注正确率。发行人已针对该项技术于 2018 年 7 月提交了“音素误标注的检测方法和装置”发明专利申请。

2、发行人利用基于语音特征的韵律预测技术，采用自动预标注，用半自动代替全人工从而提高工作效率。

由于语音韵律具有较强的差异性和个人特色，因此韵律标注需要标注人员通过语音听辨，根据韵律标注规范，判断每个语法词的韵律边界类型。



如上图所示，韵律标注一般有两种方法：

①全人工韵律标注：即上图“1、人工标注韵律”，由标注人员通过听辨语音直接在原始文本上标注韵律符号；

②半自动韵律标注：即先使用算法预测韵律信息（“3、自动预测韵律”），而后再“2、人工校对韵律”，由标注人员结合语音听辨在预测结果上进行修改。

发行人所掌握的基于语音特征的韵律预测技术，可以利用语音识别模型对音素进行自动切分并预测韵律边界，修正基于文本的韵律预测结果，从而提高韵律预标的准确率，有效减少了标注过程中约 25%的人工标注工作量。发行人已针对该项技术于 2018 年 8 月提交了“韵律标注方法、装置和设备”发明专利申请。

（2）结合不同项目特征和语言类型的项目情况，进一步列示单位价格、单位成本中采集和标注服务费的变动情况并对毛利率波动进行分析

报告期内，发行人主营业务的毛利率情况如下：

项目	2018年度			2017年度			2016年度		
	收入占比	毛利率	综合毛利率贡献	收入占比	毛利率	综合毛利率贡献	收入占比	毛利率	综合毛利率贡献
数据资源定制服务	64.20%	45.64%	29.30%	52.89%	47.87%	25.32%	53.15%	40.44%	21.49%
数据库产品	34.27%	100.00%	34.27%	46.10%	100.00%	46.10%	45.48%	100.00%	45.48%
数据资源相关的应用服务	1.53%	78.70%	1.20%	1.01%	82.90%	0.84%	1.36%	79.27%	1.08%
综合毛利率			64.77%			72.26%			68.06%

注：综合毛利率贡献=收入占比*毛利率

（一）综合毛利率变动分析

报告期内，发行人综合毛利率水平分别为 68.06%、72.26%和 64.77%，2017 年比 2016 年上升 4.20%，2018 年比 2017 年下降 7.39%。

发行人综合毛利率的变动影响因素可分解为收入结构的变化及各业务类型毛利率的变化。其中，数据资源定制服务和数据库产品是发行人综合毛利的主要组成部分，其在收入结构中占比的变化和自身各年毛利率的变动，是造成发行人

综合毛利率变动的主要因素。数据资源相关的应用服务收入和毛利规模较小，对发行人综合毛利率水平不构成实质性影响。

发行人数据库产品业务类型的毛利率为 100% 保持不变，收入占比会发生变化；数据资源定制服务的毛利率、收入占比均会发生变化。因此，发行人综合毛利率的变动主要取决于发行人收入结构的变化和数据资源定制服务自身的毛利率水平变动：

1、2017 年综合毛利率较 2016 年增长 4.21%，主要是数据资源定制服务和数据库产品的收入占比无明显变化的情况下，定制服务的毛利率增长 7.43% 所致；

2、2018 年综合毛利率较 2017 年下降 7.49%，主要是数据资源定制服务毛利率基本稳定的情况下，其收入占比提高 11.31% 所致。该收入结构的变化，是由于发行人部分客户随着其人工智能业务的深入发展，数据资源定制化服务需求增长迅速，发行人对数据资源定制复核和数据库产品开发进行更有针对性和更有效的资源分配，以更好的满足客户的各种需求。

以上着重分析了定制服务和数据库产品收入占比结构的变化对综合毛利率的影响；下文重点分析数据资源定制服务的毛利率变动情况。

（二）数据资源定制服务毛利率分析

报告期内，发行人数据资源定制服务收入主要来源于智能语音领域，其收入占比分别为 83.97%、77.72% 和 79.95%，因此主要分析智能语音数据资源定制服务毛利率的变动。发行人智能语音数据资源定制服务各业务细分类别的毛利率情况如下：

业务细分类别	年度	收入（万元）	成本（万元）	收入占比	毛利率
有采集的单一化标注语音	2018 年度	2,809.18	1,606.83	28.41%	42.80%
	2017 年度	1,451.45	741.48	29.66%	48.91%
	2016 年度	1,155.84	685.67	30.78%	40.68%
无采集的单一化标注语音	2018 年度	2,960.54	2,131.45	29.94%	28.00%
	2017 年度	1,409.18	979.43	28.79%	30.50%
	2016 年度	1,610.33	1,159.85	42.89%	27.97%
多样化标注语音	2018 年度	4,119.96	1,584.05	41.66%	61.55%

	2017 年度	2,033.35	821.23	41.55%	59.61%
	2016 年度	988.81	426.01	26.33%	56.92%

报告期内，发行人智能语音类数据资源定制服务的毛利率分别为 39.51%、48.06%和 46.18%，2017 年比 2016 年明显提升，2018 年比 2017 年稳中有降，其主要原因是该类业务中各细分业务方向特点不同，自身的毛利率处于不同的水平；随着整体行业的发展和下游客户的需求变化，不同细分业务方向呈现了一定的结构性变化，促使智能语音类数据资源定制服务的总体毛利率在报告期内产生波动。

总体上看，与 2016 年相比，2017 年毛利率较低的无采集的单一化标注语音收入占比较 2016 年明显下降、毛利率较高的多样化标注语音收入占比明显升高，使得 2017 年智能语音类数据资源定制服务毛利率比 2016 年明显提升。有采集的单一化标注语音收入占比较为稳定，由于 2017 年发行人开展的该类业务特殊定制化项目毛利率较高，从而使得在收入结构未发生较大变化的情况下，2017 年毛利率略高于 2018 年。

下文首先分析各不同业务细分类别之间毛利率差异的原因，再结合单位价格、单位数据服务费以及语言类型等因素，对有采集的单一化标注语音、无采集的单一化标注语音、多样化标注语音三类项目的毛利率波动进行进一步分析，具体如下：

1、不同业务细分类别毛利率差异的原因

(1) 有采集的单一化标注语音项目：涉及采集环节，需要对文本、采集设备及软件、发音人特征分布等要素进行有效的设计，才能保证数据应用于下游客户算法模型的训练效果，是毛利率居中的项目类型，毛利率约为 40-50%。

(2) 无采集的单一化标注语音项目：不涉及采集环节且标注环节类型较为单一，相对复杂度较低，是毛利率较低的项目类型，毛利率约为 30%左右。

(3) 多样化标注语音项目：涉及到高素质发音人及音素、韵律、词性等多种标注类型，相对复杂度较高，所需技术处理环节较多，是毛利率较高的项目类型，毛利率约为 60%左右。

2、发行人智能语音类定制服务毛利率波动分析

(1) 有采集的单一化标注语音项目

该业务类型在报告期收入占比较为稳定，2017年毛利率较高。2017年行业下游各大公司在当年普遍加大投入，开发符合自身定位的智能音箱等产品，产生了较多基于特殊设备的数据需求。具体情况请参见第二轮审核问询函回复问题10(1)。

(2) 无采集的单一化标注语音项目、多样化标注语音项目

这两种业务类型在报告期毛利率相对稳定，而收入占比结构发生了变化。随着智能语音领域人工智能技术的快速进步和相关产品在众多应用场景实现落地，2017年和2018年，毛利率较高的多样化标注语音和毛利率较低的单一化标注语音的收入占比较2016年分别出现明显上升和下降，带动了2017年和2018年整体毛利率水平较2016年有较大幅度提升。

(三) 数据资源定制服务毛利率分析——进一步区分语言类型

报告期内，发行人中文业务毛利率整体上呈现下降状态，反映中文语音数据资源服务的市场竞争程度有所加强，外语种业务毛利率整体上呈现上升状态，体现资源稀缺程度所带来的溢价和发行人多语种覆盖能力所形成的竞争优势。具体分析如下：

数据服务费是数据资源定制服务的主要成本构成，并对数据资源定制服务的定价起到重要的影响作用。报告期内，发行人智能语音三种业务细分类别的单位价格、单位数据服务费、毛利率情况如下：

业务细分类别	语种	2018年度			2017年度			2016年度		
		单位价格(元)	单位数据服务费(元)	毛利率	单位价格(元)	单位数据服务费(元)	毛利率	单位价格(元)	单位数据服务费(元)	毛利率
有采集的单一化标注语音	中文	355.14	140.43	41.14%	448.14	168.71	50.00%	553.11	229.44	50.43%
	外语种	722.76	334.30	44.36%	872.46	374.83	47.75%	716.31	351.05	36.24%
	小计	481.51	207.07	42.80%	585.57	235.47	48.91%	655.82	305.97	40.68%
无采集的	中文	302.46	206.17	26.55%	312.41	189.59	31.02%	291.28	172.80	28.36%

业务细分	语种	2018 年度			2017 年度			2016 年度		
		单一化标注语音	外语种	955.13	496.72	41.28%	572.92	369.64	24.15%	653.94
	小计	324.40	215.94	28.00%	323.70	197.39	30.50%	306.96	183.59	27.97%
多样化标注语音	中文	17,110.43	5,287.70	53.66%	27,392.84	8,868.14	55.01%	20,957.44	6,688.98	57.21%
	外语种	51,796.84	12,919.86	65.36%	25,327.38	7,514.23	62.98%	21,508.92	5,878.10	56.62%
	小计	31,204.04	8,388.76	61.55%	26,161.01	8,060.67	59.61%	21,228.26	6,290.78	56.92%

1、有采集的单一化标注语音

中文业务 2016 年和 2017 年毛利率较为平稳，2018 年毛利率有所下降。由于 2017 年存在一些定制化程度较高的项目对毛利率有一定的提升作用（具体情况请参见第二轮审核问询函回复问题 10（1）），因此毛利率较 2016 年未出现明显下降。扣除该类特殊项目的影响后，2017 年毛利率为 46.15%，报告期内毛利率呈现逐年下降的趋势。该类趋势主要由于：尽管一方面资源丰富和业务成熟带来成本的下降，但另一方面市场竞争导致中文业务的单位价格下降更快，在一定程度上压缩了利润空间，导致了毛利率的下降。

外语种业务 2017 年毛利率较 2016 年明显上升，一方面是由于发行人 2016 年加强外语种业务拓展，对销售价格进行了适当控制，导致毛利率相对较低；而 2017 和 2018 年随着发行人该类外语种业务的逐渐成熟，形成了较为稳定的毛利率水平；另一方面是由于发行人在 2017 年开展了一些低噪声环境、车载环境等场景以及法语、德语、非母语的英语的项目，项目难度和资源稀缺度较高，属于毛利率较高的项目，从而提升了整体毛利率水平。

2、无采集的单一化标注语音

中文业务毛利率 2016 年和 2017 年总体稳定，2018 年有所下降的原因是随着智能客服等领域的需求增加，发行人当年长句音频项目较多，长句通常语速较快且变化更为复杂，人物角色和领域较为多样，因而成本较高，但发行人出于扩大业务规模的考虑未相应提高价格，因此降低了毛利率水平。

外语种业务毛利率 2016 年和 2017 年较为平稳，2018 年显著提升，是由于发行人当年首次开展印度及巴基斯坦地区语种项目，由于语种稀缺，成本提高的同时发行人也具备较强的销售议价能力，因而毛利率较高，从而提升了外语种业

务的毛利率水平。

3、多样化标注语音

中文业务毛利率逐年小幅下降，主要受到中文数据资源日益丰富和市场竞争日趋激烈的影响。外语种业务毛利率逐年持续提升，是由于 2016 年发行人处于外语种多样化标注语音业务的发展初期，发音人资源有限，成本相对较高而价格相对较低，通过报告期内的经营积累，发行人有效扩展了发音人渠道，掌握了该类业务的主要市场资源，提升了议价能力，价格涨幅超过成本，且外语种业务相较于中文业务更为稀缺，下游需求的持续增长也带动了毛利率的逐步提升。

请保荐机构和申报会计师进行核查并发表明确意见。

针对上述事项，保荐机构、申报会计师执行的核查程序：

1、查阅了发行人多样化标注语音的业务介绍文档，了解了发行人多样化标注语音的项目特点、操作流程以及发行人相关辅助数据处理技术；

2、查阅并复核了发行人按照不同项目特征和语言类型划分的业务类型的单位价格、单位成本中采集和标注服务费、收入和毛利、毛利率等相关数据。

3、访谈发行人高管和财务人员，了解发行人业务开展情况以及相关财务指标的变动情况。

经核查，保荐机构、申报会计师认为：

发行人多样化标注语音由于其标注内容复杂，环节较多，因此标注产出比远高于其他项目，具备合理性；2018 年多样化标注语音单位产出的标注量下降的原因是随着发行人音素边界自动预测、基于语音特征的韵律预测等高效数据处理技术的成熟和普遍应用，提高了发行人数据处理工作效率，有效减少了多样化标注语音标注过程中的人工工作量，使得单位产出的标注量有所下降；发行人智能语音类数据资源定制服务的单位价格、单位数据服务费以及毛利率变动情况合理，不存在异常情况。

（本页无正文，为北京海天瑞声科技股份有限公司对《关于北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市申请文件的第三轮审核问询函的回复》之签章页）

北京海天瑞声科技股份有限公司



2019年6月11日

（本页无正文，为华泰联合证券有限责任公司对《关于北京海天瑞声科技股份有限公司首次公开发行股票并在科创板上市申请文件的第三轮审核问询函的回复》之签章页）

保荐代表人：



葛 青



姚玉蓉

华泰联合证券有限责任公司



保荐机构总经理关于审核问询函回复报告的声明

本人已认真阅读北京海天瑞声科技股份有限公司本次审核问询函回复报告的全部内容，了解报告涉及问题的核查过程、本公司的内核和风险控制流程，确认本公司按照勤勉尽责原则履行核查程序，审核问询函回复报告不存在虚假记载、误导性陈述或者重大遗漏，并对上述文件的真实性、准确性、完整性、及时性承担相应法律责任。

保荐机构总经理：



江 禹

华泰联合证券有限责任公司

2019年6月11日